



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# **ANÁLISIS DE LA ANCESTRÍA EN POBLACIONES DE ORIGEN LINGÜÍSTICO CHIBCHA, CARIBE Y ARAWAK DE LA REGIÓN CARIBE COLOMBIANA**

**NELLY DAYANA SUÁREZ MEDELLÍN**

Universidad Nacional de Colombia  
Facultad de Ciencias, Departamento de Biología  
Bogotá D.C., Colombia  
2019

# **ANALISIS DE LA ANCESTRÍA EN POBLACIONES DE ORIGEN LINGÜÍSTICO CHIBCHA, CARIBE Y ARAWAK DE LA REGION CARIBE COLOMBIANA**

**NELLY DAYANA SUAREZ MEDELLIN**

Tesis de Doctorado presentada como requisito parcial para optar al título de:  
**Doctorado en Ciencias-Biología**

Director

WILLIAM USAQUEN MARTINEZ. MSc PhD.  
Profesor Asociado  
Universidad Nacional de Colombia

Línea de Investigación Genética de Poblaciones  
Grupo de Genética de Poblaciones e Identificación

Universidad Nacional de Colombia  
Facultad de Ciencias, Departamento de Biología  
Bogotá D.C., Colombia  
2019

## *Dedicatoria*

*A mis padres por su amor incondicional, su esfuerzo, su sacrificio y dedicación; a mi abuela por sus enseñanzas y su amor y porque aún en su ausencia me hace una mejor persona; a mis hermanas y mis sobrinas por su amor, su compañía, su comprensión.*

# Agradecimientos

Al profesor William Usaquen, por su orientación y apoyo.

Andrea Casas, Fernanda Mogollón, Nicolás Lalinde, Jorge Gómez, Julie Moncada, Diana Roa, Sara Zea, Fredy Rodriguez, Claudia Bonilla y a todos los compañeros del Grupo de Genética de Poblaciones e Identificación

A los hospitales y centros de salud de las poblaciones donde se realizó el muestreo y a todos los participantes del estudio.

Al profesor Fernando Parra y a Henry Flores y su familia por su inmensa colaboración en la etapa del muestreo.

Al profesor Angel Carracedo, Raquel Cruz, Maria Torres, Ana Mosquera, Vanessa Alvarez, Maria Victoria Lareau y a todos los demás compañeros del Laboratorio de Genética Forense y de CE-GEN por su colaboración y apoyo durante mi pasantía en la Universidad de Santiago de Compostela.

A Maika por abrirme las puertas de su hogar y acogerme en su familia cuando estuve lejos de la mía.

A Andres Gutierrez y Diana Pulido por su incondicional apoyo, colaboración y preocupación.

Al doctor Alejandro Giraldo y a la Fundación Gillow por su colaboración y apoyo

A la profesora Indiana Bustos, a Mauricio Rey, Blanca Shroeder y Rita Baldrich del Instituto de Genética de la Universidad Nacional

A mis amigas Jenny, Angela, Manuela, Suly, Lili, Yaneth, por su apoyo y su fe

A mis padres por su amor incondicional, su apoyo, su gran esfuerzo y sacrificio para lograr mis metas educativas.

A mis hermanas y mis sobrinas por su incondicionalidad, su apoyo, su paciencia, su sacrificio y amor, por sostenerme siempre y no dejarme decaer.

A mi querida abuela, llena de valores, mujer luchadora, comprensiva, honesta, amorosa, la que me enseñó letras, números, canciones, y sobre todo el valor de la vida y la familia. A ella, que aún en su ausencia está conmigo cada día en mi corazón.

# Resumen

## ANÁLISIS DE LA ANCESTRÍA EN POBLACIONES DE ORIGEN LINGÜÍSTICO CHIBCHA, CARIBE Y ARAWAK DE LA REGIÓN CARIBE COLOMBIANA

Las poblaciones americanas contemporáneas se han definido como poblaciones triétnicas, por la mezcla histórica de los pueblos originarios nativoamericanos y los linajes europeos y africanos posteriores al descubrimiento de América. En Colombia, la proporción en la que esta mezcla de linajes se dió en los años posteriores a la llegada de los europeos, depende en gran medida de las dinámicas de poblamiento, actividades económicas, supervivencia de los pueblos nativos y el desarrollo cultural y social alrededor de estas realidades. En el interés por realizar el análisis de la ancestría de la región Caribe Colombiana, originalmente habitada por pueblos de filiación lingüística Chibcha, Caribe y Arawak, se han explorado varios niveles históricos, con base en los resultados obtenidos empleando marcadores STRs de Cromosoma Y, SNPs Autosomicos y Secuencias Hipervariable I y II del ADN mitocondrial. Se realizó la comparación entre paneles AIMs y de SNPs de Identificación para el análisis de ancestría y se sugiere una combinación de marcadores como punto de partida para mejorar la resolución a nivel de muestras de poblaciones indígenas suramericanas. Se analizan también los resultados de la estructura genética de las poblaciones colombianas desde marcadores STRs de Cromosoma Y en el contexto de las dinámicas de mezcla, la historia y la cultura, y se realiza un análisis final considerando los linajes nativoamericanos de ADN mitocondrial y de cromosoma Y, desde el punto de vista filogenético en el marco de la geografía y la lingüística. Adicionalmente se reporta el hallazgo de los Subhaplogrupos de ADNmt A2ac, A2q, A2al, C1c3 y D1f y se analiza su implicación en el marco de las hipótesis del poblamiento americano.

**Palabras Clave:** Marcadores informativos de ancestría, Marcadores de identificación forense, STRs Cromosoma Y, Haplogrupos ADN mitocondrial, Ancestría Nativo Americana Colombia.

# Abstract

## ANALYSIS OF ANCESTRY IN POPULATIONS OF CHIBCHA, CARIBBEAN AND ARAWAK LINGUISTIC ORIGIN IN THE COLOMBIAN CARIBBEAN REGION

Contemporary American populations have been defined as triethnic populations, due to the historical mixture of native Native American peoples and European and African lineages after the discovery of America. In Colombia, the proportion in which this mixture of lineages occurred in the years after the arrival of Europeans, depends largely on the dynamics of settlement, economic activities, survival of native peoples and the cultural and social development around of these realities. In the interest of carrying out the analysis of the ancestry of the Colombian Caribbean region, originally inhabited by peoples of Chibcha, Caribbean and Arawak linguistic affiliation, several historical levels have been explored, based on the results obtained using markers STRs of Chromosome Y, SNPs Autosomal and Hypervariable Sequences I and II of mitochondrial DNA. The comparison between AIMs and Identification SNPs panels was made for the ancestry analysis and a combination of markers is suggested as a starting point to improve the resolution at the level of samples from South American indigenous populations. The results of the genetic structure of the Colombian populations are also analyzed from STRs markers of Y Chromosome in the context of mixing dynamics, history and culture, and a final analysis is carried out considering the Native American lineages of mitochondrial DNA and Y chromosome, from the phylogenetic point of view within the framework of geography and linguistics. Additionally, the finding of mtDNA Subhaplogroups A2ac, A2q, A2al, C1c3 and D1f is reported and their implication is analyzed within the framework of the American settlement hypotheses.

**Key Words:** Ancestry Informative Markers, Forensic Identification Markers, STRs Y Chromosome, Mitochondrial DNA Haplogroups, Native American ancestry Colombia

# Contenido

Objetivos .....	10
Objetivo General.....	10
Objetivos Específicos.....	10
Introducción General.....	11
1. Capítulo 1. Poblamiento Americano: ¿Qué nos dice el uso de los diferentes marcadores moleculares? .....	14
<b>1.1 Introducción</b> .....	14
<b>1.2 Marcadores uniparentales</b> .....	16
1.2.1    DNA mitocondrial.....	16
1.2.2    Cromosoma Y .....	25
<b>1.3.    Marcadores Autosómicos</b> .....	31
1.3.1    Genética y lenguaje.....	35
<b>1.4.    Historia de mezcla en Iationamérica</b> .....	36
<b>1.5.    Perspectivas</b> .....	41
<b>1.6.    Conclusiones</b> .....	43
2.    Capítulo 2. Comparación de un panel de SNPs autosómicos de ancestría frente a uno de identificación y los diferentes modelos de estructura poblacional .....	46
<b>2.1 Introducción</b> .....	46
<b>2.2 Materiales y Métodos</b> .....	48
2.2.1.    Poblaciones y encuestas .....	48
2.2.2.    Extracción de ADN y genotipificación de SNPs .....	49
2.2.3.    Análisis genético poblacional y estadístico .....	50
<b>2.3 Resultados y Discusión</b> .....	52
2.3.1    Análisis de estructura genética e inferencia de ancestría.....	52

2.3.2	Análisis de Ancestría individual .....	57
2.3.3	Análisis de Componentes Principales y Análisis Discriminante de Componentes Principales .....	59
2.3.4	Evaluación de Informatividad de marcadores y análisis de clasificación Bayesiano .....	62
2.3.5	Comparación de métodos de asignación .....	64
2.3.6	Relación con datos históricos y migratorios .....	65
<b>2.4</b>	<b>Conclusiones</b> .....	<b>67</b>
3.	Capítulo 3. Análisis de la ancestría nativoamericana de poblaciones del norte de Suramérica .....	70
<b>3.1</b>	<b>Introducción</b> .....	<b>70</b>
<b>3.2</b>	<b>Materiales y Métodos</b> .....	<b>72</b>
3.2.1	Poblaciones y encuestas.....	73
3.2.1	Extracción de ADN y genotipificación de SNPs.....	74
3.2.1	Análisis genético poblacional y estadístico .....	74
<b>3.2</b>	<b>Resultados y discusión</b> .....	<b>75</b>
3.3.1	Estructura genética .....	75
3.3.2	Mejores SNPs para análisis de ancestría nativoamericana .....	88
3.3.4	Relación genética .....	93
3.3.5	Análisis en el marco del poblamiento Suramericano .....	96
<b>3.3</b>	<b>Conclusiones</b> .....	<b>99</b>
4.	Capítulo 4. Diversidad genética y ancestría de los linajes masculinos en el noroccidente de Suramérica .....	101
<b>4.1</b>	<b>Introducción</b> .....	<b>101</b>
<b>4.2</b>	<b>Materiales y métodos</b> .....	<b>103</b>
4.2.1.	Poblaciones y encuestas.....	103
4.2.2.	Análisis molecular.....	104
4.2.3	Análisis genético poblacional y estadístico .....	104
<b>4.3</b>	<b>Resultados y discusión</b> .....	<b>110</b>
4.3.1	Diversidad genética.....	110
4.3.1	Estructura Poblacional y Relaciones Genéticas .....	132
4.3.3	Exploración de datos de posible ancestría .....	135
<b>4.4</b>	<b>Conclusiones</b> .....	<b>139</b>



5. Capítulo 5. Linajes de ADN Mitocondrial y de Cromosoma Y nativoamericanos, revelan la historia de poblamiento del norte de Suramérica.....	143
<b>5.1 Introducción</b> .....	143
<b>5.2 Materiales y Métodos</b> .....	145
5.2.1 Poblaciones y encuestas.....	145
5.2.2 Análisis Molecular.....	145
5.2.3 Análisis genético poblacional y estadístico .....	146
<b>5.3 Resultados y Discusión</b> .....	148
5.3.1 Cromosoma Y .....	148
5.3.2. ADN Mitocondrial.....	157
<b>5.4 Conclusiones</b> .....	191
6. Conclusiones y Recomendaciones.....	193
<b>6.1 Conclusiones</b> .....	193
Material Suplementario Capitulo 2.....	203
Bibliografía.....	227

# Objetivos

## Objetivo General

Caracterizar las relaciones genético-estructurales entre poblaciones de origen indígena del Caribe Colombiano provenientes de diferentes orígenes lingüísticos, desde una perspectiva de ancestría con relación a las teorías sobre el poblamiento suramericano.

## Objetivos Específicos

- Determinar la variación genética en marcadores autosómicos, de ADN mitocondrial y de cromosoma Y en poblaciones de origen indígena del Caribe Colombiano de origen lingüístico Chibcha, Caribe y Arawak.
- Establecer relaciones genético-estructurales entre las poblaciones estudiadas y compararlas con otras poblaciones de referencia suramericanas.
- Comparar los patrones de migración matrilineales y patrilineales.
- Contrastar el análisis demográfico y genealógico con el genético en las comunidades incluidas en el estudio.

# Introducción General

Las preguntas acerca del origen del hombre, las rutas de migración, la dinámica de expansión, las características y el origen de las poblaciones fundadoras fueron abordadas por mucho tiempo analizando yacimientos arqueológicos, restos de asentamientos urbanos, vocablos y registros pictográficos o escritos, que por mucho tiempo fueron la única fuente de información para encontrar las respuestas a estas preguntas.

En los años sesenta se reconstruyeron las primeras filogenias de poblaciones humanas empleando datos genéticos de grupos. Con base en estos datos se propuso un modelo de expansión por reemplazo de poblaciones preexistentes, donde a partir de una población núcleo que se divide, se da origen a dos poblaciones que evolucionan independientemente y posteriormente se vuelven a separar, pero también es posible que nuevas poblaciones se formen a partir de mezclas de poblaciones parentales de las cuales evolucionan también en forma independiente. Según el autor, “el tiempo de los pasos de expansión puede ser muy útil para calibrar el proceso de diferenciación genética y examinar la constancia de las tasas evolutivas en nuestro árbol” (Cavalli-Sforza *et al.*, 1992, Cavalli-Sforza *et al.*, 1988).

La forma como las fuerzas evolutivas han labrado los cambios desde las primeras poblaciones humanas hasta hoy aun no está clara, pero la genética de poblaciones ofrece un conjunto de conceptos que permiten un acercamiento a la resolución de estos cuestionamientos, estudiando la evolución de las poblaciones humanas desde la perspectiva de la mutación, la selección natural, la deriva genética y la migración entre otros.

Así la diversidad humana es producto no solo de los aspectos biológicos, sino que su historia y su cultura son parte fundamental de esa diferenciación. Se favorece la interdisciplinariedad y se permite la integración del estudio de la estructura genética de las

poblaciones humanas actuales desde una perspectiva de ancestría, en convergencia con los análisis genealógicos y demográficos, la filología de las lenguas y las teorías basadas en hallazgos arqueológicos, como un importante elemento que contribuye con la resolución de las preguntas acerca de la historia de la humanidad. (Cavalli-Sforza *et al.*, 1992, Cavalli-Sforza *et al.*, 1988).

Parte del proceso de entendimiento de la diversidad humana está en conocer los modelos de migración y poblamiento de la Tierra, un fenómeno que comenzó hace alrededor 100.000- 50.000 años en Africa.(Goldstein & Chikhi, 2002, Pitblado, 2011, Schurr, 2004). Sobre el poblamiento Americano existen dos teorías la del poblamiento temprano o pre-clovis y la del poblamiento tardío (López & Ranere, 2008, Pitblado, 2011, Schurr, 2004), en ambos casos se considera la posibilidad de que los desplazamientos sucedieran siguiendo la costa Pacífica, vía acuática o terrestre, o caminando por los territorios del interior. Colombia debió ser paso obligado de los primeros pobladores que no conocían la navegación y que tuvieron que atravesar la región del Darién para entrar a Sur América, y aquellos que realizaban sus recorridos en bote por la ruta del Pacífico, pudieron tocar las costas colombianas y posteriormente desplazarse por tierra hacia la costa norte colombiana, en busca de mejores condiciones de vida y acceso a los recursos.

Con base en los hallazgos arqueológicos, la región Caribe fue un foco de dispersión cultural. Actualmente, las comunidades indígenas presentes en esta zona del país corresponden a las familias lingüísticas Chibcha, Caribe y Arawak. Los primeros con mayor número se encuentran distribuidos en una amplia extensión de la región, pero los grupos Caribe en Colombia solo son 2: los Yuko/Yukpa, ubicados en la Serranía del Perijá y los Karijona en la región Amazónica. En el caso de las tribus de lengua Arawak, la mayoría de concentra en las regiones del Amazonas y Orinoco y solamente los Wayuu se encuentran asentados en la península de La Guajira. Bajo esta perspectiva es de interés conocer los procesos de migración y asentamiento de las comunidades precolombinas a través del análisis de ancestría de los pueblos actuales que representan las tres familias lingüísticas.

Con el presente trabajo, se espera analizar marcadores tipo SNPs (Single Nucleotide Polymorphism) autosómicos de identificación y de ancestría, así como STRs (Short Tandem Repeats) de cromosoma Y y las regiones hipervariables I y II de ADN mitocondrial de una

muestra de población del Caribe Continental Colombiano, con el fin de estudiar su estructura genética en busca de los rastros de los grupos indígenas precolombinos, poder establecer relaciones con otras comunidades colombianas y suramericanas documentadas, comparar los patrones de migración matrilineales y patrilineales y con esta información evaluar la presencia de variantes genéticas específicas para los marcadores de ancestría en las poblaciones estudiadas y explorar el concepto de aislado genético. Adicionalmente, con el uso de una herramienta diseñada para obtener datos demográficos y genealógicos de los participantes, se realizará la comparación de esta información con el resultado de los análisis genéticos. Finalmente se espera poder presentar una propuesta sobre modelos de migración y poblamiento considerando aspectos genéticos, con relación a datos históricos, demográficos y genealógicos.

# **1. Capítulo 1. Poblamiento Americano: ¿Qué nos dice el uso de los diferentes marcadores moleculares?**

## **1.1 Introducción**

Las diferentes hipótesis acerca del poblamiento americano se soportan en evidencias que provienen de disciplinas como la arqueología, antropología, etnografía, lingüística y más recientemente, la genética. Aunque inicialmente se plantearon distintas rutas migratorias, en la actualidad parece haber acuerdo con relación a que los primeros pobladores de América entraron al continente por el norte, a través del estrecho de Bering (Fiedel, 2000; Schurr, 2004; Pitblado, 2011). Las controversias hoy en día giran principalmente en torno a preguntas tales como el tiempo en que esta migración se produjo, cuántas migraciones sucedieron, cuál era el origen de los primeros pobladores y cuál era el tamaño de las poblaciones migrantes, entre otras.

Los primeros acercamientos desde el punto de vista genético a este problema se hicieron utilizando marcadores proteicos y grupos sanguíneos para caracterizar las poblaciones (Cavalli-Sforza, et.al. 1994). El desarrollo de la tecnología de ADN favoreció el uso de polimorfismos de longitud como los STR, de secuencia y los polimorfismos binarios. Cada uno de estos tipos de marcadores permite dibujar un panorama diferente con relación al tipo de herencia, y explorar así la historia de las poblaciones humanas en los contextos migratorios y demográficos.

Para estudiar las etapas tempranas del poblamiento americano se recurre a poblaciones nativoamericanas en las que se espera encontrar menor grado de mezcla y por tanto hallar marcadores específicos que puedan ser trazados hasta los grupos que originalmente migraron por el estrecho de Bering, siguiendo el recorrido hacia el sur y su distribución en Norte, Centro y Suramérica.

Múltiples estudios se enfocan en el análisis de marcadores uniparentales con el objetivo de determinar los linajes fundadores. Tras el análisis de polimorfismos binarios, aproximadamente 600 marcadores o eventos de mutación clasificados en más de 300 haplogrupos constituyen hoy el árbol de cromosoma Y (Karafet, et.al., 2008; van Oven, et.al., 2014). En cuanto a los linajes maternos la clasificación más general de los haplogrupos de ADN mitocondrial puede realizarse por técnicas de RFLP's. La secuenciación directa de las regiones HVS-I y II y más recientemente la mitogenómica y la comparación con la secuencia de referencia de Cambridge, permitieron la construcción también un árbol de haplogrupos con sus respectivas ramificaciones (van Oven, et.al. 2009). El haplogrupo Q de cromosoma Y y los linajes A, B, C, D y X de ADN mitocondrial se han encontrado representativamente en las poblaciones nativas americanas (Torroni, et.al.,1993), (Santos et.al., 1999), (Lell, et.al. 2002), (Bortolini, et.al., 2003), (Shurr, 2004),(Pitblado, 2011), (Dulik, et.al., 2012), (Regueiro, et.al.; 2013). Diferentes tipos de análisis multivariados, redes de subhaplogrupos específicos y distribución de frecuencias entre otras estrategias, se han empleado para reconstruir la historia de los linajes paternos y maternos. A este tipo de estudios se suman los de marcadores autosómicos tipo STRs, SNP's o INDELS para estudiar la divergencia genética de las poblaciones nativoamericanas ó la ancestría en poblaciones mezcladas y tiempos desde la mezcla (Wang, et.al., 2007), (Kosoy,et.al; 2009), (Galanter, et.al., 2012), (Pereira, et.al.,2012).

Al realizar la caracterización genética de las poblaciones empleando los diferentes tipos de marcadores, el hallazgo de alelos o haplogrupos privados permite contrastar las hipótesis acerca de las primeras migraciones humanas hacia América y su consiguiente dispersión en el continente, planteadas con base en evidencias diferentes a las de tipo genético, como por ejemplo con datos obtenidos desde la clasificación lingüística (Cavalli-Sforza, et.al., 1992 y 1997; Hunley, et.al., 2008; Guerra, et.al. 2013; Roewer et.al., 2013), lo cual permite interesantes y múltiples perspectivas hacia el futuro de la investigación interdisciplinaria.

## **1.2 Marcadores uniparentales**

El uso de los marcadores uniparentales cuenta una historia de poblamiento o de divergencia que depende de aspectos como el modo de herencia, el contexto histórico, cultural, demográfico y geográfico, así como la interacción entre las fuerzas de cambio evolutivo, entre otros aspectos. Para el continente americano en particular, eventos de deriva genética y efecto fundador, extinción y colonización y circunstancias históricas como la llegada de los europeos y africanos al continente, así como los patrones de migración y las costumbres culturales para encontrar pareja, modelaron la estructura genética de las poblaciones, que actualmente muestran un sesgo ligado al sexo que se evidencia en los resultados de diferentes estudios realizados con marcadores de ADN Mitocondrial, de Cromosoma Y y autosómicos (Bolnick, et.al.; 2006; Hunley, et.al, 2011; Bisso-Machado, et.al., 2012; Salzano & Sanz, 2014). A continuación se describen los hallazgos con las diferentes clases de evidencia genética.

### **1.2.1 DNA mitocondrial**

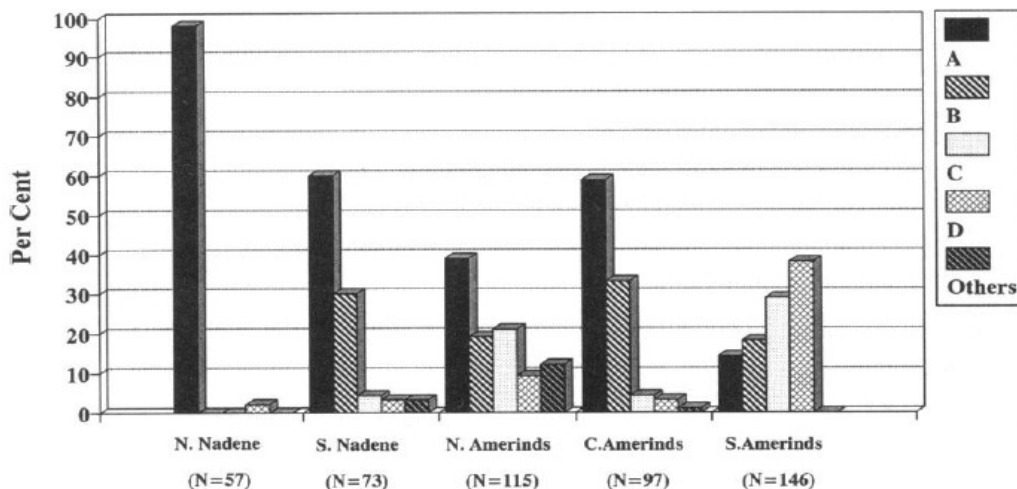
#### **1.2.1.1. Diversidad Genética**

Después de que Anderson publicara la secuencia completa del DNA mitocondrial (Anderson et.al., 1981), los estudios sobre la diversidad de las poblaciones humanas empleando RFLPs, mostraron la existencia de “tipos” (hoy haplogrupos) restringidos a determinados grupos étnicos (caucásicos, asiáticos, africanos y nativos americanos) y la estimación de distancias genéticas resaltó que la relación de los nativos americanos (representados por muestras de indígenas Waorani) con las poblaciones asiáticas, y se propuso un tiempo de divergencia de 5500 años para la separación de caucásicos y Asiáticos/Nativos americanos (Johnson et.al 1983). Los estudios realizados con mapeo de restricción sobre regiones codificantes de las moléculas de DNA mitocondrial en muestras de población de diferentes orígenes, llevaron a plantear como hipótesis, que los primeros linajes de los humanos para esta molécula fueron de origen Africano, dada la mayor diversidad existente en esta población y a los resultados obtenidos para la construcción de los árboles por el método de parsimonia, (Cann, 1987).



Posteriores trabajos incluyeron un mayor número de comunidades indígenas e hicieron uso de técnicas de secuenciación en la región D loop, para estudiar la diversidad y distribución de las variantes de cada haplogrupo (Wallace & Torroni 1992), obteniendo un árbol de consenso estricto de 92 haplotipos para los 4 haplogrupos descritos hasta el momento para población nativoamericana A, B, C y D y población asiática. Clasificando los grupos estudiados de acuerdo con su filiación lingüística según Greenberg, se estableció la frecuencia de los haplogrupos para poblaciones Na Dene del norte y del sur y Amerindios de norte, centro y Suramérica (**Figura 1-1**). Con estos resultados también fue posible determinar la ausencia de ciertos haplogrupos y la presencia de “haplotipos privados” en ciertas poblaciones, así como los tiempos de divergencia entre los haplogrupos, que variaron entre 22750 y 45500 años para el haplogrupo A y 6000 a 12000 para el haplogrupo B (Torroni et.al., 1993). El haplogrupo X fue descrito en 1998 por Brown y colaboradores y a través del análisis de secuencias completas de DNA mitocondrial, fue posible establecer la existencia de no más de 5 linajes fundadores en América, incluyendo el haplogrupo X y la definición de subhaplogrupos (Bandelt, 2003)

**Figura 1-1.** Tomada de Torroni et.al., 1993. Distribución de haplogrupos de DNAmT en Na-Dene del Norte (Canadá), Na-Dene del Sur (Atapascanos del Sur), y Amerindios Norte, Centro y Suramericanos. “Otros” indica haplotipos de probable origen europeo.



Análisis de RFLPs y la secuenciación de las regiones HVS I y II en poblaciones nativas americanas, muestran diferencias en el grado de diversidad o la presencia de marcadores

restringidos a poblaciones específicas, y escasos linajes no nativos. La distribución de frecuencias de los haplogrupos respecto de la ubicación geográfica muestra que la frecuencia del haplogrupo A disminuye de norte a sur siendo mayor en núcleos claros en norte, centro y Suramérica y la de C y D aumenta de Sur a Norte, siendo la más alta la frecuencia del haplogrupo C en la región noroccidental, sur de Brasil y norte de Argentina, mientras el haplogrupo B está ausente en Norteamérica. Las altas frecuencias de los haplogrupos B y D permiten observar una clara separación entre oriente y occidente del continente y la presencia de haplotipos privados para algunas tribus, es resultado de efecto fundador y deriva genética, debido a la fijación y extinción de haplotipos (Shurr, 2004) (Bisso-Machado et.al., 2012).

Según Shurr (1999 citado en Shurr, 2004), el haplogrupo B no se encuentra en poblaciones actuales del nororiente de Siberia, debido a un probable reemplazo de la población tras el inicio de la migración inicial, pero si está presente en Mongoles, Siberianos y en la región Panamericana, lo cual está en contra de la idea de una migración tardía por la costa del Pacífico. Sin embargo, esto podría indicar que se dio una divergencia de este haplogrupo en Asia oriental, posterior a la divergencia de A,C y D, y ya que la edad aproximada del haplogrupo B es de 13590 años antes del presente (aap), es compatible con la datación para los sitios Clovis tempranos. Adicionalmente, el haplogrupo X se restringe a Norteamérica pero se ha detectado en poblaciones europeas, lo que podría ser resultado de flujo genético entre poblaciones paleolíticas de la región de Trans-Baikalia y pueblos de origen Mongol, de migraciones tempranas desde la cordillera del Caúcaso y los montes Urales o producto de las invasiones de los Turco-Mongoles, lo que también podría explicar los linajes ancestrales de cromosoma Y encontrados y que relacionan los linajes masculinos con ancestros de los europeos modernos y no con mongoloides de Asia oriental (Fiedel,2000).

Los hallazgos de la mitogenómica han permitido identificar diferentes linajes presentes en la región de Asia/Beringia y a lo largo del continente americano. A nivel Panamericano se han encontrado A2, B2, C1, D1 y D4h3a; en el sur de Suramérica B2i2, C1b13, D1g y D1j; en Estados Unidos y Canadá A2a, A2b, C4c, D4e1, X2a (Ojibwa) X2a y X2g y en familias Eskimo-aleutianas y Na-dene D2a y D3. El haplogrupo C4c se reportó en 2 individuos Ijka (Arhuaco), 1 individuo Shuswap en Canadá y en un individuo “criollo” en Argentina. Así

mismo se han detectado 1 variante de origen asiático C1a y 3 variantes de americanas C1b, C1c y C1d (un individuo C1d1) (Saint Pierre et.al. 2012), (Bodner et.al., 2012), (Bravi 2013).

Lewis y colaboradores (2007) usaron la región HVS-I en el análisis de la diversidad genética del DNA mitocondrial de poblaciones andinas y amazónicas y encontraron patrones diferentes de distribución de la variación entre y dentro de las poblaciones de ambas regiones. Las poblaciones andinas presentaron valores de diversidad genética significativamente mayores que las poblaciones no andinas y menor nivel de variación genética entre ellas que las poblaciones amazónicas ( $F_{st}$  0,06 Vs  $F_{st}$  0.35 respectivamente). A pesar de estos resultados, el análisis molecular de varianza (AMOVA) no rechazó la similitud entre poblaciones occidentales y orientales, sugiriendo que cuentan con un ancestro común. Diferentes escenarios migratorios pueden plantearse con respecto a estos resultados: uno con poblaciones originales no significativamente diferentes, uno con un flujo genético entre poblaciones orientales y occidentales suficientemente amplio como para homogenizar la variación, y un tercero en el que múltiples migraciones ocurrieron con un patrón de dispersión diferente al presentado en el estudio. Dado que los resultados de este trabajo no pueden desestimar la hipótesis de una sola población fuente, la variación dependería del aislamiento regional y de los procesos evolutivos diferenciales que generaron los patrones de variación observados entre las poblaciones del oriente y del occidente suramericano. (Lewis et.al., 2007)

Los cambios climáticos sucedidos hace 7000-10000 años, obligaron el desplazamiento de las poblaciones chibchas desde centro hacia Suramérica con un subsecuente cambio en las estrategias de subsistencia y modificaciones culturales que generaron bajo flujo genético. En el estudio de la relación de poblaciones chibchas de centro y Suramérica, la presencia del haplogrupo C en grupos chibchas de Santa Marta (ausente en chibchas del sur de Centroamérica) puede ser indicativo de deriva genética o migraciones tempranas hacia Suramérica. También se ha detectado la presencia del sitio de corte para la enzima Msp I en chibchas de Santa Marta, que está ausente en Centroamérica, lo que podría indicar que la pérdida del sitio de corte en esas poblaciones sucedió después de la separación de los grupos pertenecientes a la subfamilia Magdalénica. (Melton et.al. 2007)

### 1.2.1.1. Migración y rutas de poblamiento

Los resultados obtenidos a través de estos estudios moleculares han sido empleados para elaborar hipótesis de migraciones ancestrales y de la dispersión de las poblaciones en los eventos de colonización, con base en la observación de haplogrupos compartidos con posibles poblaciones de origen y en la detección de variantes específicas para América, lo que coincide con la idea de que una vez han divergido las poblaciones, cada rama sufre cambios independientes en un proceso de diferenciación de generación en generación con respecto de las poblaciones ancestrales (Cavalli-Sforza et al., 1994); (He et al., 2012; Henn et al., 2012).

Considerando una tasa de mutación estable, los datos sobre la divergencia de secuencias acumulada de los haplogrupos con respecto a asiáticos y siberianos (de acuerdo con Torroni y sus colaboradores (1993), para los haplogrupos A, C, y D se calcula entre 0,053% y 0,096%, y la del haplogrupo B en 0,024%), permiten plantear hipótesis sobre las rutas migratorias, considerando los patrones de dispersión geográfica de los diferentes linajes y estimar los tiempos de separación de los mismos. Wallace propone el poblamiento americano en dos fases, una primera migración donde se dispersaron numerosas variantes de los haplogrupos A, C y D que con el paso de las generaciones derivaron también en haplotipos únicos para la región y una segunda oleada sucedida después de que sumergiera el puente terrestre en el norte del continente, que permitió la entrada de poblaciones con haplogrupo B empleando una ruta costera por vía marítima, que luego se distribuyeron entre las poblaciones ya existentes (Wallace, 1994). El análisis de RFLP's de autores como Brown et al., 1998 (citado en Bandelt et al., 2003), Shurr, 2004 y Torroni et al., 1994, han sugerido un único haplotipo fundador para cada uno de los haplogrupos A, B, C, D y X. Estos planteamientos son coherentes con los tiempos de migración propuestos por Torroni y su equipo (Torroni et al., 1994), de 22000 a 29000 años antes del presente (aap) para los haplogrupos A, C y D y de 8000 a 11000 (aap) para el haplogrupo B, con base en una tasa de evolución estimada de 0,022 a 0,029% por 10000 años, a partir de muestras de una población de origen lingüístico Chibcha en Panamá. El estudio de grupos lingüísticos Na-dene de Canadá Noroccidental y Alaska, encontró solamente haplogrupo A con una diversidad estimada de 0,021, por lo que se propone la llegada de una migración posterior a la del haplogrupo B antes mencionada (Wallace, 1994). Otros

autores como Bailliet et.al 1994, Easton et. al. 1996, Merriwether et. al. 1994 y 1995, Rickards et al. 1999 y SE Santos et. al. 1996, citados por Shurr (2004), apoyan la idea de más de un haplotipo fundador. A pesar de las dificultades para determinar los haplotipos fundadores que han perdido polimorfismos clave desde los linajes asiáticos (Ejemplo: mutaciones recurrentes en la región control), el análisis de la región HVS-I sugiere que los haplotipos fundadores de estos haplogrupos en América, son idénticos a los de Asia y Siberia. (Shurr, 2004).

Los análisis de RFLP's han permitido obtener la información demográfica útil para estudiar fenómenos de expansión demográfica que pueden ser fechados mediante tasas de mutación conocidas, determinando la dispersión de haplogrupos en Siberia o Eurasia entre 25000 a 18000 aap y en el nuevo mundo entre 18000 y 12000 aap, por lo que un tiempo aproximado de 35000 a 20000 aap refleja diversidad genética acumulada en América. (Shurr, 2004).

Con el uso de herramientas bayesianas como el gráfico de línea de horizonte bayesiano (Bayesian skyline plot), considerando tasas de sustitución diferenciales para regiones codificantes y no codificantes en el análisis de las regiones HVS I y II del DNAm, se ha propuesto un modelo de poblamiento en "3 estaciones": la primera considerando la divergencia de los ancestros amerindios desde el pool genético asiático, la segunda correspondiente a un largo periodo de aislamiento (de 7 a 15 mil años de duración) en el cual se generaron las variantes alélicas específicas del nuevo mundo, y la tercera etapa relacionada con un periodo de expansión en América (16000 años antes del presente) coincidente con la apertura del Corredor Libre de Hielo y acompañada de un incremento poblacional. Esta hipótesis se soporta con datos obtenidos a partir de fósiles de plantas y polen que describen la región de Beringia como una pradera con las condiciones apropiadas para mantener la vida mamífera, y los hallazgos arqueológicos y etnográficos que muestran a las primeras poblaciones humanas americanas como grupos adaptados al incremento de la población como hábiles cazadores. Adicionalmente, se estima un grupo aproximado de migrantes de 1000 a 2000 individuos. (Mulligan, et.al. 2008).

Las tecnologías de secuenciación de ADN se han empleado para estudiar con mayor profundidad las hipótesis sobre la llegada y dispersión de los diferentes haplogrupos al

continente americano. Así, la secuenciación de 46 genomas mitocondriales mostró que los subhaplogrupos B2i2 y C1b13 se encuentran restringidos a Chile y Argentina, pero no cuentan con la misma distribución. Así, B2i2 se encuentra presente en mayores frecuencias en grupos nativos que habitan la región centro sur de Chile y en Argentina (la mayoría pertenecientes a la familia de lenguaje Araucano) y está ausente en Norteamérica y otros lugares de Suramérica y en Tierra del Fuego y el haplogrupo B2 no se ha encontrado en restos antiguos de la zona. Mientras el haplogrupo C1b13 está ampliamente distribuido en el norte y sur del cono sur de Suramérica. Según la edad de coalescencia de los subhaplogrupos B2i2 y C1b13 ( $10.8 \pm 3.8$  y  $12 \pm 3.3$  miles de años respectivamente) se infiere que se originaron posiblemente a la llegada de los primeros pobladores al cono sur de Suramérica, cuando iniciaron los procesos de tribalización y con ellos, la diferenciación del lenguaje. Se presume que la mutación en B2 que dio origen al subhaplogrupo B2i tuvo lugar antes o poco después de la entrada al norte de Suramérica. (Saint Pierre et.al. 2012).

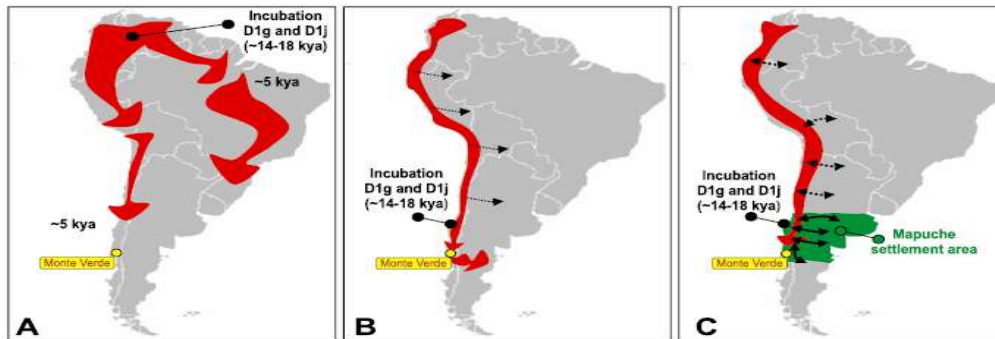
Del mismo modo, con el fin de evaluar la posibilidad de que la dispersión del hombre en América, se haya dado en dos olas migratorias, la primera con individuos portadores de los haplogrupos A, B, C y D y la segunda con portadores del haplogrupo X, Fagundes y colaboradores realizan en 2008 el análisis de 86 secuencias completas de DNA mitocondrial, incluyendo representantes de los cinco linajes fundadores americanos, y empleando una estimación de edad de coalescencia con un método bayesiano. Los resultados obtenidos indican que las edades de coalescencia de todos los linajes, incluido el X, se encuentra entre 20.680 y 22.140 años. De acuerdo con estos datos, se plantea que la distribución limitada de los haplogrupos Xa2 actuales, puede estar relacionado con la posición de los portadores en la ola migratoria o con una baja frecuencia inicial del haplogrupo entre los primeros pobladores. Propone como posible modelo una colonización inicial del territorio de Beringia por una población fundadora con un posterior desplazamiento por vía costera del Pacífico hacia el sur de América y una dispersión más reciente y extensa hacia el interior del continente. (Fagundes et.al. 2008). El análisis filogeográfico de dos subhaplogrupos raros D4h3 y X2a, a partir del estudio de genomas mitocondriales completos apoya el uso simultáneo de dos rutas de entrada y dispersión por el continente americano, donde poblaciones provenientes de Beringia se desplazaron hacia el sur por la costa pacífica de Norteamérica portando el subhaplogrupo D4h3 y otros grupos

migraron al interior por el corredor libre de hielo, creado entre las placas de hielo Laurentina y de la Cordillera, portando el haplogrupo X2a (Perego, et.al. 2009).

También el análisis más específico de los linajes D1g correspondientes a una transición C16187T con al menos 6 ramas identificadas (D1g1, D1g2, D1g3, D1g4, D1g5 y D1g6) con mayor frecuencia en Chile (16.6%), y D1j correspondiente a un motivo mutacional T152C-C16242T-T16311C con dos linajes informados (D1j1a1 y D1j1a2), con mayor frecuencia en Argentina (9.3%), se realizó por la secuenciación completa de 45 genomas mitocondriales en los que se observó la formación clara de 2 clados con una distribución geográfica específica. Considerando que las edades de coalescencia calculados para estos dos últimos linajes son de  $18.3 \pm 2.4$  y  $13.9 \pm 2.9$  miles de años, puede pensarse que los primeros pobladores ya contaban con ellos, pero el recorrido de norte a sur del continente americano debió suceder en aproximadamente 2000 años. Esta idea es respaldada por los resultados de tres simulaciones realizadas con base de demografía y ecología, que indicaron un tiempo aproximado de 1.5 miles de años para llegar a la parte norte de Suramérica y 2.5 miles de años para llegar a Monte Verde. Una simulación basada en datos genéticos (Fix, 2005 citado en Bodner, et.al. 2012) arrojó un resultado de 3000 años para la colonización de norte a sur a lo largo de la costa. De esta forma se proponen tres posibles escenarios de poblamiento al interior de Suramérica: 1. Una única población de origen establecida inicialmente al norte de Suramérica, con una separación tardía en un grupo que tomó la ruta costera y otro que migró por el interior, que habría sucedido después de la diversificación completa de todos los subhaplogrupos de D1g y D1j. Este escenario implica un inicio reciente de la migración hacia el sur y por tanto no apoya la idea de la colonización humana en Monte Verde hace 14000 años, ni explica la ausencia de los linajes D1g y D1j en el norte de Suramérica. 2. Una migración costera de norte al sur de Suramérica, con posterior migración transandina hacia el interior con flujo genético limitado a lo largo de la cordillera. 3. Una colonización costera de sur a norte de Suramérica con una posterior migración transandina y continuo intercambio a través de la Cordillera de los Andes, especialmente al sur de Suramérica. (**Figura 1-2**). Los resultados de este estudio favorecen la idea de las migraciones “trans-andinas”, que explicarían tanto la diversidad del linaje D1g a ambos lados de los Andes como los resultados del estudio en población Mapuche, ya que esta región geográfica corresponde con el territorio habitado por los ancestros de este grupo indígena, que mantuvo su aislamiento de los conquistadores europeos por 3 siglos.

La hipótesis de la colonización costera con posterior flujo genético a través de la Cordillera de los Andes, es apoyada al considerar que la supervivencia en la zona de la costa requería menos adaptación facilitando la rápida migración y que la fase de calentamiento pudo abrir pasos a través de las montañas a la altura de 2000 a 2500 msnm hace aproximadamente 14.6 – 14.3 miles de años (Bodner, et.al. 2012).

**Figura 1-2.** Tomada de Bodner, et.al. 2012. Modelos de migración dentro del continente americano.



El posible hallazgo de un individuo con linaje D1j en República Dominicana y de unos restos humanos precolombinos correspondientes a la etnia Tainos (extintos por la colonización europea en la zona Caribe) podrían estar relacionados con la hipótesis del origen de este grupo por desplazamientos de población desde la región de la Orinoquía hacia las islas del Caribe hace aproximadamente 3000 años, aunque también se ha planteado la posible participación de individuos provenientes de Norte y Centroamérica. (Bodner, et.al. 2012)

### 1.2.1.2. Genética, geografía y lenguaje

Los resultados del estudio realizado con al menos dos poblaciones representantes de los lenguajes Tupi, Macro-Ge, Carib, Chibchan, Mataco-Guaicuru, Arawakan, Araucanian, Choco, Chon, Tucanoan, Aymaran, Barbacoan, Guahiban, Witotoan, Salivan, Quechuan y 21 poblaciones con otros lenguajes, solo un 7% de la variabilidad de la frecuencia de los haplogrupos mitocondriales hallados dentro de las poblaciones puede ser atribuible al lenguaje. La influencia de la distribución geográfica y del lenguaje parece ser similar en el DNA mitocondrial (Bisso-Machado et.al., 2012)



## 1.2.2 Cromosoma Y

### 1.2.2.1. Diversidad Genética

En el estudio de la caracterización genética de las poblaciones empleando marcadores de cromosoma Y, la presencia de alelos, haplotipos o haplogrupos específicos de ciertos grupos de población, ha sido útil en la discusión de hipótesis de ancestría, migración y poblamiento, desde la historia patrilineal humana. En estudios con poblaciones nativas americanas se observa que el linaje de mayor frecuencia es Q, definido por el marcador QM242, que se reporta como mutación ancestral aQM3 por Seielstad en 2002 y Bortolini 2003 (Bravi, 2013). Los linajes considerados autóctonos americanos son Q1a3a (QM3), Q1a3\* (M346) y C3 (Bravi, 2013).

El marcador QM3 (definido por DYS199-T), tiene alta frecuencia en nativos americanos y su incremento clinal de Norte a Sur, muestra diferencias significativas entre Norte, Centro y Suramérica, en un patrón de distribución similar al evidenciado con Gamma Globulinas, HLA y datos genéticos autosómicos (Shurr, 2004). El alelo DYS199T es producto de una mutación por transición en el alelo original DYS199C (con origen en Beringia hace alrededor de 30.000 años durante una fase estacional) (Lell et.al.,1997), que se encuentra presente en todas las muestras estudiadas pertenecientes al antiguo continente, así como en chimpancés y orangutanes. En las muestras provenientes de América, además del alelo C, también se encontró el alelo DYS199-T en el 90% de los individuos estudiados de Centro y Suramérica y aproximadamente en el 50% en Norteamérica y esporádicamente encontrado en poblaciones siberianas como Evens, Chukchi y Esquimales ubicados en Chukotka (Siberia) (Underhill, 1996), (Lell et.al.,1997), (Bravi, 2013). La presencia de los alelos C y T en poblaciones Eskimo y Navajo, sugiere que la mutación que dio origen al alelo T, pudo haberse presentado antes de la dispersión hacia Centro y Suramérica (Underhill, 1996). El haplotipo fundador determinado con los marcadores 92R7, DYS199, SRY1532, Tat, YAP, ah y pYal (presente en poblaciones a ambos lados de Beringia), se encontró en el 90% de 306 muestras provenientes de las poblaciones de nativos americanos, y junto con los hallazgos anteriores se consideran como evidencia a favor de una única migración colonizadora hacia América, proveniente de Siberia Central y que durante su permanencia en Beringia en el Pleistoceno, sufrió una diferenciación que dio

origen a un nuevo linaje de cromosoma Y, caracterizado por la mutación en DYS199 (Santos et.al., 1999).

El estudio sobre la diversidad de cromosoma Y de nativos americanos (incluyendo poblaciones Na Dene y poblaciones sur América) y Mongoles, propone que el linaje caracterizado por la mutación PM45 (carente del alelo derivado T en DYS199) es ancestral al linaje determinado por la mutación QM242 (Haplogrupo Q) (Bortolini et.al., 2003). La existencia de 2 linajes para PM45: a y b, donde PM45a un linaje más cercano a QM3 presente en poblaciones de Siberia Centro-Sur y Norte de Suramérica, y PM45b con un haplotipo diferente a QM3 portando la mutación M173 presente en poblaciones de Siberia Oriental y Norteamérica (Chipewayan) (Lell et.al., 2002), respaldan la teoría de 2 migraciones iniciales para el poblamiento Americano. Sin embargo Tarazona-Santos considera que PM45b presente en Chipewayan no puede considerarse ancestral porque también se encuentra en Europa, por lo que su presencia se relacionaría con mezcla reciente, pero dada su alta frecuencia (63%) en estos nativos americanos, para Bortolini et.al. considerar la mezcla reciente sería inconsistente con la preservación cultural de la identidad europea. Aunque el haplotipo 3 del haplogrupo PM45 solo está presente en Chipewayan, el haplotipo 2, con el que tiene solo una diferencia se encuentra presente en Asia, por lo que Lell et.al. (2002) consideran que el haplotipo 3 se originó en Asia y no es producto de mezcla reciente. (Bortolini et.al., 2003).

En la recopilación de Bisso-Machado, Bortolini y Salzano (2012) el linaje Q1a3a\* (M3) presenta la mayor frecuencia y los sublinajes Q1a3a1 (M19), Q1a3a2 (M194), Q1a3a3 (M199) y Q1a3a4 (SA01), muestran localizaciones regionales restringidas. El linaje Q1a3\*(M346) reportado por Bailliet en 2009, corresponde a un grupo parafilético a M3, presente en una frecuencia de 7%, encontrado tanto en grupos indígenas como en población mestiza de Argentina, Chile, Paraguay, Bolivia, Ecuador, Estados Unidos, China, Asia Central, Rusia, Pakistán, India y Arabia Saudita (Bravi, 2013), es el segundo grupo más frecuente y su distribución generalizada sugiere su existencia en las poblaciones ancestrales iniciales. (Bisso-Machado, et.al., 2012).

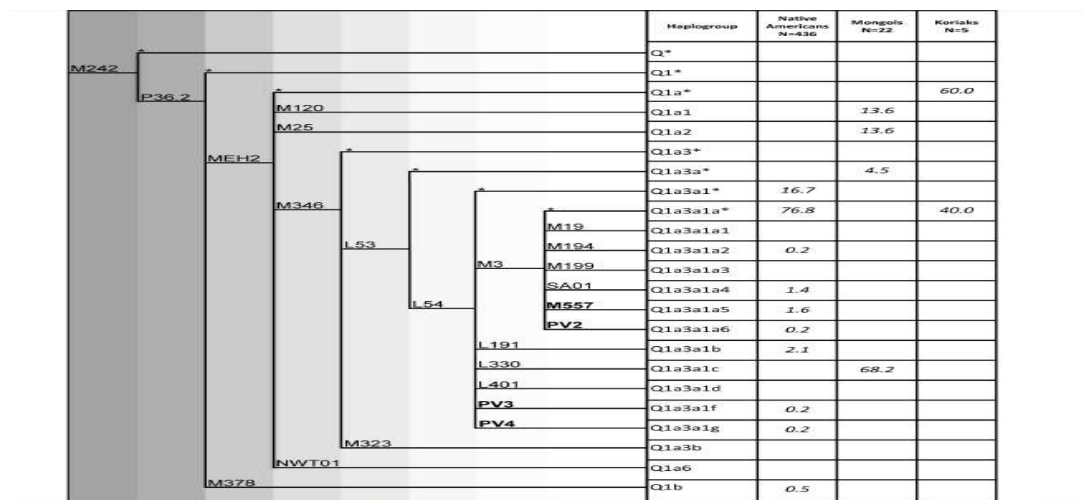
Battaglia y colaboradores (2013) realizaron un estudio de haplogrupos de cromosoma Y con 463 muestras de nativos americanos clasificadas como pertenecientes al haplogrupo

Q (**Figura 1-3**). En esta investigación el 76.8% de las muestras fueron clasificadas en el paragrafo Q1a3a1a\* (M3\*), incluyendo 11 individuos del sitio Tompullo y 2 en Peru, y el 16.7% de las muestras se incluyeron en el paragrafo Q1a3a1\* (L54\*). En total, se encontró que el 99.5% de los nativos americanos fueron portadores de la mutación L54, que incluye los paragrafos Q1a3a1\*(L54) en la zona de Mesoamérica, Q1a3a1a\* (M3\*) y los subhaplogrupos Q1a3a1b (M194), Q1a3a1a2 (M199), Q1a3a1a4 (SA01) y los 4 nuevos linajes Q1a3a1a5 (M557) (1.6%), Q1a3a1a6 (PV2) (0.2%), Q1a3a1f (PV3) (0.2%) y Q1a3a1g (PV4) (0.2%) descritos por este autor. Los subhaplogrupos definidos por las mutaciones M194, SA01, M557 y PV2 fueron encontrados en poblaciones Andinas. Los subhaplogrupos Q1a3a1a1 (M19), Q1a3a1a3 (M199), Q1a3a1c (L330), Q1a3a1d (L401), también caracterizados por esta mutación, no fueron encontrados en la muestra analizada. El 50% de los individuos pertenecientes a las poblaciones propuestas como origen de los primeros migrantes desde Asia (Mongoles y Koriaks), fueron clasificados en el subhaplogrupo Q1a3a1c (L330) y el paragrafo Q1a3a1a\* (M3\*). El subhaplogrupo Q1a3a1b (mutación L191) reportado generalmente en poblaciones de la costa occidental americana, fue encontrado en nativos americanos de México y Perú y el tiempo estimado desde su expansión fue de  $3600 \pm 1600$  años, correspondiente con el tiempo en que vivieron los antiguos Mayas. Se encontró el subhaplogrupo Q1a3a1f (mutación PV3) asociado al alelo 13.2 del Sistema STR DYS19, encontrado en la base de datos YHRD (<http://www.yhrd.org>) en Mayas de Guatemala y 2 Hispanos de Estados Unidos. (Battaglia et.al., 2013)

La presencia del haplogrupo C se observa en Norteamérica, Asia Centro y Sur Oriental, Oceanía y Siberia. (Bravi, 2013). El subhaplogrupo C3\* hallado en Asia oriental con mayor frecuencia en Kamchatka (38% en Koryaks, 36% en Mongolia exterior y 38% en Mongolia Interior), pero no en Norte o Centroamérica, se encontró en 14 individuos de la muestra de Ecuador y correspondieron al 26% de la población Quechua y 7.5% de la población Waorani. En el análisis Median Joining Networks incluyendo 396 individuos previamente reportados, los 14 indígenas del Ecuador se ubicaron lejos del centro de la red, indicando que el último contacto entre poblaciones nativas americanas y las otras poblaciones del análisis (China, Corea, Mongolia, Indonesia, Siberia, Tibet, Japón, Altai, Rusia), sucedió a la llegada de los primeros pobladores a América. El haplotipo más frecuente en los individuos de Ecuador fue parcialmente compartido por 2 individuos de Kamchatka y en

general se relacionaron de forma distante con los haplotipos provenientes de población Wayuu en Colombia. (Roewer et.al., 2013).

**Figura 1-3.** Tomada de Battaglia et.al., 2013. Arbol filogenético del haplogrupo Q de cromosoma Y. Nomenclatura y frecuencia (%) de ramas de haplogrupos en las poblaciones analizadas. La nomenclatura está en acuerdo con la propuesta por CCY 2002, la negrilla indica los marcadores nuevos reportados por el autor y los \* señalan los paragrupos.



### 1.2.2.2. Migración y rutas de poblamiento

De acuerdo con los datos recopilados, la entrada de los primeros migrantes a América se produjo por Beringia, donde permanecieron por un tiempo como consecuencia de las dificultades de desplazamiento producidas por el último periodo glacial. Posteriormente, estos portadores se trasladaron hacia América a través del Corredor Libre de Hielo en dirección al sur y quedaron aislados de la población de origen cuando desapareció el corredor, dando origen a la población Amerindia. Una vez retirada la glaciación, nuevos migrantes que habían permanecido en Beringia, moviéndose desde Siberia Central o Mongolia, pudieron desplazarse hacia el continente americano, dando origen a los grupos de lenguaje Na-dene y Eskimo portadores tanto de la mutación DYS199-T (QM3\*), como del linaje original DYS199-C. (Santos et.al., 1999)

De acuerdo con Roewer y colaboradores (2013) es poco probable la llegada del subhaplogrupo C3\* con los primeros pobladores y su posterior pérdida por deriva genética para explicar la ausencia de evidencia actual del linaje C3\* en norte y Centroamérica, ya que se requerirían tasas de migración aproximada de 2.5%. El resultado aproximado de

6000 años como Tiempo al Ancestro Común Más Reciente (TACMR) permite proponer que se presentaron olas migratorias posteriores que no dejaron rastro en Norte y Centroamérica, o que se presentó contacto posterior entre grupos asentados en Ecuador y grupos provenientes de Asia oriental por mar vía costera occidental, por Sur América occidental o a través del pacífico. (Roewer et.al., 2013).

Varios hallazgos apoyan la hipótesis de que los primeros migrantes provenían de la región Centro Sur de Siberia, siendo portadores del haplogrupo Q, que sufrió un proceso de diferenciación inicial en Beringia. Empleando análisis de redes se observa una amplia distribución en todo el continente americano de los subhaplogrupos definidos por la mutación QM3\* (lo que se considera compatible con una entrada temprana del subhaplogrupo al continente), con difusión de norte a sur, con una posterior expansión local en Canadá, México y región Andina, asociada a pérdida de varianza y diversidad haplotípica. Los individuos portando el marcador QM3, probablemente llegaron a México junto con grupos con la mutación L54\*. Tras sucesivos eventos de aislamiento y posterior expansión demográfica facilitada por el uso de la agricultura y el establecimiento de los grandes imperios, sucedió la diversificación local, como se puede evidenciar en las muestras de los nativos americanos de la zona Andina y de México. (Battaglia et.al., 2013)

El hallazgo del linaje definido por QM19 solo en población Ticuna y Wayuu, de origen reciente, con un tiempo estimado muy cercano al de QM3\*, sugiere una separación temprana poco después del inicio de la dispersión humana. Esta teoría se apoya con los resultados de DNA mitocondrial y es consistente con la fuerte deriva genética en nativos americanos y podría explicar la gran diversidad de lenguaje indígena. (Bortolini et.al., 2003). Dado que este grupo fue encontrado también en la población Toba de Argentina, es posible inferir que la colonización inicial no se dio únicamente por la vía del Pacífico y tuvo un periodo de colonización trans andina. (Battaglia et.al., 2013).

El análisis de los haplotipos obtenidos con los marcadores microsátélites (DYS19, DYS389A, DYS389B, DYS390, DYS391 y DYS393) y el alelo DYS199-T, y los resultados de distancias genéticas y análisis de autocorrelación, apoyan la idea de que las poblaciones Andinas tienen mayor diversidad genética y menor distancia genética que las poblaciones Amazónicas, como consecuencia de que las poblaciones Occidentales o Andinas fueron

originadas primero a partir de movimientos migratorios diferentes a los que generaron las poblaciones Amazónicas ó, a que la estructura genética de los distintos grupos de población, podría ser consecuencia de eventos de cuellos de botella sufridos durante la dispersión o de diferenciación evolutiva posterior, con muy bajo flujo genético entre las poblaciones Andinas y Amazónicas, e igualmente sugieren que esta situación está relacionada con tamaños efectivos de población relativamente grandes. (Tarazona-Santos et.al., 2001).

#### **1.2.2.3. Tiempo de migración**

El análisis de las mutaciones determinó a QM3 con una edad de 7570 años y a QM19 con una edad de 7972 años y QM242 como haplogrupo fundador con una edad estimada de 13611 en Amerindios y 15416 en Mongoles, lo que sugiere una pronta entrada a América, luego de la mutación y posterior cuello de botella. Por la frecuencia y diversidad haplotípica de PM45 en Siberia Centro-Sur, se propone ésta como la región de origen de la migración inicial a América. (Bortolini et.al., 2003). El linaje Q1a3a1a (M3) es aparentemente más diferenciado en la región andina. El subhaplogrupo correspondiente a la mutación SA01 (Q1a3a1a4), reportado en un individuo peruano y uno boliviano tiene una edad estimada de  $7.2 \pm 2.6$  ka, lo que coincide con la aparición de la agricultura andina temprana y el subhaplogrupo Q1a3a1a5 (mutación M557) se ha calculado por tener  $2.6 \pm 1.4$  ka. (Battaglia et.al., 2013)

#### **1.2.2.4. Geografía y lenguaje**

5 grandes categorías de lenguaje: Tupi, Carib, Macro-Ge, Quechua, Mataco-Guaicuru y un sexto grupo integrado por las lenguas aisladas se tuvieron en cuenta en el análisis de marcadores STR de cromosoma Y. Las divisiones geográficas incluyeron Amazonia/Central Meseta Brasileña, Sur de Brasil, Chaco y Andes. La mayor parte de la variación por marcador fue atribuida a las diferencias intrapoblacionales, tanto para clasificación lingüística como para distribución geográfica, excepto para DYS437. Con relación a la variación entre subdivisiones, se detectaron resultados significativos atribuibles a la distribución geográfica en 5 de seis marcadores. La variabilidad interpoblacional dentro de las subdivisiones, se atribuyó en promedio el 16% a la geografía y el 21% al lenguaje. La

influencia de la distribución geográfica es aparentemente mayor que la del lenguaje (Bisso-Machado, et.al., 2012).

Un estudio posterior de la variación geográfica y de lenguaje empleando 1011 muestras de 81 localidades y 51 grupos étnicos, mostró que la variación genética entre sitios de muestreo pudo atribuirse entre un 21% y un 28% al set de marcadores tipo STR y el 85% al conjunto de 16 SNPs de Cromosoma Y. El Análisis de Autocorrelación Espacial (AAE) mostró correlación positiva entre haplotipos de STR de Cromosoma Y únicamente para sitios separados por unos cientos de kilómetros. El análisis de redes para el haplogrupo Q tampoco mostró correlación entre la clasificación lingüística ni geográfica. Sin embargo el análisis de variación haplotípica y lenguaje llevado a cabo bajo dos formas de clasificación lingüística, en los 26 grupos de Campbell y en las 4 clases de Ruhlen, determinó una asociación significativa para haplogrupos, aunque los resultados del AMOVA no indican que la variación genética para los STR de cromosoma Y puedan ser explicados por la variación lingüística de acuerdo con las clases de lenguaje, mientras las diferencias entre grupos explican entre el 8% y el 12% de la variación en el haplotipo. El análisis de escalamiento multidimensional, no mostró relación entre los resultados genéticos y las clases de lenguaje, excepto ligeramente para Chibcha-Paez, para el que se estimó un TACMR de 168 a 206 generaciones (con 30 años por generación indicaría 5040 a 6180 años). (Roewer et.al., 2013)

### **1.3. Marcadores Autosómicos**

Los primeros acercamientos a este problema desde el punto de vista genético se hicieron a través del uso de marcadores proteícos y los grupos sanguíneos para caracterizar las poblaciones. El análisis de los sistemas ABO, Diego, Duffy, MNS, P1, Kell, RH y otros marcadores como Fosfatasa Acida, Adenilato kinasa 1, Esterasa D, Glioxilasa 1, Proteína de enlace a la vitamina D, Haptoglobina, Transferrina y HLA entre otros., apoyan la hipótesis de 3 oleadas de poblamiento planteadas por las evidencias lingüísticas: Amerindio, Na-Dene y Eskimo y no descartan el aporte de más de una población en la primera oleada. Los datos genéticos también soportan las evidencias lingüísticas que muestran los grupos Na-Dene divididos en norte y sur, éstos últimos relacionados con los grupos amerindios. En Suramérica la estructura genética muestra división de tribus y el inicio de nuevos grupos

con pequeños números de fundadores apoyado por el efecto de deriva y en congruencia con las múltiples ramificaciones de los lenguajes de origen amerindio. Sin embargo, los datos genéticos obtenidos a través de estas estrategias no fueron suficientes para aclarar y soportar hipótesis de poblamiento más específicas dadas las incongruencias con las evidencias geográficas y lingüísticas. (Cavalli-Sforza et.al., 1994).

El análisis de 678 microsatélites autosómicos en 29 poblaciones nativas americanas y dos siberianas estableció que la heterocigocidad disminuye de norte a sur y de occidente a oriente, encontrándose los valores más bajos en poblaciones aisladas del Amazonas y del Oriente Suramericano. El único alelo privado detectado en todas las poblaciones de nativos americanos muestreadas con una frecuencia de 31.6%- 36.4% corresponde a un alelo "regional" de 275 pares de bases en el locus D9S1120 (Shroeder et.al., 2007) (Wang et.al., 2007), no observado en las otras poblaciones estudiadas, excepto en 2 poblaciones del límite oriental de Siberia, lo cual se relaciona con resultados obtenidos a través de estudios de polimorfismos de DNA mitocondrial y de cromosoma Y y podría explicarse considerando una sola migración de norte a sur con efecto fundador serial. Los resultados de sus estudios de correlación entre heterocigocidad y distancia geográfica y la evaluación de distintos escenarios de rutas de migración costeras versus por el interior, favorecen la idea de la preferencia de las costas para el desplazamiento inicial de los primeros pobladores. La ruta más probable de acuerdo con estos resultados es a través de la costa pacífica (y no se da soporte a la idea de las migraciones a lo largo de los ríos), lo cual es apoyado por la similitud entre las poblaciones mesoamericanas y andinas (**Figura 4-1**). De igual manera, los resultados obtenidos muestran un posible origen de los pueblos orientales a partir de los occidentales en Suramérica y el árbol neighbor-joining de las poblaciones nativas americanas muestra una posible relación entre el cluster Chibcha-Paez y las poblaciones del sur de Mesoamérica Cabecar y Guaymi, que podrían originarse a partir de las primeras. Así mismo, la formación de un cluster entre poblaciones de lengua Chibcha-Paez y Tucano ecuatorial sugiere un origen a partir de una migración separada de la de los andes. (Wang, et.al. 2007).

En 2012, He y colaboradores proponen una medida de la divergencia de poblaciones con base en el análisis de sistemas STR de Wang et.al., 2007. El modelo planteado tiene como base la relación entre la varianza en el tamaño del STR y el tiempo de coalescencia entre



pares de linajes, considerando la acumulación de mutaciones sufridas posteriormente a la divergencia de las poblaciones y el tamaño efectivo de las poblaciones ancestrales. Al calcular los tiempos de divergencia genética con base en los datos de Wang et.al., 2007, se encuentran los mayores valores para la población Chipewayan y entre las poblaciones norteamericanas estudiadas la más reciente sería Cree. Las poblaciones Arhuaco y Guaraní mostraron los mayores tiempos de divergencia en Centro y Suramérica. En general el mayor tiempo de divergencia se observó entre Chipewayan (Norteamérica) y Mixe (Centroamérica) con 697 generaciones y el menor entre Quechua y Aymara (141 generaciones) ambos ubicados en la región occidental de Suramérica y su lenguaje perteneciente a la rama Andina de la familia amerindia. (He et.al., 2012).

**Figura 1-4.** Tomado de Wang et.al., 2007. Vías de menor dificultad en el escenario de migración costera



El modelo migratorio planteado con base en este análisis indica que el inicio de la colonización tuvo lugar hace aproximadamente 556 generaciones por Norteamérica con algunas poblaciones migrando hacia Centroamérica. Después de un periodo estacionario

de al menos 162 generaciones, un segundo grupo de migrantes se movió hacia el sur de Norteamérica y algunos de ellos se desplazaron hasta llegar a Suramérica por vía terrestre sucedió hace aproximadamente 394 generaciones, colonizando este territorio en un tiempo aproximado de 23 generaciones moviéndose especialmente a través de los Andes, sin considerar flujo genético entre las poblaciones ni una gran contribución europea, por lo que podría haberse desestimado el tiempo de divergencia, y aunque el tiempo de separación calculado entre las poblaciones de Norte y Suramérica de 556 generaciones (11120-13900 años) “no permite soportar la hipótesis de que sus ancestros llegaron en tiempos significativamente diferentes”, tampoco se elimina la posibilidad de mezcla o de reemplazo poblacional, ni de más de una ola migratoria. (He et.al., 2012).

La construcción de un árbol neighbor-joining basado en distancias  $F_{st}$  entre pares de poblaciones nativas americanas y siberianas genotipificadas con 364470 SNP, mostró concordancia con la distribución geográfica y categorías lingüísticas (Rulhen, 1991 citado en Reich et.al., 2012). Se observó además una separación temprana entre Asiáticos y Nativos americanos y de Siberianos nororientales, éstos últimos agrupados en una rama separada con Eskimo-aleutianos y separados a su vez de los nativos americanos. En América se observa una división en grupos consistente con poblamiento de norte a sur, lo que se apoya en la correlación negativa con la heterocigocidad respecto a la distancia al estrecho de Bering. Esta correlación también se incrementa a menor distancia de la costa. Este modelo no permite considerar flujo genético después de la separación de las poblaciones. Usando el Test Hotelling-T, se infirió que el número mínimo de eventos migratorios requeridos para explicar los datos obtenidos era 3, consistentes con la clasificación de lenguajes de Greenberg y Rulhen. El modelo de expansión fue propuesto mediante el uso de “admixture graph”, que considera posibilidades de flujo genético unidireccional. Este análisis permite soportar las hipótesis del retorno de poblaciones Eskimo-aleutianas a Siberia Nororiental y que los nativos de baja Centroamérica no son descendientes directos de los primeros pobladores sino producto de una regresión de población por Panamá. (Reich et.al., 2012)

Según este modelo, la mayoría de los nativos americanos provienen de la primera ola migratoria procedente de Asia que cruzaron por el estrecho de Bering hace más de 15000 años. Adicionalmente, ha sido posible reconocer al menos otros 2 eventos migratorios

relacionados por un lado con Eskimo aleutianos y por el otro con grupos Chipewayan hablantes del lenguaje Na-dene. Sin embargo, estos autores reportan un alto grado de mezcla de las poblaciones de los primeros americanos y posteriores migraciones de asiáticos que no son consideradas por el modelo empleado, por ejemplo al considerar que tanto Eskimo-Aleutianos como Na-dene, tienen la mitad de su ancestría proveniente de los primeros americanos. (Reich et.al., 2012)

### **1.3.1 Genética y lenguaje**

Las conclusiones acerca de la correlación entre diversidad genética y lenguaje hechas considerando la clasificación de Greenberg (1987) y Ruhlen (1991), mostraron bajos valores de correlación de Mantel para distancias genéticas y lingüísticas por lo cual Wang y colaboradores consideran que una correlación entre estos dos factores solo puede establecerse hasta cierto grado de distancia lingüística. (Wang et.al., 2007).

Los resultados del estudio de la relación entre genoma y lingüística con base en coalescencia, a través de una aproximación bayesiana que involucra varios marcadores genéticos y que considera factores como la varianza del tamaño efectivo de las poblaciones en el tiempo entre demes, el flujo genético y posibles eventos de fisión, se emplearon para el análisis las clasificaciones lingüísticas de Loukotka, Greenberg y Campbell y para el análisis genético 381 de los 678 marcadores tipo STRs empleados por Wang et.al. en 2007, en 5 poblaciones representando los 5 principales grupos lingüísticos: Andino, Arawak, Chibcha-Paez, Macro-je y Tupi. La más alta probabilidad posterior para cada modelo se obtuvo con la clasificación de Greenberg, seguida por la de Campbell. La clasificación de Loukotka no contó con soporte frente a los datos genéticos. De acuerdo con los resultados, la divergencia entre Tupí y Arawak sucedió con mayor probabilidad 2812 años antes del presente, y el tiempo de la primera separación de los lenguajes suramerindios hace 3094 años. Estos resultados presentan una cierta concordancia entre los datos genéticos y lingüísticos, aunque los intervalos de confianza resultantes son muy amplios y los datos lingüísticos no se obtuvieron con estrictos criterios estadísticos. (Guerra et.al., 2013)

## 1.4. Historia de mezcla en Latinoamérica

Los grupos nativoamericanos que habitaban el continente al momento del descubrimiento, tuvieron distintos patrones de dispersión y de acuerdo con la evidencia establecieron entre ellos relaciones comerciales, culturales, de poder e incluso flujo genético. En la historia del poblamiento americano se plantean hipótesis que sugieren que acontecimientos como la separación de un grupo de migrantes hacia la región Amazónica desde los Andes Orientales al llegar al norte de Suramérica, la expansión del imperio Inca, la migración de las tribus nativas fuera de la región del Orinoco (Bolnick et.al., 2016), la relación ancestral de poblaciones nativas de Venezuela con pueblos mesoamericanos y la relación entre etnias hablantes Tukano-ecuatorial del Amazonas con las Antillas apoyando la hipótesis del origen de los Tainos en Suramérica (Moreno et.al., 2013) y la migración separada de las poblaciones andinas con respecto de las poblaciones de lengua Chibcha-Paez y Tukano-Ecuatorial (Wang et.al., 2007) moldearon la estructura genética de las poblaciones nativoamericanas ancestrales.

La estructura genética de la población latinoamericana actual es producto de un complejo patrón de mezcla interétnica que comenzó con el contacto con los europeos en el siglo XV. Diversos factores como el origen, número y sexo de los migrantes, el número de generaciones transcurridas, las disposiciones de la corona frente al matrimonio con los locales, las epidemias, el uso de la fuerza para dominar a los habitantes que causó la muerte a miles de nativos y la esclavitud de otros, así como la llegada de grupos de africanos para reemplazar la fuerza de trabajo indígena, entre otros aspectos, influenciaron la forma y grado en el que este proceso sucedió en diferentes zonas del continente (Salzano & Sanz, 2014). Tanto la mezcla entre algunos grupos nativoamericanos antes de la conquista como la sucedida con los europeos posteriormente, dificultan encontrar claras evidencias acerca de las diferencias entre los primeros pobladores de América y las relaciones que se establecieron entre los pueblos durante la dispersión en el continente (Hunley & Healy, 2011).

Los datos históricos, los patrones de migración y el análisis de diferentes tipos de marcadores muestran un claro sesgo sexual en el proceso de mezcla post contacto (Bryc et.al., 2010; Adhikari, et.al. 2016; Salzano & Sanz, 2014; Ibarra et.al., 2014; Eyheramendy

et.al., 2015). Las proporciones de mezcla son variables entre regiones y están influenciadas por patrones complejos de movimiento posterior de los grupos, por lo que el análisis de ancestría a partir de marcadores autosómicos en poblaciones mezcladas, puede ser apropiado para evitar la influencia del sesgo sexual (Adhikari, et.al. 2016), lo cual además de permitir corroborar hipótesis históricas y antropológicas, es de gran importancia en la actualidad en el campo clínico, para evitar el sesgo en los estudios de asociación (casos y controles) por estratificación poblacional y en el campo forense en lo relacionado con establecimiento de ancestría biogeográfica.

Las diferentes estrategias de estudio de mezclas “genome-wide” y con marcadores autosómicos como SNPs, Indels y STRs se sustentan en la idea de que el análisis de ancestría biogeográfica permite orientar el probable origen geográfico de un individuo con base en los marcadores informativos de ancestría (AIMs), es decir de alelos heredados de sus antecesores tanto de poblaciones ancestrales como de poblaciones contemporáneas que pueden ser rastreados mediante análisis genómicos (Phillips et.al, 2015a; Bolnick et.al., 2016). Los resultados de estas investigaciones han establecido que la amplia variabilidad de los patrones de mezcla en América Latina puede no ser suficientemente explicable a través de modelos únicos (Bryc et.al., 2010).

Dadas las limitaciones técnicas y de presupuesto para que un laboratorio pueda implementar herramientas de análisis genome-wide, varios investigadores enfocan sus esfuerzos en establecer paneles eficientes de Marcadores Informativos de Ancestría (AIMs) con estrategias NGS, Electroforesis Capilar, SnapShot ó microarrays (Santos et.al., 2015a; Phillips et.al., 2016). Una comparación de los distintos tipos de marcadores puede verse en la **Tabla 1-1** (Santos, 2015b). Pardo-Seco y colaboradores (2014) al evaluar la precisión de los sets AIMs para cuantificar la ancestría de los genomas, sugieren usar sets de ancestría con más de 400 marcadores para evitar sesgo y tener precaución en inferencias realizadas con paneles pequeños, lo que también es sugerido por Eyheramendy et.al. (2015) y De la Puente, et.al. (2016). Aunque autores como Tsai 2005 y Turakulov y Easteal 2003 (Citados en Russo et.al., 2016) proponen 100 y 65 como el número mínimo de marcadores para estimar ancestría individual, en población Argentina se concluyó 30 como el mínimo aceptable, coincidiendo con resultados de Ruiz Narvaez 2010 (Russo et.al., 2016).

Grandes conjuntos de marcadores STR autosómicos han sido empleados para estudiar la diversidad genética y composición ancestral de poblaciones nativoamericanas, estableciendo hipótesis de rutas de poblamiento y relaciones entre los diferentes grupos en comparación con la clasificación lingüística (Wang et.al., 2007; Guerra et.al., 2013). De igual forma plataformas multiplex con números variables de marcadores tanto SNPs como INDELS que se listan en la **Tabla 1-2** se utilizan para diferenciar poblaciones a nivel continental. De acuerdo con Phillips, 2015b, los marcadores autosómicos permiten estimar frecuencias con un número bajo de muestras (30 a 40). El estudio de marcadores para diferenciar la filiación étnica fue sugerido inicialmente por Shiver en 1997 pero su utilidad forense fue puesta en uso en 2003 con un panel de 178 AIMs SNPs de la compañía DNAprint. Los marcadores ideales son aquellos cuya variación en las frecuencias alélicas permite diferenciar poblaciones y su selección se basa en medidas de diferenciación tales como  $F_{st}$ ,  $\delta$ , la medida de informatividad de asignación  $I_n$  de Rosenberg o de diversidad de Shannon (Phillips, 2015b y Phillips et.al., 2016). Un punto básico en la selección de los marcadores es obtener el balance en el grado de informatividad para evitar el sesgo, usando el Valor Acumulado de Divergencia Específico de Población (Population-Specific Divergence PSD), con el fin de obtener el ajuste de la proporción de SNPs. De acuerdo con Ding et.al. 2011, el  $I_n$  es el valor que estima la proporción de ancestría con menor sesgo. Un ejemplo de la detección de sesgo con diferentes tipos de marcadores es reportado por Phillips et.al. (2016), al indicar que la contribución europea está sobrestimada (20 a 40%) en el análisis ancestría nativoamericana en poblaciones Bolivianas con el set de 46 INDELS (Taboada et.al., 2013). Uno de los trabajos para estudiar la mezcla en Latinoamérica más destacados por realizar una estricta selección de marcadores AIMs, con base en la longitud de rama locus específica (Locus-Specific Branch Length) para el  $I_n$ , fue publicado por Galanter y colaboradores en 2012 y permite inferir con menor sesgo el origen geográfico de los individuos a nivel continental.

**Tabla 1-1.** Comparación de tipos de marcadores. Adaptado de Santos, 2015b.

	STRs	SNPs	INDELs
<b>Número de sitios</b>	1 por cada 10 Kb	1 por cada 100 pb	1 por cada 1000 pb
<b>Tasa de mutación</b>	$10^{-3}$ a $10^{-4}$	$1.2 \times 10^{-8}$	$2.3 \times 10^{-9}$
<b>Alelos por marcador</b>	5 a 50	Usualmente 2 algunos trialélicos	Usualmente 2
<b>Amplificación de Artefactos</b>	Si	Si	Si
<b>Información fenotípica</b>	No	Es posible predecir algunos caracteres	No
<b>Información de Ancestría</b>	Limitada	Si	Si
<b>Principales ventajas para uso forense</b>	Altamente informativos Marcadores multialélicos que facilitan detección de mezclas Bases de datos establecidas	Amplificados en pequeños amplicones (Mayor tasa de éxito de amplificación en muestras degradadas)  Bajas tasas de mutación  Predicción de fenotipo y ancestría	Tipificación por electroforesis capilar  Amplificados en pequeños amplicones (Mayor tasa de éxito de amplificación en muestras degradadas)  Bajas tasas de mutación  Predicción de ancestría
<b>Principales limitaciones para uso forense</b>	Amplificados en grandes amplicones (Menor tasa de éxito de amplificación en muestras degradadas)	Baja informatividad para identificación (se requiere mayor número para alcanzar poder de discriminación)  Mezclas son difíciles de detectar	Baja informatividad para identificación (se requiere mayor número para alcanzar poder de discriminación)

**Tabla 1-2.** Paneles de ancestría desarrollados con diferentes tipos de AIMs

Tipo de marcador AIM	No. De Marcadores	Poblaciones de referencia	Referencia
SNPs	34	Europeos, Nativoamericanos, Africanos, Asiáticos Orientales, Oceanía	Phillips, 2007 (Revisado en Fondevila, 2013)
SNPs	128	Europeos, Nativoamericanos, Africanos occidentales, Asiáticos Orientales	Kosoy, 2011
SNPs	446	Europeos, Nativoamericanos, Africanos	Galanter, 2012
INDELs	46	Europeo, Nativoamericano, Africano y Asiático oriental	Pereira, 2012
SNPs	23	Europeos, Medio orientales, Asia Sur Central, Asiático orientales	Phillips, 2013a
STRs	12	Europeos, Africanos, Asiáticos orientales, Nativo Americanos, y Oceanía	Phillips, 2013b
SNPs	128	Europeos, Nativoamericanos, Africanos, Asiático orientales, y Oceanía	Phillips, 2014
SNPs	55	Europeos, Nativoamericanos, Africanos, Asiático y Oceanía	Kidd, 2014
SNPs	31	Europeos, Nativoamericanos, Africanos, Asiático orientales, y Oceanía	De la Puente, 2016

Diferentes estudios muestran que también hay heterogeneidad (Eyheramendy et.al., 2015) y una posible subestructura dentro de las poblaciones nativoamericanas, por ejemplo en un grupo que incluye poblaciones norte, centro y suramericanas y otro de solo norteamericanas, o dividiendo las poblaciones andinas y las amazónicas o por clusters que se alinean parcialmente con la filiación lingüística, entre otras (Bryc et.al., 2010, Jhonson et.al., 2011, Reich et.al., 2012, Bolnick et.al., 2016).

Diversas investigaciones con AIMS en poblaciones mezcladas de Centro y Suramérica demuestran la variabilidad en la proporción de ancestría global estimada en distintos países. Brasil, Argentina, Uruguay y Colombia tienen menor ancestría Nativoamericana en comparación con Chile, México, Guatemala, Ecuador, Bolivia y Perú y países como República Dominicana, Jamaica y Haití tienen mayor proporción de ancestría afroamericana que otros del continente (Bryc et.al., 2010; Galanter et.al., 2012; Manta et.al., 2013; Ruiz-Linares et.al., 2014; Salzano & Sanz, 2014; Eyheramendy et.al., 2015). Sin embargo, estudios a nivel local muestran también que los componentes de la mezcla pueden variar aun en distintas regiones del mismo país. Por ejemplo en Colombia, el análisis de una muestra de la población de Antioquia indica que el componente mayoritario es el Europeo (Bedoya et.al., 2006; Price et.al., 2007), mientras la ancestría africana es mayor en las costa Pacífica y la nativoamericana predomina en la región suroccidental y amazónica (Ibarra et.al., 2014; Ossa et.al., 2016). De igual forma, al evaluar el componente nativoamericano en población indígena se encuentra un grado de variabilidad encontrando comunidades como los Nasa, Pastos, Awa, Coyaima, Bari/Motilon, Embera Chamí y Guainía con valores superiores a 0.8 y Zenú, por debajo de este valor (Criollo, 2012; Galanter et.al., 2012 y Ossa et.al., 2016).

La estructura de las poblaciones actuales es consecuencia de una dinámica poblacional repetitiva donde un grupo pequeño de individuos se separa de una población mayor creando un efecto de deriva genética cuya fuerza modeladora interactúa con la selección, la migración y los procesos posteriores de mezcla y la influencia de la introgresión arcaica (Phillips, 2015b). Este paisaje está influenciado tanto por las características geográficas que generan barreras entre los pueblos, como por las características culturales y dinámicas de mezcla dirigidas por acontecimientos históricos como los diferentes periodos de tiempo



en los que avanzó la colonización europea que comenzó en las zonas insulares del caribe donde arribó Cristóbal Colón y posteriormente las costas continentales y finalmente las regiones del interior (Salzano & Sanz, 2014), el sistema de castas para regular el matrimonio durante el siglo XVIII (Ruiz-Linares et.al., 2014), el aislamiento poscolonial, la densidad de la población nativa para la mezcla y de la cantidad y volumen de las olas migratorias posteriores entre otros (Bedoya et.al., 2006), lo cual demuestra la importancia de desarrollar investigaciones en marcos interdisciplinarios que permitan la integración de herramientas de análisis genético poblacional, genealógico, demográfico e histórico que permitan establecer modelos e hipótesis más precisas.

## **1.5. Perspectivas**

En el contexto científico y tecnológico actual, permanentemente se desarrollan nuevas metodologías que mejoran el alcance y eficacia de las herramientas disponibles. En el marco del proyecto genográfico con la colaboración de diferentes grupos de investigación, se desarrolló la estrategia de microarreglos conocida como “genochip” con la finalidad de estudiar las migraciones humanas y la expansión de la especie tras la salida de Africa para su aplicación en antropología genética. En él se incluye aproximadamente 1 SNP por cada 100 kilobases, en regiones no codificantes sin asociación funcional conocida (130000 autosómicos y de cromosoma X, 12000 de cromosoma Y y 3300 de DNA mitocondrial), correspondientes a marcadores de ancestría (Ancestral Informative Markers) (AIMs) seleccionados a partir del análisis de más de 450 poblaciones en 15 estudios ya publicados, mostró durante su validación su capacidad para diferenciar entre subpoblaciones usando datos globales. (Elhaik et.al., 2013). Actualmente, un método con aplicación biogeográfica desarrollado con base en análisis de mezclas, llamado “Geographic Population Structure” (GPS), emplea la información obtenida por la genotipificación de más de 100000 AIM's (Ancestral Informative Markers) autosómicos, que hacen parte del GenoChip (Elhaik et.al., 2014) a través del cual es posible inferir con mayor precisión, la región geográfica de origen de una muestra desconocida. El uso de estas herramientas con muestras de nativos americanos, puede generar una colección de datos que permitan complementar los análisis ya desarrollados.

De acuerdo con los resultados obtenidos del análisis de los diferentes tipos de marcadores, es posible plantear futuros análisis, por ejemplo de los linajes caracterizados por la mutación L54 que presenten el alelo 6 del sistema STR DYS391, encontrado en un individuo colombiano y uno panameño y reportado por otros autores en 5 panameños y dos nicaragüenses, por ser una variante que parece no haberse distribuido hacia el sur del continente. (Battaglia et.al., 2013).

Adicionalmente, pueden considerarse tanto el estudio de poblaciones nativas americanas actuales en la región noroccidental de Suramérica, empleando una amplia batería de marcadores tanto uniparentales como autosómicos, así como estudios comparativos entre los datos obtenidos del estudio de poblaciones indígenas actuales, poblaciones mezcladas y aparentes aislados genéticos con resultados de genotipificaciones de restos antiguos datados apropiadamente y en contextos, lo que podría ser útil en la evaluación de las diferentes hipótesis de poblamiento americano.

## 1.6. Conclusiones

Tanto los marcadores uniparentales como los autosómicos presentan evidencia a favor del origen de los primeros pobladores de América en grupos originarios de Asia, especialmente de la región centro sur de Siberia, en una o más migraciones; sin embargo solo Reich et.al., 2012 apoya la hipótesis de 3 migraciones, en forma concordante con la propuesta de Greenberg, hecha con base en la clasificación lingüística. Por su parte, los resultados presentados por algunos autores tanto con marcadores de DNAm<sub>t</sub> como de cromosoma Y apoyan la idea de un ingreso por Beringia, con un periodo estacional que permitió la diferenciación de los linajes y una posterior etapa de rápida dispersión de norte a sur del continente, considerando en algunos casos la antigüedad de las evidencias arqueológicas del sitio Monte Verde en Chile.

En cuanto al tiempo de entrada a América y dispersión los cálculos realizados con los diferentes tipos de marcadores para DNA mitocondrial, las estimaciones oscilan alrededor de 30.000 a 15.000 años para la entrada y 18.000 a 12.000 para la expansión en América. Para cromosoma Y origen del linaje caracterizado por la mutación QM242 calculado alrededor de 14000 años

Los resultados para marcadores autosómicos se presentan en términos de generaciones, por lo que la estimación en años es variable dependiendo del tiempo considerado por generación (que puede variar entre 20 y 30 años), que está sujeto a las condiciones culturales como respuesta a las condiciones de vida de cada población en determinado momento de la historia en un contexto geográfico, climático y de acceso a los recursos específico. Así, considerando un tiempo por generación de 20 a 25 años, el tiempo calculado de entrada a América sería de 11.120-13.900 años, la llegada a Suramérica hace 7800 -9850 años y la dispersión realizada en aproximadamente 460 a 575 años. El tiempo de divergencia entre población Chipewayan (Norteamérica) y Mixe (Centroamérica) sugiere una separación de linajes entre 13940 y 17425 años.

La discrepancia entre datos de cromosoma Y y de DNA mitocondrial, puede deberse a la gran reducción de población masculina a la llegada a América o la diferencia en la tasa de mutación de los 2 marcadores, lo cual altera el reloj molecular. (Bortolini et.al., 2003)

Adicionalmente, el análisis de estos marcadores soporta las hipótesis del retorno de poblaciones Eskimo-aleutianas a Siberia Norotiental y del origen de los nativos de baja Centroamérica a partir de una regresión de población por Panamá. Igualmente, marcadores mitocondriales apoyan la idea del poblamiento de las islas del Caribe como producto de una regresión de poblaciones que habitaban la región de la Orinoquía en Suramérica.

La evaluación de la diversidad genética de las poblaciones nativas americanas muestra con todos los marcadores una tendencia a disminuir de norte a sur y de occidente a oriente, lo que ha generado una especial atención en cuanto a las hipótesis de poblamiento a nivel de Suramérica, en relación con la región de los andes, las zonas costeras, amazónicas y en general fuera de la cordillera.

Al respecto, los datos con DNAm<sub>t</sub> muestran una mayor diversidad entre poblaciones andinas que no andinas y algunos proponen escenarios de colonización transandina con flujo genético bidireccional; mientras que los resultados obtenidos con análisis de marcadores de cromosoma Y aunque reportan mayor diversidad en poblaciones andinas que en las amazónicas, consideran poco probable el flujo genético entre ellas, lo que podría eventualmente relacionarse con un patrón de patrilocalidad.

Por otro lado, los marcadores autosómicos apoyan la hipótesis de un probable origen de los pueblos orientales a partir de los occidentales en Suramérica, pero el modelo propuesto no permite considerar flujo genético después de la separación de las poblaciones. Con base en las relaciones establecidas entre poblaciones de lengua Chibcha-Paez y Tucano Ecuatorial se sugiere un origen a partir de una migración separada de la de los andes y muestra similitud entre las poblaciones mesoamericanas y andinas.

Respecto a la relación entre datos genéticos y geográficos y lingüísticos, los estudios con marcadores de DNA mitocondrial muestran una baja correlación y con cromosoma Y: la variación genética actual para marcadores de cromosoma Y no permite establecer

relaciones claras con la dispersión geográfica y clasificación lingüística. Con marcadores autosómicos, aunque se encuentra un acuerdo con la distribución geográfica, esta correlación solo puede establecerse hasta cierto grado de distancia lingüística, y debe considerarse para profundizar en este tipo de análisis, la amplitud de los intervalos de confianza, los criterios estadístico con relación a la obtención de datos lingüísticos e incluso, la exactitud de la clasificación de las lenguas en las poblaciones nativas americanas (Roewer et.al., 2013).

En la actualidad existen a disposición de los investigadores distintas herramientas para el estudio de la diversidad genética de las poblaciones, que pueden ser apropiadamente aplicadas para el desarrollo de objetivos específicos en el campo forense ó en el área de la genética de poblaciones. Por un lado puede mostrarse su “imagen actual”, proporcionando el escenario para el análisis forense con poblaciones de referencia apropiada y suficientemente descritas, y por otro, hacer uso de marcadores de ancestría para trazar relaciones entre poblaciones e hipótesis de poblamiento. Cada tipo de marcador proporciona perspectivas diferentes que deben ser analizadas en el marco de la dinámica poblacional resultado de la historia de cada una de ellas y de la interacción cultural. En este contexto la relación entre la genética y el lenguaje ha resultado de gran interés, por considerarse que las formas de divergencia entre poblaciones con respecto a estos dos elementos deben ser concordantes al menos hasta cierto punto de la historia. En la actualidad, el mapa genético de América continua en construcción y detectar qué fenómenos poblacionales y demográficos explican mejor la actual diversidad genética y si estos soportan o no los patrones de diversidad lingüística y de distribución geográfica, exige el desarrollo de estrategias interdisciplinarias con diseños metodológicos pertinentes y apoyados en análisis estadísticos que generen modelos de poblamiento planteados con base en evidencias objetivas.

## **2. Capítulo 2. Comparación de un panel de SNPs autosómicos de ancestría frente a uno de identificación y los diferentes modelos de estructura poblacional**

### **2.1 Introducción**

El estudio de la estructura genética de las poblaciones humanas ha sido el soporte para el planteamiento de hipótesis en diferentes campos del conocimiento. Preguntas relacionadas con el origen del hombre, la dispersión de las poblaciones en los continentes y la posterior dinámica de poblamiento y de mezcla en los períodos subsiguientes a la colonización, se han analizado empleando marcadores moleculares de cromosoma Y, cromosoma X y ADN mitocondrial dependiendo de los objetivos propuestos. Actualmente, paneles de marcadores autosómicos recomendados para el estudio de poblaciones mezcladas, en especial los compuestos por Single Nucleotide Polymorphisms (SNPs), se utilizan en análisis de ancestría biogeográfica, identificación forense y evaluación de estratificación poblacional en estudios de asociación. El diseño de estos paneles se ha centrado en lograr metodologías eficientes con un número de marcadores viable en términos de costos y soporte tecnológico para los grupos de investigación, y con la informatividad suficiente para alcanzar los objetivos propuestos y del uso final para el que son elaborados.

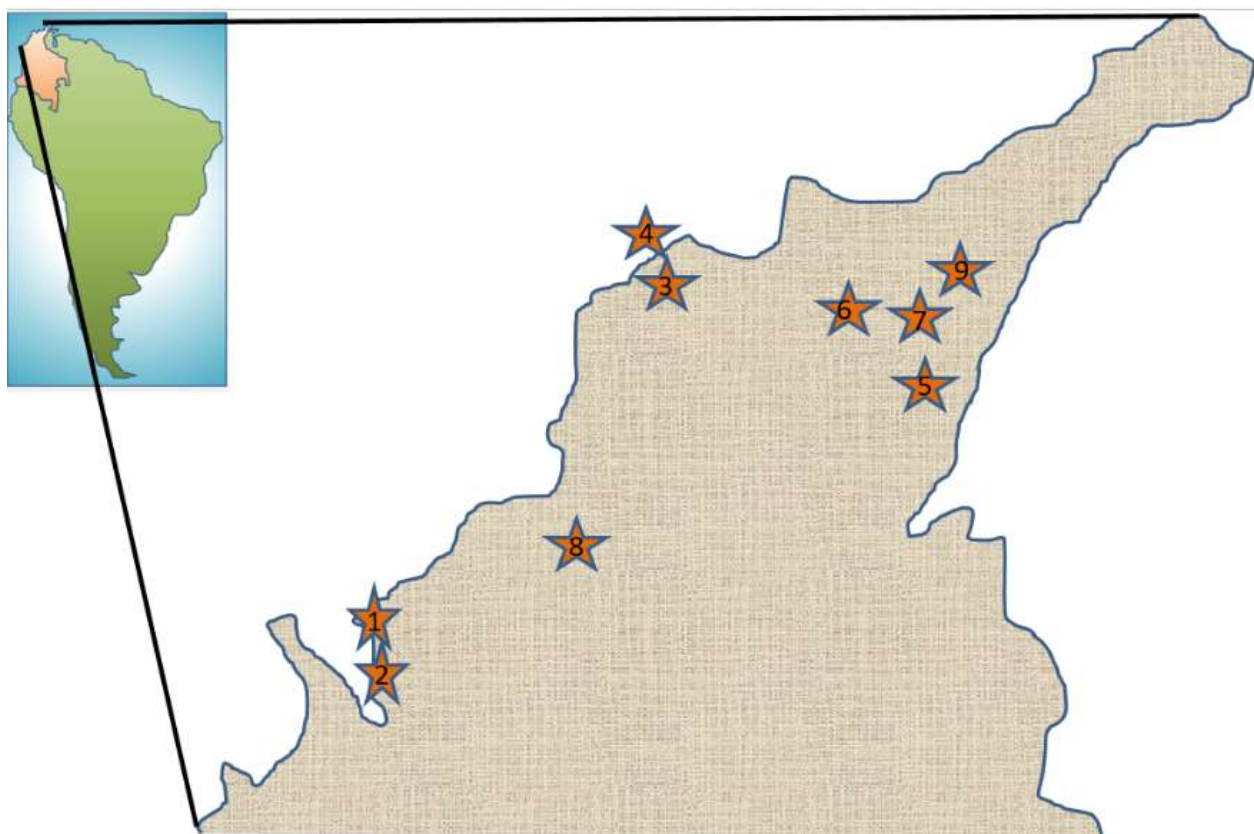
Varias publicaciones (Nassir et.al., 2009; Kossoy, 2009; Phillips et.al., 2007; Phillips, et.al., 2012; Phillips et.al., 2014; Kidd et.al., 2011; Kidd et.al, 2014; Pereira et.al, 2012; Galanter et.al., 2012) muestran paneles de marcadores informativos de ancestría, Ancestry

Informative Markers (AIMs), tipo SNPs o INDELS diseñados para diferentes niveles jerárquicos de la estructura de las poblaciones, desde lo continental a lo regional, haciendo uso de muestras de “referencia” provenientes de poblaciones no mezcladas. Con este mismo principio, grupos de marcadores tipo SNPs contruidos para aplicación en identificación humana como el set *SNPforID 52plex*, (Sanchez et.al., 2006), son en algunos casos empleados para realizar análisis de mezclas de poblaciones e inferencias de ancestría (Ibarra et.al., 2014; Moreno et.al., 2014; Ruiz et.al., 2012). La eficiencia de estos paneles para la asignación de ancestría y la construcción de sets con menor número de marcadores aplicables a diferentes tipos de tecnologías, capaces de cumplir eficazmente con este objetivo, han sido evaluadas para encontrar la mejor combinación de SNPs (Fondevila et.al, 2013; Taboada et.al., 2013; De la Puente et.al., 2016) que permitan obtener informatividades (In) equilibradas y estimar probabilidades de membresía más precisas, disminuyendo los errores y los sesgos de asignación en muestras de poblaciones con aportes ancestrales multiétnicos (Rosenberg et.al, 2003; Phillips, 2015b).

El trabajo presentado en esta publicación, incluye el análisis de poblaciones de diferentes orígenes étnicos de la región de la región Caribe Continental Colombiana, que han sido seleccionadas tras un análisis previo que incluyó datos históricos y demográficos. Se tipificó un subconjunto de 79 SNPs de ancestría (AIMs) del panel LACE (Galanter et.al., 2012) y el panel de identificación *SNPforID 52plex* (Sanchez et.al, 2006, Porras et.al., 2009), se compararon los modelos de estructura poblacional y el análisis de ancestría proporcionado por los dos tipos de datos. El análisis de resultados se enfocó en el marco de la historia de poblamiento de la región, en especial en lo relacionado con la ancestría nativoamericana, evaluando la eficacia de la estrategia de selección de la muestra a conveniencia y de los marcadores empleados para las genotipificaciones, así como el uso de herramientas de apoyo para la toma de datos en campo en el momento de la recolección de muestras, que faciliten el análisis interdisciplinario de tipo genealógico y demográfico.

## 2.2 Materiales y Métodos

**Figura 2-1.** Mapa de la región Caribe Continental de Colombia. Se indican los nombres y ubicaciones aproximadas de las poblaciones de muestreo: 1. Necoclí (12), 2. Turbo, (14) 3. Baranoa (28), 4. Tubará (16), 5. Becerril (24), 6. Pueblo Bello (33), 7. Valledupar (40), 8. Tuchín-San Andrés de Sotavento (54), 9. San Juan del Cesar (37). Entre paréntesis se indica el número de muestras tomadas en cada localidad.



### 2.2.1. Poblaciones y encuestas

La etapa de muestreo se llevó a cabo en 9 poblaciones de la Costa Caribe Colombiana, que fueron seleccionadas con base en la revisión de datos históricos sobre los asentamientos indígenas precolombinos y las actuales zonas de resguardo. De acuerdo con la información demográfica obtenida de los datos del último censo nacional de población realizado por el Departamento Nacional de Estadísticas de Colombia (DANE), la



mayoría de estas cabeceras municipales, presentan mayor tasa de autodeterminación indígena (**Figura 2-1**) y **Tabla suplementaria 2-1**. En el municipio de San Andrés de Sotavento se tomaron sólo 3 muestras, por lo que se analizaron junto con los datos de Tuchín, debido a la cercanía geográfica de los dos municipios y su origen ancestral común (Etnia Zenú). Las muestras de 194 mujeres y 64 hombres voluntarios mayores de edad residentes en la localidad durante 5 años o más, fueron tomadas por hisopado de células de mucosa oral o por venopunción (posteriormente depositadas sobre tarjetas FTA), previa firma del consentimiento informado y con aplicación de una encuesta para recolectar información genealógica y demográfica de los participantes, incluyendo lugares de nacimiento y de residencia tanto del participante como de sus padres, abuelos y bisabuelos tanto por línea materna como paterna, entre otros datos.

### **2.2.2. Extracción de ADN y genotipificación de SNPs**

La extracción de ADN se llevó a cabo empleando el PrepFiler Express™ Forensic DNA Extraction Kit empleando el AutoMate Express™ Forensic DNA Extraction System de Applied Biosystems, siguiendo las instrucciones del fabricante. Se utilizó un subset de 79 marcadores (**Tabla suplementaria 2-2A**) tipo SNPs de ancestría AIMs (Ancestry Informative Markers) incluidos tanto en el panel LACE de 446 como en el set de 128 AIMs publicados en Galanter et.al., 2012 y Marcheco et.al., 2014 respectivamente. Los marcadores de identificación correspondieron al 52 SNP-plex assay (SNPforID group) publicado por Sanchez, et.al. 2006. (**Tabla suplementaria 2-2B**). Los productos amplificados de ambos paneles fueron tratados con SAP y posteriormente sometidos a una reacción de extensión para finalmente ser genotipificados usando la plataforma Sequenom MassArray System para análisis de espectrometría de masas con la técnica MALDI-TOF. El análisis de resultados se realizó mediante el software “Typer Analyzer”.

### 2.2.3. Análisis genético poblacional y estadístico

Con el Software Arlequín 3.5.2.2 (Excoffier & Lischer, 2010) se estimaron las frecuencias alélicas y se realizó el análisis molecular de varianza (AMOVA), que consideró además de las poblaciones, un nivel jerárquico superior de grupos definidos de acuerdo con la división político-administrativa de Colombia en departamentos, donde el grupo uno reunió las poblaciones de Necoclí y Turbo, (Departamento Antioquia), el grupo dos corresponde a las poblaciones de Baranoa y Tubará (Departamento Atlántico), el grupo tres incluye Becerril, Pueblo Bello y Valledupar (Departamento Cesar) , el grupo cuatro corresponde a Tuchín (Departamento Córdoba) y el grupo cinco a San Juan del Cesar (Departamento La Guajira). Además se construyó la matriz de  $F_{st}$  pareados con los valores de  $p$  se estimados por 10000 permutaciones.

Para el Análisis de Componentes Principales (ACP) se usó el software MVSP 3.22 y el paquete SNPassoc (Gonzalez, J. et.al., 2007) en el ambiente R (<http://cran.r-project.org>). Con esta misma plataforma se realizó el Análisis Discriminante de Componentes Principales (DAPC) con el paquete Adegenet (Jombart, 2008 y Jombart & Collins, 2015). Tanto para el ACP como para el DAPC se incluyeron adicionalmente como poblaciones de referencia 716 muestras de población europea, 103 de población africana, 127 nativoamericanos (**Tabla suplementaria 2-3A**) procedentes de la base de datos del proyecto HAPMAP genotipificadas con el panel de ancestría Genome-Wide de LACE (Galanter et.al., 2012), aportados por el Centro Nacional de Genotipado (CEGEN) en la Fundación Pública Gallega de Medicina Genómica. DAPC es una técnica multivariante que combina PCA y DA (discriminant analysis) para maximizar la diferenciación entre grupos que pueden ser predefinidos por el usuario o escogidos en un paso previo realizando un k-means clustering (opción find clusters). Para el análisis del panel SNPforID 52plex, los genotipos de las poblaciones de referencia empleadas se obtuvieron de las bases de datos 1000 genomas fase I y HapMam, desde SNP forID Browser en la herramienta SPSmart v5.1.1 (<http://spsmart.cesga.es/snpforid.php>) incluyendo 396 Europeos, 340 Africanos y 219 nativoamericanos (**Tabla suplementaria 2-3B**). Adicionalmente también se incluyeron 33 genotipos colombianos de la etnia Pastos, publicados en Ibarra et.al., 2014. Estas mismas poblaciones de referencia se emplearon con el software STRUCTURE y con la herramienta SNIPPER.

El software STRUCTURE 2.3.4 (Pritchard et.al., 2000) se empleó primero para la evaluación del  $K$  óptimo usando solo las poblaciones de referencia (NO POPFLAG) con un periodo de burn-in de 50000 y 100000 repeticiones MCMC, evaluando 6 iteraciones desde  $K=3$  hasta  $K=6$ . El valor óptimo de  $K$  se estimó con Structure Harvester (Earl, D & vonHoldt, B., 2012). El análisis de ancestría se realizó con  $K=3$ , bajo el modelo de mezcla con frecuencias de alelos correlacionadas, con un periodo de burn-in de 100000 y 100000 repeticiones MCMC.

Con la herramienta en línea SNIPPER 2.5 App suite (<http://mathgene.usc.es/snipper/>) se realizó la validación cruzada del set de poblaciones de referencia y a partir de los valores de divergencia se estimó el valor de informatividad de asignación de Rosenberg (Rosenberg et.al., 2003) para cada marcador (In) multiplicando por 0.693 (Phillips, 2015b). También se calcularon mediante un análisis bayesiano los valores likelihood ratio (LR) para la clasificación poblacional de individuos a partir de la genotipificación con marcadores AIMs y SNPforID 52plex usando el set de poblaciones de referencia descrito arriba (Phillips, 2015b).

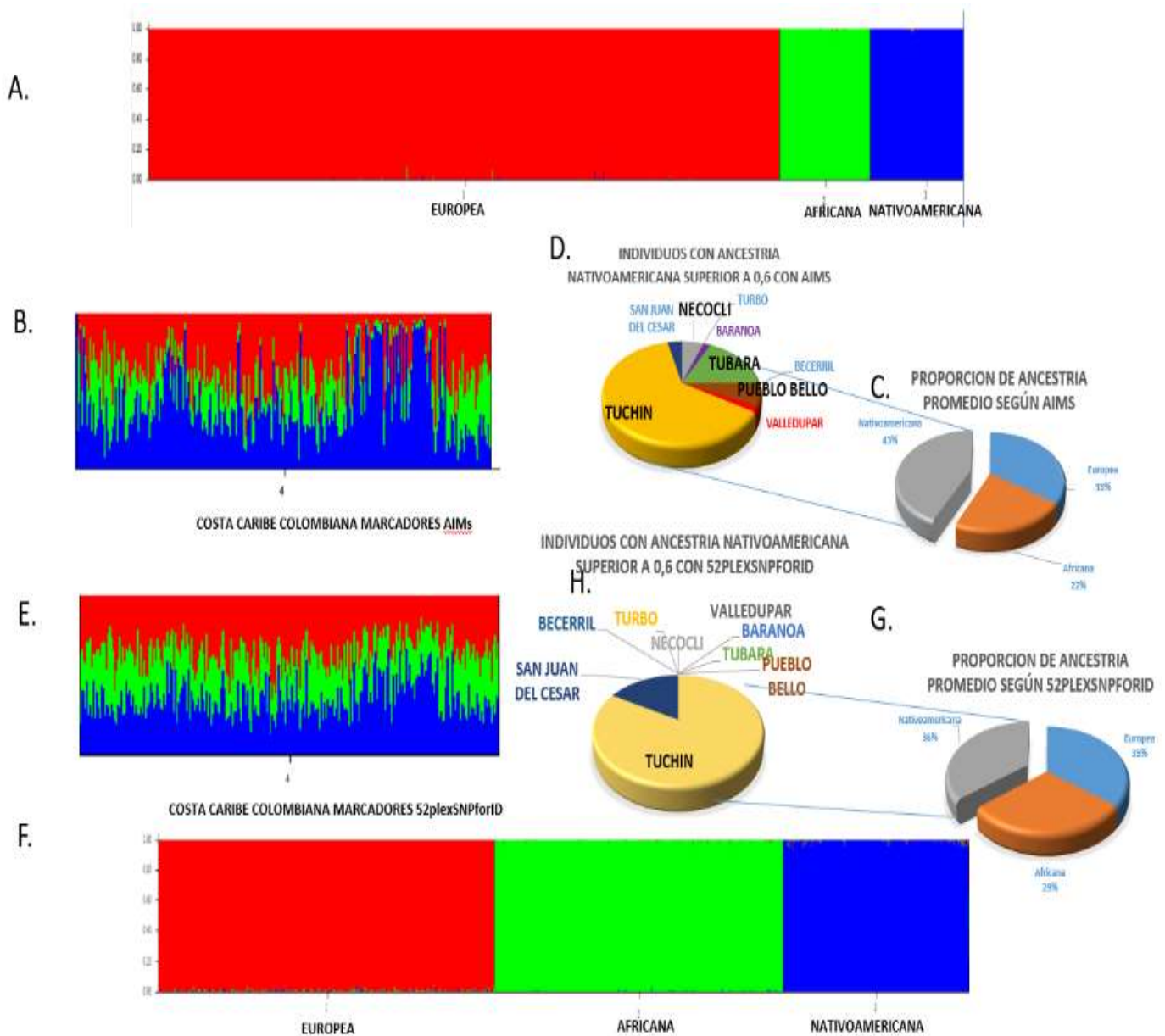
## 2.3 Resultados y Discusión

Tanto para los 79 marcadores tipo AIMs y para los del panel *SNPforID 52plex* se estimaron las frecuencias genotípicas (**Tabla suplementaria 2-4**) (se destaca que en la población de Tubará se encontró solo el alelo G para el marcador (AIM) rs7927234) y se evaluó la estructura poblacional mediante análisis molecular de varianza (AMOVA). Los dos tipos de marcadores muestran que el mayor porcentaje de variación se da entre los individuos dentro de las poblaciones y que existen diferencias significativas entre las poblaciones dentro de los departamentos para los marcadores de ancestría (FSC= 0,01518 con p-valor= 0,0002±0,00001) y entre los departamentos para los marcadores de identificación (FCT= 0,01597 con p-valor= 0,00623±0,00024). Estos resultados se corroboran con los valores de p estimados para la comparación por pares de poblaciones de los Fst con AIMs, donde se observa que las muestras de Tubará, Pueblo Bello y Tuchín presentan diferencias significativas con todas las demás poblaciones, mientras con el panel *SNPforID 52plex* para estas mismas localidades, solo se observan diferencias con 5, 3 y 7 poblaciones respectivamente. Se destaca que no hay un valor significativo de p entre la población de Tuchín y la de Tubará. (**Tabla suplementaria 2-5**). Este análisis permite evidenciar que existen diferencias entre las poblaciones, por lo que para ambos tipos de marcadores se acepta la hipótesis de subestructura.

### 2.3.1 Análisis de estructura genética e inferencia de ancestría

Salzano y Sans (2014), reúnen datos de varias publicaciones sobre la ancestría de distintas poblaciones en Latinoamérica. En este trabajo se resalta que países como México, Guatemala, Perú y Ecuador muestran una importante ancestría nativoamericana y países como Cuba, Puerto Rico, Nicaragua, Costa Rica, Argentina y Uruguay presentan mayores proporciones de ancestría Europea, mientras en Colombia se detecta una gran variabilidad. Para comparar los resultados del análisis de la estructura genética y la inferencia de ancestría empleando tanto los marcadores AIMs como el panel *SNPforID 52plex*, se utilizó el software STRUCTURE usando  $K=3$  para estudiar las muestras de la región Caribe Continental Colombiana.

**Figura 2-2.** Análisis de STRUCTURE para la población mezclada de la Región Caribe Continental Colombiana. A, B, C y D hacen referencia a AIMs. Bar Plot a partir del análisis de las poblaciones de referencia (A) y de las muestras de la región Caribe de Colombia (B), Proporción de contribución ancestral promedio (C.) y número de individuos con ancestría nativoamericana igual o superior a 0.6 por cada subpoblación de la región Caribe Colombiana (D.). E, F, G y H corresponden a SNPforID 52plex. Bar Plot a partir del análisis de las poblaciones de referencia (E) y de las muestras de la región Caribe de Colombia (F), proporción de contribución ancestral promedio (G.) y número de individuos con ancestría nativoamericana igual o superior a 0.6 por cada subpoblación de la región Caribe Colombiana (H.)

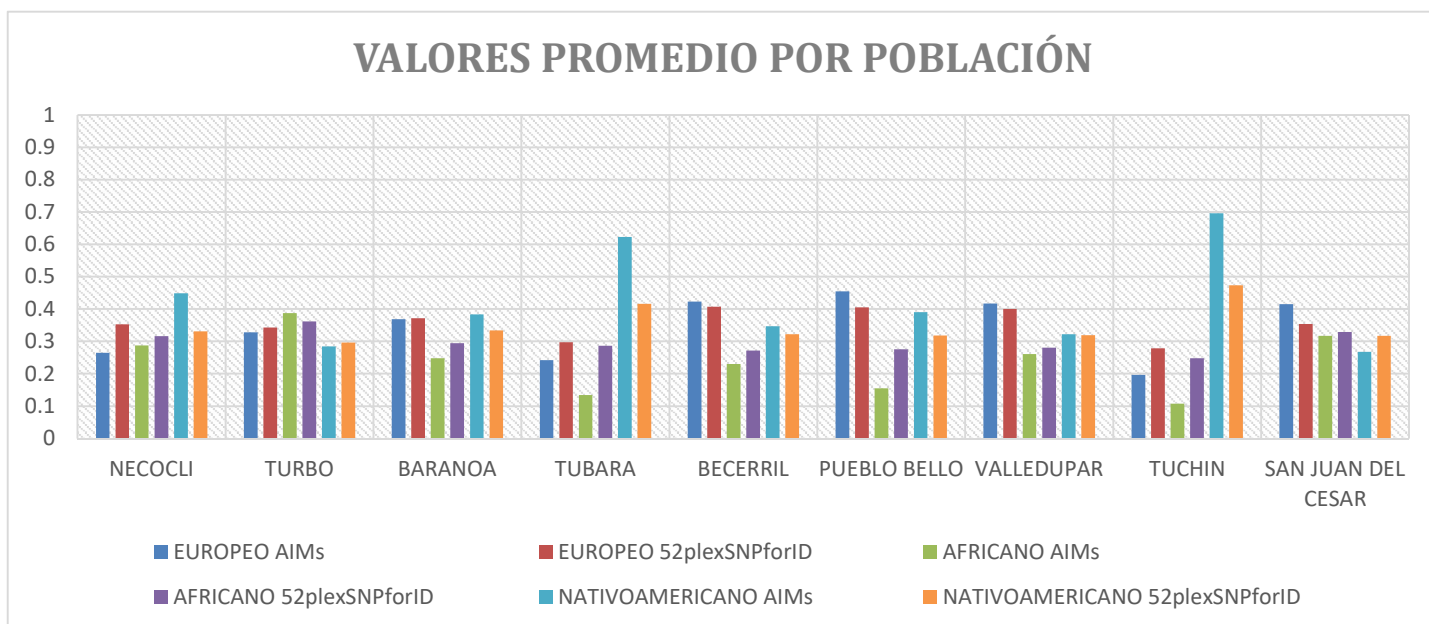


La **Figura 2-2** muestra la estructura poblacional en la que se diferencian los clusters correspondientes a las poblaciones de referencia (europeos, africanos y nativoamericanos) tanto con marcadores de ancestría, como de identificación. La proporción de membresía a cada clúster de las muestras de la región Caribe Continental Colombiana con panel AIMs fue de 0.348 a la población europea, 0.218 a la población africana y 0.434 al cluster Nativoamericano, y con el set SNPforID 52plex, de 0.356 a la población europea, 0.287 a la población africana y 0.358 al cluster Nativoamericano. Para la región Caribe Continental Colombiana, un 67% del total de individuos analizados con AIMs y un 62% de los individuos analizados con SNPforID 52plex, mostraron una proporción de ancestría nativoamericana superior a 0.3. Este porcentaje de población baja al 22% con AIMs y apenas al 5% con SNPforID 52plex, si se considera un umbral de ancestría nativoamericana más alto ( $>0.6$ ). La distribución de estos dos últimos grupos se observa en las figuras 2D y 2H.

Los resultados de Structure con los marcadores AIMs en este estudio, muestran que todas las poblaciones tienen 3 componentes ancestrales, siendo el de mayor proporción promedio el Nativoamericano para las poblaciones de Necoclí, Tubará, y Tuchín. La proporción de ancestría africana promedio fue más alta en Turbo y los valores más bajos se observan en Tuchín, Tubará y Pueblo Bello. (**Figura 2-3 y tabla suplementaria 2-6**). El promedio de proporción de ancestría nativoamericana total es de 0.434, con valores que varían entre 0.27 y 0.70, estimaciones más altas comparadas con las obtenidas en otras investigaciones realizadas en otras poblaciones mezcladas en Colombia, que han presentado resultados promedio entre 0.22 y 0.39 (**Tabla Suplementaria 2-7**), pero similares a los encontrados para poblaciones mexicanas, chilenas o bolivianas (0.36-0.86) (Wang et.al. 2008, Silva-Zolezzi et.al., 2009, Heinz et.al., 2013).

Tanto en la estimación de la proporción promedio de cada contribución ancestral, como en la determinación del componente mayoritario por población (**Tabla suplementaria 2-6 y Tabla 2-1**), al emplear el set de marcadores SNPforID 52plex se sobreestima del componente africano en Necoclí, Baranoa, Tubará, Pueblo Bello y Tuchín y se subestima el componente nativoamericano en Necoclí, Turbo, Tubará y Tuchín, con respecto a los resultados obtenidos con AIMs.

**Figura 2-3.** Valores promedio de ancestría por población estimados por STRUCTURE para los marcadores AIMs y SNPforID 52plex para la población mezclada de la Costa Caribe Colombiana



**Tabla 2-1.** Análisis de muestras por población. Porcentaje de individuos con componente de ancestría mayoritario Europeo, Africano y Nativoamericano para el panel AIMs y 52plexSNPforID

POBLACION	Componente Mayoritario Europeo		Componente Mayoritario Africano		Componente Mayoritario Nativoamericano	
	AIMs	52plex	AIMs	52plex	AIMs	52plex
<b>NECOCLI</b>	25%	25%	25%	42%	50%	33%
<b>TURBO</b>	28.6%	31%	42.86%	54%	28.5%	15%
<b>BARANOA</b>	50%	47%	7.14%	21%	42.86%	32%
<b>TUBARA</b>	0%	19%	0%	12%	100%	69%
<b>BECERRIL</b>	58.33%	62%	20.83%	17%	20.83%	21%
<b>PUEBLO BELLO</b>	63.64%	52%	6.1%	15%	30.3%	33%
<b>VALLEDUPAR</b>	60%	57%	20%	23%	20%	20%
<b>TUCHIN</b>	18.52%	17%	0%	6%	81.48%	77%
<b>SAN JUAN DEL CESAR</b>	59.46%	46%	27.03%	32%	13.51%	22%

En una observación más detallada de la ancestría en la población del Caribe Continental Colombiano en este estudio, según las encuestas aplicadas los habitantes de la localidad

de Tuchín, son en su mayoría personas que se auto reconocen como pertenecientes a la etnia indígena Zenú. De acuerdo con los resultados de este trabajo, la muestra correspondiente a esta población presenta un patrón de mezcla principalmente nativoamericano/europeo con una contribución africana menor. La probabilidad estimada de ancestría nativoamericana promedio con AIMS fue de 0.7 (con valores individuales oscilando entre 0.155 y 0.966 siendo el 50% superior a 0.8), mientras Criollo (2012) reportó un valor de 0.48 para esta misma etnia, lo cual se explicaría porque las zonas muestreadas por este autor corresponden a los municipios de Montelíbano y Sahagun, que según datos del Censo Nacional de Población 2005, tienen un bajo porcentaje personas que se autorreconocen como indígenas (3% y 4.5% respectivamente) y tienen una historia de mezcla reciente alta (Criollo, 2012).

En comparación otras poblaciones, Galanter et.al., 2012 muestra valores de ancestría nativoamericana para grupos indígenas entre 0.8 y 1, con excepción de la etnia Wichi de Argentina que presenta una mayor proporción de mezcla europea. En estudios nacionales, Criollo 2012 también estimó este valor en poblaciones nativas colombianas (Nasa, Pijao, Embera) entre 0.84 y 0.96, e Ibarra et.al., 2014 con *SNPforId* 52plex encontró diferencias significativas de los nativoamericanos colombianos Awa, Coyaima, Emberá y Pijao con todas las demás poblaciones estudiadas por *Fst* pareados. Esta última autora estima una proporción de membresía nativoamericana de 0.923 para la etnia Pastos, e indica que en las poblaciones mezcladas de la región andina colombiana (incluyendo el estudio de Porras et.al., 2009), prevalecen los componentes europeo y nativoamericano con valores promedio entre 0.4-0.47 y 0.34-0.4 respectivamente, resultados que son similares a los encontrados en este estudio para población de la región Caribe Continental Colombiana (componentes europeo con valores promedio entre 0.28-0.41 y nativoamericano entre 0.3-0.47). Respecto a las poblaciones afrodescendientes, en Ibarra 2014 la población mezclada del Chocó y población de Mulaló (Porras et.al., 2009) la proporción de ancestría africana es mayor (valores promedio entre 0.47-0.54). Para las muestras de este estudio, la población con mayor ancestría africana promedio fue Turbo (0.37), una localidad que tiene un alto flujo migratorio desde el departamento del Chocó y del centro y sur de Antioquia que favorece la mezcla interétnica, reflejada en la diferencia de estimaciones entre estos dos trabajos.



### 2.3.2 Análisis de Ancestría individual

Al realizar una comparación para cada uno de los individuos, se observa que en el 41% de los casos (15% nativoamericano, 19% europeo y 7% africano) independientemente del valor, el componente más alto de la mezcla indicado por el análisis con los marcadores AIMs, no es concordante con el indicado por los marcadores del SNPforId 52plex. En 87 de los 258 casos, usando los marcadores de identificación se estimó entre un 31% y un 92% del valor del componente ancestral nativoamericano indicado por el análisis con AIMs y en 46 casos se sobreestimó el componente ancestral africano entre el 101% y el 622%.

Para los siguientes análisis se consideró 0.6 como el valor límite inferior para la presencia predominante de un componente en la mezcla, por acercarse a los 2/3 bajo la apreciación histórica de mezcla triétnica de la población colombiana, y 0.1 la cifra mínima requerida para interpretar la presencia de aporte a la mezcla, siguiendo con las recomendaciones de Phillips (2015b) para los análisis de ancestría biogeográfica.

Adaptando los conceptos de la epidemiología clínica, se emplearon los criterios para establecer la validez analítica de las interpretaciones realizadas con el set SNPforId 52plex (test alternativo), considerando el panel AIMs como test de referencia. Se compararon las interpretaciones sobre el aporte ancestral mayoritario sin importar el valor numérico indicado por el análisis de STRUCTURE para cada componente Vs los demás (Europeo Vs No Europeo, Africano Vs No Africano y Nativoamericano Vs No Nativoamericano) y también comparaciones pareadas. En el primer análisis se observa que el uso del set SNPforId 52plex muestra mayor especificidad que sensibilidad, la proporción de falsos negativos es mayor que la de falsos positivos y la prevalencia (frecuencia) aparente es considerablemente menor que la prevalencia (frecuencia) real. Así por ejemplo, el valor predictivo positivo del test alternativo, entendido como la probabilidad a posteriori que un individuo tenga el componente ancestral mayoritario Africano dado que el test con SNPforID 52plex lo ha indicado así, es de 0.35 y el valor predictivo negativo, entendido como la probabilidad a posteriori de que en un individuo el componente ancestral mayoritario no sea Africano, dado que el test alternativo así lo indica, es 0.91, lo que quiere decir que el test alternativo es más eficiente para detectar individuos cuyo componentes ancestral mayoritario no es africano que para detectar individuos que si tienen este componente mayoritario. En las comparaciones pareadas se obtienen mejores valores predictivos

positivos y disminuyen los valores predictivos negativos, a excepción de la comparación entre ancestría mayoritaria Africana Vs ancestría mayoritaria nativoamericana donde sucede lo contrario. (Tabla 2-2).

**Tabla 2-2.** Evaluación analítica de la determinación del componente ancestral mayoritario en población mezclada de la Región Caribe Continental Colombiana, usando como método de referencia el uso de marcadores AIMs y como test alternativo el uso de los marcadores SNPforID 52plex. (Eur= Europeo, Afr= Africano, Nam=Nativoamericano).

INDICADOR	EVALUACION GENERAL			EVALUACION PAREADA		
	Eur Vs No Eur	Afr Vs No Afr	Nam Vs No Nam	Eur Vs Afr	Eur Vs Nam	Afr Vs Nam
<b>Sensibilidad</b>	0.563	0.500	0.651	0.716	0.724	0.750
<b>Especificidad</b>	0.718	0.844	0.797	0.600	0.711	0.885
<b>Cociente Falso Positivo</b>	0.282	0.156	0.203	0.400	0.289	0.115
<b>Cociente Falso Negativo</b>	0.438	0.500	0.349	0.284	0.276	0.250
<b>VPP</b>	0.612	0.346	0.697	0.840	0.692	0.667
<b>VPN</b>	0.675	0.911	0.761	0.419	0.742	0.920
<b>Prevalencia real</b>	0.441	0.142	0.417	0.746	0.473	0.235
<b>Prevalencia aparente</b>	0.248	0.071	0.272	0.534	0.342	0.176

Posteriormente este análisis se repitió en dos subgrupos de individuos: seleccionando solamente individuos con valores iguales o superiores a 0.6 del componente ancestral mayoritario usando el panel AIMs, sin límite del valor obtenido con el test alternativo y seleccionando individuos con componente ancestral mayoritario nativoamericano con valor superior a 0.8 según el test de referencia. En general se observa que al establecer un límite ascendente, las diferencias entre la prevalencia aparente y la prevalencia real disminuyen considerablemente y se obtienen mejores estimados de sensibilidad y especificidad, con valores más reducidos en la proporción de falsos positivos y falsos negativos. También se observa que el valor predictivo positivo tiende a 1 en las comparaciones generales, lo que significa que la probabilidad de que el componente ancestral mayoritario detectado con el panel de identificación sea el mismo detectado con el panel de ancestría, dado resultado del test alternativo, es más eficiente en poblaciones con menor grado de mezcla. Sin embargo, con relación específicamente al valor predictivo negativo en el análisis general y en el valor predictivo positivo de ancestrías pareadas para el componente nativoamericano, no se observan mejores estimas, lo que sugiere como en otros aspectos de esta investigación, que las interpretaciones realizadas con los resultados del análisis de STRUCTURE del

panel SNPforID 52plex en poblaciones mezcladas, deben ser cuidadosamente analizadas. Estos resultados se detallan en las **Tablas suplementarias 2-8 y 2-9**.

### **2.3.3 Análisis de Componentes Principales y Análisis Discriminante de Componentes Principales**

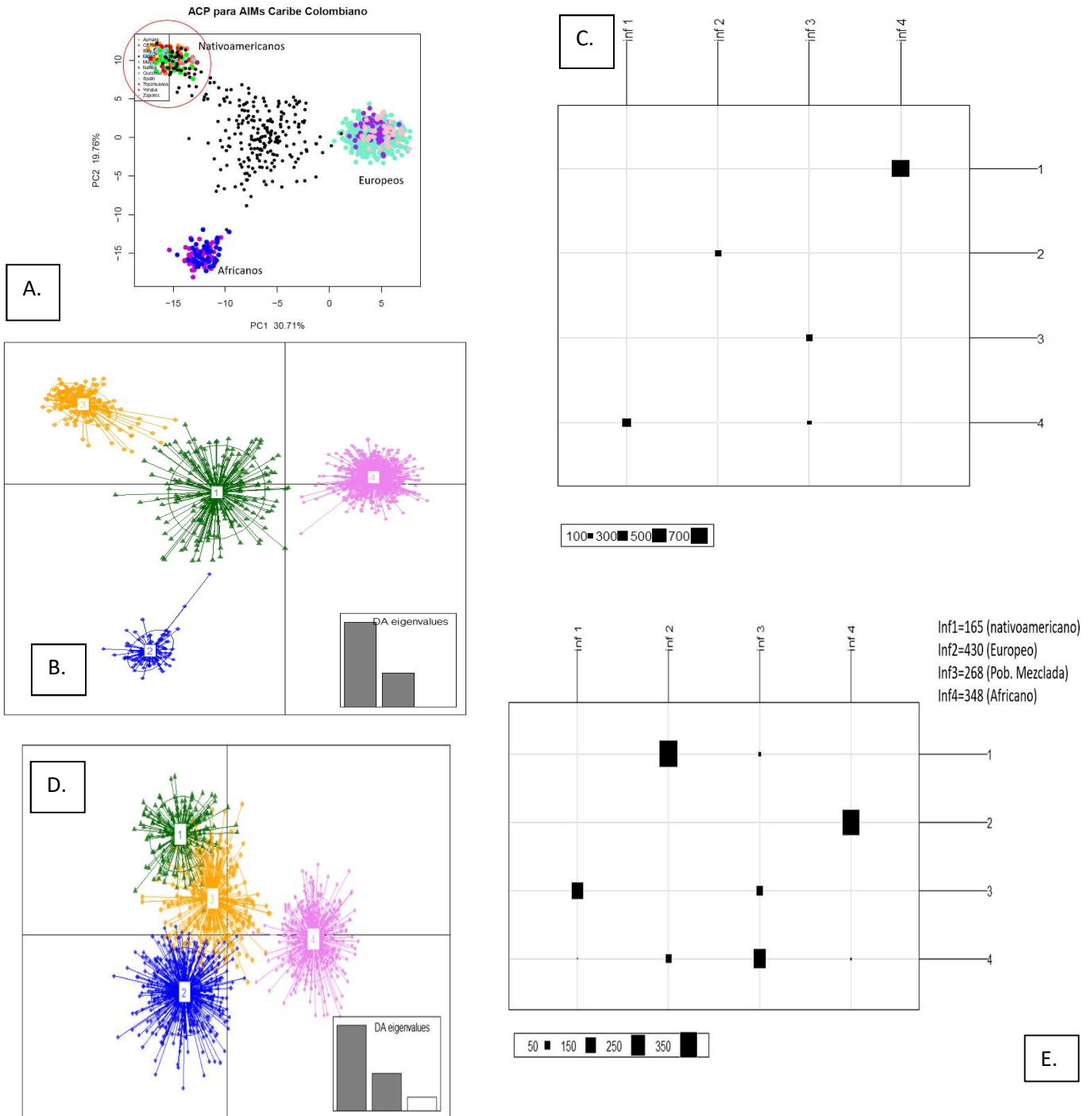
Para evaluar la diferenciación entre las poblaciones y la contribución de los marcadores en la formación de los agrupamientos, se emplearon estrategias de análisis multivariado. En el análisis de componentes principales realizado con los marcadores AIMS, los dos primeros componentes explican el 50,47% de la variación observada y en conjunto separan las muestras de referencia de la población Europea de las de Natioamericanos y Africanos mientras el segundo componente principal muestra la divergencia entre estos dos últimos grupos. Las muestras de estudio, de acuerdo con su origen triétnico se distribuyen en su mayoría de forma relativamente equidistante a las de referencia, aunque se observa un set de puntos que se separa con base en los dos componentes (en su mayoría muestras provenientes de la población de Tuchín), solapándose con las muestras de nativoamericanos (**Figura 2-4A**). Al analizar solamente las muestras colombianas usando el software MVSP, el plot obtenido corrobora los resultados anteriores (**Figura Suplementaria 2-1**), donde se observa un conjunto de puntos que se separa del resto con relación al segundo componente principal. La gráfica biplot de MVSP permitió identificar los 8 marcadores (rs132663, rs4812381, rs485645, rs2736340, rs17157450, rs11743823, rs2187239 y rs11960137) que están influyendo en la constitución de este grupo, que por sus frecuencias, se encuentran entre los que tienen mayor informatividad para ancestría nativoamericana dentro del panel LACE (Galanter et.sl., 2012) (**Figura Suplementaria 2-2**). Sin embargo, en el plot del ACP realizado únicamente con las muestras de estudio para SNPforID 52plex, no se observan agrupaciones de individuos generadas por la influencia específica de una variable, por lo que se considera que no hay individuos sustancialmente diferentes. (**Figura suplementaria 2-3**).

El análisis discriminante de componentes principales DAPC, que se llevó a cabo inicialmente usando poblaciones Europeas, Africanas, Natioamericanas y las poblaciones mezcladas del Caribe Continental Colombiano (**Figura suplementaria 2-4**), mostró

resultados similares a los del ACP con los marcadores AIMs (**Figura 2-4A**), donde un conjunto de las muestras de población colombiana se solapan con las muestras de nativoamericanos.

De acuerdo con los resultados del análisis, los SNPs que contribuyen de forma más significativa en este modelo son rs1809119 y rs862500. La mayor variación genética entre grupos fue determinada por las dos primeras funciones discriminantes con ambos tipos de marcadores, sin embargo con los marcadores de identificación se observa un aparente menor grado de divergencia entre las poblaciones. Al emplear el algoritmo K-means, DAPC identificó 4 grupos (clusters inferidos: Inf1, Inf2, Inf3, Inf 4) útiles para describir los datos (**Figura 2-4B y 2-4D**) tanto con marcadores AIMs como con SNPforID 52plex. Con los marcadores de ancestría la comparación de estos resultados con los clusters predefinidos agrupando poblaciones de ancestría común (1. Europeos, 2. Africanos, 3. Nativoamericanos 4. Colombianos (este estudio)) se observa en la **Figura 2-4C**, que muestra la membresía de los individuos agrupados en cada uno de los clusters originales con respecto a los clusters inferidos. De esta manera, se observa concordancia total en la asignación de las muestras de referencia y con respecto a las 258 muestras totales de este estudio (si bien no se aprecia con detalle en la figura 4C), 200 fueron asignadas como miembros del cluster inf1 (Mezclados), 1 muestra al cluster Inf2 (Africa), 52 muestras al cluster Inf3 (Nativoamericano) y 5 muestras al cluster Inf4 (Europa). En los dos tipos de análisis las dos primeras funciones discriminantes reúnen la mayor cantidad de variación entre los grupos. El mismo análisis realizado con los marcadores de SNPforID 52plex muestra diferencias en la asignación de membresía en las poblaciones de referencia (**Figura2-4E**), lo que sugiere un posible desbalance en la informatividad de los SNPs seleccionados de acuerdo con las frecuencias alélicas en las poblaciones de referencia, por lo que el grado de diferenciación de que se logra con SNPforID 52plex, es menor que el que se alcanza con el panel AIMs.

**Figura 2-4. A.** Análisis de componentes principales para los marcadores AIMs. con las poblaciones de referencia en plataforma R **B** . Gráfica de DAPC con marcadores AIMs mostrando los clusters inferidos: Inf1 (Población mezclada), Inf2 (Africanos), Inf3 (Nativoamericano) e Inf4 (Europeos). **C.** Membresía de los individuos de los grupos originales (1. Europeos, 2. Africanos, 3. Nativoamericanos y 4. Colombianos costa Caribe) a los clusters inferidos con AIMs. **D.** Gráfica de DAPC con marcadores SNPforID 52plex mostrando los clusters inferidos: Inf1 (Nativoamericano), Inf2 (Europeo), Inf3 (Población mezclada) e Inf4 (Africanos). **E.** Membresía de los individuos de los grupos originales (1. Europeos, 2. Africanos, 3. Nativoamericanos y 4. Colombianos costa Caribe) a los clusters inferidos con SNPforID 52plex

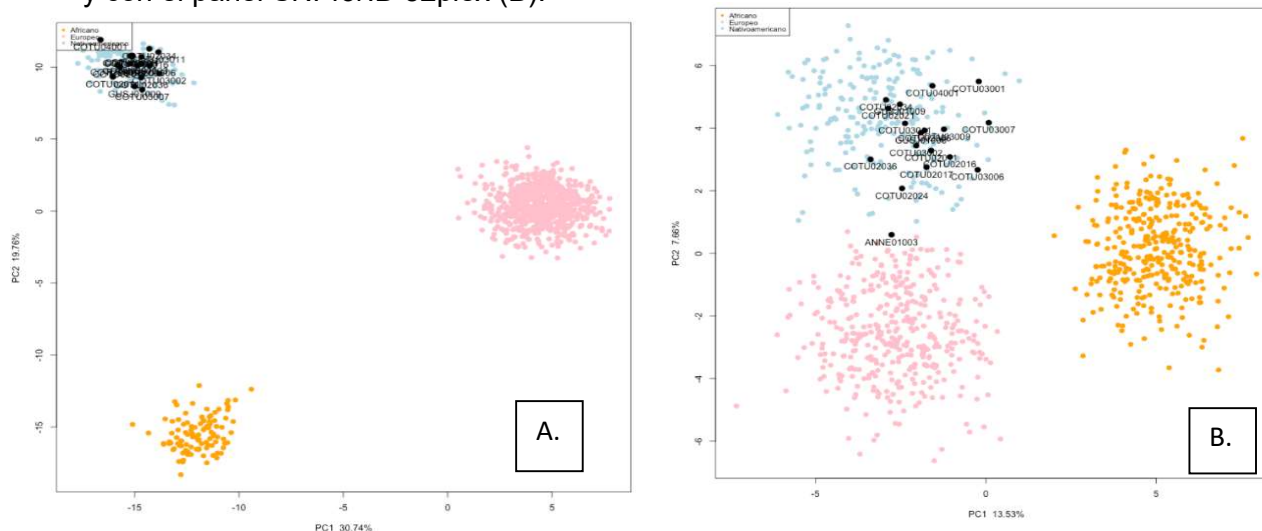


### 2.3.4 Evaluación de Informatividad de marcadores y análisis de clasificación Bayesiano

Teniendo en cuenta los resultados de la comparación de la inferencia de ancestría con el software Structure y los obtenidos con la función “find clustrers” con  $K=4$  con DAPC empleando los marcadores AIMS y los de identificación, se utilizó la herramienta en línea Snipper para hacer la validación cruzada de los datos de referencia del SNPforID 52plex, y evaluar la informatividad de los estos marcadores para las muestras del caribe colombiano. Los resultados que se observan en la **Tabla suplementaria 2-10** muestran un valor de In acumulado de 2.6132 para población Europea, 4.0218 para población Africana y 3.3829 para población Nativoamericana. Aunque los 3 valores no son notablemente diferentes, los SNPs que tienen mayor informatividad para población europea en su mayoría no tienen valores de divergencia superiores a 0.1. Si se evalúan las divergencias pareadas se observa que es menor la informatividad de los marcadores para el análisis de las poblaciones europea y nativoamericana (3.4355) con respecto a los valores para la población Africana Vs Europea (5.6117) y Africana Vs Nativoamericana (5.6117), y en la mayoría de los marcadores que son informativos para para poblaciones europeas se observa que los valores de divergencia son inferiores a 0.1. Estos datos se soportan en que varios de los marcadores incluidos en el panel SNPforID 52plex, no muestran diferencias sustanciales en las frecuencias alélicas entre las poblaciones de referencia (**Figura suplementaria 2-5**).

Para comparar los resultados de la asignación con los dos sets de marcadores, con esta misma herramienta mediante un Algoritmo Bayesiano de Clasificación, se determinaron los valores  $-\text{Log}(\text{Likelihood})$  y los Likelihoods ratios para cada muestra con el fin de clasificar cada uno de los perfiles genéticos con marcadores AIMS con valor de ancestría nativoamericana superior a 0.9 según el resultado de STRUCTURE **Tabla 2-3**. Los resultados obtenidos muestran que valores  $-\text{Log}(\text{Likelihood})$  son mejores (más pequeños) para los resultados con los AIMS y considerando un umbral límite de 100 en el valor Likelihood Ratio como se recomienda en Phillips et.al., 2013a y Bulbul et.al., 2016, los Likelihoods son mucho menores con los marcadores de SNPforID 52plex, e incluso se observa la clasificación errónea de una muestra ( $\text{LR} = 9.38$ ). Los mismos resultados pueden verse en los plots de los ACP para cada tipo de panel. **Figura 2-5**.

**Figura 2-5.** Análisis de componentes principales con la herramienta en línea SNIPPER, usando las muestras de la región Caribe Continental Colombina con aporte de ancestría nativoamericana superior a 0.9 según el resultado de STRUCTURE. Con el panel AIMs (A) y con el panel SNPforID 52plex (B).



**Tabla 2-3.** Resultados de la asignación hecha por la herramienta en línea Snipper (con marcadores de ancestría y de identificación) de las muestras de la región Caribe de Colombia con aporte de ancestría nativoamericana igual o superior a 0.9 (con marcadores AIMs) con el software Structure.

MUESTRA	AIMs			SNPforID 52plex		
	-Log (likelihood)	Likelihood ratio	Ancestría más probable	-Log (likelihood)	Likelihood ratio	Ancestría más probable
ANNE01003	44.7732	1.0 E09	Nativoamericano	50.2267	<b>9.38</b>	<b>Europeo</b>
COTU03001	33.9003	1.0 E09	Nativoamericano	42.1275	429443181	Nativoamericano
COTU03002	39.7637	1.0 E09	Nativoamericano	42.5512	92614	Nativoamericano
COTU03006	45.1366	1.0 E09	Nativoamericano	50.0253	9425	Nativoamericano
COTU03007	40.8777	1.0 E09	Nativoamericano	47.6257	78889	Nativoamericano
COTU03009	36.6621	1.0 E09	Nativoamericano	47.7728	429038	Nativoamericano
COTU03011	38.695	1.0 E09	Nativoamericano	44.6386	754115	Nativoamericano
COTU04001	21.6868	1.0 E09	Nativoamericano	39.3977	18740547	Nativoamericano
COTU02006	41.5328	1.0 E09	Nativoamericano	42.9403	153660	Nativoamericano
COTU02011	33.3789	1.0 E09	Nativoamericano	43.3876	151131	Nativoamericano
COTU02016	37.6887	1.0 E09	Nativoamericano	42.5637	5335	Nativoamericano
COTU02017	41.093	1.0 E09	Nativoamericano	54.8724	246	Nativoamericano
COTU02021	36.122	1.0 E09	Nativoamericano	43.8065	8333549	Nativoamericano
COTU02024	29.7268	1.0 E09	Nativoamericano	58.2691	104	Nativoamericano
COTU02034	37.0879	1.0 E09	Nativoamericano	41.9958	2527649	Nativoamericano
COTU02036	41.2717	1.0 E09	Nativoamericano	51.7156	9364	Nativoamericano
GUSJ01006	43.926	1.0 E09	Nativoamericano	42.064	980580	Nativoamericano
GUSJ01009	39.8387	1.0 E09	Nativoamericano	45.2566	11210933	Nativoamericano

Dados estos resultados, debe realizarse con precaución la interpretación sobre proporciones de mezcla en poblaciones latinoamericanas empleando SNPforId 52plex, ya que de acuerdo con los valores  $I_n$  estimados usando los datos de las poblaciones de referencia, las frecuencias alélicas para los marcadores en conjunto no son suficientemente divergentes para permitir valores equilibrados de informatividad en las tres poblaciones de referencia, por lo que algunos resultados pueden sesgar las interpretaciones. Es importante considerar entonces el alcance del uso de estos marcadores de identificación para análisis de ancestría, evaluar otras combinaciones de marcadores o aumentar los tamaños de muestra de las poblaciones de referencia nativoamericanas para obtener mayor precisión.

### 2.3.5 Comparación de métodos de asignación

Los resultados del ACP de los marcadores AIMs (**Figura 2-4A y Figura Suplementaria 2-1**) se analizaron con más detalle para los individuos que se separan del resto de las muestras, los cuales se citan en la **tabla suplementaria 11**, donde se comparan los datos de autodeterminación y de asignación genealógica obtenidos mediante el uso de la encuesta y los resultados de estimación de proporción de ancestría de acuerdo con el análisis de STRUCTURE, asignación por DAPC y clasificación bayesiana con SNIPPER. Los resultados correspondientes al análisis de concordancia simple estimando la probabilidad de acuerdo observado para estas herramientas de asignación, se presentan en la **Tabla 2-4**. Los valores individuales de componente ancestral nativoamericano estimados por Structure para este grupo de muestras fue superior a 0,6. Como se observa para estos métodos la concordancia observada supera el 80%.

**Tabla 2-4.** Probabilidad de acuerdo observado para la asignación de ancestría nativoamericana para las 48 muestras de la región caribe colombiana que se separan de las demás en la gráfica ACP (Figura Suplementaria 1)

METODO	Autodeterminación	Genealogía	Structure	Snipper
<b>Genealogía</b>	0.875			
<b>Structure</b>	0.833333333	0.833333333		
<b>Snipper</b>	0.833333333	0.833333333	1	
<b>DAPC</b>	0.8125	0.85416667	0.97916667	0.97916667



### 2.3.6 Relación con datos históricos y migratorios

Las muestras de la población de Tubará y de Tuchín tienen la mayor proporción promedio de mezcla nativoamericana, y se encuentra que el 100% de los individuos en Tubará y el 81.48% en Tuchín tienen componente nativoamericano mayoritario. Los datos del DANE muestran que estas dos poblaciones tienen un alto porcentaje de personas que se autodeterminan indígenas (DANE, 2010a; DANE, 2010b). Las etnias históricamente relacionadas con estos municipios son Mokaná para Tubará y Zenú para Tuchín. El pueblo Mokaná asentado desde la época precolombina en el hoy municipio de Tubará, se consideró desaparecido por muchos años hasta que los líderes de la comunidad propendieron a partir de los años 90, por la reagrupación y recuperación de su identidad cultural, constituyéndose incluso las tierras de resguardo en esa región del departamento del Atlántico (Arango & Sanchez, 2006). El 82% de los autodeterminados Mokaná mostraron un mayor componente nativoamericano. Por otro lado, en la población de Tuchín se ubica una de las zonas de resguardo indígena de la etnia Zenu, cuyos ancestros habitaron las tierras ubicadas entre los ríos Sinú, San Jorge y Cauca antes de la llegada de los españoles. En ninguna de las dos etnias se conserva la lengua ancestral, pero se relaciona a los Zenues con la familia lingüística Chibcha y a los Mokaná con la familia lingüística Arawak. (Arango & Sánchez, 2006; Baquero & De la Hoz, 2010; Chaves, 1979; López, 2007a).

Las diferentes proporciones de mezcla estimadas en cada localidad son respaldadas por varios eventos históricos y demográficos. Turbo es una ciudad inicialmente formada por personas provenientes del Chocó, Panamá y Cartagena. Hasta 1909 la explotación del caucho fue fuente de empleo para muchos migrantes de otras regiones del país y en 1926 la apertura de la carretera al mar facilitó la comunicación con Antioquia y el interior dando paso a la colonización antioqueña. (Osorio, 2006). La bonanza bananera en los 60's atrajo de nuevo la llegada de personas de otras regiones pero la violencia desatada en la zona de Urabá generó el desplazamiento y un gran número de muertes y desapariciones. En Becerril, la inmigración entre los años 60 y 70 sucede como consecuencia de la bonanza algodонера y cultivos de palma recibiendo migrantes desde el interior del país y varias zonas de la costa caribe, igualmente se desata la violencia con el consabido desplazamiento de la población civil y el recrudecimiento del conflicto armado alrededor del

año 2000. Igualmente la inmigración reciente por minería de carbón y los más de 10000 muertos entre los años 2000 y 2006 y despojo de tierras de más de 250 familias en la región comprendida entre las poblaciones de Valledupar y Becerril, fueron eventos fundamentales en la conformación de la estructura genética de los habitantes de esta región (Gutiérrez, 2012). Pueblo Bello recibió inmigrantes Europeos que huían de las 2 guerras mundiales, especialmente alemanes, y por sus condiciones climáticas se convirtió en el sitio vacacional de habitantes de Barranquilla y otras ciudades de la Costa Atlántica muchos de los cuales eligieron esta localidad como residencia permanente. En los años 40 se reciben inmigrantes del interior desplazados por la violencia colonizando las tierras bajas para sembrar café, que se incrementan en número con la comunicación terrestre con Valledupar. De igual forma, el desplazamiento de los miembros de la etnia Arhuaca por la violencia y la escasez de alimentos en la Sierra Nevada de Santa Marta, han obligado a los indígenas a vivir o trabajar en esta y otras poblaciones aledañas (Datos recolectados por entrevistas a los participantes) (Zambrano, 2000).

## 2.4 Conclusiones

Varios grupos de investigación han demostrado la composición triétnica de las poblaciones en América usando distintos tipos de marcadores genéticos. La proporción de mezcla observada depende de las diferencias en los eventos históricos y demográficos, de tipo climático, obstáculos geográficos, agotamiento de los recursos, guerras locales y procesos de colonización entre otros, que modelan los patrones y rutas de migración ancestral y reciente. De igual forma, el tiempo transcurrido desde la mezcla y el tamaño efectivo de población en cada generación, mediadas por múltiples circunstancias de orden económico, social, cultural e incluso de salud, han cambiado entre las generaciones desde la llegada a América de los primeros pobladores, y el impacto de estas circunstancias en cada población es variable. Por su posición geográfica, Colombia, han sido estratégica durante los diferentes eventos de colonización, tanto ancestral (poblamiento americano, colonización europea, desplazamiento del ejército libertador) como reciente (llegada de Judíos, Turcos, Sirios, Libaneses, Palestinos, desplazamiento por bonanzas agrícolas y mineras por la guerra de los mil días y la violencia liberal y conservadora y más recientemente por el conflicto armado entre guerrilla, paramilitarismo y fuerzas del estado), lo que probablemente ha favorecido mayormente la mezcla interétnica que ha dado como resultado la actual composición genética de su población, con menores proporciones de mezcla nativoamericana que en países como México, Perú y Bolivia, pero con mayor proporción que la indicada por estudios previos en población mezclada.

Los resultados obtenidos en este estudio indican que realizar un muestreo no probabilístico a conveniencia, diseñado con base en un modelamiento previo de la región de interés, con base en datos estadísticos poblacionales e históricos, y respaldado por el uso de una encuesta que contenga datos genealógicos y demográficos de los participantes, puede favorecer el hallazgo de muestras con ancestría nativoamericana en una proporción mayor del 30% en poblaciones mezcladas, facilitando la búsqueda de los haplotipos ancestrales con diferentes clases de marcadores sin tener que recurrir al muestreo en comunidades “aisladas”, y permitiendo explorar de forma más precisa las relaciones genéticas entre poblaciones descendientes de pueblos ancestrales.

Los resultados de este trabajo muestran, que la evaluación de la ancestría de poblaciones mezcladas de origen triétnico en el norte de Sur América, con los marcadores del set SNPforID 52plex, es sesgada en comparación con la asignación hecha a partir del subset de 79 marcadores AIMs del panel Lace (Galanter, et.al., 2012), por lo que requieren una interpretación cuidadosa de los resultados generados con el panel de identificación y la búsqueda de marcadores que permitan un mejor balance de la divergencia en las poblaciones de referencia. Se observa sin embargo, que a medida que el componente ancestral mayoritario aumenta hay mayor concordancia entre los resultados obtenidos con el set de marcadores de ancestría y el de identificación en especial en los casos en los que el componente ancestral mayoritario (nativoamericano) es superior a 0.8, de acuerdo con la evaluación analítica realizada.

Aunque los marcadores empleados han sido útiles para determinar diferencias entre grupos a nivel continental, es importante explorar nuevos SNPs que permitan niveles de análisis de ancestría más finos, aumentar los tamaños de muestra para corroborar las observaciones y proponer la construcción de un panel informativo para grupos étnicos nativoamericanos, con el fin de evaluar su ancestría. Dados los resultados de la evaluación de  $K$  de STRUCTURE con el set SNPforID 52plex que mostró un posible  $K=4$  dividiendo las poblaciones nativoamericanas, y los posteriores resultados de DAPC para poblaciones nativoamericanas de referencia donde se separan poblaciones, cobra importancia profundizar en estos resultados.

Teniendo en cuenta que una gran parte de la violencia por casos de desaparecidos y fosas comunes incluyen poblaciones de la región Caribe Continental Colombiana y grupos étnicos que por muchos años han sido altamente vulnerables, el desarrollo de este tipo de trabajos contribuye con la implementación de herramientas genético-poblacionales para la determinación de la ancestría biogeográfica y sus aplicaciones en el campo forense, en el marco de la actual transición de Colombia hacia el manejo del posconflicto.

Por otro lado, perspectivas de trabajos interdisciplinarios en cuanto a la historia del poblamiento americano y la relación por filiación lingüística de los pueblos indígenas del continente, se abren con la búsqueda de paneles de SNPs que permitan discriminar más detalladamente la variación intracontinental empleando herramientas de análisis

multivariado y técnicas de muestreo a conveniencia con modelamientos poblacionales previos.

## **3. Capítulo 3. Análisis de la ancestría nativoamericana de poblaciones del norte de Suramérica**

### **3.1 Introducción**

Alrededor de la historia del poblamiento americano se han planteado diversas hipótesis construidas con base en evidencias aportadas por la arqueología, la antropología, la lingüística y más recientemente por la genética (Dillehay, 2008; Rothammer & Dillehay, 2009; Pitblado, 2011; He et.al., 2012; Reich et.al., 2012). La idea de mayor aceptación entre la comunidad científica hace referencia a que el poblamiento de América comenzó hace aproximadamente 15000 a 18000 años, con pueblos provenientes de Asia Sur Centro que atravesaron el Estrecho de Bering. Respecto a la llegada y posterior dispersión de los pueblos desde el norte de Suramérica, varias hipótesis incluso consideran la particular geografía del subcontinente como un factor que influyó los patrones migratorios y la posibilidad de establecer relaciones culturales, económicas y de pareja entre los grupos habitantes de la época (Wang et.al., 2007; Rothammer & Dillehay, 2009; Reich et.al., 2012). Una vez iniciada la colonización europea tuvieron lugar diversos procesos de reducción y en muchos casos extinción de la población Nativoamericana, así como de mezcla entre las culturas ya establecidas con europeos y africanos, originando a una compleja estructura poblacional que hoy es objeto de varios tipos de estudios en el campo de la genética de poblaciones.

Los resultados del análisis de la proporción de ancestría en la región usando marcadores uniparentales, muestran una clara proporción mayoritaria de linajes paternos europeos y de linajes maternos nativoamericanos y las contribuciones africanas son variables

dependiendo de las regiones específicas analizadas (Bryc et.al., 2010; Ruiz-Linares et.al. 2014; Salzano & Sanz, 2014; Adhikari et.al, 2016; Adhikari, et.al., 2017). Estos trabajos han considerado tanto poblaciones mezcladas como poblaciones con aparentes grados de mezcla mínimos, con claros orígenes nativoamericanos o afrodescendientes. Sin embargo, se ha recomendado el uso de marcadores autosómicos para el análisis de la ancestría de poblaciones mezcladas y varios trabajos involucran STRs, SNPs o Indels, como una mejor forma de estimar los valores de proporción de ancestría (Wang, et.al. 2007; Galanter, et.al. 2012; Pereira, et.al, 2012). Salsano y Sanz en 2014, reúnen datos de este tipo con diversos marcadores y muestran que países como México, Guatemala, Peru y Ecuador, tienen alta proporción de ancestría nativoamericana en rangos que superan el 50% e incluso alcanzan el 98% en algunas localidades, mientras para el norte de Suramérica como Colombia y Venezuela, esta proporción resulta menor. De acuerdo a las zonas específicas muestreadas, estas proporciones tienen un grado de variabilidad que puede relacionarse con las historias particulares de poblamiento posteriores a la conquista incluyendo los aportes diferenciales de los genomas africanos, dependiendo del manejo particular de la esclavitud en distintas zonas del subcontinente, aspectos que son menos estudiados, aunque no menos interesantes que los análisis continentales.

Este trabajo plantea cómo un enfoque de selección de la muestra basado en datos históricos, demográficos y genealógicos, puede ser de utilidad para encontrar genotipos con ancestría nativoamericana en poblaciones mezcladas, permitiendo analizar relaciones entre etnias a nivel subcontinental y propone un set de marcadores SNPs autosómicos que han sido seleccionados a partir de marcadores AIMs (Galanter, et.al., 2012) y de identificación (SNPforID 52plex, Sanchez et.al., 2006, Porras et.al., 2009), para ser probados en el análisis de la ancestría nativoamericana en Sur y Centroamérica, en el marco de las hipótesis de poblamiento.

### 3.2 Materiales y Métodos

Figura 3-1. Etnias indígenas incluidas en los análisis. Se señalan con ● las etnias muestreadas en este trabajo. Figura adaptada de Wang et.al., 2007





### 3.2.1 Poblaciones y encuestas

La etapa de muestreo se llevó a cabo en 9 poblaciones de la Costa Caribe Colombiana, que fueron seleccionadas con base en la revisión de datos históricos sobre los asentamientos indígenas precolombinos y sus movimientos posteriores hasta el establecimiento de las actuales zonas de resguardo, eligiendo preferiblemente aquellos municipios que de acuerdo con la información demográfica obtenida de los datos del último censo nacional de población realizado por el Departamento Nacional de Estadísticas de Colombia (DANE), presentaban mayores valores de componente indígena por autodeterminación. (**Figura 3-1 y Tabla suplementaria 2-1**). Las muestras de personas mayores de edad residentes en la localidad por al menos 5 años fueron tomadas por hisopado de células de mucosa oral o por venopunción (posteriormente depositadas sobre tarjetas FTA), previa firma del consentimiento informado y con aplicación de una encuesta para recolectar información genealógica y demográfica incluyendo otros aspectos como lugares de nacimiento y de residencia tanto del participante como de sus padres, abuelos y bisabuelos tanto por línea materna como paterna. En total se recolectaron 258 muestras, 194 femeninas y 64 masculinas. Las poblaciones de referencia empleadas para el análisis de AIMs corresponden a 716 muestras de población europea (616 España, 56 residentes de Utah con ancestría de norte y occidente europeo, 44 Italia), 103 de población africana (50 Luhya y 53 Yoruba), 127 nativoamericanos (25 Aymara, 22 Maya, 14 Nahua, 23 Quechua, 22 Tepehuanos, y 21 Zapoteca) procedentes de la base de datos del proyecto HAPMAP genotipificadas con el panel de ancestría Genome-Wide de LACE (Galanter, et.al., 2012). Además se utilizaron los genotipos de las muestras de referencia nativoamericanas de las etnias Awa (2), Coyaima (14), Panare (7), Pastos (6), Venezuela Re (Pemón) (3) y Venezuela W (Warao) (8). Estos datos fueron aportados por el Centro Nacional de Genotipado (CEGEN) en la Fundación Pública Gallega de Medicina Genómica, Santiago de Compostela, España. Para el análisis del panel SNPforID 52plex, los genotipos de las poblaciones de referencia empleadas se obtuvieron de las bases de datos 1000 genomas fase I y HapMap, desde SNP forID Browser en la herramienta SPSmart v5.1.1 (<http://spsmart.cesga.es/snpforid.php>), incluyendo 396 Europeos (16 Británicos, 88 Italianos, 125 España Noroccidental, 167 Portugal), 340 Africanos (65 Angola, 20 C. African Republic Biaka Pygmy, 12 D. R. of Congo Mbuti Pygmy, 9 Kenya Bantu N.E., 133 Mozambique, 6 Namibia San, 19 Nigeria Yoruba, 10 Nigerian\_(resident\_in\_UK), 19 Senegal Mandenka, 7 South Africa Bantu S.E., 40 Uganda), y 219 nativoamericanos (14 Brasil

Karitiana, 6 Brasil Surui, 19 Colombia Awa, 6 Colombia CEPH (Piapoco-Curripaco), 27 Colombia Coyaima, 4 Colombia Embera, 8 Colombia Pijao, 33 Colombia Pastos, 76 Ecuador Kichwa, 16 Mexico Maya, 10 Mexico Pima). Adicionalmente, también se incluyeron 33 genotipos colombianos de la etnia Pastos, publicados en Ibarra, et.al., 2012. Estas mismas poblaciones de referencia se emplearon con el software Structure 2.3.4 y con la herramienta Snipper 2.5.

### 3.2.1 Extracción de ADN y genotipificación de SNPs

La extracción de ADN se llevó a cabo empleando el PrepFiler Express™ Forensic DNA Extraction Kit empleando el AutoMate Express™ Forensic DNA Extraction System de Applied Biosystems, siguiendo las instrucciones del fabricante. Un set de 79 marcadores tipo SNP's de ancestría AIMs (Ancetry Informative Markers) fueron seleccionados del panel LACE de 446 AIMs publicado en Galanter, et.al., 2012. Los marcadores seleccionados se listan en la **Tabla suplementaria 2-2**. Los marcadores de identificación correspondieron al SNPforID 52plex assay (SNPforID group) publicado por Sánchez, et.al. 2006. **Tabla suplementaria 2-2**. Los productos amplificados de ambos paneles fueron tratados con SAP y posteriormente sometidos a una reacción de extensión para finalmente ser genotipificados usando la plataforma Sequenom MassArray System para análisis de espectrometría de masas con la técnica MALDI-TOF. El análisis de resultados se realizó mediante el software "Typer Analyzer".

### 3.2.1 Análisis genético poblacional y estadístico

Las frecuencias alélicas se estimaron usando la librería GenABEL en plataforma R (<http://cran.r-project.org>) (Aulchenko, 2014). Arlequín 3.5.2.2 (Excoffier, et.al. 2010) se usó para calcular la heterocigocidad observada, el Análisis Molecular de Varianza (AMOVA) agrupando las poblaciones como Mesoamericanas, Andinas, Colombianas y Brasileñas (para SNPforID 52plex) o Venezolanas (para AIMs) y para construir las matrices de valores Fst pareados. El análisis de componentes principales (ACP) fue hecho con el software MVSP 3.22 incluyendo la gráfica biplot para determinar las variables que influyen en la distribución de los puntos. El análisis discriminante de componentes principales (DAPC) se realizó con el paquete Adegenet (Jombart, 2008) en el ambiente R (<http://cran.r-project.org>). Las poblaciones de referencia nativoamericanas y las muestras tomadas en la región

Caribe Continental Colombiana que por el análisis de ancestría realizado en el capítulo anterior, presentan un valor de ancestría nativoamericana superior a 0.7 fueron analizadas con el software STRUCTURE 2.3.4 (Pritchard et.al., 2000) para la evaluación del  $K$  óptimo (NO POPFLAG) con un periodo de burn-in de 100000 y 100000 repeticiones MCMC, evaluando 6 iteraciones desde  $K=1$  hasta  $K=6$ . El valor óptimo de  $K$  se estimó con STRUCTURE HARVESTER (Earl, D & vonHoldt, B., 2012). Con la herramienta en línea Snipper 2.5 App suite (<http://mathgene.usc.es/snipper/>) se realizó la validación cruzada del set de poblaciones y a partir de los valores de divergencia se estimaron los valores de informatividad de asignación de Rosenberg (Rosenberg et.al., 2003) para cada marcador ( $\ln$ ) multiplicando por 0.693 (Phillips, 2015b) y se calcularon los valores likelihood ratio (LR) mediante un análisis bayesiano para evaluar el éxito de la clasificación de individuos sin predefinir poblaciones. Los árboles de poblaciones fueron construidos con el método Neighbor-Joining (Saitou, N. & Nei, M. 1987) con base en la matriz de distancias de Fst pareados de Arlequín (Excoffier, et.al. 2010) usando MEGA 7 (Kumar et.al. 2016). Para realizar comparaciones también se construyeron árboles con la misma metodología usando Distancia de Cuerda (Cavalli-Sforza, L & Edwards, A., 1967) y distancia Estandar de Nei (Nei, M., 1972), en el software Populations 1.2.28 (Langella, O. 1999).

## 3.2 Resultados y discusión

El análisis de la encuesta aplicada muestra que el 92% de los participantes nacieron en la región Caribe continental Colombiana y la asignación realizada por los investigadores con base en la información genealógica indicó que el 30% de las muestras corresponden a personas con ancestría nativoamericana al menos dos generaciones principalmente habitantes de las localidades de Tubará, Pueblo Bello y Tuchín, lo que coincide con las proporciones de autodeterminación de pertenencia étnica indígena del DANE 2005 (**Tabla suplementaria 2-1**).

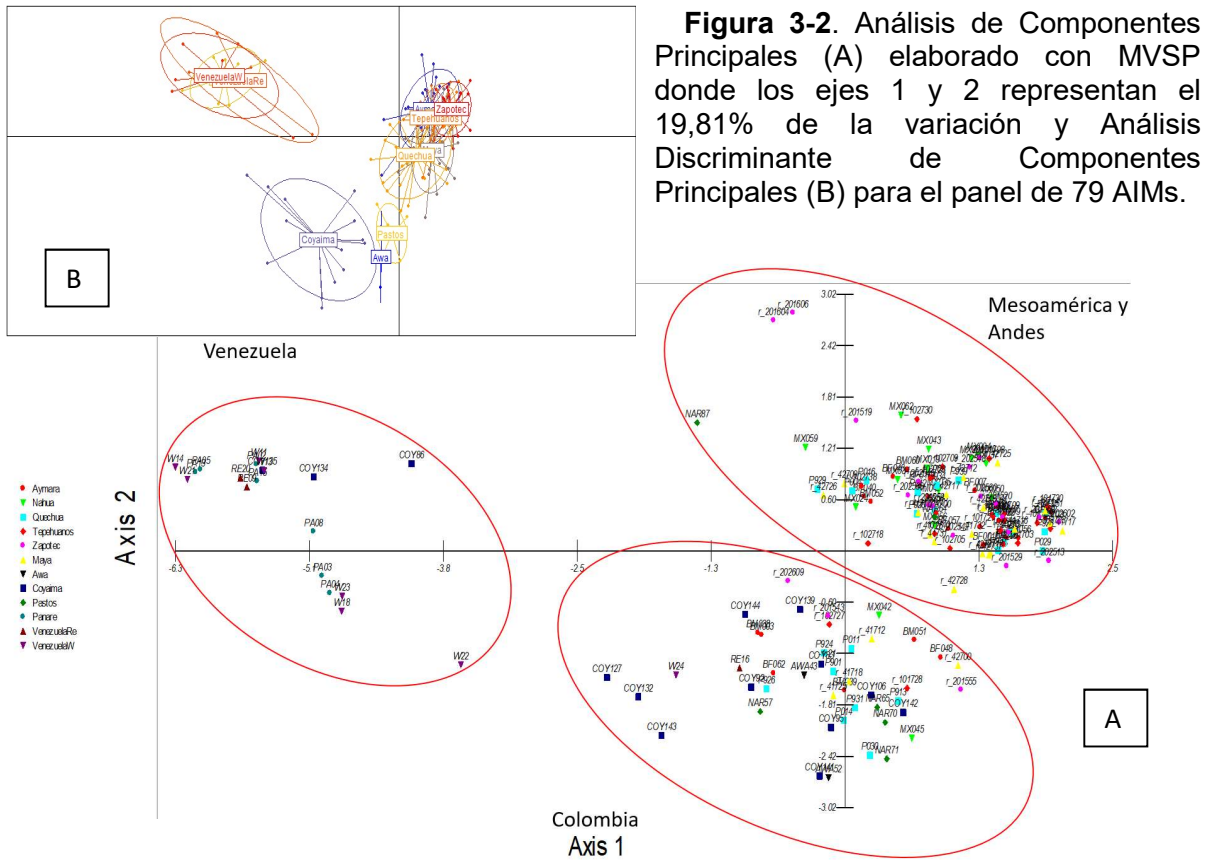
### 3.3.1 Estructura genética

Las frecuencias alélicas y los resultados del análisis de subestructura del conjunto completo de datos de la región Caribe Continental de Colombia con los dos sets de marcadores

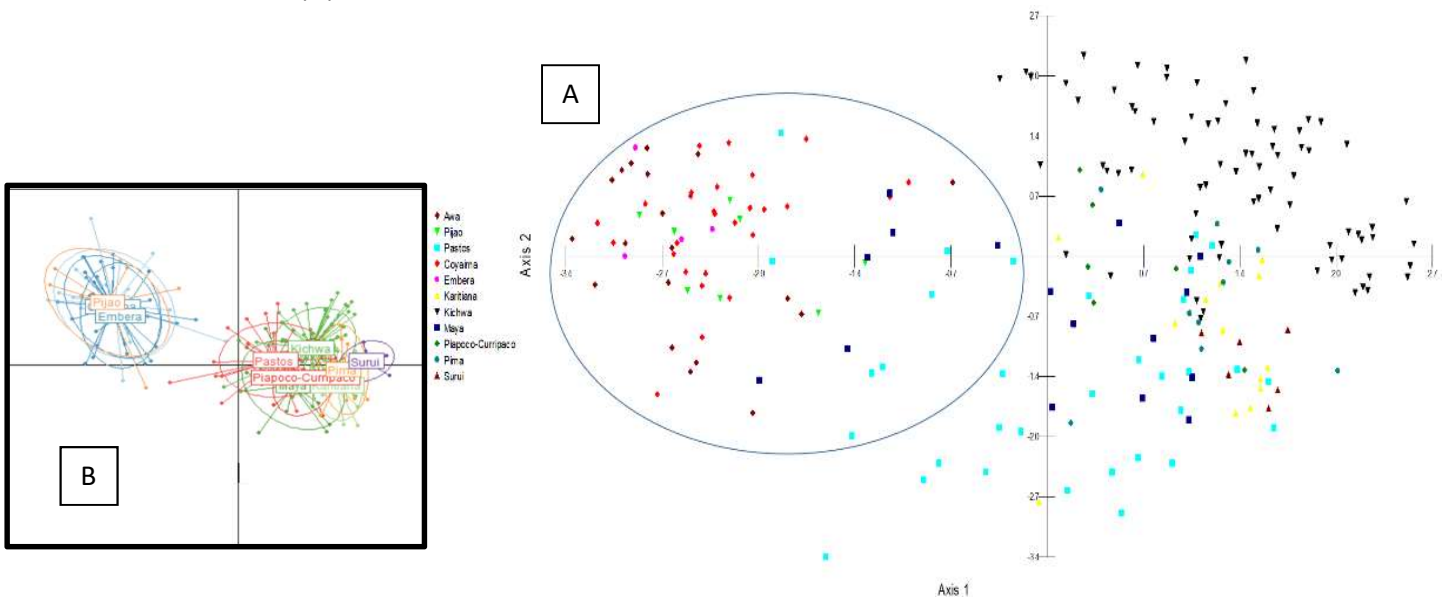
SNPs (AIMs y SNPforID 52plex), fueron presentados en el capítulo anterior, donde se estimó un 43% de ancestría Nativoamericana para el total de la muestra del Caribe Continental Colombiano, empujando el subpanel de 79 AIMs, con las poblaciones de Tubará y Tuchín que presentaron los valores promedio más altos (62% y 70% respectivamente) y detectaron en ambas localidades que para cerca 60% de los individuos muestreados presentan una contribución de ancestría nativoamericana individual superior a 0.6

Para evaluar la utilidad de los marcadores tanto AIMs como SNPforID 52plex en el análisis de la ancestría nativoamericana a nivel de subpoblaciones, se utilizaron solamente las muestras de referencia para realizar en primer lugar un ACP que mostró tanto con los marcadores AIMs como con SNPforID 52plex, una clara separación del grupo de poblaciones colombianas. Sin embargo, se observa con marcadores AIMs individuos Quechua y con SNPforID 52plex individuos Pastos que se dispersan. Con los marcadores de ancestría se aprecia además un tercer grupo conformado por las muestras de etnias nativoamericanas habitantes del sur de Venezuela. El DAPC con clusters predefinidos soporta los resultados del ACP con AIMs y con los marcadores de SNPforID 52plex, mostrando un grupo conformado por las etnias colombianas Pijao, Embera, Awa y Coyaima y un segundo conjunto que reúne Karitiana, Surui, Quechua, Maya, Pima, Pastos y Piapoco-Curripaco (**Figuras 3-2 y 3-3**).

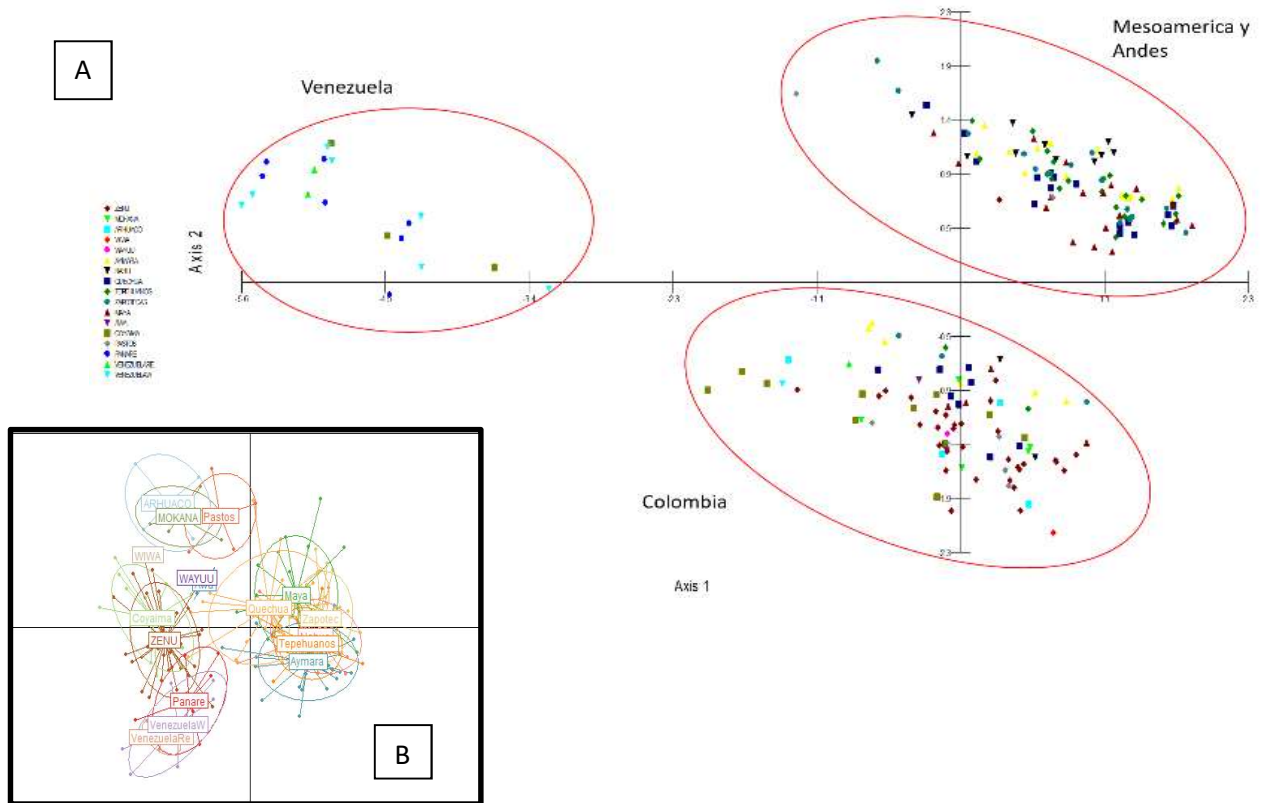
Para el análisis de la ancestría nativoamericana de la región Caribe Continental Colombiana, se seleccionaron solo aquellos individuos cuyo análisis de ancestría individual con marcadores AIMs usando el software Structure, mostró una proporción de ese componente ancestral mayor o igual a 0.7, y que adicionalmente por sus genealogías registradas a través de la encuesta, pudieran ser asignados como pertenecientes a una determinada etnia independientemente del lugar de muestreo. Así, se incluyeron en el análisis 4 Arhuacos, 5 Mokaná, 1 Wiwa, 1 Wayuu, 33 Zenues. El ACP no mostró cambios sustanciales respecto al análisis solo con las muestras de referencia. Al realizar un DAPC usando los grupos predefinidos, se observó la agrupación de las poblaciones de referencia Quechua, Tepehuanos, Zapoteca, Aymara, Nahua y Maya y un segundo conjunto con las demás poblaciones de referencia y las etnias de este estudio, que aparenta subdividirse (**Figura 3-4**).



**Figura 3-3.** Análisis de Componentes Principales (A) elaborado con MVSP donde los ejes 1 y 2 representan el 21,81% de la variación y Análisis Discriminante de Componentes Principales (B) para el panel SNPforID 52plex.

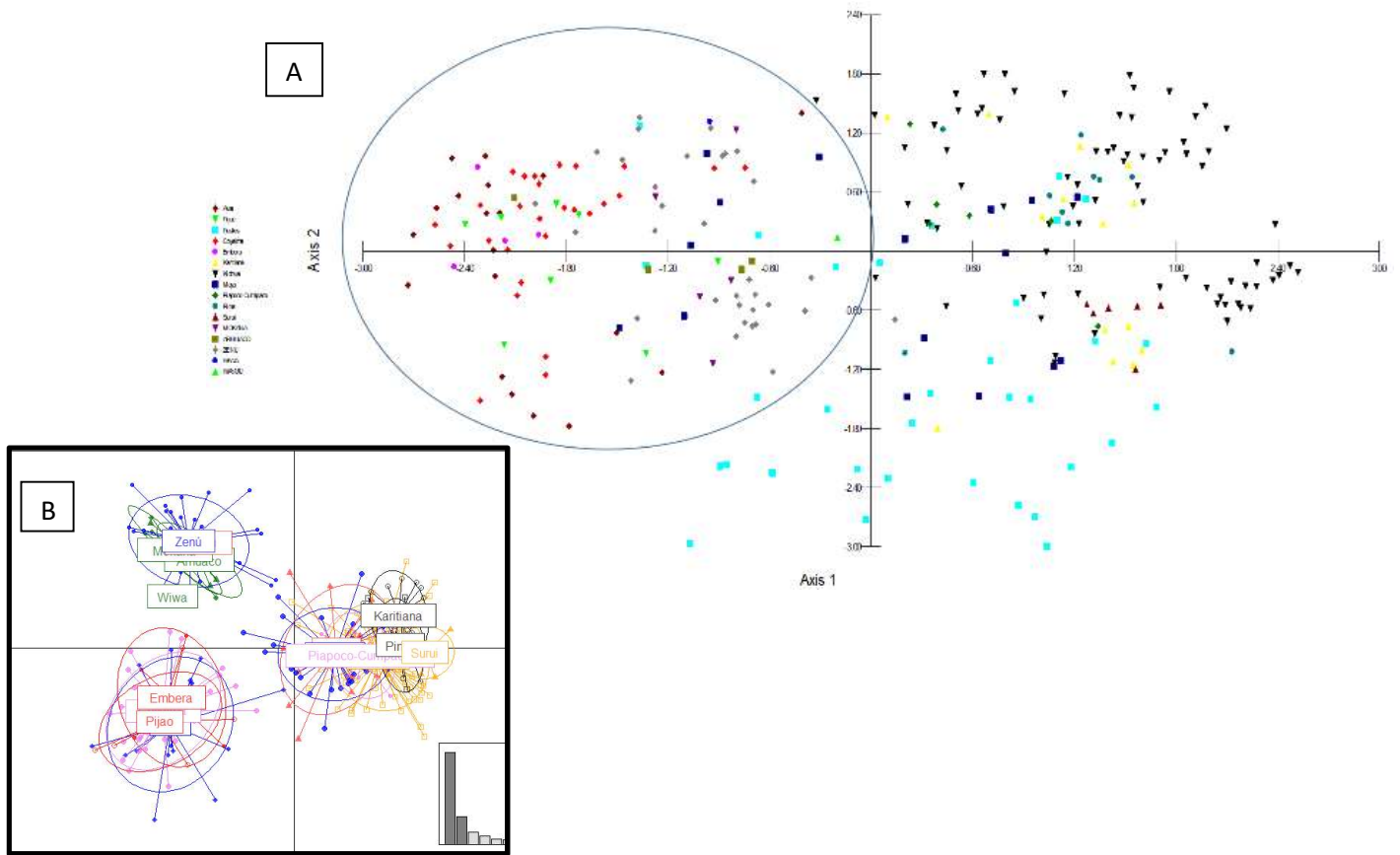


**Figura 3-4.** ACP (A) y DAPC (B) con panel de 79 AIMs incluyendo las muestras de la región Caribe Continental Colombiano por etnias. AIMs Assign.prop 0,649 ACP Eje 1 20.71 Eje 2 8.48



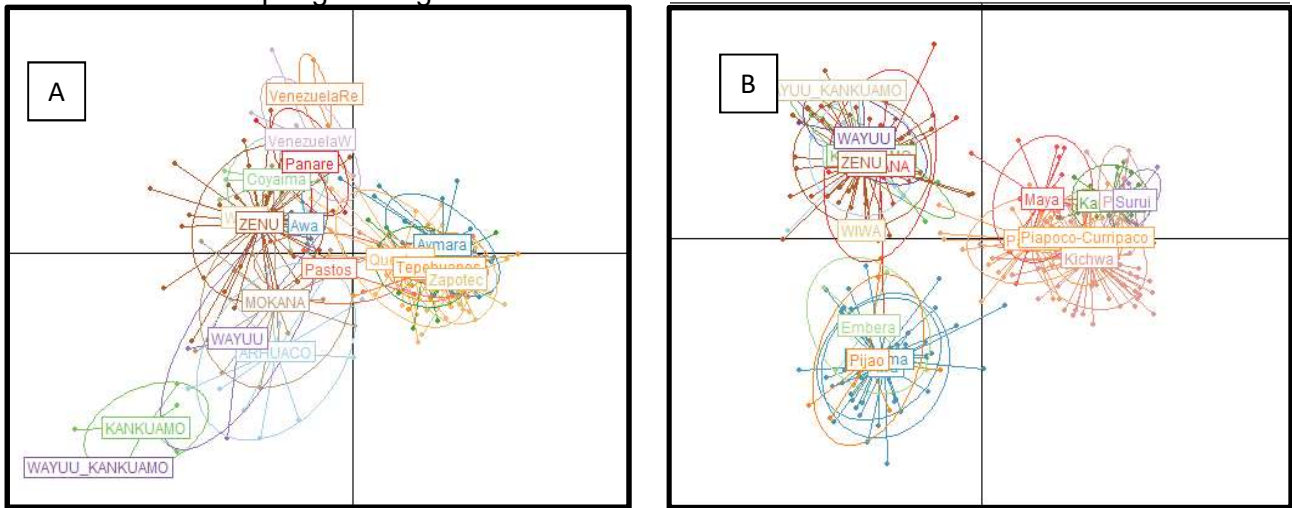
Con los marcadores del set *SNPforID 52plex*, la inclusión de las muestras del Caribe Continental Colombiano con ancestría nativoamericana mayor a 0.7 de acuerdo las etnias, no generó cambios en el ACP, sin embargo, estas muestras forman un grupo separado de los demás en el DAPC y se observa una subdivisión de las muestras de referencia nativoamericanas en un grupo conformado por Pijao, Embera, Coyaima y Awa (todas poblaciones colombianas) y otro por Pastos, Piapoco-Curripaco (etnias colombianas), Karitiana, Surui, Maya, Pima, y Quechua (**Figura 3-5**).

**Figura 3-5.** ACP donde los ejes 1 y 2 representan el 19,77% (A) y DAPC (B) para el set de marcadores SNPforID 52plex incluyendo las muestras de Caribe Continental Colombiano con ancestría nativoamericana mayor a 0.7 por etnias.



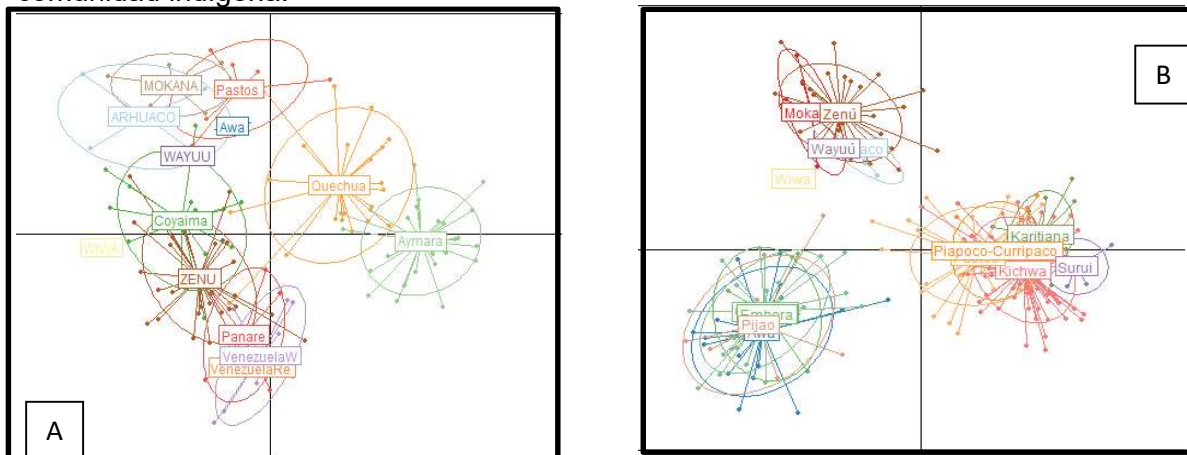
Adicionalmente, se incluyeron en el análisis las muestras correspondientes a los individuos que genealógicamente pertenecían a una etnia indígena, es decir aquellos que de acuerdo con los datos de la cuesta pertenecen a una etnia indígena por línea paterna y materna, sin importar el valor de ancestría nativoamericana estimada. En total se analizaron 11 Arhuacos, 5 Kankuamos, 15 Mokaná, 4 Wayuu, 1 Wiwa, 47 Zenues y una persona perteneciente por línea materna a la etnia Kankuamo y por línea paterna a la etnia Wayuu. Los resultados obtenidos son consistentes con el análisis anterior. **(Figura 3-6).**

**Figura 3-6** Análisis Discriminante de Componentes Principales con marcadores AIMs (A) y SNPforID 52plex (B), usando las muestras de Caribe Continental Colombiano seleccionadas por genealogía.



Al analizar las muestras del Caribe Continental Colombiano con ancestría nativoamericana superior a 0.7 organizadas por etnias, usando solamente las muestras de referencia Suramericanas con ambos tipos de marcadores, se observa en las gráficas DAPC que existe una tendencia a la formación de núcleos, más claramente definidos con el set SNPforID 52plex. Aunque no todas las poblaciones son compartidas con los dos sets de marcadores, se aprecia que las muestras del Caribe Continental Colombiano se agrupan lejos de Quechuas, y su relación con Coyaimas varía según el conjunto de marcadores. (Figura 3-7)

**Figura 3-7.** DAPC con marcadores AIMs (A) y SNPforID 52plex (B), con las etnias Suramericanas incluyendo las muestras de la región Caribe Continental Colombiana con ancestría nativoamericana igual o mayor a 0.7 con autodeterminación de pertenencia a una comunidad indígena.



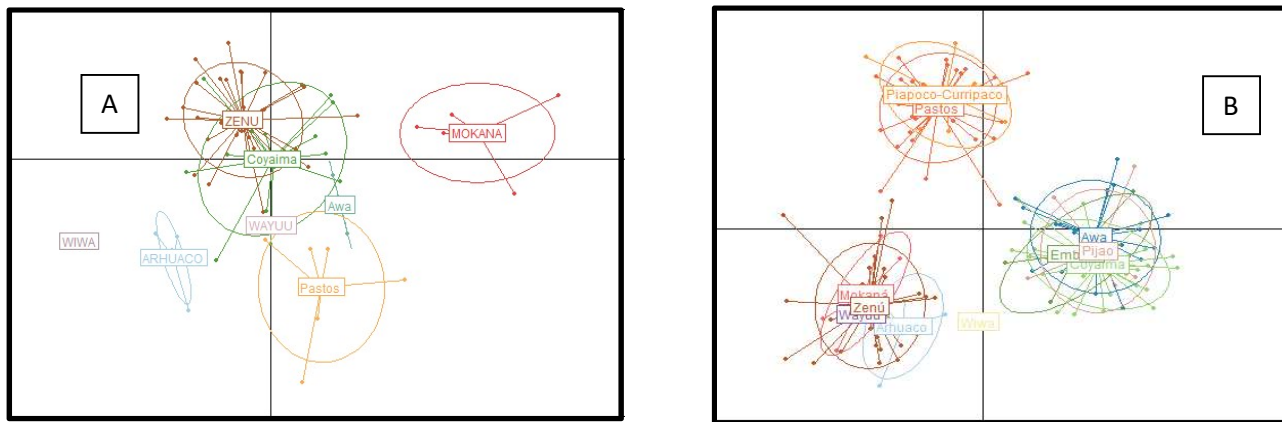
AIMs Assign.prop 0,79

AIMs Assign.prop 0,74

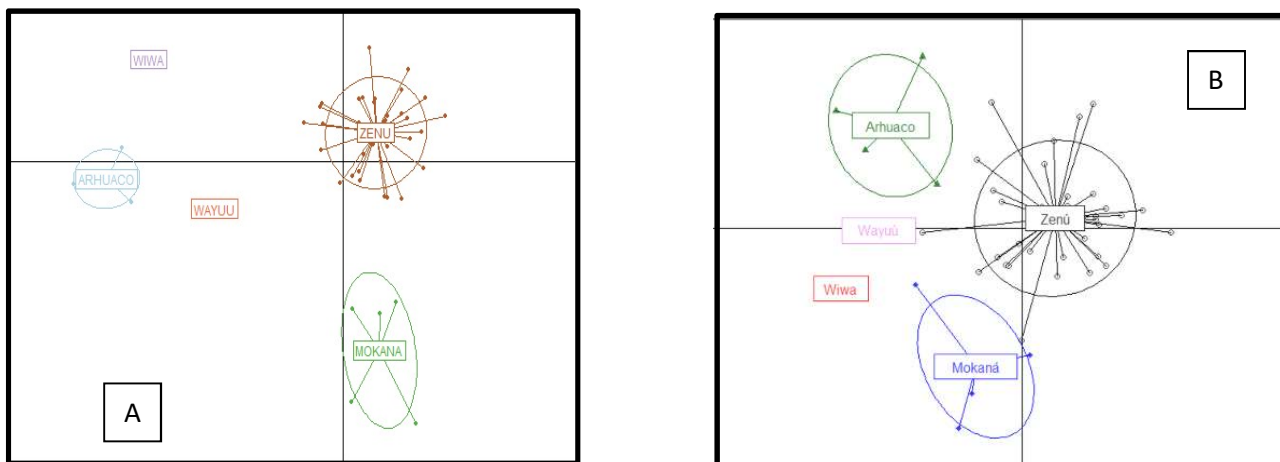


Al limitar el análisis solamente a las etnias Colombianas en las mismas condiciones de los análisis anteriores, consistentemente, el set SNPforID 52plex parece ser más eficiente en la separación de clusters. Se observa que todas las muestras del Caribe Continental Colombiano se separan de las demás etnias, destacándose la agrupación de Pastos y Piapoco-Curripaco, mientras los Awa, que geográficamente se ubican más cerca a los Pastos se agrupan con las etnias Pijao, Coyaima y Embera. (Figura 3-8). El DAPC realizado solo para las etnias del Caribe Continental Colombiano es menos informativo, aunque con los marcadores AIMs las etnias Arhuaco, Wiwa y Wayuu tienden a agruparse. (Figura 3-9)

**Figura 3-8** DAPC con marcadores AIMs (A) y SNPforID 52plex (B), con las etnias Colombianas incluyendo las muestras de la región Caribe Continental Colombiana con ancestría nativoamericana igual o mayor a 0.7 con autodeterminación de pertenencia a una comunidad indígena.



**Figura 3-9** DAPC con marcadores AIMs (A) y SNPforID 52plex (B), solo con las etnias de la región Caribe Continental Colombiana con ancestría nativoamericana igual o mayor a 0.7 con autodeterminación de pertenencia a una comunidad indígena.

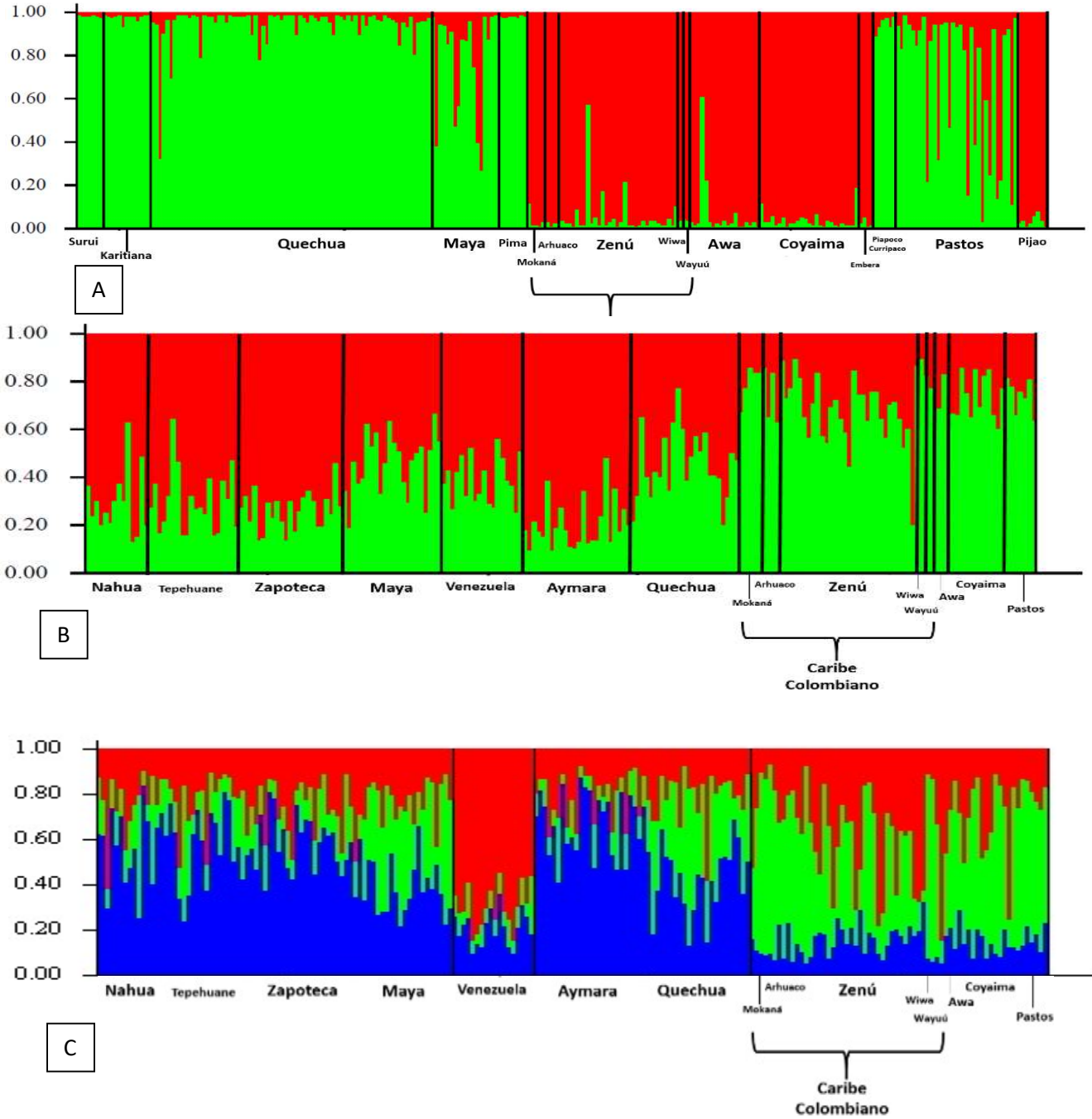


El análisis con el software STRUCTURE para los dos sets de marcadores en el conjunto total de muestras nativoamericanas, es decir incluyendo las de poblaciones de referencia y las ya seleccionadas del Caribe Continental Colombiano, mostró un  $K$  óptimo de 2 en ambos casos. (**Figura 3-10**). Con los marcadores AIMs se aprecia que un patrón de mezcla es compartido entre las etnias colombianas, y otro entre las etnias mesoamericanas y andinas. Las etnias del sur de Venezuela se aprecian en el plot  $K=2$  con un patrón similar a este último, sin embargo, cabe resaltar que el patrón obtenido con  $K=3$  diferencia estas poblaciones de los otros dos grupos, lo que puede tenerse en cuenta para futuras utilidades de los marcadores AIMs en la determinación de ancestría biogeográfica a nivel subcontinental.

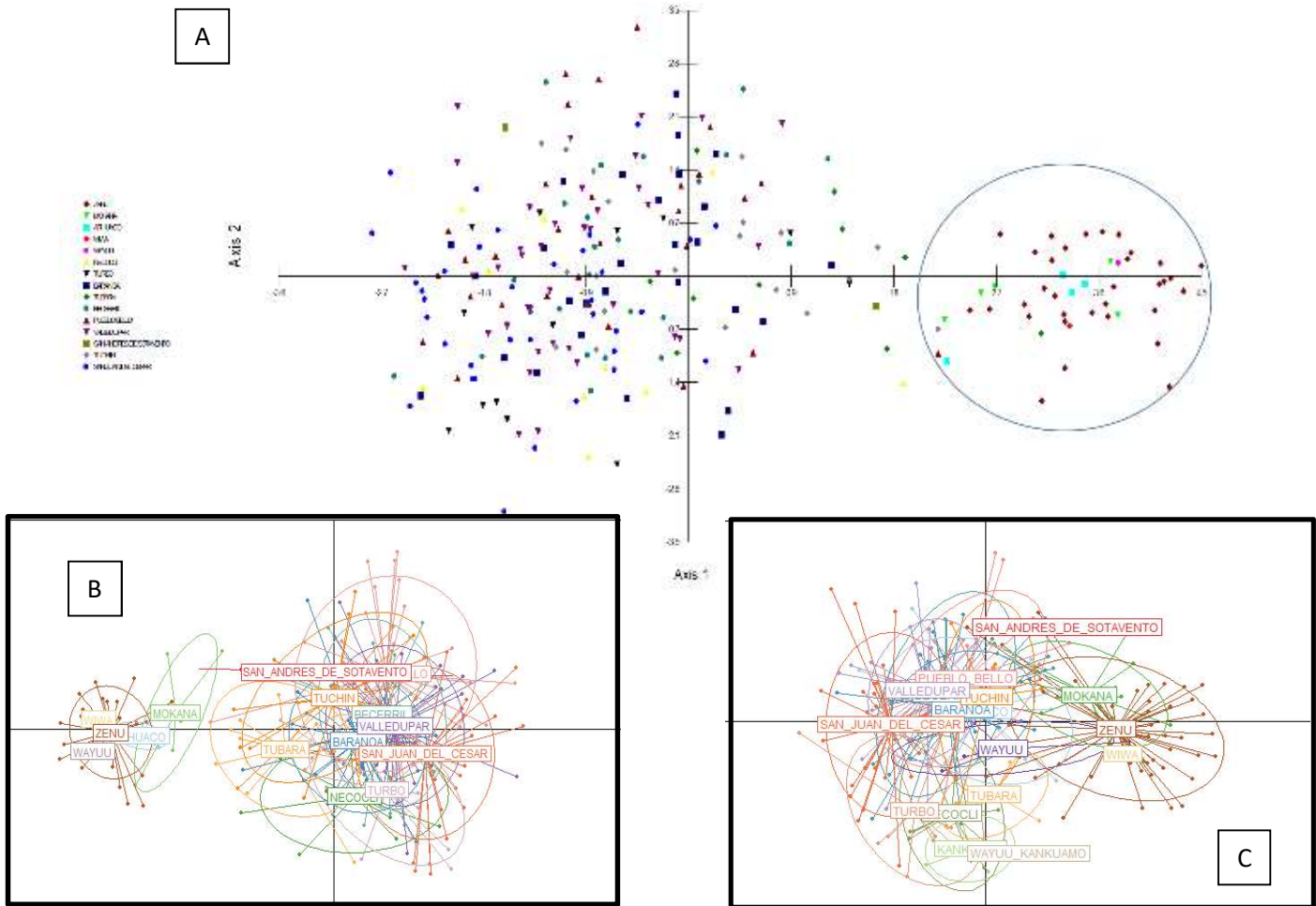
De igual forma, el AMOVA indicó que la mayor fuente de variación corresponde a las diferencias entre grupos (FCT= 0.04479, p valor= 0.00142+/- 0.00012), seguido por las diferencias entre poblaciones dentro de los grupos (FSC= 0.03084, p valor= 0.00000+/- 0.00000). Para los marcadores SNPforID 52plex, el análisis de STRUCTURE distingue un grupo formado por las etnias Mesoamericanas y Brasileñas que además incluye las etnias colombianas Pastos y Piapoco-Curripaco y un segundo grupo conformado por las restantes etnias Colombianas. Para estos marcadores de acuerdo con el AMOVA, la variación más alta corresponde a las diferencias entre poblaciones dentro de los grupos (FSC= 0.08204, p valor= 0.00000+/-0.00000).

256 muestras del Caribe Continental Colombiano se estudiaron usando el conjunto total de 131 SNPs. Tanto en el ACP como en el DAPC las muestras con contribución de ancestría nativoamericana superior a 0.7, organizadas de acuerdo con la pertenencia étnica, se separan de las restantes de la población mezclada. Al asignar por genealogías la pertenencia a un grupo nativoamericano, sin tener restricción en el valor de este componente estimado individualmente por STRUCTURE, la eficiencia de los marcadores para separar las poblaciones en el DAPC disminuye, como consecuencia del aumento de la proporción de los componentes europeo y africano en la ancestría de la muestras clasificadas como nativoamericanas, haciéndolas genéticamente más similares a las poblaciones mezcladas. (**Figura 3-11**).

**Figura 3-10** Análisis de STRUCTURE con  $K=2$  para marcadores SNPforID 52plex (A) y AIMS (B), con las etnias Suramericanas, incluyendo el Caribe Colombiana con ancestría nativoamericana igual o mayor a 0.7, con autodeterminación de pertenencia a una comunidad indígena. (C) Plot para  $K=3$  con AIMS.

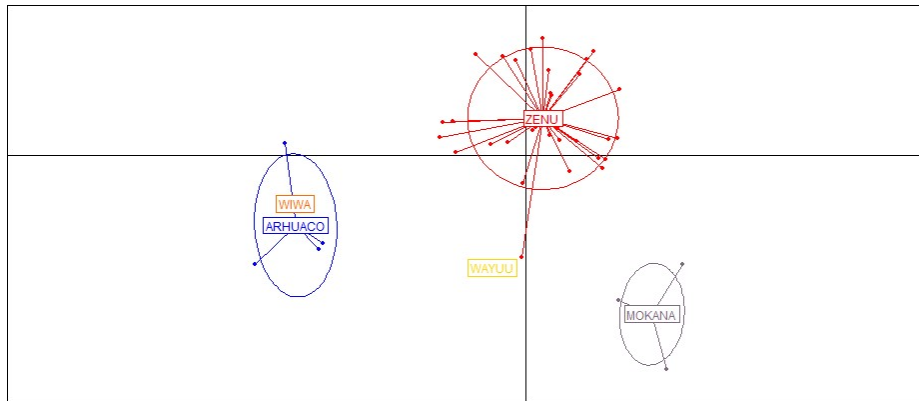


**Figura 3-11** ACP donde los ejes 1 y 2 representan el 12.55% (A) y DAPC (B) para el set completo de 131 SNPs para 256 muestras del Caribe Continental Colombiano. Las muestras dentro del círculo en la gráfica ACP corresponden a los individuos con ancestría nativoamericana mayor a 0.7 que se autodeterminan como indígenas. (C) Gráfico DAPC asignando pertenencia a un grupo nativoamericano a partir de las genealogías



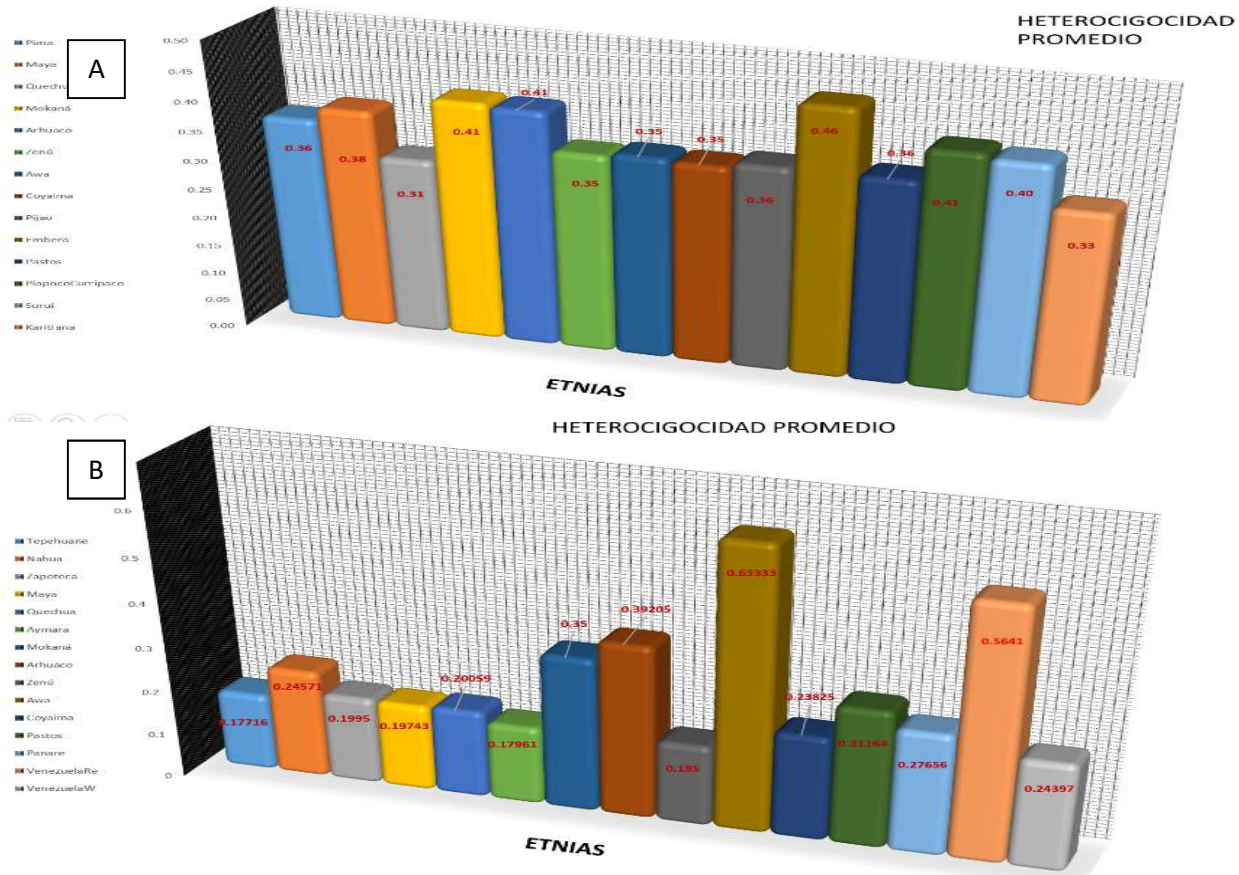
Al analizar solamente las muestras con ancestría nativoamericana mayor a 0.7 que se autodeterminan como indígenas por separado, la muestra Wiwa se agrupó con las muestras de Arhuacos, lo cual concuerda con su origen ancestral común por ser ambas etnias pertenecientes al gran grupo Tayrona habitantes de la Sierra Nevada de Santa Marta y de las que se conoce mantienen en la actualidad flujo genético. Adicionalmente, la muestra Wayuú se ubica de forma equidistante a Zenú y Mokaná, al respecto a lo cual se realiza un análisis más adelante. **(Figura 3-12)**

**Figura 3-12.** DAPC con el conjunto completo de 131 SNPs, sólo con las etnias de la región Caribe Continental Colombiana con ancestría nativoamericana igual o mayor a 0.7, con autodeterminación de pertenencia a una comunidad indígena.



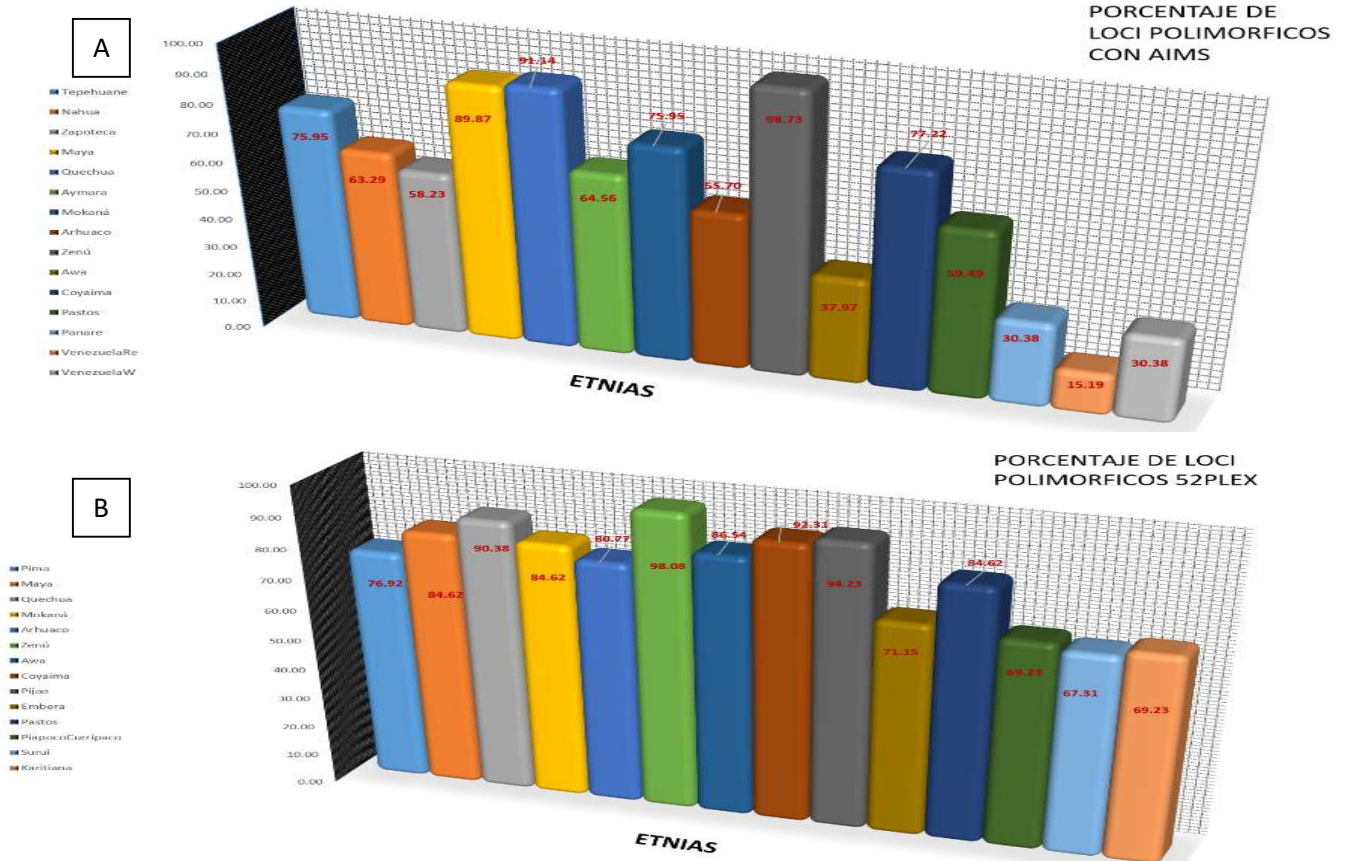
La Heterocigocidad promedio (H) y el Porcentaje de Loci Polimórficos (PLP) se emplearon para explorar la diversidad genética de las poblaciones analizadas. Aunque no se observa disminución de H hacia el sur del continente como en Wang et.al., 2007, para los marcadores SNPforID 52plex los datos de H y de PLP son más homogéneos entre las etnias que los mismos índices con marcadores AIMs, aunque en ambos casos, menores en comparación con los datos de marcadores STRs analizados por Wang y colaboradores (2007), quien reportó valores de H entre 0.482 y 0.79. Sin embargo, las poblaciones Surui y Karitiana que conservan un modo de subsistencia de cazadores y recolectores (He et.al., 2012) y que presentaron los más bajos valores de H en el trabajo de Wang et.al., 2007, tienen también en este trabajo el más bajo PLP con los marcadores SNPforID 52plex. La misma observación se hace para las poblaciones del sur de Venezuela con los marcadores AIMs, lo que podría semejar un patrón decreciente de disminución de occidente a oriente en concordancia con Wang, et.al., 2007, representando a través de esta medida, la disminución de la diversidad genética, considerando que los marcadores son bialélicos. (Figuras 3-13 y 3-14)

**Figura 3-13** Valores de Heterocigicidad promedio por población con AIMs (A) y SNPforID 52plex (B)

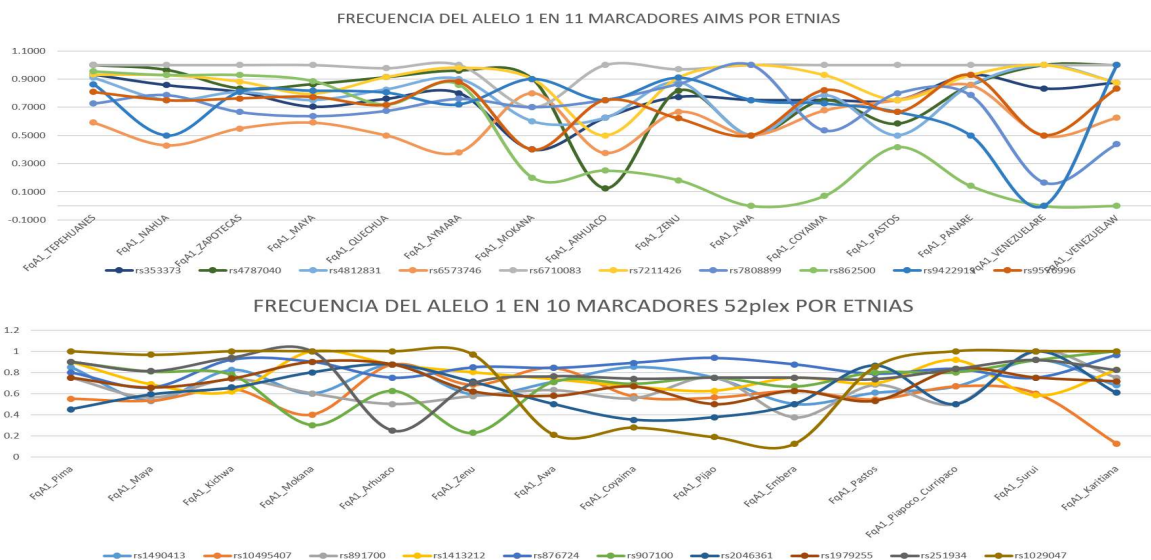


Por otro lado, con ambos sets de marcadores se reconoce un patrón de menor rango de variabilidad de las frecuencias alélicas, en el núcleo formado por las etnias Mesoamericanas y de los Andes, y un rango más amplio en las etnias Colombianas a excepción del caso de los Pastos, lo cual puede estar relacionado con eventos de flujo genético reciente con la población mezclada de las áreas de domicilio, producto de una disminución en las barreras geográficas y culturales, así como con el desplazamiento de las comunidades de sus áreas ancestrales de habitación y/o de subsistencia por efecto de la violencia en las diferentes localidades. Este mismo patrón se observa en el agrupamiento realizado en la construcción de los árboles que se mostrará más adelante. **(Figura 3-15)**. Las frecuencias alélicas por cada marcador para cada una de las etnias se presentan en el **Anexo 4**.

**Figura 3-14** Valores de Porcentaje de Loci Polimórficos por población con AIMS (A) y SNPforID 52plex (B)



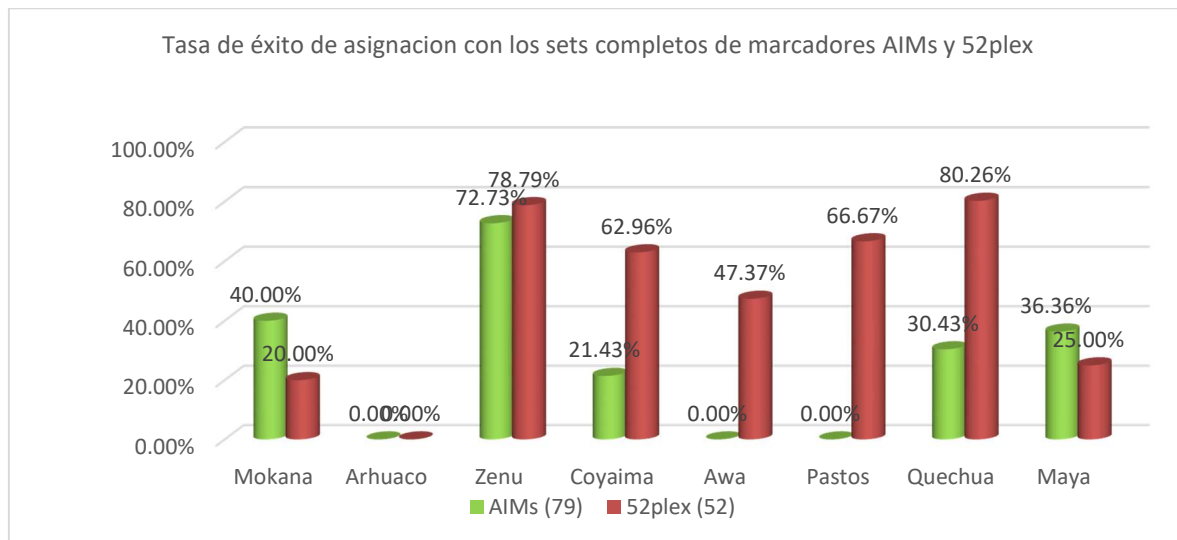
**Figura 3-15** Frecuencia del alelo mayoritario de 11 marcadores AIMS (A) y de 10 marcadores SNPforID 52plex (B) de los subsets seleccionados por diferencia de frecuencias, como candidatos para ser empleados en el análisis de ancestría biogeográfica a nivel subcontinental. Los marcadores que no se muestran presentan la misma tendencia en el comportamiento de los datos.



### 3.3.2 Mejores SNPs para análisis de ancestría nativoamericana

Se evaluó a utilidad de los sets de marcadores completos para clasificar muestras apropiadamente dentro de las etnias nativoamericanas de referencia, mediante la herramienta en línea para validación cruzada de Snipper, descartando las muestras de Wiwa y Wayuu por contar sólo con un individuo. Con el conjunto SNPforID 52plex, la tasa de éxito de asignación fue de 0% para Pijao, Embera, y Arhuaco, mientras para Karitiana, Quechua y Zenú, superó el 75%. Cabe destacar que de los individuos de la étnia Pijao, de los que no se tienen datos de origen de recolección de muestras por pertenecer a los grupos de referencia, el 62.5% fue clasificado como Coyaima, lo cual es concordante con los datos históricos que indican que éstos últimos junto con los Natagaimas, habitaban las riveras del río Magdalena y hacían parte del gran grupo Pijao (Cubillos, 1946, Triana 1990 citado en Pachón & Caillavet, 1996). Para el set de 79 marcadores AIMs, la tasa de éxito de asignación fue de 0% para Arhuaco, Awa, Pastos, Panare y VenezuelaRe y en ningún caso se alcanza el 75%, siendo los valores más altos para Zenú y Aymara con 72.73% y 52% respectivamente. En el caso de las etnias tipificadas con ambos tipos de marcadores, se observa que la tasa de éxito es mayor con SNPforID 52plex para Awa, Pastos, Coyaima, Quechua y Zenú, mientras es mayor con AIMs para Maya y Mokaná. (Figura 3-16).

**Figura 3-16** Comparación del éxito de asignación con los 79 marcadores AIMs con respecto a los 52 marcadores SNPforID 52plex, para las etnias Mokaná, Arhuaco, Zenú, Coyaima, Awa, Pastos, Quechua y Maya.

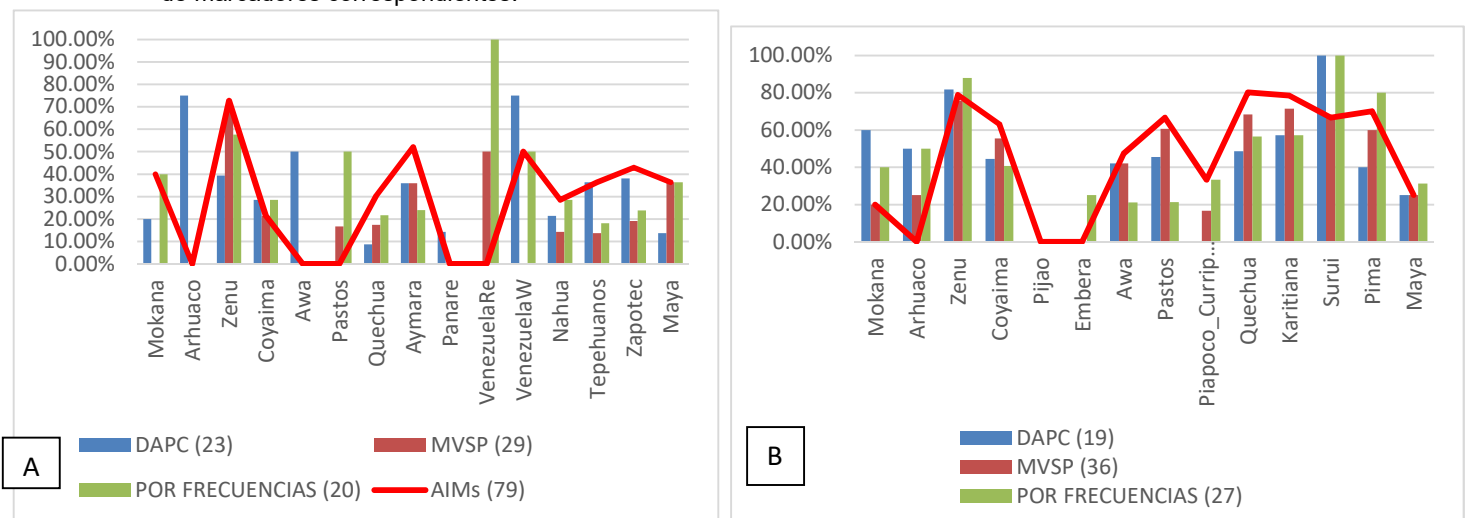




Dados estos resultados, se procedió a construir un listado de los posibles candidatos a mejores marcadores para el análisis de poblaciones nativoamericanas, empleando la función “Contribution” durante la realización del DAPC y el gráfico biplot del ACP en MVSP, eligiendo aquellos que con cada tipo de herramienta fueran más frecuentes en los distintos análisis realizados y un tercer conjunto conformado por los marcadores cuyas frecuencias alélicas diferenciales favorecen la identificación de alguno de los grupos de interés, de acuerdo con las recomendaciones de Phillips, 2013 y 2015 para la selección de marcadores para análisis de ancestría biogeográfica (**Tabla 3-1**). En los tres casos se empleó la opción de validación cruzada de la herramienta en línea Snipper, para evaluar el éxito de la clasificación.

La tasa de éxito de asignación usando el set AIMS, mejora con los marcadores seleccionados por DAPC para las etnias Arhuaco, Awa y VenezuelaW, y con los marcadores seleccionados por diferencia de frecuencias para las etnias Pastos y VenezuelaRe. El mismo efecto se obtiene usando los marcadores del set SNP*for*ID 52plex, seleccionados por DAPC para las etnias Mokane, Arhuaco y Surui, en este último caso, alcanzando el 100% y con los marcadores seleccionados por diferencia de frecuencias para las etnias Embera y Pima. Para Zenú, Coyaima, Nahua, Tepehuanos, Zapotecas y Mayas, la eficiencia es similar al usar los 79 marcadores AIMS y para las etnias Zenú, Coyaima, Awa, Pastos, Piapoco-Curripaco y Maya al usar los 52 marcadores del set SNP*for*ID 52plex. (**Figura 3-17**).

**Figura 3-17** Tasas de éxito de asignación por etnias, comparando los resultados obtenidos usando el set completo de marcadores con respecto a los subsets seleccionados por DAPC, MVSP y por diferencia de frecuencias, tanto con los AIMS (A), como con el set SNP*for*ID 52plex (B). Entre paréntesis se indica el número de marcadores correspondientes.



**Tabla 3-1.** Marcadores seleccionados con DAPC, MVSP y por diferencia de frecuencias para evaluar el éxito de asignación con AIMS, SNPforID 52plex y la unión de los dos sets.

DAPC			MVSP			Por Frecuencias		
AIMs	SNPforID 52plex	AIMs Y SNPforID 52plex	AIMs	SNPforID 52plex	AIMs Y SNPforID 52plex	AIMs	SNPforID 52plex	AIMs Y SNPforID 52plex
rs10751867	rs1015250	rs2251244	rs10145908	rs1005533	rs10145908	rs1244379	rs1490413	rs10175357
rs10946113	rs1029047	rs353373	rs10175357	rs1015250	rs10175357	rs16851773	rs10495407	rs11960137
rs11960137	rs10495407	rs4787040	rs11743823	rs1028528	rs1244379	rs17157450	rs891700	rs16851773
rs1244379	rs1454361	rs4812831	rs12380601	rs1029047	rs132663	rs1790740	rs1413212	rs1809119
rs16851773	rs1528460	rs6768750	rs1244379	rs1335873	rs1453858	rs1809119	rs876724	rs2187239
rs17150066	rs1886510	rs7664076	rs132663	rs1355366	rs17157450	rs2251244	rs907100	rs2251244
rs2187239	rs2046361	rs2040411	rs1396883	rs1357617	rs2304580	rs2304580	rs2046361	rs2304580
rs2251244	rs2076848	rs735155	rs1453858	rs1454361	rs2384982	rs2384982	rs1979255	rs314580
rs2585897	rs2107612	rs740910	rs17157450	rs1463729	rs2736340	rs305163	rs251934	rs353373
rs2736340	rs251934		rs1790740	rs1490413	rs298259	rs314580	rs1029047	rs4787040
rs298259	rs2830795		rs2304580	rs1493232	rs4787040	rs353373	rs727811	rs6573746
rs305163	rs2831700		rs2384982	rs1528460	rs485645	rs4787040	rs735155	rs6684063
rs4787040	rs354439		rs2736340	rs1886510	rs7211426	rs4812831	rs901398	rs6768750
rs4812381	rs735155		rs298259	rs2016276	rs7630522	rs6573746	rs2076848	rs7211426
rs4812831	rs740910		rs314580	rs2040411	rs878874	rs6710083	rs1335873	rs9576996
rs6684063	rs763869		rs4402506	rs2046361	rs9422913	rs7211426	rs354439	rs1024116
rs6711746	rs826472		rs4787040	rs2076848	rs1005533	rs7808899	rs2016276	rs10495407
rs6768750	rs891700		rs4812381	rs2111980	rs1015250	rs862500	rs8037429	rs1454361
rs7164838	rs914165		rs4812831	rs251934	rs1355366	rs9422913	rs1528460	rs1490413
rs7211426			rs485645	rs2831700	rs1454361	rs9576996	rs729172	rs1528460
rs7665516			rs6803290	rs354439	rs1463729		rs1382387	rs1979255
rs8041147			rs7124676	rs719366	rs1490413		rs740910	rs2040411
rs862500			rs7211426	rs722098	rs1886510		rs1493232	rs251934
			rs7630522	rs733164	rs2046361		rs1024116	rs2830795
			rs7664076	rs735155	rs2111980		rs719366	rs2831700
			rs862500	rs740910	rs251934		rs722098	rs354439
			rs878874	rs763869	rs733164		rs2831700	rs729172
			rs9422913	rs8037429	rs735155			rs735155
			rs9533623	rs826472	rs740910			rs740910
				rs891700	rs763869			rs763869
				rs901398	rs8037429			rs901398
				rs907100	rs826472			rs907100
				rs914165	rs891700			rs914165
				rs917118	rs901398			rs964681
				rs938283	rs914165			
				rs964681	rs917118			

El mismo análisis fue realizado agrupando las poblaciones en diferentes niveles. Como se observa en la **Tabla 3-2**, considerando el nivel subcontinental, se observa una mayor eficacia en la asignación para las muestras pertenecientes a la región mesoamericana usando cualquiera de las selecciones de AIMS o de SNPforID 52plex, aunque para este último caso se obtiene un 96% con el conjunto seleccionado a partir de la diferencia de frecuencias.

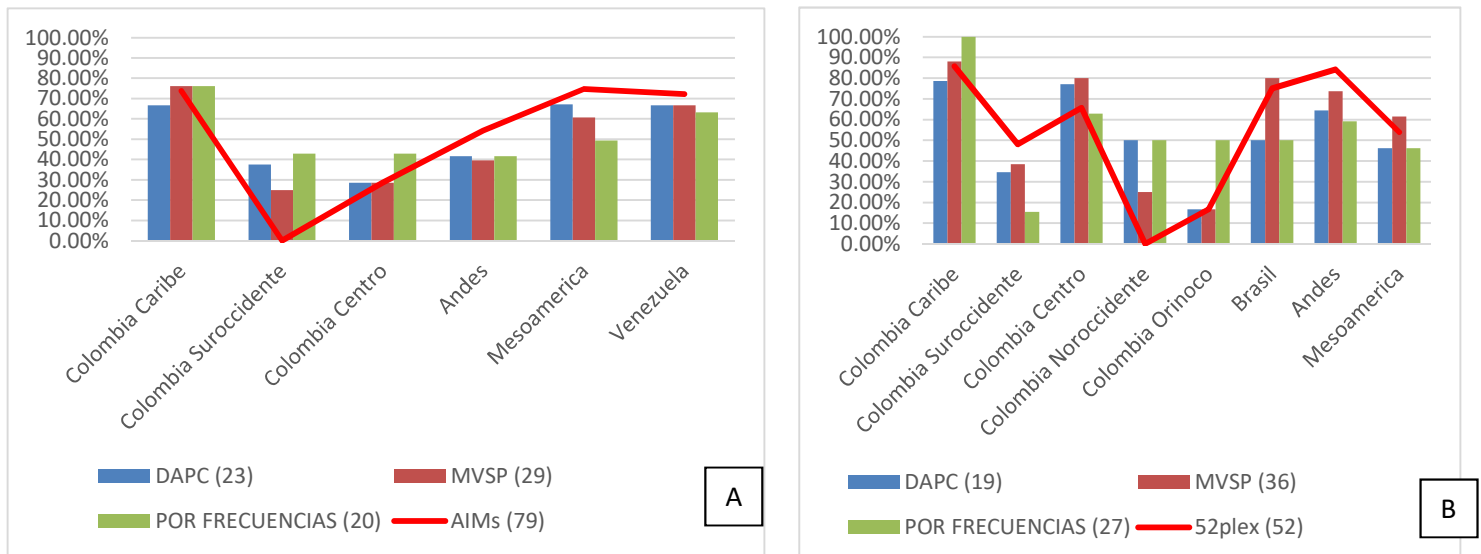
**Tabla 3-2** Comparación de los resultados de éxito de asignación agrupando las étnias en Suramérica (Mokana, Arhuaco, Zenu, Coyaima, Awa, Pastos, Quechua, Aymara, Panare, VenezuelaRe, VenezuelaW, Pijao, Embera, Piapoco\_Curripaco, Karitiana y Surui) y Mesoamérica (Nahua, Tepehuanes, Zapoteca, Maya y Pima). Entre paréntesis se indica el número de marcadores.

<b>AIMs</b>	<b>DAPC (23)</b>	<b>MVSP (29)</b>	<b>POR FRECUENCIAS (20)</b>	<b>AIMs (79)</b>
SURAMERICA	73.85%	75.38%	72.31%	76.15%
MESOAMERICA	82.28%	82.28%	82.28%	77.22%
<b>SNPforID 52plex</b>	<b>DAPC (19)</b>	<b>MVSP (36)</b>	<b>POR FRECUENCIAS (27)</b>	<b>SNPforID 52plex (52)</b>
SURAMERICA	76.60%	83.40%	50.21%	88.51%
MESOAMERICA	76.92%	73.08%	<b>96.15%</b>	69.23%

En otros niveles de análisis, ninguno de los conjuntos de marcadores seleccionados mejoró sustancialmente el éxito de asignación de los sets completos al agrupar las muestras en Colombia, Andes, Mesoamérica y Venezuela ó Brasil. Sin embargo, al agrupar las etnias Colombianas por regiones: Caribe Continental (Mokaná, Arhuaco, Zenú), Suroccidente (Pastos, Awa), Centro (Coyaima, Pijao), Noroccidente (Embera) y Orinoco(Piapoco-Curripaco), el conjunto de AIMS seleccionados por diferencia de frecuencias mejoró en un 14% el éxito de asignación para la etnia Coyaima de la zona centro colombiana. En el caso de las etnias del sur occidente colombiano, para las que este valor es del 0% usando los 79 marcadores de ancestría, con el uso de los marcadores seleccionados por DAPC, MVSP y por diferencia de frecuencias, se obtuvo un 37.5%, 25% y 42.9% de éxito de asignación. Por otro lado, con los marcadores del panel SNPforID 52plex seleccionados con DAPC y por diferencia de frecuencias, la etnia del noroccidente colombiano (Embera) obtuvo un

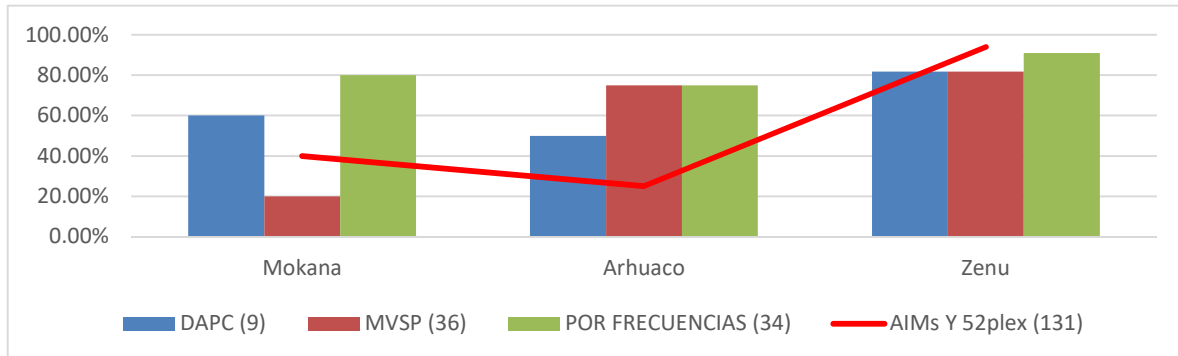
50% de éxito de asignación, respecto al 0% usando los 52 SNPs, y en las etnias del Orinoco (Piapoco-Curripaco) se triplicó la eficacia usando el set de 27 mejores marcadores elegidos por diferencia de frecuencias, mismo que permitió obtener un 100% de éxito de asignación para las etnias de la región Caribe Continental Colombiana. (Figura 3-18)

**Figura 3-18** Comparación de la tasa de éxito de asignación por regiones, usando el set completo de marcadores con respecto a los subsets elegidos con base en los resultados de mejores SNPs por DAPC, datos biplot de MVSP y por diferencia de frecuencias alélicas poblacionales. (A). Marcadores AIMs (B). Marcadores del panel SNPforID 52plex



El uso de los 131 marcadores en las muestras de las etnias Mokaná, Arhuaco y Zenú, de la zona Caribe Continental Colombiana, arrojó una tasa de éxito de asignación de 40%, 25% y 93.94% respectivamente. El set de 34 marcadores seleccionados por diferencia de frecuencias alélicas duplicó y triplicó la eficacia para las etnias Mokaná y Arhuaco y se mantuvo para la etnia Zenú. (Figura 3-19)

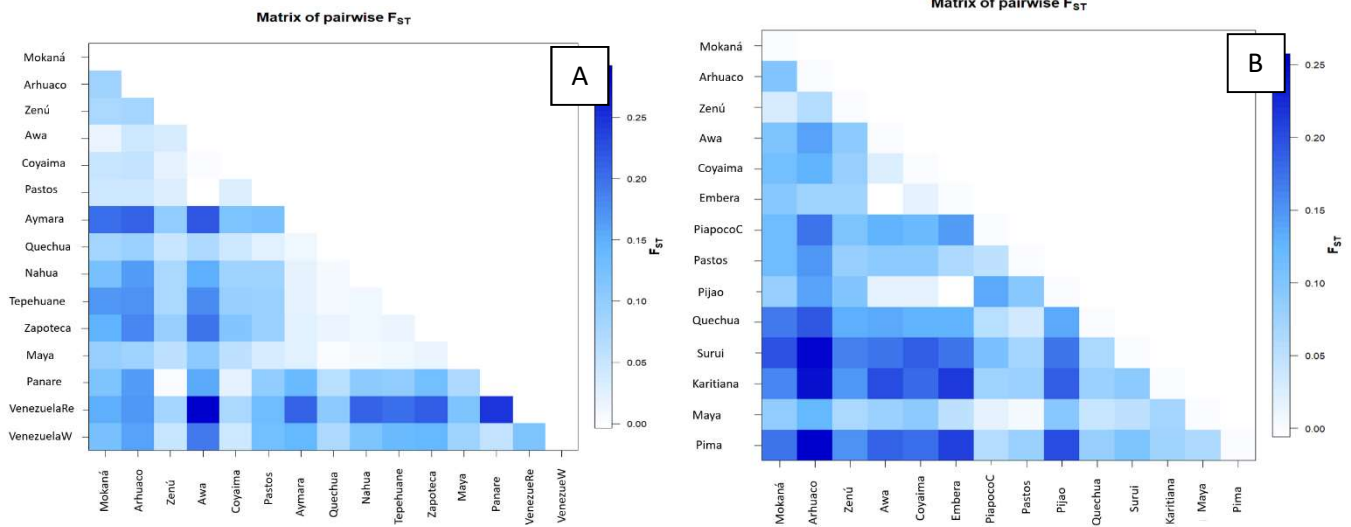
**Figura 3-19** Comparación de la tasa de éxito de asignación para las étnias del Caribe Continental Colombiano usando la combinación del set de 79 AIMs y el SNPforID 52plex, respecto a los subsets elegidos con base en los resultados de mejores SNPs por DAPC, datos biplot de MVSP y por diferencia de frecuencias alélicas poblacionales.



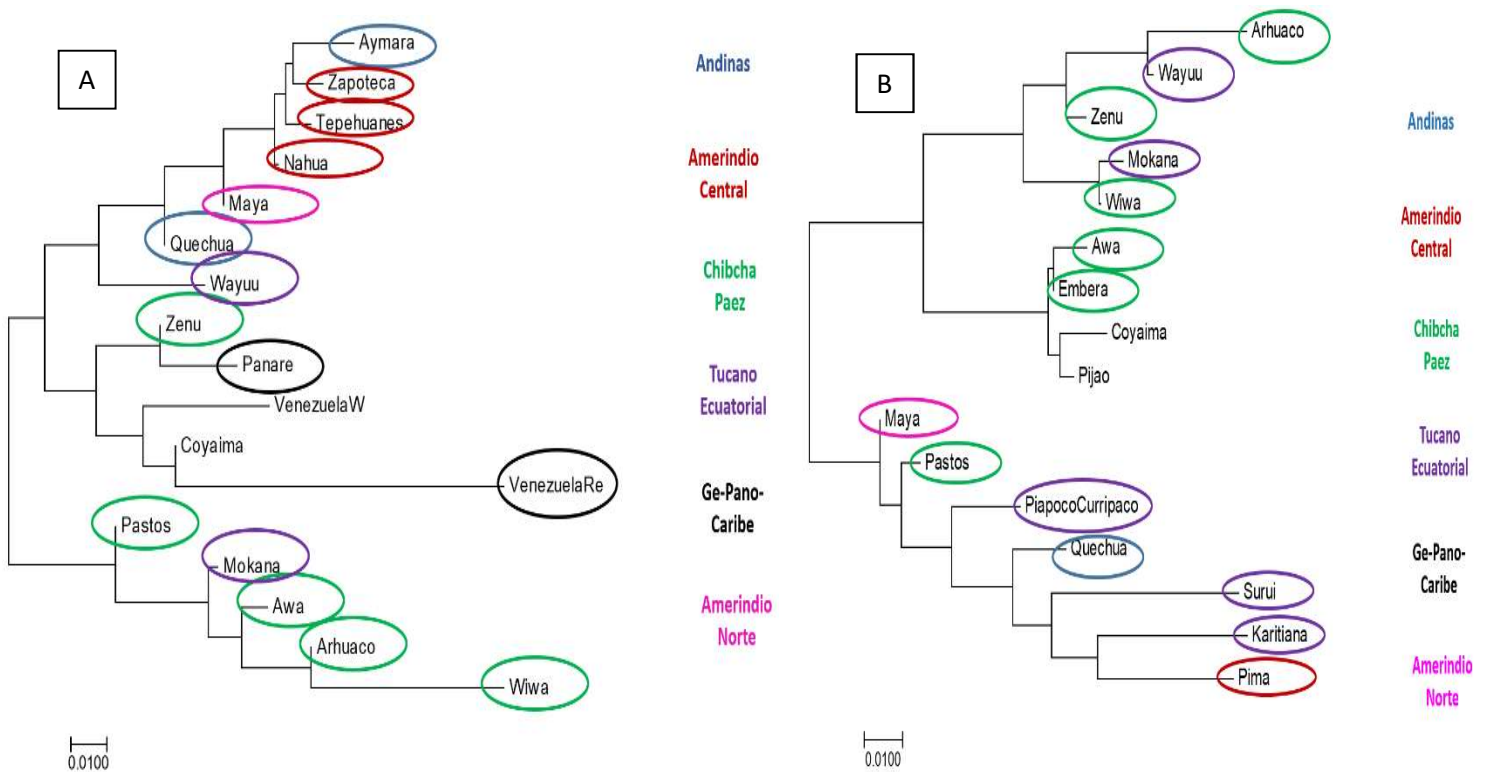
### 3.3.4 Relación genética

Las matrices de  $F_{st}$  pareados obtenidos con Arlequín (**Figura 3-20**) fueron usadas para llevar a cabo el análisis de la relación genética entre las etnias, mediante la construcción de un árbol con la metodología Neighbor Joining. Los resultados con ambos sets de marcadores coinciden con los obtenidos a través del DAPC (Figura 3-2 y 3-3). En ambos casos se identifican tres clusters: con el set 79AIMs, uno formado por las etnias Aymara, Zapoteca, Tepehuane, Nahua, Maya, Quechua y Wayuu, el segundo reuniendo Zenú, Panare, VenezuelaW, Coyaima y VenezuelaRe, y un tercero integrado por Pastos, Mokane, Awa, Arhuaco, Wiwa. Con el SNPforID 52plex se reúnen en un primer grupo las etnias del caribe continental colombiano, en el segundo Awa, Embera, Coyaima y Pijao y un tercer cluster conformado por Maya, Pastos, Piapoco-Curripaco, Quechua, Surui, Karitiana y Pima. (**Figura 3-21**)

**Figura 3-20** Matrices de  $F_{ST}$  pareados por etnias usando los sets completos de marcadores. (A) Para los marcadores AIMs, Aymara y Tepehuane muestran valores de  $p$  significativos para 15 poblaciones. Para el set SNPforID 52plex (B). Con este último conjunto se observa mayor diferenciación entre las poblaciones, con Pastos y Quechua mostrando valores de  $p$  significativos para 15 poblaciones.



**Figura 3-21** Arboles Neighbor Joining construidos a partir de la matriz de  $F_{ST}$  pairwise para etnias de Caribe Continental Colombiano y poblaciones de referencia, usando 79 AIMs (A) y SNPforID 52plex (B). Se indica la filiación lingüística de cada una según Greenberg (Greenberg, 1987; Ruhlen 1991).



Para realizar comparaciones, árboles adicionales con la misma metodología fueron contruidos a partir de distancias de Cavalli Sforza y Edwards y Distancia Estándar de Nei. Para el set AIMs, con las tres metodologías se soporta la relación entre: a. Panare, VenezuelaRe y VenezuelaW, b. entre las etnias Mesoamericanas y Andinas (Quechua y Aymara) formando un cluster separado, c. entre las etnias colombianas Pastos, Awa, Arhuaco y Wiwa, mientras la relación de Mokaná, Zenú y Coyaima muestra ligeras variaciones según la distancia genética empleada. Además, al comparar los resultados de STRUCTURE con  $K=2$ , se observa que las muestras de Coyaima se diferencian en estructura de las etnias venezolanas, que de acuerdo con los árboles Neighbor Joining contruidos a partir de  $F_{st}$  pareados y de distancia Estándar de Nei, parecen relacionarse cercanamente.

Por otro lado, el análisis con los marcadores SNPforID 52plex agrupa las etnias estudiadas en forma similar por las tres metodologías, manteniendo la relación entre Zenú, Mokaná, Arhuaco y, Wiwa y Wayuu en un cluster y Maya, Pastos, Quechua, Piapoco-Curripaco, Karitiana, Surui y Pima en otro. En este último caso, cabe resaltar que Wang y colaboradores (2007) con distancia de Nei utilizando marcadores tipo STRs, presentan una relación cercana entre las etnias Surui, Karitiana y Piapoco, y entre las etnias Zenú, Wayuu y Arhuaco. De la misma forma, el análisis de STRUCTURE apoya estos resultados. La relación establecida entre Wayuu y Zenú con los marcadores AIMs en el árbol Neighbor Joining, aunque sólo se analizó un individuo Wayuu, es congruente con el análisis de Wang, et.al. (2007) y con el de Reich et.al. (2012), donde esta etnia se asocia a los clusters formados por las poblaciones de origen lingüístico Chibcha y con el de He, et.al. (2012), quien estimó un tiempo de divergencia correspondiente a 270 generaciones entre estos dos grupos, el menor entre Zenú y las otras etnias colombianas estudiadas. Como lo indica Wang et. al. (2007), con frecuencia se observa que algunos grupos nativoamericanos forman parte de los clusters a los que pertenecen poblaciones geográficamente cercanas más que a aquellas con quienes comparten filiación lingüística.

La etnia Mokaná habitante del departamento del Atlántico en la región Caribe Continental de Colombia, no ha sido estudiada en ninguna publicación desde el punto de vista genético poblacional. Respecto a su origen, las crónicas españolas indican que venían de la zona venezolana de Maracapaná y Caracas, habitada en aquella época por indígenas caribes

según reportes etnográficos (Arango & Sanchez, 2006), aunque su filiación lingüística se ha determinado como Arawak, sin que exista evidencia consistente de ello (Baquero & de la Hoz, 2010). A pesar de contar con un bajo número de muestras Mokaná, los dos tipos de marcadores con distancias Fst pareados, relacionan esta etnia más cercanamente con las poblaciones de origen lingüístico Chibcha. Aunque para el caso del set SNPforID 52plex se cuenta con las poblaciones Surui, Karitiana y Piapoco Curripaco cuya lengua (Tupi para los dos primeros y Arawak para el último) hacen parte del gran grupo Tucano Ecuatorial (Greenberg, 1987 y Ruhlen, 1991), y considerando que el trabajo de Wang et.al. (2007) asoció estas etnias en el mismo cluster y mostró valores altos de soporte bootstrap, una ampliación del tamaño de muestra para evaluar esta hipótesis sería pertinente, teniendo en cuenta que los resultados de varios estudios no han mostrado una relación consistente entre datos genéticos y lingüísticos. Cabe anotar que con marcadores AIMs usando distancia de Cavalli-Sforza y Edwards, Mokaná y Wayuu forman un cluster, por lo que para esclarecer esta circunstancia es apropiado aumentar el tamaño de muestra o emplear una estrategia genome-wide, ya que puede presentarse el mismo fenómeno descrito para las poblaciones Wayuú anteriormente.

En el presente trabajo, la etnia Embera se agrupa junto con las etnias Awa, Coyaima y Pijao, siendo estas dos últimas cercanamente relacionadas por su origen histórico como ya se comentó. La misma cercana relación entre Emberas y Pijao fue establecida por Criollo (2012), usando un set de 100 marcadores SNP-AIMs. De acuerdo con datos de narraciones de cronistas españoles, los Pijao se establecieron en la región del Tolima Colombiano tras haber sido desplazados por los grupos nativos del norte del Chocó, zona geográfica de habitación actual de algunos grupos Embera. (Cubillos, 1946). Igualmente, de acuerdo con datos históricos, los indígenas Awa resistieron la conquista del imperio Inca y se relacionaron más cercanamente con los grupos de la costa pacífica (Riaño, 2004), lo cual eventualmente podría soportar la relación encontrada en esta investigación.

### **3.3.5 Análisis en el marco del poblamiento Suramericano**

En concordancia con el planteamiento de Cavalli-Sforza et.al., (1994) acerca de que los descendientes de las poblaciones Andinas Nativoamericanas podrían estar cercanamente



relacionadas con los primeros migrantes desplazándose hacia el sur por la parte occidental de la cordillera de los Andes como continuación de la primera oleada de poblamiento, los datos obtenidos con los sets de marcadores 79 AIMs y SNPforID 52plex, donde se relacionan las poblaciones Mesoamericanas y Andinas y se observa menor diversidad en las poblaciones Amazónicas Karitiana, Surui y las poblaciones del sur de Venezuela, representadas en el Porcentaje de Loci Polimórficos analizados. Correspondiendo con lo obtenido por Wang et.al., 2007, quien comenta que tales resultados también pueden ser consecuencia de mezcla reciente y la diferenciación entre poblaciones debido a grados variables de deriva y flujo genético. Estos datos son posteriormente analizados por Reich et.al., 2012 quien presenta un modelo que muestra pocos eventos de mezcla para los dos grupos, pero bajo efecto de deriva genética en las poblaciones Mesoamericanas y alto para Suruí, población que representa las etnias amazónicas.

El modelo de poblamiento por vía costera apoyado por los resultados de Wang et. al., 2007 y retomado por Rothhammer y Dillehay 2009, propone en el contexto del poblamiento Suramericano, que un vez llegados los migrantes a la zona del istmo de Panamá, algunos continuaron hacia el sur del continente vía Costa Pacífica o vía cordillera de los Andes, siguiendo inicialmente el cauce de los ríos Cauca y Magdalena en Colombia, mientras otros grupos se desplazaron hacia el nororiente por la Costa Caribe, hasta alcanzar Guyana y Brasil. En este último trayecto algunos migrantes se movieron en dirección sur en Venezuela, siguiendo posteriormente las riveras de los ríos de la región Amazónica y de la Orinoquía. Según López y Cano (2011), debido al descenso de las aguas del mar durante el periodo del pleniglacial, la Costa Caribe contaba con una mayor área y los valles interandinos estaban constituidos por vegetación baja que facilitaba el desplazamiento de los grupos humanos que se dirigían hacia el extremo oriental de Suramérica en el primer caso y al interior de Colombia por las riveras de los ríos en el segundo. A pesar de que la Costa Caribe Continental Colombiana se constituyó como paso obligado desde el principio del poblamiento suramericano, las evidencias más antiguas de yacimientos arqueológicos en la zona noroccidental se han encontrado en Colombia en la región andina (Pubenza y el Abra), andes suroccidentales (La Elvira) y en la región de la Amazonía (Peña Roja) y el Orinoco (Pozo Azul), y en Venezuela en los yacimientos Taima-Taima y El Jobo, con dataciones que oscilan entre 7000 y 16000 aap (años antes del presente), pero las primera evidencias de vida sedentaria aparentemente relacionadas con el uso de la cerámica de la

zona Caribe Colombiana, se encontraron en las localidades de San Jacinto y Puerto Hormiga hace alrededor de 5000 años, sin que sea claro el motivo por el cual no se formaron asentamientos o no se encuentran evidencias de ello. (Dillehay, 2008). Registros históricos muestran que al momento de la colonización europea la región costera noroccidental del continente (región Caribe Continental de Colombia) se encontraba poblada por etnias de filiación lingüística (según la clasificación lingüística de Greenberg 1.987) Chibcha-Paez, Ge-Pano-Caribe y Tucano-Ecuatorial. En actualidad, el primer grupo está constituido por los pueblos Chibcha hablantes Tayrona (Kogi, Wiwa, Arhuaco y Kankuamo), Chimila, Zenu, y Kuna, el segundo por el pueblo de lengua Caribe Yuko y el tercero por los pueblos de lengua Arawak Wayuu y Mokaná. Varios autores sugieren migraciones tardías de grupos de lengua Caribe y Arawak que llegaron a territorio Colombiano entrando por el oriente estableciendo relaciones con grupos Chibchas ya asentados en diferentes regiones, lo que inició un procesos de diferenciación del lenguaje y de divergencia genética por efecto de deriva y posterior aislamiento para algunos grupos o flujo genético con los pobladores vecinos para otros (Chaves, 1979; Rodriguez, 2007, Lleras, 1995), y podría explicar los resultados acerca de una posible relación genética más cercana entre Mokaná, Wayuú y poblaciones Chibchas, que de las primeras con otras etnias del tronco lingüístico Tucano-Ecuatorial.

### 3.3 Conclusiones

En la búsqueda de estrategias cada vez más eficientes para el análisis de la estructura de las poblaciones mezcladas, debe considerarse cómo eventos históricos antiguos y recientes moldean la forma como las poblaciones cambian en el tiempo, por lo que los grupos de investigación desarrollan estrategias de trabajo aplicables a diferentes niveles de agrupamiento de poblaciones, en especial a nivel continental. Sin embargo, dada la influencia de los cambios climáticos, geográficos, culturales e incluso de los desarrollos tecnológicos que modifican la forma en la que las poblaciones se adaptan al ambiente y con ello las dinámicas poblacionales, hacen que, a nivel subcontinental, se presenten fenómenos de divergencia aumentando la variación genética.

En este contexto, uno de los intereses actuales se relaciona con los aportes ancestrales diferenciales en poblaciones con historia de mezcla, con aplicaciones en el campo de la medicina, de las ciencias forenses y de la antropología. Este trabajo se desarrolló con el objetivo de analizar la ancestría nativoamericana de poblaciones del Noroccidente Suramericano, usando dos paneles de SNPs que se han empleado fundamentalmente para el análisis de la ancestría biogeográfica (Set de 79 marcadores AIMs contenidos en el Panel del consorcio LACE) y para la identificación humana (SNPforID 52plex). De acuerdo con Galanter, et.al., 2012, el uso del set de 446 marcadores de ancestría AIMs desarrollado para el consorcio LACE (Galanter et.al., 2012), aunque es de gran eficiencia para la estima de proporción de mezcla reciente en Latinoamérica, tiene un uso limitado a niveles más específicos, lo que es concordante con los resultados de este estudio realizado con un subset de 79 marcadores de dicho panel. Sin embargo, el análisis realizado en esta investigación, mostró resultados consistentes con varias estrategias de análisis con respecto a algunas ventajas del set SNPforID 52plex para estas aplicaciones a nivel subcontinental, e incluso permitió proponer un punto de partida para el diseño de un panel que combina marcadores de ambos sets, con valores de éxito de clasificación a nivel regional alrededor del 80%, que pueden ser mejorables con la adición de otros SNPs suficientemente informativos, usando como apoyo para su selección las herramientas útiles del ACP, del DAPC y la selección de marcadores con alta variación de frecuencias alélicas entre las poblaciones de interés, y evaluando su eficacia con estrategias de análisis bayesiano como el módulo de validación cruzada de Snipper.

Con este enfoque los resultados demuestran la existencia de un nivel subestructura a nivel de las poblaciones nativoamericanas, que usualmente son consideradas homogéneas en los análisis a nivel continental, e incluso puede notarse una posible disminución en la diversidad genética en cuanto al porcentaje de Loci Polimórficos realizados en una tendencia de occidente a oriente previamente registrada por otros autores. Los análisis del patrón de variación de las frecuencias alélicas y la formación de los clusters tras la construcción de los árboles genéticos, son consistentes con observaciones previas de otros autores en lo referente a las relaciones entre poblaciones Mesoamericanas y Andinas, o entre las etnias Zenú, Arhuaco y Wayuu, y de las etnias de la región Amazónica y de la Orinoquía, en el marco de la filiación lingüística. Se destaca el caso de la etnia Mokana, no estudiada previamente por marcadores genéticos, cuya pertenencia aparente al tronco lingüístico Arawak, pero relacionándose más cercanamente con poblaciones chibchas, permiten proponer hipótesis relacionadas con eventos de migración, interacción con las comunidades locales y flujo genético en diferentes periodos de tiempo, e incluso tiempo desde la mezcla, lo que está de acuerdo con datos históricos y que puede también evidenciarse a través de las encuestas que acompañan el proceso de muestreo, que soportan datos genealógicos y demográficos del participante, lo que constituye un punto a favor en la reconsideración del concepto de “aislados genéticos”.

## **4. Capítulo 4. Diversidad genética y ancestría de los linajes masculinos en el noroccidente de Suramérica**

### **4.1 Introducción**

Estudios acerca de origen de los primeros pobladores del continente americano han hecho uso de diferentes clases de marcadores genéticos, con el fin de contrastar las hipótesis planteadas desde el campo antropológico y lingüístico. Diversas investigaciones han asociado los linajes masculinos de los ancestros nativoamericanos, a poblaciones Asiáticas, especialmente de la región Centro-Sur y Suroriente de Siberia (Santos et.al., 1999), (Lell et.al., 2002), (Bortolini et.al., 2003), (Pitblado, 2011), (Dulik, et.al., 2012), (Regueiro, et.al., 2013), describiendo polimorfismos específicos (Underhill, 1996) y alelos privados (Shroeder et.al., 2007), y contrastando las hipótesis sobre el número de oleadas migratorias, el tiempo de la migración inicial y el número de pobladores (Shurr, 2004), (Rothhammer & Dillehay, 2009), (Roewer et.al., 2013), (Bravi, 2013), (Bataglia, et.al., 2013). La mezcla con linajes africanos y europeos tras la colonización del nuevo mundo, y la disminución de la población nativa se reflejan en los análisis de la estructura genética en poblaciones americanas tanto conservadas como mezcladas, que evidencian además la ocurrencia de eventos demográficos y migratorios posteriores (Blanco-Verea, et.al., 2010), (Ruiz-Narvaez, et.al., 2005), (Stefflova, et.al., 2011), mediados por aspectos culturales, comerciales e incluso relacionados con guerras, que dirigen los patrones complejos de deriva y flujo genético que influyen en la composición de la población (Tarazona-Santos et.al., 2001).

Aunque el norte de Suramérica, fue la zona receptora de las migraciones iniciales nativoamericanas y punto de dispersión inicial para el poblamiento del resto del continente, los diferentes eventos históricos dejaron como consecuencia poblaciones urbanas actuales con linajes masculinos fundamentalmente europeos y poblaciones rurales con una frecuencia de linajes nativoamericanos un poco mayor y adicionalmente, aportes recientes por migraciones intercontinentales, de tal forma que aunque el origen de las poblaciones actuales es triétnico, un importante porcentaje de otros haplogrupos del antiguo continente como G, I, J, L y T, integran la variabilidad actual en Centro y Suramérica (Lovo-Gomez, et.al., 2007), (Martínez, et.al., 2012), (Grugni, et.al., 2015), (Resque, et.al., 2016).

Estudios sobre los linajes masculinos en la población colombiana, muestran una amplia variabilidad que en ocasiones ha sido ligada a la división político administrativa del país (Acosta, et.al., 2009), (Rojas, et.al., 2010), (Criollo, 2012), (Noguera, et.al., 2014). Sin embargo, mecanismos de muestreo enfocados en datos genealógicos y demográficos, permiten explorar en poblaciones mezcladas los linajes ancestrales y realizar análisis enfocados en la historia de las regiones, observando contrastes con reportes previos.

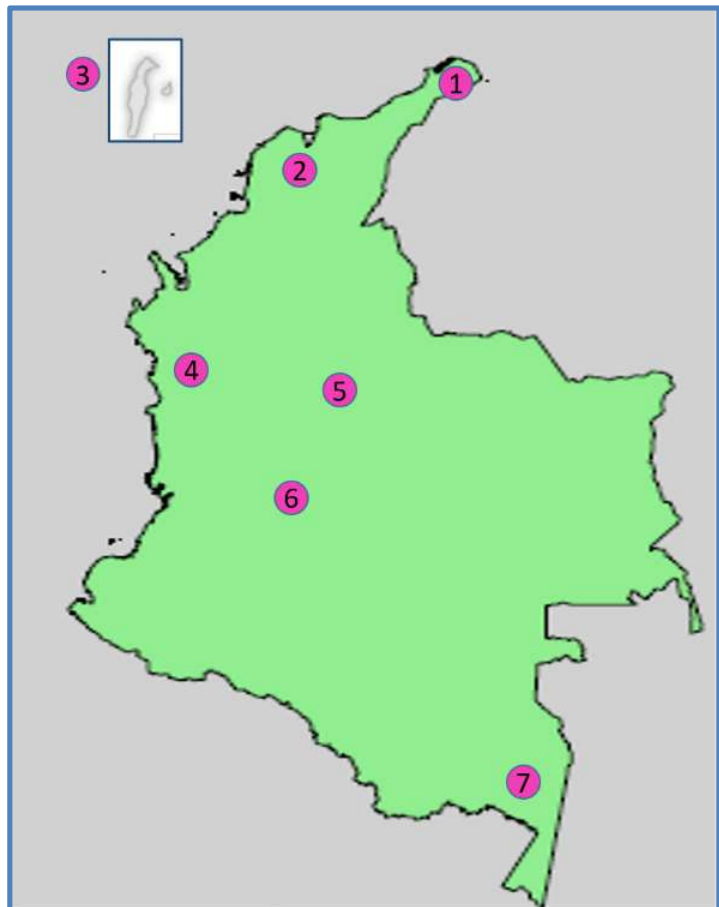
Se analizaron 17 sistemas STR de cromosoma Y en 403 muestras recolectadas en diferentes regiones del territorio Colombiano, clasificadas de acuerdo a los datos genealógicos de la encuesta aplicada. Los haplotipos en su mayoría fueron empleados para realizar la predicción del haplogrupo por métodos bayesianos de Athey, 2006. Los resultados de diversidad y relaciones genéticas con otras poblaciones, son discutidos en contextos históricos y de poblamiento.

## 4.2 Materiales y métodos

### 4.2.1. Poblaciones y encuestas

Las muestras correspondientes a las poblaciones colombianas empleadas en este estudio, han sido recolectadas durante los últimos 10 años por los investigadores de Grupo de Genética de Poblaciones e Identificación, del Instituto de Genética de Universidad Nacional de Colombia. **Figura 4-1.** En todos los casos se aplicó una encuesta para obtener información adicional de tipo genealógico y demográfico. 403 muestras masculinas de sangre total y/o de raspado de células epiteliales de cavidad oral, fueron colectadas previa firma del consentimiento informado, de personas mayores de edad residentes al menos por cinco años en la localidad de muestreo. Las muestras de Guajira corresponden a población Wayuu por genealogía. Las muestras del Archipiélago de San Andrés son las mismas reportadas en Alonso et.al., 2013.

**Figura 4-1:** Localización de los puntos de muestreo. 1. La Guajira, 2. Caribe Continental, 3. San Andrés y Providencia, 4. Chocó, 5. Región Cundiboyacense (Cundinamarca y Boyacá), 6. Tolima Grande (Tolima y Huila), 7. Amazonas.



### 4.2.2. Análisis molecular

DNA genómico fue obtenido a partir de las muestras de sangre o de células epiteliales empleando metodologías convencionales de Salting-out, Chelex y/o por el método comercial de extracción FTA de Whatman. La amplificación de 17 sistemas STRs de cromosoma Y (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 (Y GATA C4) y GATA H4), fueron realizadas empleando los kits comerciales AmpFISTR® Yfiler y Yfiler™Plus de Applied Biosystems, siguiendo las instrucciones del fabricante.

### 4.2.3 Análisis genético poblacional y estadístico

Las frecuencias haplotípicas, alélicas y de haplogrupos fueron estimadas por el método de conteo simple, y la diversidad en los tres casos, de acuerdo con Nei (1987) usando la ecuación  $DG = (n * (1 - \sum p_i^2)) / (n - 1)$ . Para calcular las frecuencias de haplogrupos se descartaron 27 muestras porque la probabilidad estaba debajo de 80%. El software Arlequín 3.5.2.2 (Excoffier, et.al. 2010) se usó para realizar el Análisis Molecular de Varianza (AMOVA) y para construir las matrices de Rst pareados a partir del análisis de los haplotipos, y para elaborar las matrices de Fst pareados a partir de los datos de frecuencia de haplogrupo de las poblaciones colombianas y de las poblaciones de referencia (**Tabla 4-1**) empleadas para las comparaciones. El análisis de Escalamiento Multidimensional (Multidimensional scaling MDS) se realizó con la herramienta en línea en la plataforma YHRD para AMOVA y MDS (<https://yhrd.org/amova>) basado en Rst (Willuweit & Roewer, 2015), sin incluir los haplotipos con alelos intermedios, duplicaciones o posibles alelos nulos. Los árboles de poblaciones fueron construidos con el método Neighbor-Joining (Saitou, N. & Nei, M. 1987) con base en la matriz de distancias de Rst pareados de Arlequín (Excoffier, et.al. 2010) usando MEGA 7 (Kumar, et.al., 2016). Para realizar comparaciones también se construyeron árboles con la misma metodología usando Distancia de Cuerda (Cavalli-Sforza & Edwards, 1967) y distancias de Nei (Nei, M. et.al., 1983), en el software Populations 1.2.28 (Langella, O. 1999).



La determinación de haplogrupos se realizó utilizando el Predictor de Haplogrupos (Athey, 2006). Solo se consideraron para los análisis aquellos haplotipos cuya probabilidad de asignación del haplogrupo señalado por el software, fue mayor o igual a 0.8.

El análisis de componentes .principales (ACP) se realizó con el software MVSP 3.22 a partir de las frecuencias alélicas, usando los haplotipos previamente publicados en (Franco & Barreto, 2017; Grugni, et.al., 2015; Purps, et.al., 2014 y Xavier, et.al. 2015), que también fueron empleados para conformar las poblaciones de referencia para el análisis de ancestría con el Software Structure 2.3.4. y los resultados fueron sometidos a evaluación de concordancia con la asignación realizada por el predictor de haplogrupos, empleando el estadístico Kappa.

**Tabla 4-1:** Poblaciones usadas para las comparaciones mediante MDS.

POBLACION	N (Haplotipos)	Referencia	Datos
Angola [African]	70	Obtenidos desde bases de datos de poblaciones YHRD	Afrodescendientes
Long Island, Bahamas [Bahamian]	41	Obtenidos desde bases de datos de poblaciones YHRD	Afrodescendientes
New Providence, Bahamas [Bahamian]	137	Obtenidos desde bases de datos de poblaciones YHRD	Afrodescendientes
Ketou, Benin [Yoruba]	54	Fortes Lima C., Brucato N., Croze M., Bellis G., Schiavinato S., Massougbodji A., Migot Nabias F., Dugoujon J. (2015), 'Genetic population study of Y-chromosome markers in Benin and Ivory Coast ethnic groups', <i>Forensic Science International: Genetics</i> 19, 232–237	Afrodescendientes
Rio de Janeiro, Brazil [African]	100	Domingues PM., Gusmão L., Silva D., Amorim A., Pereira RW., Carvalho D. (2007), 'Sub-Saharan Africa descendents in Rio de Janeiro (Brazil): population and mutational data for 12 Y-STR loci.', <i>Int J Legal Med</i> 121(3), 238-41	Afrodescendientes
Ghana [Ewe]	49	Obtenidos desde bases de datos de poblaciones YHRD	Afrodescendientes
Haiti [Haitian]	118	Obtenidos desde bases de datos de poblaciones YHRD	Afrodescendientes
Jamaica [Jamaican]	203	Obtenidos desde bases de datos de poblaciones YHRD	Afrodescendientes
Webuye, Kenya [Luhya]	42	Obtenidos desde bases de datos de poblaciones YHRD	Afrodescendientes
Ibadan, Nigeria [Yoruba]	112	Obtenidos desde bases de datos de poblaciones YHRD	Afrodescendientes
Kenya Bantu_Luhya, Other]	60	Iacovacci G., D'atanasio E., Marini O., Coppa A., Sellitto D., Trombetta B., Berti A., Cruciani F. (2017), 'Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries', <i>Forensic Science International: Genetics</i> Epub ahead of Print, In press	Afrodescendientes
London, United Kingdom [British Afro Caribbean]	1126	Aliferi A., Thomson J., Mc Donald A., Molin V., Ferguson S., Vanhinsbergh D., Syndercombe D., Ballard D. (2018), 'UK	Afrodescendientes

		and Irish Y-STR population data—A catalogue of variant alleles', <i>For Sci Int Genetics</i> Epub ahead of print, In press	
<b>United States [African American]</b>	1529	Budowle B., Ge J., Aranda XG., Planz JV., Eisenberg AJ., Chakraborty R. (2009), 'Texas population substructure and its impact on estimating the rarity of Y STR haplotypes from DNA evidence*', <i>J Forensic Sci</i> 54(5), 1016-21 Coble MD., Hill CR., Butler JM. (2013), 'Haplotype data for 23 Y-chromosome markers in four U.S. population groups.', <i>Forensic Sci Int Genet</i> 7(3), e66-8	Afrodescendientes
<b>Zimbabwe [Zimbabwean]</b>	54	Obtenidos desde bases de datos de poblaciones YHRD	Afrodescendientes
<b>Buenos Aires, Argentina [Admixed]</b>	562	Obtenidos desde bases de datos de poblaciones YHRD	Población Mezclada
<b>Salta, Argentina [Chiriguano]</b>	14	Toscanini U., Gusmão L., Berardi G., Amorim A., Carracedo A., Salas A., Raimondi E. (2008), 'Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data.', <i>Forensic Sci Int Genet</i> 2(4), 274-80	Nativoamericano
<b>Tucumán, Argentina [Kolla]</b>	14	Toscanini U., Gusmão L., Berardi G., Amorim A., Carracedo A., Salas A., Raimondi E. (2008), 'Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data.', <i>Forensic Sci Int Genet</i> 2(4), 274-80	Nativoamericano
<b>Rio Negro, Argentina [Mapuche]</b>	11	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Chubut, Argentina [Tehuelche]</b>	10	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Nativoamericano Chaco, Argentina [Toba]</b>	46	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>33 Haplotypes from Formosa, Argentina [Wichi]</b>	33	Roewer L., Nothnagel M., Gusmão L., Gomes V., González M., Corach D., Sala A., Alechine E., Palha T., Santos N., Ribeiro Dos Santos A., Geppert M., Willuweit S., Nagy M., Zweynert S., Baeta M., Núñez C., Martínez Jarreta B., González Andrade F., De F., Silva D., Builes JJ., Turbón D., Parra L., Arroyo Pardo E., Toscanini U., Borjas L., Barletta C., Ewart E., Santos S., Krawczak M. (2013), 'Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans.', <i>PLoS Genet</i> 9(4), e1003460	Nativoamericano
<b>Bolivia [Mestizo]</b>	44	Obtenidos desde bases de datos de poblaciones YHRD	Población Mezclada
<b>Beni, Bolivia [Moxenos]</b>	10	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Rio de Janeiro, Brazil [Admixed]</b>	1630	Obtenidos desde bases de datos de poblaciones YHRD	Población Mezclada
<b>Pará, Brazil [Arara]</b>	20	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Pará, Brazil [Asurini]</b>	15	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Rondonia, Brazil [Karitiana]</b>	34	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Rondonia, Brazil [Surui]</b>	24	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Amazonas, Brazil [Ticuna]</b>	23	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Roraima, Brazil [Yanomami]</b>	10	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Santiago de Chile, Chile [Chilean]</b>	186	Obtenidos desde bases de datos de poblaciones YHRD	Población Mezclada
<b>Guajira, Colombia [Mestizo]</b>	15	Romero RE., Briceño I., Rdel L., Willuweit S., Roewer L., Gómez A. (2008), 'A Colombian Caribbean population study of 16 Y-chromosome STR loci.', <i>Forensic Sci Int Genet</i> 2(2), e5-8	Población Mezclada

<b>Costa Rica [Mestizo]</b>	160	Obtenidos desde bases de datos de poblaciones YHRD	Población Mezclada
<b>Ecuador [Ecuadorian]</b>	401	Gaviria A., Sánchez ME., Morejón G., Vela M., Aguirre V., Burgos G., Zambrano AK., Leone PE., Paz Y Mino C. (2013), 'Characterization and Haplotype analysis of 11 Y-STR loci in Ecuadorian population', <i>Forensic Science International: Genetics Supplementary Series</i> 4(1), e310–e311	Población Mezclada
<b>Pastaza, Ecuador [Kichwa]</b>	42	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Cauca/Nariño/Chocó, Colombia [Guambiano,Coconuco,Na sa,Yanacona, Pastos,Embera]</b>	31	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Guatemala [Maya]</b>	93	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Guatemala [Mestizo]</b>	111	Martínez González LJ., Saiz M., Alvarez Cubero MJ., Gómez Martín A., Alvarez JC., Martínez Labarga C., Lorente JA. (2012), 'Distribution of Y chromosomal STRs loci in Mayan and Mestizo populations from Guatemala.', <i>Forensic Sci Int Genet</i> 6(1), 136-42	Población Mezclada
<b>Yucatán, Mexico [Maya]</b>	72	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Guadalajara, Mexico [Mestizo]</b>	239	Salazar Flores J., Dondiego Aldape R., Rubi Castellanos R., Anaya Palafox M., Nuño Arana I., Canseco Avila LM., Flores Flores G., Morales Vallejo ME., Barojas Pérez N., Muñoz Valle JF., Campos Gutiérrez R., Rangel Villalobos H. (2010), 'Population structure and paternal admixture landscape on present-day Mexican-Mestizos revealed by Y-STR haplotypes.', <i>Am J Hum Biol</i> 22(3), 401-9	Población Mezclada
<b>Nicaragua [Mestizo]</b>	148	Nuñez C., Baeta M., Sosa C., Casalod Y., Ge J., Budowle B., Martínez Jarreta B. (2010), 'Reconstructing the population history of Nicaragua by means of mtDNA, Y-chromosome STRs, and autosomal STR markers.', <i>Am J Phys Anthropol</i> 143(4), 591-600	Población Mezclada
<b>Amazonas, Peru [Cajamarca]</b>	21	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Amazonas, Peru [Chachapoya]</b>	106	Guevara EK., Palo JU., Guillén S., Sajantila A. (2016), 'MtDNA and Y-chromosomal diversity in the Chachapoya, a population from the northeast Peruvian Andes-Amazon divide', <i>Am J Hum Biol.</i> 28(6), 857-67	Nativoamericano
<b>Cuzco, Peru [Chumbivilca]</b>	10	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Peru [Mestizo]</b>	218	Iannacone GC., Tito RY., Lopez PW., Medina ME., Lizarraga B. (2005), 'Y-chromosomal haplotypes for the PowerPlex Y for twelve STRs in a Peruvian population sample.', <i>J Forensic Sci</i> 50(1), 239-42	Población Mezclada
<b>Lima, Peru [Quechua]</b>	11	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Central Alaska, United States [ Athabaskan Indian]</b>	141	Davis C., Ge J., Chidambaram A., King J., Turnbough M., Collins M., Dym O., Chakraborty R., Eisenberg AJ., Budowle B. (2011), 'Y-STR loci diversity in native Alaskan populations.', <i>Int J Legal Med</i> 125(4), 559-63	Nativoamericano
<b>Northern Alaska, United States [Inupiat]</b>	145	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Phoenix, United States [Pima]</b>	22	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Western Alaska, United States [Yupik]</b>	139	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Venezuela [Bari Boxi]</b>	16	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano
<b>Venezuela [Wayuu]</b>	17	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericano

<b>Kazakhstan [Kazakh]</b>	305	Zhabagin M., Sarkyrbayeva A., Tazhigulova I., Yerezhpeov D., Li S., Akilzhanov R., Yeralinov A., Sabitov Z., Akilzhanova A. (2018), 'Development of the Kazakhstan Y-chromosome haplotype reference database: analysis of 27 Y-STR in Kazakh population', <i>Int J Legal Med</i> Epub ahead of print, In press	OJOOJO
<b>Central Mongolia, Mongolia [Khalkh]</b>	72	Obtenidos desde bases de datos de poblaciones YHRD	
<b>Northern Mongolia, Mongolia [Khalkh, Darkhad, Uriankhai]</b>	23	Obtenidos desde bases de datos de poblaciones YHRD	
<b>Western Mongolia, Mongolia [Khalkh, Durvud, Bayad, Uriankhai, Khazak, Torguud, Uuld, Zakhchin]</b>	31	Obtenidos desde bases de datos de poblaciones YHRD	
<b>Salzburg, Austria [Austrian]</b>	193	Pickrahn I., Müller E., Zahrer W., Dunkelmann B., Cemper Kiesslich J., Kreindl G., Neuhuber F. (2016), 'Yfiler® Plus amplification kit validation and calculation of forensic parameters for two Austrian populations.', <i>Forensic Science International Genetics</i> 21, 90-94	Europeooriental
<b>Paris, France [French]</b>	89	Roewer L., Croucher P.J., Willuweit S., Lu T.T., Kayser M., Lessig R., Knijff D., Jobling M.A., Tyler Smith C., Krawczak M. (2005), 'Signature of recent historical events in the European Y-chromosomal STR haplotype distribution.', <i>Hum Genet</i> 116(4), 279-91 [PubMed] [DOI]. Ramos Luis E., Blanco Vereá A., Brión M., Huffel V., Sánchez Diz P., Carracedo A. (2014), 'Y-chromosomal DNA analysis in French male lineages.', <i>Forensic Sci Int Genet</i> 9, 162-8 [PubMed] [DOI]	Europeo
<b>Berlin-Brandenburg, Germany [German]</b>	185	Purps J., Siegert S., Willuweit S., Nagy M., Alves C., Salazar R., Angustia S.M., Santos L.H., Anslinger K., Bayer B., Ayub Q., Wei W., Xue Y., Tyler Smith C., Bafalluy M.B., Martínez Jarreta B., Egyed B., Balitzki B., Tschumi S., Ballard D., Court D.S., Barrantes X., Bäßler G., Wiest T., Berger B., Niederstätter H., Parson W., Davis C., Budowle B., Burri H., Borer U., Koller C., Carvalho E.F., Domingues P.M., Chamoun W.T., Coble M.D., Hill C.R., Corach D., Caputo M., D'amato M.E., Davison S., Decorte R., Larmuseau M.H., Ottoni C., Rickards O., Lu D., Jiang C., Dobosz T., Jonkisz A., Frank W.E., Furac I., Gehrig C., Castella V., Grskovic B., Haas C., Wobst J., Hadzic G., Drobnic K., Honda K., Hou Y., Zhou D., Li Y., Hu S., Chen S., Immel U.D., Lessig R., Jakovski Z., Ilievská T., Klann A.E., García C.C., Knijff D., Kraaijenbrink T., Kondili A., Miniati P., Vouropoulou M., Kovacevic L., Marjanovic D., Lindner I., Mansour I., Al Azem M., Andari A.E., Marino M., Furfuro S., Locarno L., Martín P., Luque G.M., Alonso A., Miranda L.S., Moreira H., Mizuno N., Iwashima Y., Neto R.S., Nogueira T.L., Silva R., Nastainczyk Wulf M., Edelmann J., Kohl M., Nie S., Wang X., Cheng B., Núñez C., Pancorbo M.M., Olofsson J.K., Morling N., Onofri V., Tagliabracci A., Pamjav H., Volgyi A., Barany G., Pawlowski R., Maciejewska A., Pelotti S., Pepinski W., Abreu Glowacka M., Phillips C., Cárdenas J., Rey Gonzalez D., Salas A., Brisighelli F., Capelli C., Toscanini U., Piccinini A., Pignionica M., Baldassarra S.L., Ploski R., Konarzewska M., Jastrzebska E., Robino C., Sajantila A., Palo J.U., Guevara E., Salvador J., Ungria M.C., Rodríguez J.J., Schmidt U., Schlauderer N., Saukko P., Schneider P.M., Sinker M., Shin K.J., Oh Y.N., Skitsa I., Ampati A., Smith T.G., Calvit L.S., Stenzl V., Capal T., Tillmar A., Nilsson H., Turrina S., Leo D., Verzeletti A., Cortellini V., Wetton J.H., Gwynne G.M., Jobling M.A., Whittle M.R., Sumita D.R., Wolańska Nowak P., Yong R.Y., Krawczak M., Nothnagel M., Roewer L. (2014), 'A global analysis of Y-chromosomal haplotype diversity for 23 STR loci.', <i>Forensic Sci Int Genet</i> 12, 12-23	Europeooriental

<b>147 Haplotypes from Italy [Italian]</b>	147	Rapone C., D'atanasio E., Agostino A., Mariano M., Papaluca MT., Cruciani F., Berti A. (2016), 'Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler®) Plus kit) in an Italian population sample.', <i>Forensic Sci Int Genet</i> 21, e1-5	Europeo
<b>Friesland, Netherlands [Dutch]</b>	94	Obtenidos desde bases de datos de poblaciones YHRD	Europeo
<b>Poland [Polish]</b>	486	Obtenidos desde bases de datos de poblaciones YHRD	Europeooriental
<b>Central Portugal, Portugal [Portuguese]</b>	372	González Neira A., Gusmão L., Brión M., Lareu MV., Amorim A., Carracedo A. (2000), 'Distribution of Y-chromosome STR defined haplotypes in Iberia.', <i>Forensic Sci Int</i> 110(2), 117-26 [Pubmed] Carvalho M., Anjos MJ., Andrade L., Lopes V., Santos MV., Gamero JJ., Real C., Vide MC. (2003), 'Y-chromosome STR haplotypes in two population samples: Azores Islands and Central Portugal.', <i>Forensic Sci Int</i> 134(1), 29-35	Europeo
<b>Basque Country, Spain [Basque]</b>	300	García O., Yurrebaso I., Mancisidor ID., López S., Alonso S., Gusmão L. (2015), 'Data for 27 Y-chromosome STR loci in the Basque Country autochthonous population', <i>Forensic Science International: Genetics</i> Epub ahead of print, in press	Europeo
<b>Andalusia, Spain [Spanish]</b>	149	Obtenidos desde bases de datos de poblaciones YHRD	Europeo
<b>Aragon, Spain [Spanish]</b>	45	Obtenidos desde bases de datos de poblaciones YHRD	Europeo
<b>1383 Haplotypes from Catalonia, Spain [Spanish]</b>	1383	Obtenidos desde bases de datos de poblaciones YHRD	Europeo
<b>Galicia, Spain [Spanish]</b>	89	Obtenidos desde bases de datos de poblaciones YHRD	Europeo
<b>Granada, Spain [Spanish]</b>	238	Ambrosio B., Novelletto A., Hernandez C., Dugoujon JM., Fortes Lima C., Rodriguez JN., Calderon R. (2012), 'Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain).', <i>Forensic Sci Int Genet</i> 6(2), e66-71	Europeo
<b>Valencia, Spain [Spanish]</b>	288	Rodríguez V., Tomás C., Sánchez JJ., Castro JA., Ramon MM., Barbaro A., Morling N., Picornell A. (2009), 'Genetic sub-structure in western Mediterranean populations revealed by 12 Y-chromosome STR loci.', <i>Int J Legal Med</i> 123(2), 137-41 [Pubmed] [DOI] Aler M., Salas A., Sánchez Diz P., Murcia E., Carracedo A., Gisbert M. (2001), 'Y-chromosome STR haplotypes from a Western Mediterranean population sample.', <i>Forensic Sci Int</i> 119(2), 254-7	Europeo
<b>London, United Kingdom [British European]</b>	1196	Aliferi A., Thomson J., Mc Donald A., Molin V., Ferguson S., Vanhinsbergh D., Syndercombe D., Ballard D. (2018), 'UK and Irish Y-STR population data—A catalogue of variant alleles', <i>For Sci Int Genetics</i> Epub ahead of print, In press	Europeo

## 4.3 Resultados y discusión

### 4.3.1 Diversidad genética

#### 4.3.1.1. Diversidad por locus

En el análisis de la diversidad genética de la población colombiana por locus, se encontraron 9 variantes alélicas (**Tabla 4-2**) correspondientes a alelos intermedios, 6 de las cuales pertenecen al sistema DYS458, que de acuerdo con la literatura ([https://yhrd.org/pages/resources/mutation\\_rates](https://yhrd.org/pages/resources/mutation_rates)), cuenta con una de las tasas de mutación más alta y tiene el mayor número de variantes alélicas reportadas (Myres, et.al., 2009), ([https://strbase.nist.gov/var\\_DYS458.htm](https://strbase.nist.gov/var_DYS458.htm)). El alelo 20.4 no ha sido reportado en dicha base de datos. También se encontró una duplicación 13,14 en el sistema DYS393 reportada previamente en STRBase y un posible alelo nulo en el sistema DYS448, ambos en muestras del Caribe Continental Colombiano. En una muestra de la región Cundiboyacense se encontró el alelo 10.2 en DYS385a/b. De acuerdo con Myres et.al., 2009, las variantes alélicas con 2 bases adicionales en este marcador, se presentan con mayor frecuencia en población del Congo y se han reportado en varias poblaciones centro y suramericanas pero no colombiana. Las frecuencias alélicas y diversidad genética para las 10 regiones de Colombia analizadas se observan en la **Tabla 4-3**.

**Tabla 4-2:** Variantes alélicas encontradas para los 17 Sistemas STRs de Cromosoma Y

REGION	SISTEMA	VARIANTE ALELICA	Frecuencia en YHRD	Ancestría Asociada más probable	Match Colombia
Amazonas	DYS458	18.2	1360 (0.009)	Semita Haplogrupos I/J	12 (0.008)
Caribe	DYS458	22.2	12 (0.00008)	Semita Haplogrupo J	0
Caribe	DYS448	Posible alelo nulo	452 (0.003)	Euroasiático Haplogrupo O/R/C	0
Caribe	DYS393	13-14	7 (0.00004)	Urálico/Nativoamericano Haplogrupo O/Q	0
Chocó	DYS458	17.2	732 (0.0047)	Afroasiático/Afrocaribe/ Afroamericano Haplogrupo J/R/E	20(0.014)
Chocó	DYS458	17.2	732 (0.0047)	Afroasiático/Afrocaribe/ Afroamericano Haplogrupo J/R/E	20(0.014)
Cundiboyacense	DYS448	20.4	2 (0.00001)	Mezclada	2(0.001)
Cundiboyacense	DYS458	16.2	107 (0.0007)	Semita/Eurooccidental Haplogrupo J/R	1 (0.0007)
Cundiboyacense	DYS385a/b	10.2-12	1 (0.000006)	Chino (Han)	0
Tolima grande	DYS458	18.2	1360 (0.009)	Semitas Haplogrupo I/J	12 (0.008)
Tolima grande	DYS458	20.2	176 (0.001)	Semita	3 (0.002)

**Tabla 4-3: Frecuencias alélicas de 15 sistemas STRs de Cromosoma Y en Colombia.**

MARCADOR	ALELO	Amazonas	Caribe	Cauca-Nariño	Chocó	Cundiboyacense	La Guajira	Archip. San Andres	Tolima Grande	Zona Cafetera
DYS19	13	0.5962	0.3478	0.6000	0.0784	0.1400	0.5185	0.1111	0.2128	0.1429
	14	0.3077	0.3913	0.4000	0.2353	0.6800	0.1482	0.3704	0.6064	0.2857
	15	0.0577	0.2391	0.0000	0.3333	0.1800	0.2222	0.1482	0.1170	0.4286
	16	0.0192	0.0217	0.0000	0.2549	0.0000	0.0741	0.2222	0.0426	0.0000
	17	0.0192	0.0000	0.0000	0.0980	0.0000	0.0370	0.1482	0.0213	0.1429
	<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>
<b>D.G.</b>	<b>0.5566</b>	<b>0.6831</b>	<b>0.6000</b>	<b>0.7678</b>	<b>0.4955</b>	<b>0.6781</b>	<b>0.7715</b>	<b>0.5772</b>	<b>0.8095</b>	
DYS389I	12	0.0577	0.1522	0.0000	0.2353	0.1600	0.0370	0.2222	0.1383	0.2857
	13	0.7115	0.5652	0.8000	0.5490	0.7000	0.5556	0.7407	0.6277	0.7143
	14	0.2308	0.2826	0.2000	0.2157	0.1400	0.4074	0.0370	0.2234	0.0000
	15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0106	0.0000
	<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>
	<b>D.G.</b>	<b>0.4457</b>	<b>0.5903</b>	<b>0.4000</b>	<b>0.6086</b>	<b>0.4743</b>	<b>0.5442</b>	<b>0.4081</b>	<b>0.5427</b>	<b>0.4762</b>
DYS389II	27	0.0000	0.0000	0.0000	0.0000	0.0200	0.0000	0.0000	0.0213	0.0000
	28	0.1154	0.1556	0.0000	0.0980	0.1200	0.0370	0.1296	0.1383	0.2857
	29	0.3077	0.1778	0.4000	0.2549	0.4800	0.3333	0.2963	0.4468	0.1429
	30	0.3077	0.2444	0.2000	0.2941	0.3400	0.1482	0.4074	0.3298	0.4286
	31	0.1923	0.3111	0.4000	0.2941	0.0400	0.2963	0.1482	0.0426	0.1429
	32	0.0769	0.1111	0.0000	0.0392	0.0000	0.1482	0.0185	0.0213	0.0000
	33	0.0000	0.0000	0.0000	0.0196	0.0000	0.0370	0.0000	0.0000	0.0000
	<b>N</b>	<b>52</b>	<b>45</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>
<b>D.G.</b>	<b>0.7692</b>	<b>0.7929</b>	<b>0.8000</b>	<b>0.7655</b>	<b>0.6506</b>	<b>0.7835</b>	<b>0.7205</b>	<b>0.6770</b>	<b>0.8095</b>	
DYS390	21	0.0385	0.1087	0.0000	0.4706	0.0200	0.0000	0.4815	0.0213	0.1429
	22	0.0385	0.0652	0.0000	0.1569	0.1000	0.1111	0.0370	0.0532	0.1429
	23	0.2885	0.2391	0.4000	0.0784	0.2200	0.3333	0.1296	0.2447	0.4286
	24	0.4808	0.3478	0.6000	0.2157	0.5400	0.4444	0.2222	0.5851	0.1429
	25	0.1539	0.2391	0.0000	0.0784	0.1000	0.1111	0.1296	0.0957	0.1429
	26	0.0000	0.0000	0.0000	0.0000	0.0200	0.0000	0.0000	0.0000	0.0000
	<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>
<b>D.G.</b>	<b>0.6719</b>	<b>0.7652</b>	<b>0.6000</b>	<b>0.7090</b>	<b>0.6522</b>	<b>0.6923</b>	<b>0.6967</b>	<b>0.5916</b>	<b>0.8571</b>	
DYS391	6	0.0000	0.0000	0.0000	0.0000	0.0200	0.0370	0.0000	0.0638	0.0000
	8	0.0000	0.0000	0.0000	0.0000	0.0200	0.0370	0.0000	0.0000	0.0000
	9	0.0385	0.0222	0.2000	0.0588	0.0200	0.0370	0.0000	0.0532	0.0000
	10	0.6923	0.5111	0.6000	0.6863	0.2600	0.7407	0.7222	0.4362	0.2857
	11	0.2692	0.4444	0.2000	0.2549	0.6600	0.1482	0.1852	0.4362	0.7143
	12	0.0000	0.0222	0.0000	0.0000	0.0200	0.0000	0.0926	0.0106	0.0000
	<b>N</b>	<b>52</b>	<b>45</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>
<b>D.G.</b>	<b>0.4555</b>	<b>0.5525</b>	<b>0.7000</b>	<b>0.4698</b>	<b>0.5053</b>	<b>0.4416</b>	<b>0.4437</b>	<b>0.6191</b>	<b>0.4762</b>	
DYS392	7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0370	0.0000	0.0000	0.0000
	9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0741	0.0106	0.0000
	10	0.0000	0.0217	0.0000	0.0000	0.0000	0.0000	0.0000	0.0106	0.0000
	11	0.1346	0.2826	0.6000	0.7255	0.2600	0.2963	0.5185	0.1915	0.4286
	12	0.0962	0.0000	0.2000	0.0392	0.0400	0.0741	0.0185	0.0213	0.0000
	13	0.2115	0.3696	0.2000	0.2157	0.6000	0.2222	0.3704	0.5957	0.4286
	14	0.4231	0.3044	0.0000	0.0000	0.0400	0.3333	0.0000	0.1277	0.1429
	15	0.1346	0.0217	0.0000	0.0196	0.0600	0.0370	0.0000	0.0426	0.0000
	17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0185	0.0000	0.0000
	<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>
<b>D.G.</b>	<b>0.7451</b>	<b>0.7053</b>	<b>0.7000</b>	<b>0.4337</b>	<b>0.5771</b>	<b>0.7721</b>	<b>0.5989</b>	<b>0.5960</b>	<b>0.7143</b>	
DYS393	11	0.0000	0.0217	0.0000	0.0000	0.0200	0.0000	0.0000	0.0000	0.0000
	12	0.0962	0.1739	0.0000	0.1373	0.1600	0.3462	0.0741	0.0745	0.1429
	13	0.7692	0.5652	1.0000	0.5098	0.7000	0.5000	0.5926	0.8192	0.7143
	14	0.0962	0.1957	0.0000	0.3333	0.1000	0.0769	0.2963	0.0851	0.1429
	15	0.0385	0.0435	0.0000	0.0196	0.0200	0.0769	0.0370	0.0213	0.0000
	<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>26</b>	<b>54</b>	<b>94</b>	<b>7</b>
<b>D.G.</b>	<b>0.3959</b>	<b>0.6232</b>	<b>0.0000</b>	<b>0.6220</b>	<b>0.4833</b>	<b>0.6431</b>	<b>0.5646</b>	<b>0.3191</b>	<b>0.5238</b>	

**Tabla 4-3:** Frecuencias alélicas de 15 sistemas STRs de Cromosoma Y en Colombia.  
(Continuación)

MARCADOR	ALELO	Amazonas	Caribe	Cauca-Nariño	Chocó	Cundiboyacense	La Guajira	Archip. San Andres	Tolima Grande	Zona Cafetera
DYS437	13	0.0000	0.0217	0.0000	0.0588	0.0000	0.0000	0.0000	0.0106	0.1429
	14	0.7500	0.5870	0.6000	0.6078	0.3800	0.6667	0.5926	0.3511	0.1429
	15	0.2308	0.2826	0.4000	0.2353	0.5000	0.3333	0.3889	0.5426	0.5714
	16	0.0192	0.1087	0.0000	0.0588	0.1200	0.0000	0.0000	0.0851	0.1429
	17	0.0000	0.0000	0.0000	0.0392	0.0000	0.0000	0.0000	0.0106	0.0000
	<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>
<b>D.G.</b>	<b>0.3914</b>	<b>0.5758</b>	<b>0.6000</b>	<b>0.5780</b>	<b>0.6033</b>	<b>0.4615</b>	<b>0.5070</b>	<b>0.5811</b>	<b>0.7143</b>	
DYS438	8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0370	0.0000	0.0000	0.0000
	9	0.0962	0.1522	0.0000	0.0392	0.1400	0.1482	0.0370	0.0745	0.1429
	10	0.1154	0.1957	0.6000	0.1373	0.1200	0.1852	0.1111	0.1915	0.2857
	11	0.5385	0.2826	0.0000	0.5882	0.1400	0.4815	0.4444	0.1915	0.1429
	12	0.2308	0.3478	0.4000	0.2353	0.5800	0.1482	0.3889	0.5213	0.4286
	13	0.0192	0.0217	0.0000	0.0000	0.0200	0.0000	0.0185	0.0213	0.0000
<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>	
<b>D.G.</b>	<b>0.6463</b>	<b>0.7536</b>	<b>0.6000</b>	<b>0.5898</b>	<b>0.6220</b>	<b>0.7151</b>	<b>0.6492</b>	<b>0.6559</b>	<b>0.8095</b>	
DYS439	10	0.0769	0.0217	0.0000	0.0392	0.0600	0.0385	0.1111	0.0851	0.0000
	11	0.2885	0.2609	0.2000	0.1961	0.3400	0.4615	0.1111	0.2979	0.4286
	12	0.4615	0.5000	0.6000	0.6275	0.4400	0.3462	0.5926	0.4468	0.4286
	13	0.1539	0.1739	0.2000	0.1177	0.1600	0.1539	0.1852	0.1170	0.1429
	14	0.0192	0.0435	0.0000	0.0196	0.0000	0.0000	0.0000	0.0426	0.0000
	16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0106	0.0000
<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>26</b>	<b>54</b>	<b>94</b>	<b>7</b>	
<b>D.G.</b>	<b>0.6870</b>	<b>0.6638</b>	<b>0.7000</b>	<b>0.5631</b>	<b>0.6751</b>	<b>0.6677</b>	<b>0.6010</b>	<b>0.6962</b>	<b>0.7143</b>	
DYS448	17	0.0000	0.0000	0.0000	0.0000	0.0200	0.0000	0.0000	0.0000	0.0000
	18	0.0385	0.0870	0.0000	0.0588	0.2400	0.0385	0.0556	0.0745	0.1429
	19	0.3654	0.3696	0.4000	0.3333	0.4400	0.2308	0.3148	0.5957	0.4286
	20	0.5000	0.4130	0.6000	0.0980	0.2000	0.4615	0.1482	0.2660	0.0000
	21	0.0962	0.1304	0.0000	0.5098	0.0800	0.0769	0.4630	0.0532	0.2857
	22	0.0000	0.0000	0.0000	0.0000	0.0200	0.1923	0.0185	0.0000	0.0000
23	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0106	0.1429	
<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>26</b>	<b>54</b>	<b>94</b>	<b>7</b>	
<b>D.G.</b>	<b>0.6177</b>	<b>0.6831</b>	<b>0.6000</b>	<b>0.6282</b>	<b>0.7159</b>	<b>0.7169</b>	<b>0.6737</b>	<b>0.5720</b>	<b>0.8095</b>	
DYS456	12	0.0000	0.0000	0.0000	0.0196	0.0000	0.0000	0.0000	0.0000	0.0000
	13	0.0192	0.0217	0.0000	0.0000	0.0000	0.0000	0.0000	0.0106	0.0000
	14	0.0385	0.1087	0.0000	0.0588	0.1000	0.0741	0.0185	0.0957	0.4286
	15	0.5000	0.5000	0.6000	0.5882	0.4000	0.4074	0.4630	0.4468	0.4286
	16	0.2885	0.2174	0.4000	0.2941	0.4000	0.3333	0.3333	0.3830	0.1429
	17	0.1539	0.1522	0.0000	0.0392	0.0600	0.1111	0.1111	0.0426	0.0000
18	0.0000	0.0000	0.0000	0.0000	0.0400	0.0370	0.0741	0.0213	0.0000	
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0370	0.0000	0.0000	0.0000	
<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>	
<b>D.G.</b>	<b>0.6538</b>	<b>0.6821</b>	<b>0.6000</b>	<b>0.5733</b>	<b>0.6784</b>	<b>0.7293</b>	<b>0.6688</b>	<b>0.6490</b>	<b>0.7143</b>	
DYS458	13	0.0000	0.0435	0.0000	0.0000	0.0200	0.0000	0.0000	0.0000	0.0000
	14	0.0192	0.0000	0.0000	0.0392	0.0000	0.0000	0.0185	0.0000	0.0000
	15	0.0577	0.1304	0.4000	0.1569	0.1600	0.2593	0.0185	0.1064	0.0000
	16	0.4231	0.1739	0.2000	0.2353	0.2000	0.1852	0.4259	0.2872	0.5714
	17	0.2692	0.4130	0.2000	0.3726	0.3800	0.2963	0.2593	0.2872	0.2857
	18	0.1731	0.2174	0.2000	0.1373	0.2000	0.0741	0.1667	0.2872	0.1429
19	0.0192	0.0217	0.0000	0.0392	0.0400	0.1482	0.1111	0.0213	0.0000	
20	0.0192	0.0000	0.0000	0.0196	0.0000	0.0370	0.0000	0.0106	0.0000	
21	0.0192	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>	
<b>D.G.</b>	<b>0.7277</b>	<b>0.7488</b>	<b>0.9000</b>	<b>0.7741</b>	<b>0.7633</b>	<b>0.8120</b>	<b>0.7240</b>	<b>0.7486</b>	<b>0.6667</b>	



**Tabla 4-3:** Frecuencias alélicas de 15 sistemas STRs de Cromosoma Y en Colombia.  
(Continuación)

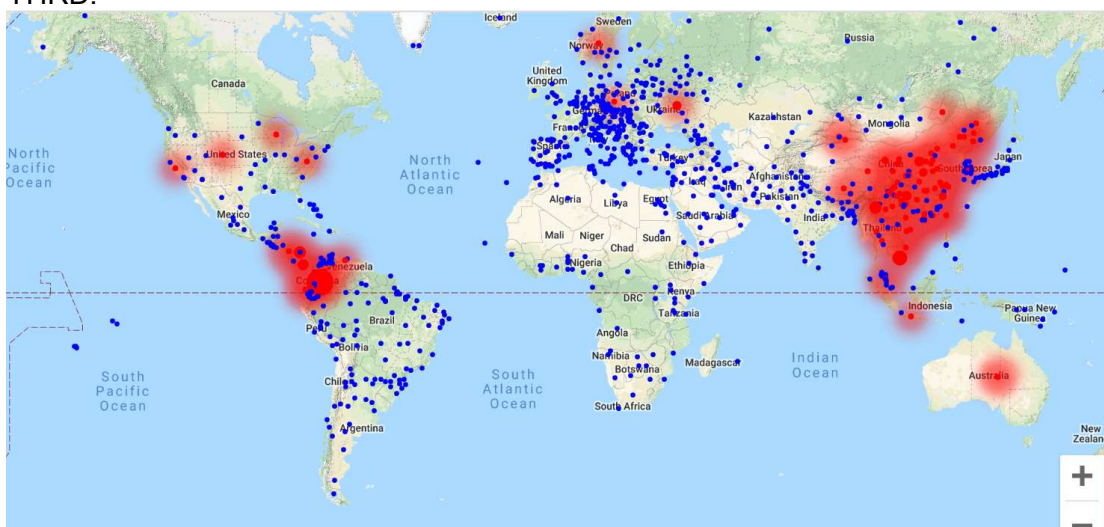
MARCADOR	ALELO	Amazonas	Caribe	Cauca-Nariño	Chocó	Cundiboyacense	La Guajira	Archip. San Andres	Tolima Grande	Zona Cafetera
DYS635	17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0370	0.0000	0.0000	0.0000
	19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0185	0.0000	0.0000
	20	0.0000	0.0222	0.4000	0.0784	0.0000	0.1111	0.1667	0.0213	0.0000
	21	0.1154	0.2222	0.2000	0.3333	0.1600	0.1111	0.2037	0.1596	0.1429
	22	0.5385	0.1556	0.0000	0.2157	0.1800	0.4444	0.1482	0.1702	0.2857
	23	0.2692	0.4667	0.2000	0.3333	0.5600	0.2963	0.3333	0.5106	0.5714
	24	0.0769	0.1333	0.2000	0.0392	0.0600	0.0000	0.1111	0.1064	0.0000
	25	0.0000	0.0000	0.0000	0.0000	0.0400	0.0000	0.0185	0.0106	0.0000
	26	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0213	0.0000
	<b>N</b>	<b>52</b>	<b>45</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>
<b>D.G.</b>	<b>0.6305</b>	<b>0.7061</b>	<b>0.9000</b>	<b>0.7380</b>	<b>0.6359</b>	<b>0.7151</b>	<b>0.7994</b>	<b>0.6797</b>	<b>0.6667</b>	
YGATAH4	9	0.0192	0.0217	0.0000	0.0196	0.0400	0.1111	0.0000	0.0000	0.0000
	10	0.0577	0.0435	0.6000	0.0392	0.0400	0.0741	0.0000	0.0106	0.0000
	11	0.4231	0.3696	0.0000	0.4902	0.3800	0.3704	0.4444	0.2872	0.1429
	12	0.4039	0.5435	0.4000	0.4118	0.5200	0.2593	0.4074	0.6702	0.7143
	13	0.0962	0.0217	0.0000	0.0392	0.0200	0.1852	0.1482	0.0213	0.1429
	14	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0106	0.0000
	<b>N</b>	<b>52</b>	<b>46</b>	<b>5</b>	<b>51</b>	<b>50</b>	<b>27</b>	<b>54</b>	<b>94</b>	<b>7</b>
	<b>D.G.</b>	<b>0.6576</b>	<b>0.5778</b>	<b>0.6000</b>	<b>0.5984</b>	<b>0.5935</b>	<b>0.7721</b>	<b>0.6261</b>	<b>0.4727</b>	<b>0.5238</b>

En el sistema DYS391 se detecta el alelo 6 en la región Cundiboyacense (2%), en la Guajira (3.7%) y en Tolima Grande (6.4%). De acuerdo con Alonso et.al., 2018, este alelo puede ser empleado para caracterizar los linajes Q1a2-M346, y a Noviembre de 2018 se ha reportado en el 0.19% de las muestras incluidas en la base de datos YHRD, principalmente en poblaciones Asiáticas y Latinoamericanas (493 match en 265324 haplotipos mínimos, 22 de ellos en Colombia, otros en Ecuador, Nicaragua, Panamá y Venezuela; la frecuencia en nativoamericanos es de 0.08%, 0.12% en Altaicos, 0.67% en China-Han) **Figura 4-2**. En 1011 haplotipos pertenecientes a los haplogrupos Q y C3 de 50 poblaciones Nativos Suramericanos, Roewer et.al., 2013 no reporta este alelo, al igual que Xavier et.al., 2015 en 53 individuos Q en 5 etnias en Colombia. Sin embargo, Franco y Barreto 2017 realizan este reporte en un individuo Nasa y 5 Coyaimas y Grugni et.al. 2015 en 6 haplotipos de panameños, 1 de ellos autorreconocido como perteneciente al grupo étnico Terebe, pero en ningún Bribe, Buglé, Kuna o Ngabe. Por su parte, Purps et.al. 2014 encontró este alelo en 4 individuos Han y 2 Kinh (Vietnam).

El alelo 19 en DYS456 y el alelo 17 en DYS635 fueron encontrados en una frecuencia de 0.037 en la población Colombiana de La Guajira y en las poblaciones de referencia solamente en Europa Oriental con frecuencias de 0.003 y 0.005 respectivamente. El alelo

7 en DYS392 que solo fue encontrado en Guajira en 3.7%, se encuentra reportado en la base de datos YHRD en un 2.2% en China, 0.8% en Kasahjstan, 0.9% en Mongolia, y en Purps, 2014 en 0.4% en Beijing, 0.7% en Hungría y 0.2% en Afroamericanos de Estados Unidos. De igual forma, el alelo 16 en DYS439 está presente en Tolima Grande en 1% y se reporta en 0.6% en Mecklenburg-Vorpommern, Alemania en Purps et.al. 2014 y en YHRD en frecuencias muy bajas de 0.08% en Bangladesh y 0.06% en Colombia y en menores frecuencias en Croacia, Australia, Alemania y China.

**Figura 4-2.** Distribución mundial del alelo 6 del sistema DSY391. Tomado de base de datos YHRD.

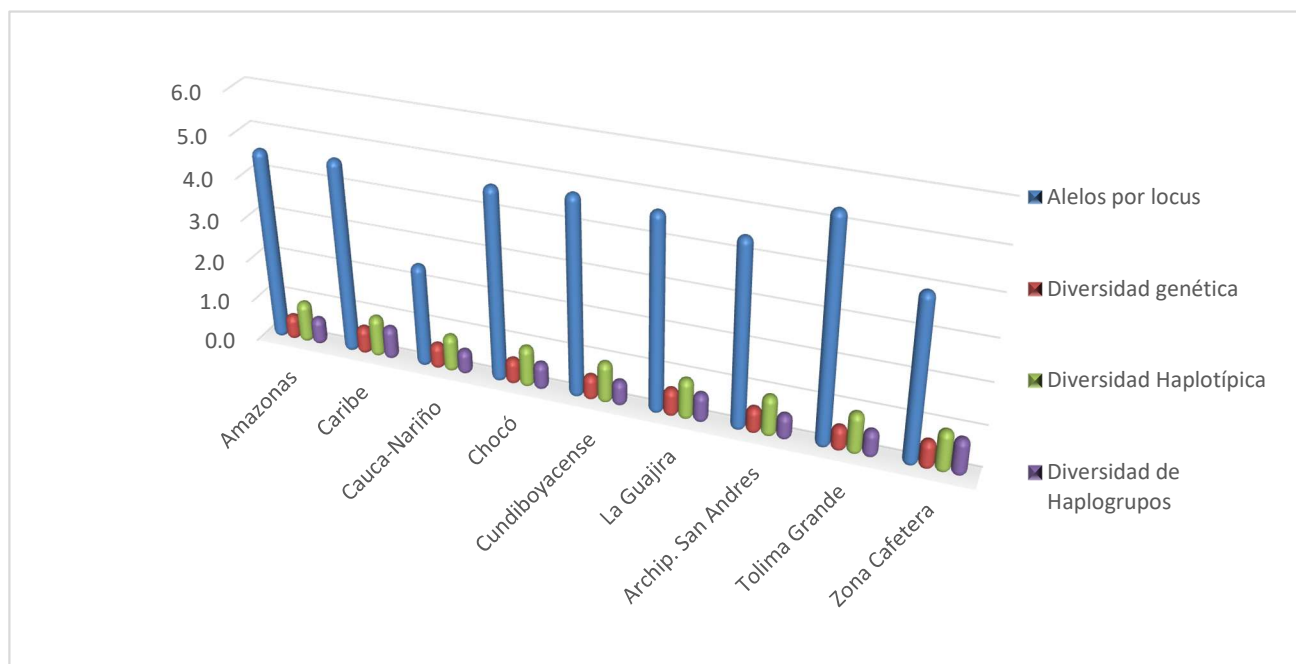


Los valores promedio de índices de diversidad genética y número de alelos por locus para cada una de las regiones se muestran en la **Figura 4-3 y Tabla 4-4**. El mayor número de alelos promedio se encuentra en la región Tolima grande, sin embargo, esta región cuenta con la más baja diversidad genética.

**Tabla 4-4:** Diversidad genética para 15 sistemas STRs de Cromosoma Y. N entre ( )

Promedios	Amazonas (52)	Caribe (46)	Cauca-Nariño (5)	Chocó (51)	Cundi-boyacense (50)	La Guajira (27)	Archipiélago San Andres (54)	Tolima Grande (94)	Zona Cafetera (7)
Alelos por locus	4.6	4.7	2.5	4.7	4.8	4.7	4.5	5.4	4.0
Diversidad genética	0.6035	0.6736	0.6200	0.6280	0.6084	0.6763	0.6302	0.5985	0.6857
Diversidad Haplotípica	0.9978	0.9970	0.9000	0.9971	0.9992	0.9739	0.9860	0.9976	1.0000
Diversidad de Haplogrupos	0.6622	0.8044	0.6000	0.6650	0.6165	0.7095	0.6147	0.6544	0.8901

**Figura 4-3.** Diversidad genética para 15 sistemas STRs de Cromosoma Y



#### 4.3.1.1 Diversidad por haplotipos

La diversidad haplotípica estimada para Colombia fue de 0.99943. En las regiones de Valle del Cauca y Zona Cafetera, todos los haplotipos son únicos, sin embargo, debe considerarse que este resultado pueda relacionarse con el tamaño de muestra. Rojas 2010, empleando 6 marcadores STRs de cromosoma Y encontró una diversidad haplotípica de 0.96973255 en el Valle del Cauca y 0.9446464 en la Zona Cafetera (Cálculos realizados por nuestro grupo de investigación a partir de los datos originales en el material suplementario de Rojas et.al. 2010). La menor diversidad haplotípica se observa en las regiones de Cauca-Nariño que tiene un bajo tamaño de muestra, La Guajira y el archipiélago de San Andrés (**Tabla 4-4**). Parte de estos resultados puede relacionarse con que las muestras de La Guajira corresponden al grupo indígena Wayuu, quienes de acuerdo con los resultados de Rojas, et.al. 2013, en un 80% prefieren establecerse con una pareja de la misma etnia. Por otro lado, la estructura poblacional isleña Raizal del Archipiélago de San Andrés, se ha estudiado en detalle en Alonso y Usaqué, (2013a), proponiendo que la

diversidad haplotípica más baja puede relacionarse con el hecho de que pocos individuos compartan un origen común o con un pequeño tamaño efectivo de población que genera un efecto de deriva genética.

Cinco haplotipos que se encontraron tres veces en la población colombiana, fueron revisados en la base de datos YHRD que cuenta actualmente con 156181 haplotipos con los marcadores del kit comercial AmpFISTR® Yfiler y Yfiler™Plus de Applied Biosystems, para revisar coincidencias. A continuación se describen los resultados (DYS385, DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, YGATAH4): el haplotipo 15-15,15, 13, 29, 23, 10, 11, 12, 14, 9, 11, 20, 15, 19, 23, 11 se encontró 3 veces en la población de La Guajira y no hubo coincidencia en la base de datos YHRD ni se encontraron haplotipos vecinos con diferencia de un paso. Los haplotipos 11-14, 14, 12, 28, 25, 12, 13, 13, 15, 12, 12, 19, 15, 17, 24, 12 ; 17-17, 17, 13, 30, 21, 10, 11, 14, 14, 11, 13, 21, 16, 16, 22, 11 y 15- 19, 16, 13, 30, 21, 10, 11, 14, 14, 11, 12, 21, 15, 19, 20, 11 se encontraron 3 veces cada uno en la población de San Andrés. Solo el segundo ha sido reportado en un hombre de Reino Unido (Base de datos YHRD). El haplotipo 11-15, 14, 13, 30, 24, 11, 13, 13, 15, 12, 13, 19, 16, 16, 23, 12 se encontró 2 veces en la región Tolima grande y una en la región Caribe, con coincidencia en YHRD con 10 vecinos de un paso.

#### 4.3.1.3 Diversidad por haplogrupos

La diversidad de haplogrupos estimada sobre 372 individuos de la población colombiana fue de 0.7702, siendo los menores valores los de Cauca-Nariño y el Archipiélago de San Andrés y Providencia (**Tabla 4-4**). De acuerdo con la distribución frecuencias en las regiones (**Tabla 4-5 y Figura 4-4**), es posible distinguir al menos tres modelos de estructura poblacional, donde se destaca la similitud entre la zona Cundiboyacense y Tolima grande, con más del 55% de las muestras predichas por pertenecer al haplogrupo de origen europeo occidental R1b, entre el 10 y 15% al haplogrupo Q de origen nativoamericano y una presencia mínima de haplogrupos de origen Africano. Otros haplogrupos asociados a Oriente Medio y Europa como G, I, J, R1a, R2 y T (Otros haplogrupos Euroasiáticos), se encuentran individualmente en muy baja frecuencia, pero en conjunto pueden sumar entre el 20 y el 25% del total de individuos. Otro modelo de estructura corresponde al compartido

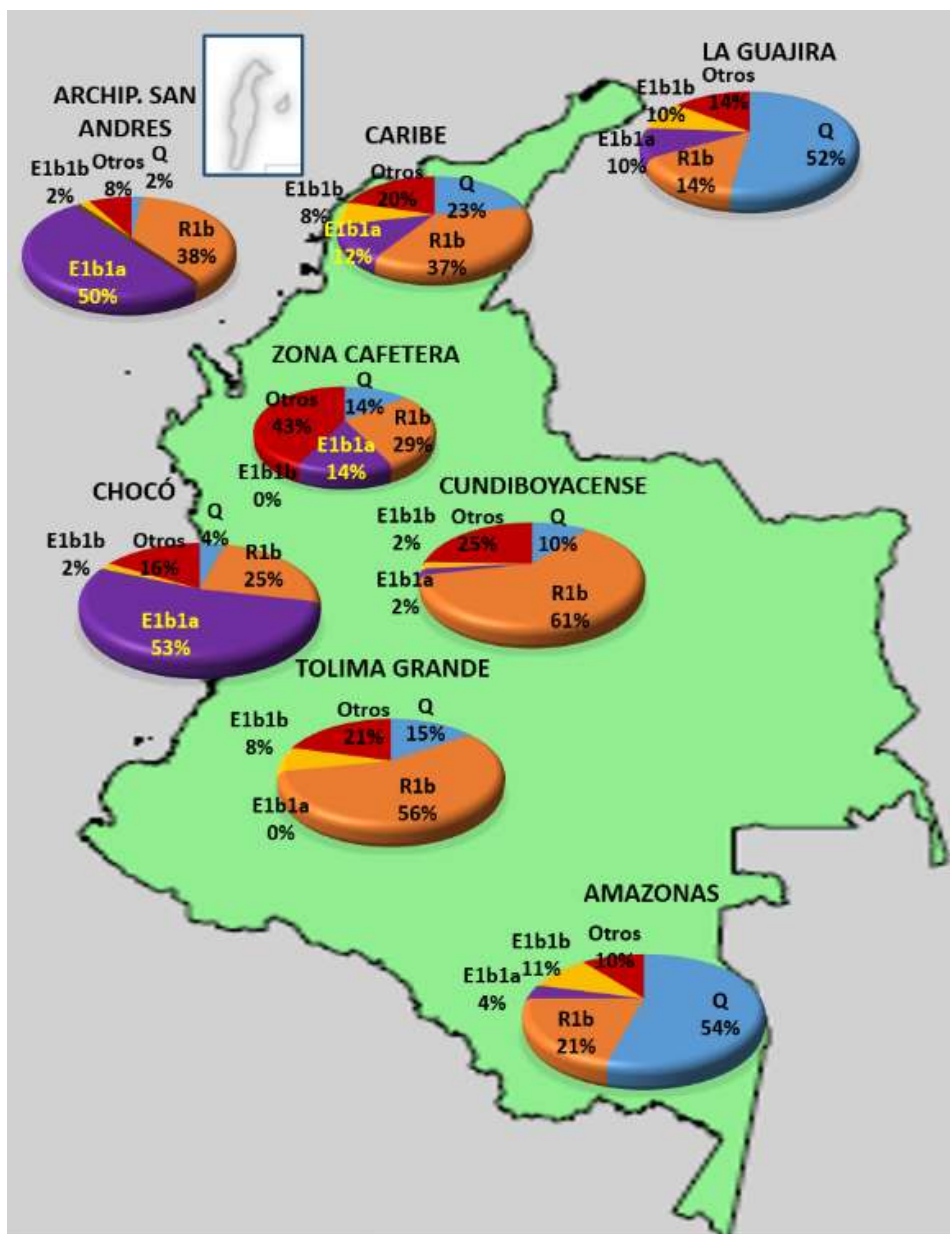
por las poblaciones colombianas de origen africano Chocó y San Andrés, caracterizadas por la presencia mayoritaria de haplogrupo E1b1a. Se diferencian sin embargo, en que en las muestras del Chocó se encuentra una mayor proporción de haplogrupos Q, lo que es consecuencia de la relación de las poblaciones afrodescendientes de esta región del Pacífico Colombiano, con las etnias indígenas habitantes de la misma región, mientras en las muestras del Archipiélago de San Andrés se nota una presencia importante del haplogrupo R1b, lo cual también es concordante con la historia de mezcla entre los colonos ingleses, holandeses y otros navegantes, con los esclavos transportados desde Jamaica y otras Antillas (Alonso y Usaquén, 2013a; Alonso, et.al., 2018) (Eastman, 1992).

**Tabla 4-5:** Diversidad de haplogrupos determinados usando el predictor de haplogrupos, a partir de 15 sistemas STRs de Cromosoma Y. Se resalta el haplogrupo más frecuente.

HAPLOGRUPO	Caribe	Amazonas	Cauca-Nariño	Chocó	Cundiboyacense	La Guajira	Tolima Grande	Valle del Cauca	Zona Cafetera	Archipiélago San Andrés	Colombia
E1b1a	0.1224	0.0417	0.0000	<b>0.5306</b>	0.0204	0.0952	0.0000	0.0000	0.1429	<b>0.5000</b>	0.1694
E1b1b	0.0816	0.1042	<b>0.6000</b>	0.0204	0.0204	0.0952	0.0761	0.5000	0.0000	0.0200	0.0672
G2a	0.0000	0.0000	0.0000	0.0204	0.0000	0.0000	0.0435	0.0000	0.1429	0.0000	0.0161
I1	0.0204	0.0208	0.0000	0.0204	0.0612	0.0000	0.0326	0.0000	0.0000	0.0200	0.0269
I2a1	0.0000	0.0000	0.0000	0.0000	0.0204	0.0000	0.0217	0.0000	0.1429	0.0000	0.0108
I2b1	0.0204	0.0208	0.0000	0.0000	0.0000	0.0952	0.0000	0.0000	0.0000	0.0200	0.0134
J1	0.0204	0.0208	0.0000	0.0408	0.0408	0.0000	0.0109	0.0000	0.0000	0.0000	0.0188
J2a1	0.0000	0.0000	0.0000	0.0204	0.0204	0.0000	0.0109	0.0000	0.0000	0.0000	0.0081
J2a4	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0027
J2a1b	0.0000	0.0000	0.0000	0.0000	0.0408	0.0000	0.0000	0.0000	0.0000	0.0000	0.0054
J2a4b	0.0204	0.0000	0.0000	0.0000	0.0000	0.0476	0.0000	0.0000	0.0000	0.0000	0.0054
J2b	0.0408	0.0000	0.0000	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0081
L	0.0204	0.0208	0.0000	0.0408	0.0204	0.0000	0.0217	0.0000	0.0000	0.0000	0.0188
Q	0.2245	<b>0.5417</b>	0.0000	0.0408	0.1020	<b>0.5238</b>	0.1522	0.5000	0.1429	0.0200	0.1935
R1a	0.0000	0.0208	0.0000	0.0000	0.0000	0.0000	0.0109	0.0000	0.0000	0.0400	0.0108
R1b	<b>0.3673</b>	0.2083	0.4000	0.2449	<b>0.6122</b>	0.1429	<b>0.5652</b>	0.0000	<b>0.2857</b>	0.3800	<b>0.3978</b>
R2	0.0204	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0027
T	0.0204	0.0000	0.0000	0.0000	0.0408	0.0000	0.0543	0.0000	0.1429	0.0000	0.0242
<b>N</b>	<b>49</b>	<b>48</b>	<b>5</b>	<b>49</b>	<b>49</b>	<b>21</b>	<b>92</b>	<b>2</b>	<b>7</b>	<b>50</b>	<b>372</b>
<b>Diversidad</b>	<b>0.8044</b>	<b>0.6622</b>	<b>0.6000</b>	<b>0.6650</b>	<b>0.6165</b>	<b>0.7095</b>	<b>0.6544</b>	<b>1.0000</b>	<b>0.9524</b>	<b>0.6147</b>	<b>0.7702</b>

Tanto en la región Cafetera como en el Caribe Continental Colombiano, la proporción de haplogrupos de origen africano disminuye, y la de haplogrupos Q aumenta, mientras la de Euroasiáticos y R1b muestra frecuencias contrarias en ambas regiones.

**Figura 4-4.** Frecuencias de los diferentes haplogrupos por regiones



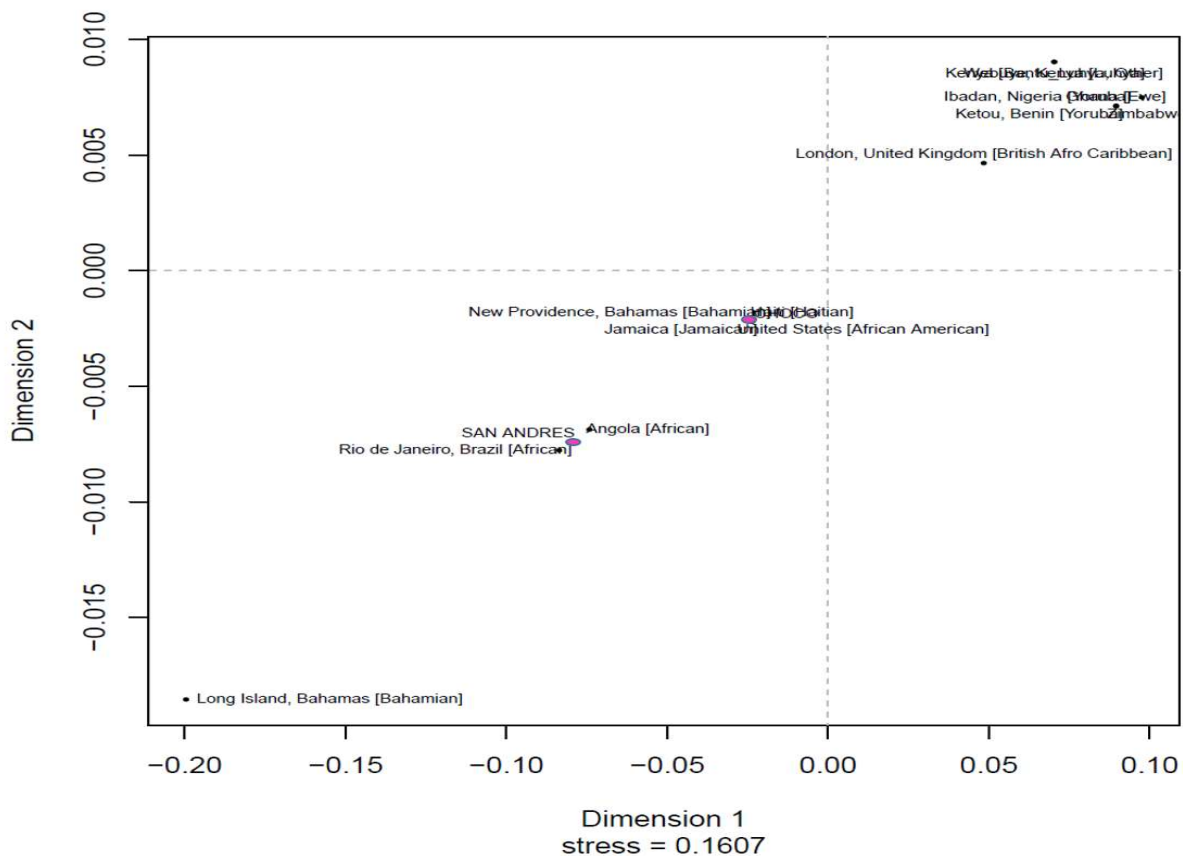
Finalmente, en la región de la Guajira y en Amazonas se observa una proporción mayoritaria (superior al 50%) de haplogrupos Q, entre el 14% y el 21% de los R1b y entre el 4% y el 11% de E1b1a y E1b1b. Cabe anotar que, así como la totalidad de las muestras de la región de la Guajira fueron clasificadas genealógicamente como pertenecientes a la etnia Wayuu, en las de Amazonas, un 48% fueron genealógicamente clasificadas como pertenecientes a una etnia de la región (Ticuna, Miraña, Huitoto, Cocama, Yagua y Bora,

entre otras) al menos por un linaje (Moncada, 2018). Los datos de la región Cauca-Nariño y Valle no se consideran por su tamaño de muestra.

#### 4.3.1.3.1 Diversidad Regional

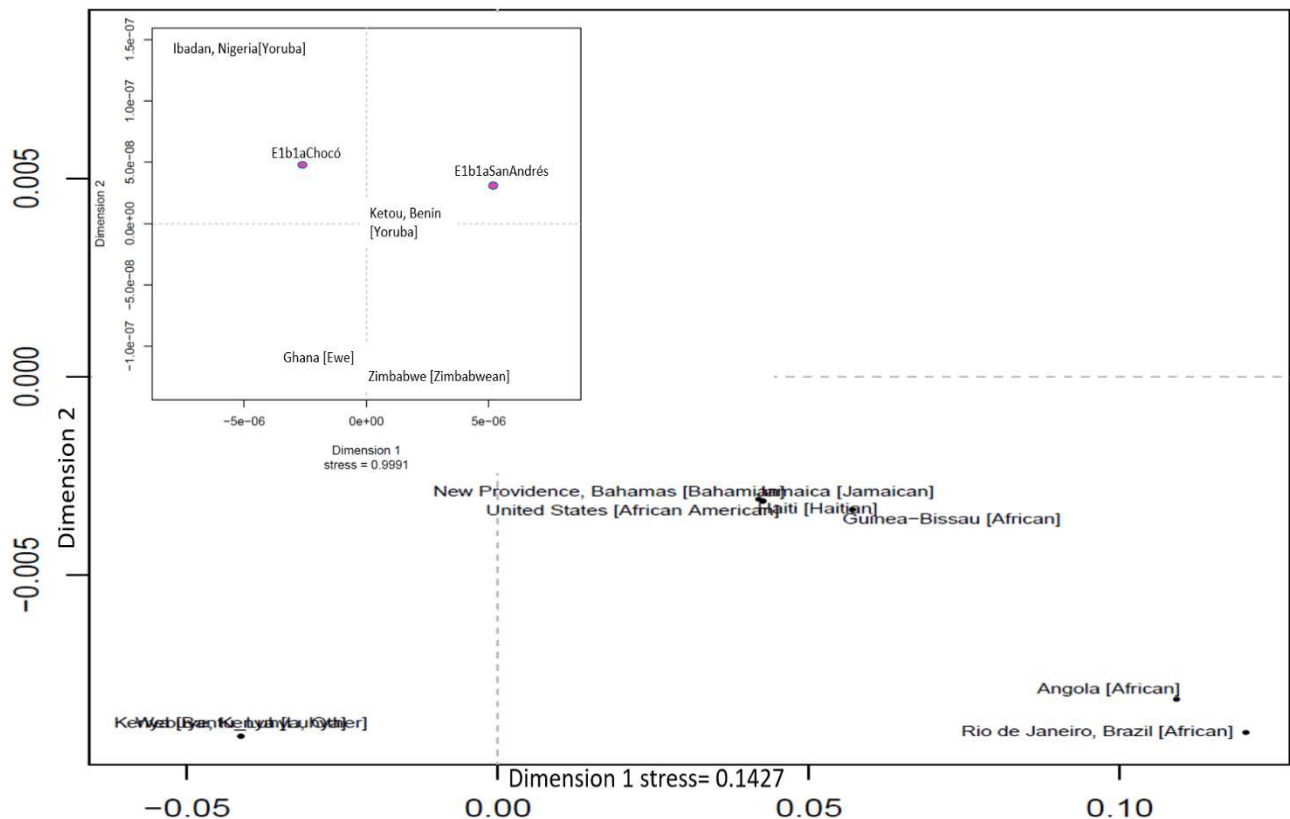
Un AMOVA y MDS para las muestras completas de la región de San Andrés y de Chocó se realizó utilizando 3805 Individuos de poblaciones de referencia afrodescendientes, listadas en la tabla 4-1 de materiales y métodos. La población de San Andrés se ubica cercana a las poblaciones afrodescendientes de Rio de Janeiro, Brasil y Angola en Africa, mientras la población de Chocó se aproxima a las afroamericanas de Estados Unidos y a las afrocaribeñas de New Providence (Bahamas), Jamaica y Haití (**Figura 4-5**).

**Figura 4-5.** MDS muestras de San Andrés y Chocó y poblaciones de referencia afrodescendientes.



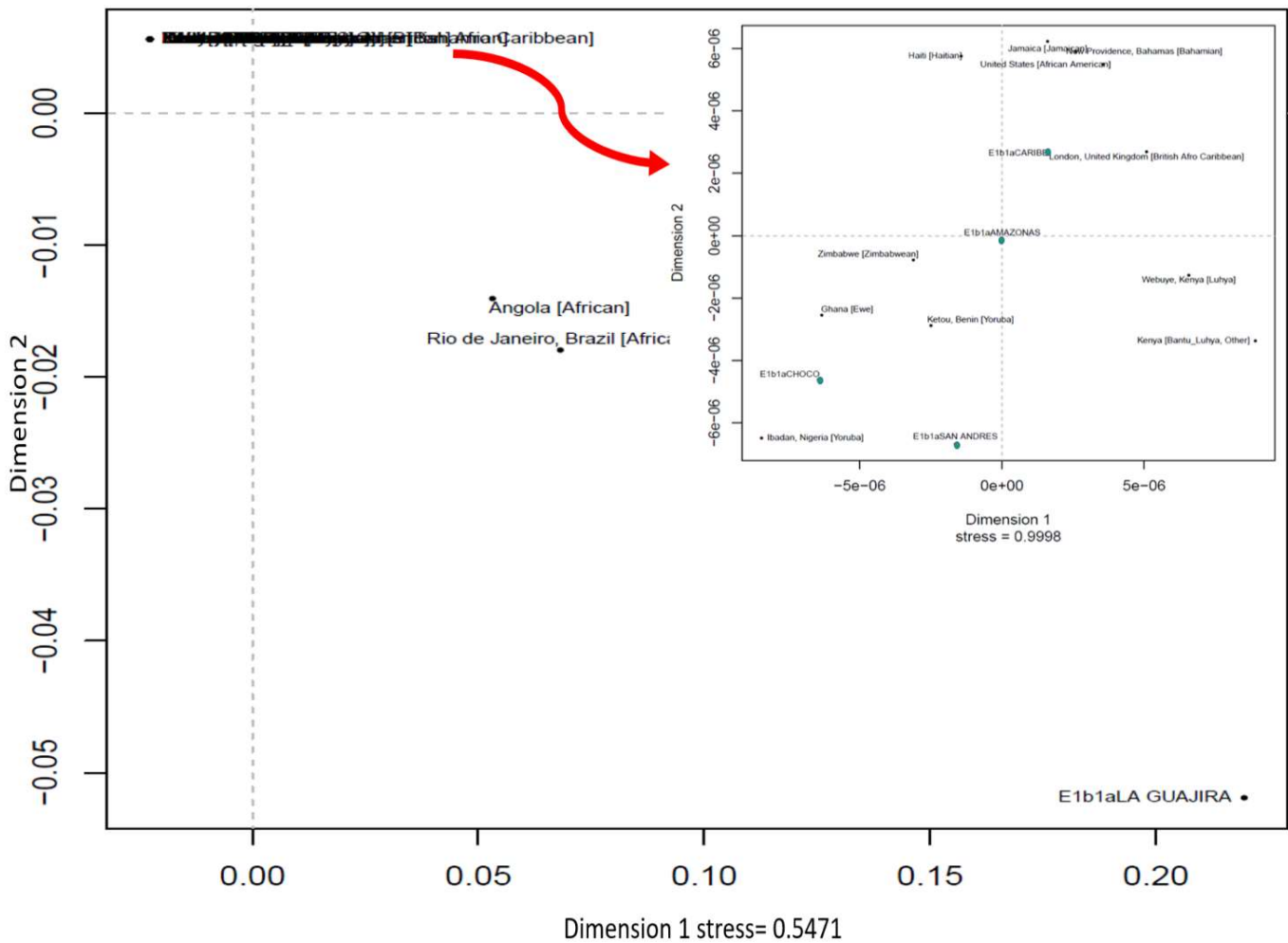
En el análisis considerando solo los individuos clasificados como pertenecientes al haplogrupo E1b1a, con 3848 individuos de poblaciones de referencia afrodescendientes, las poblaciones de San Andrés y Chocó aparecen más cercanamente relacionadas con las poblaciones Yoruba de Nigeria y Benin en Africa, que con las poblaciones afrocaribeñas, afrobrasileñas o afroamericanas y más lejos de poblaciones de Angola, de lengua Bantú y de Guinea (Figura 4-6). Al incluir en el análisis las muestras de los individuos predichos como pertenecientes a este mismo haplogrupo en La Guajira, Caribe Continental Colombiano y Amazonas, se observa que los individuos de La Guajira no se agrupan con ninguna otra población, mientras los individuos de Caribe Continental se acercan más a los grupos Afrocaribeños y afroamericanos, Chocó y San Andrés se agrupan con las poblaciones africanas y Amazonas se ubica en un punto medio con tendencia a acercarse con las poblaciones Africanas (Figura 4-7).

**Figura 4-6.** MDS para individuos E1b1a de San Andrés y Chocó con poblaciones de referencia afrodescendientes





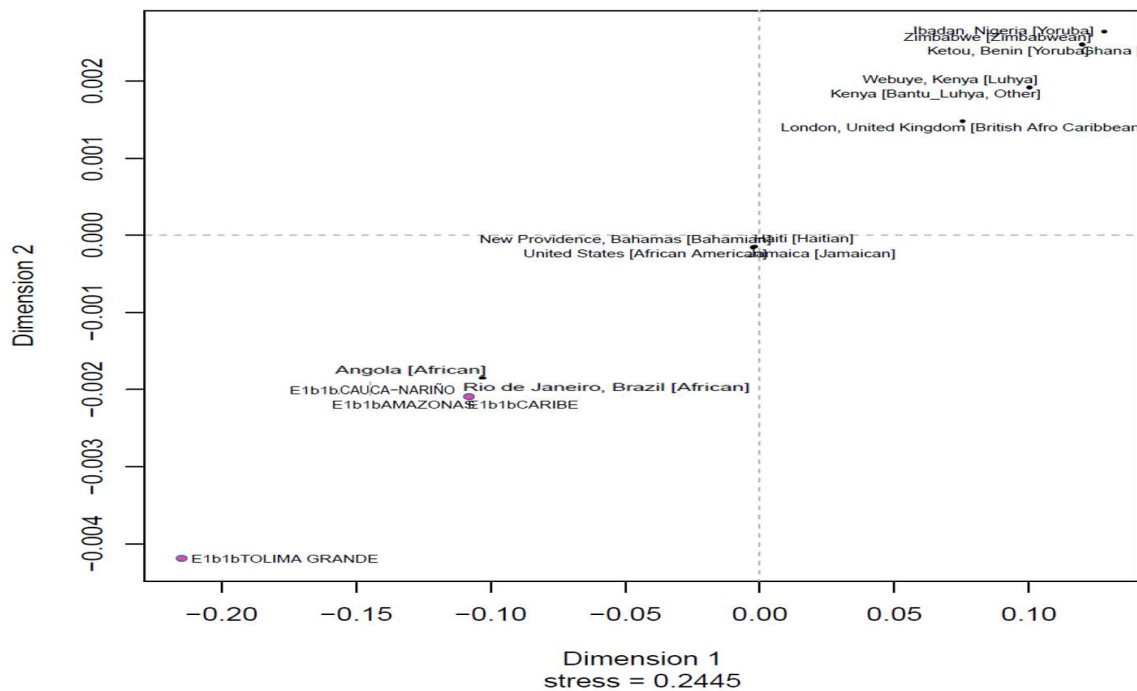
**Figura 4-7.** MDS para individuos E1b1a de todas las poblaciones colombianas con poblaciones de referencia afrodescendientes.



Para el análisis del haplogrupo de origen en Africa oriental E1b1b (Bajic, et.al., 2018), el MDS de las muestras de los individuos clasificados con el predictor de haplogrupos como pertenecientes a este conjunto de la Cauca-Nariño, Tolima grande, Caribe Continental y Amazonas comparándolo con 3761 datos de referencia (**Figura 4-8**), se observa que los individuos de Tolima grande no se agrupan ni con las otras poblaciones colombianas ni con las poblaciones de referencia, mientras las muestras de las otras 3 zonas se agrupan con las de Angola en Africa y las de afrodescendientes de Brasil. Otras dos agrupaciones se observan: una incluyendo Afroamericanos de Estados Unidos, y afrocaribeños de Bahamas, Haití y Jamaica y otro conformado por las poblaciones Africanas y Afrocaribeños británicos. Este haplogrupo, que tiene una baja frecuencia en todas las poblaciones

estudiadas, tiene su origen en Africa Oriental (Bajic et.al., 2018), sin embargo, se encuentra presente en poblaciones de la península Ibérica y el sur de Italia, haciendo parte del legado masculino de la ocupación musulmana por alrededor de 8 siglos, donde un gran componente del ejército invasor provenía del norte de Africa, previamente invadido por los árabes (Capelli, 2009), por lo que es más probable que la presencia de estos cromosomas en Colombia y Brasil se deba a que corresponden a linajes portados por los europeos colonizadores del siglo XV. De acuerdo con Capelli et.al., 2009, en una muestra de portugueses y españoles de Galicia, Cantabria y Andalucía, este haplogrupo se detectó entre el 4% y el 13% de los individuos estudiados. La cercanía con muestras de Angola, podría eventualmente ser explicada por la migración de los linajes originales africanos de este haplogrupo a la zona sur de Africa hace 2000 años, que ha sido descrita por Bajic, 2018. Por otro lado, no es posible descartar un aporte por migraciones más recientes.

**Figura 4-8.** MDS para individuos E1b1b de todas las poblaciones colombianas con poblaciones de referencia afrodescendientes.



En general se observa que los haplogrupos E1b1a son más frecuentes en las zonas costeras colombianas y disminuye en dirección al interior del país. De acuerdo con los datos históricos, los primeros africanos en América llegaron como acompañantes de sus amos en

el periodo de la conquista, y durante las primeras etapas del periodo de colonización algunos pocos eran transportados para cumplir fundamentalmente tareas domésticas o labores agrícolas (Navarrete, 2005). El tráfico directo de grandes números de esclavos desde Africa fue oficial desde 1518, tras diezmar la población indígena local. La población africana traída a América por los europeos provenía fundamentalmente en la costa occidental de Africa, entre los deltas del Rio Senegal y del Rio Níger y más adelante de la región de Angola, aunque los orígenes específicos fueron en ocasiones difíciles de establecer, debido a la costumbre de los traficantes de nombrar como lugar de procedencia, el sitio de captura o de embarque y no el verdadero origen de las personas. (Navarrete, 2005). La necesidad de mayor número de esclavos en las zonas de Mexico, Perú y Nueva Granada se dio a finales del siglo XVI. En el norte de Suramérica, el puerto de Cartagena era la principal zona de entrada de los esclavos africanos, muchos de los cuales eran enviados a la región pacífica para realizar labores de explotación minera, por lo que el censo de esta población en la época de la colonia, era mucho mayor en estas áreas, y una menor proporción de estos grupos era enviado hacia el interior por vía del río Magdalena (McFarlane, 1991). De acuerdo con un memorial presentado al consejo de Indias a mediados del siglo XVII, sobre Nueva Granada y Venezuela, se contaban en Cartagena 12.000 esclavos y un número equivalente en la costa de Tierra Firme, en Santa Marta, la Grita, Caracas, Cumaná, Margarita y Guayana; en Popayán 5000, el río grande de la Magdalena, Zaragoza 12000, en Remedios 8000, Antioquia y Anserma 2500, Mariquita 4000, Tocaima, Bagre, Muzo y Xalma 2000 (Navarrete, 2005). De acuerdo con el censo de población de 1778/1780, la población de la Nueva Granada estaba conformada por un 44% de "Mestizos", 31% de "Blancos", 18% de "Indios" y 7% de Esclavos africanos. Por regiones el porcentaje de población esclava era variable siendo menor del 5% en el interior, alrededor del 10% en la costa caribe y entre el 20 y hasta el 70% en varias poblaciones de la región del Pacífico colombiano (McFarlane, 1991).

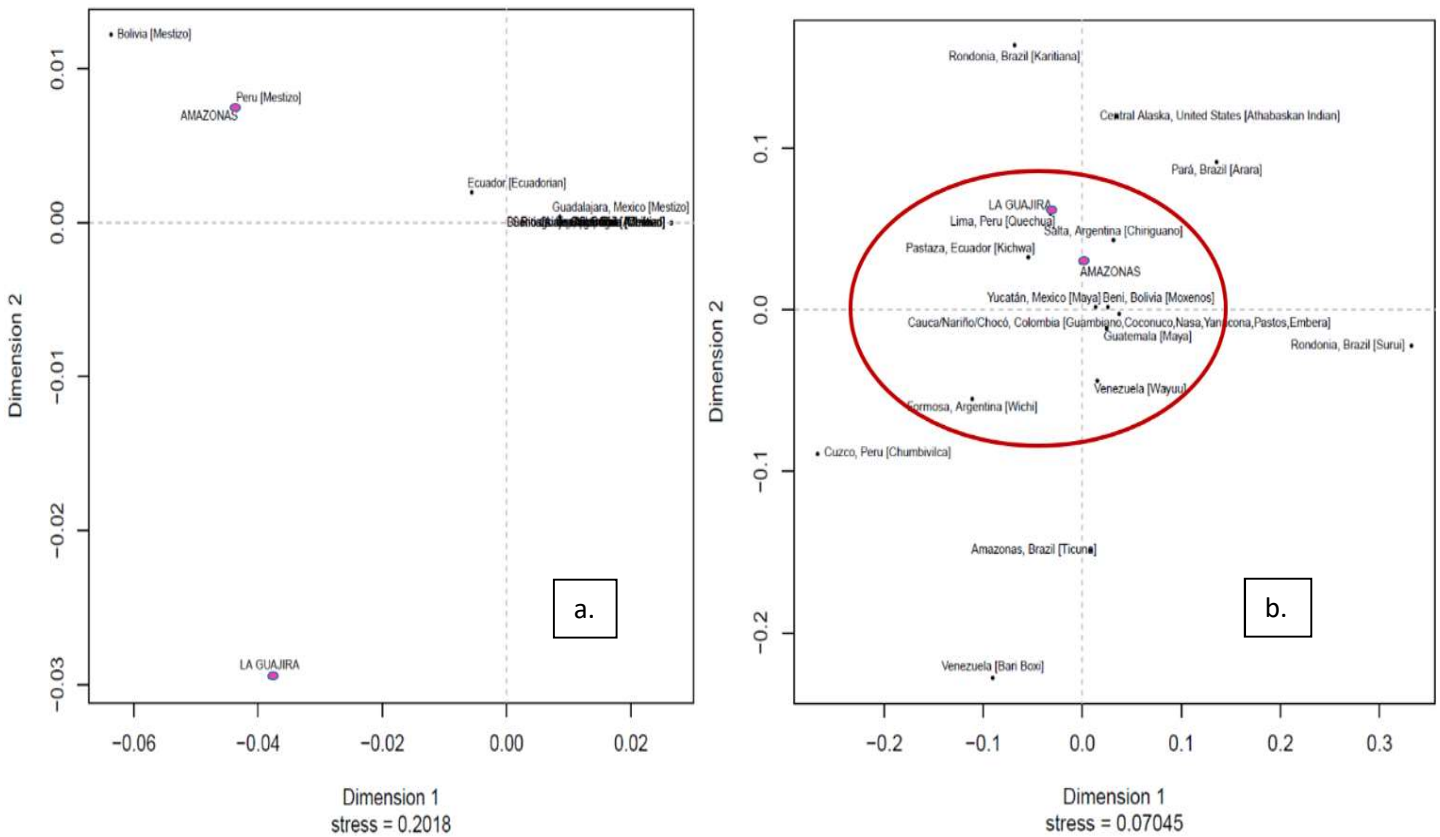
Diferentes acontecimientos históricos, sociales y culturales favorecieron tanto el aislamiento de las comunidades afrodescendientes en los llamados Palenques como producto del cimarronaje desde el siglo XVI, como los procesos de mezcla con poblaciones especialmente indígenas como sucedió en los "sitios de libres" y en las zonas mineras del Chocó, donde ya en el siglo XVII se dejaron registros de las uniones entre los dos grupos (Machado, 2007). Desde la abolición de la esclavitud en el siglo XIX, el abandono del

estado, las precarias políticas educativas, económicas y de salud en estas regiones y la más reciente historia de violencia y desplazamiento forzado, han generado una mayor migración hacia las grandes ciudades y hacia el interior del país y con ello eventos de mezcla. Es así como la amplia interacción económica y cultural del departamento del Valle del Cauca y del Chocó con la Zona Cafetera, genera patrones de migración reciente que se correlacionan con la presencia de un 20% de haplogrupos E, a diferencia de otras regiones del interior como la región Cundiboyacense y de Tolima grande, donde la frecuencia de estos haplogrupos es casi de la mitad.

En el Archipiélago de San Andrés y Providencia de acuerdo con el Perfil del censo general 2005 del DANE, el 0.1% de la población se autorreconoce indígena, el 17,6% afrodescendiente y el 39,4% raizal. Estas proporciones se correlacionan con la ausencia del haplogrupo Q en la muestra de esta región insular. Según datos históricos, en la época prehispánica las islas se constituían como corredor de comunicación en entre las zonas continentales, y se eran visitadas por indígenas Miskitos de origen chibcha habitantes de las costas de Centroamérica. La colonización de la isla por Ingleses y Holandeses y la posterior llegada de los esclavos africanos en 1635, así como la escasez de mujeres entre los colonizadores, generaron procesos de mezcla con las mujeres Miskitas (Meisel & Aguilera, 2016), por lo que probablemente se generó un marcado evento de deriva genética del cromosoma Y nativoamericano.

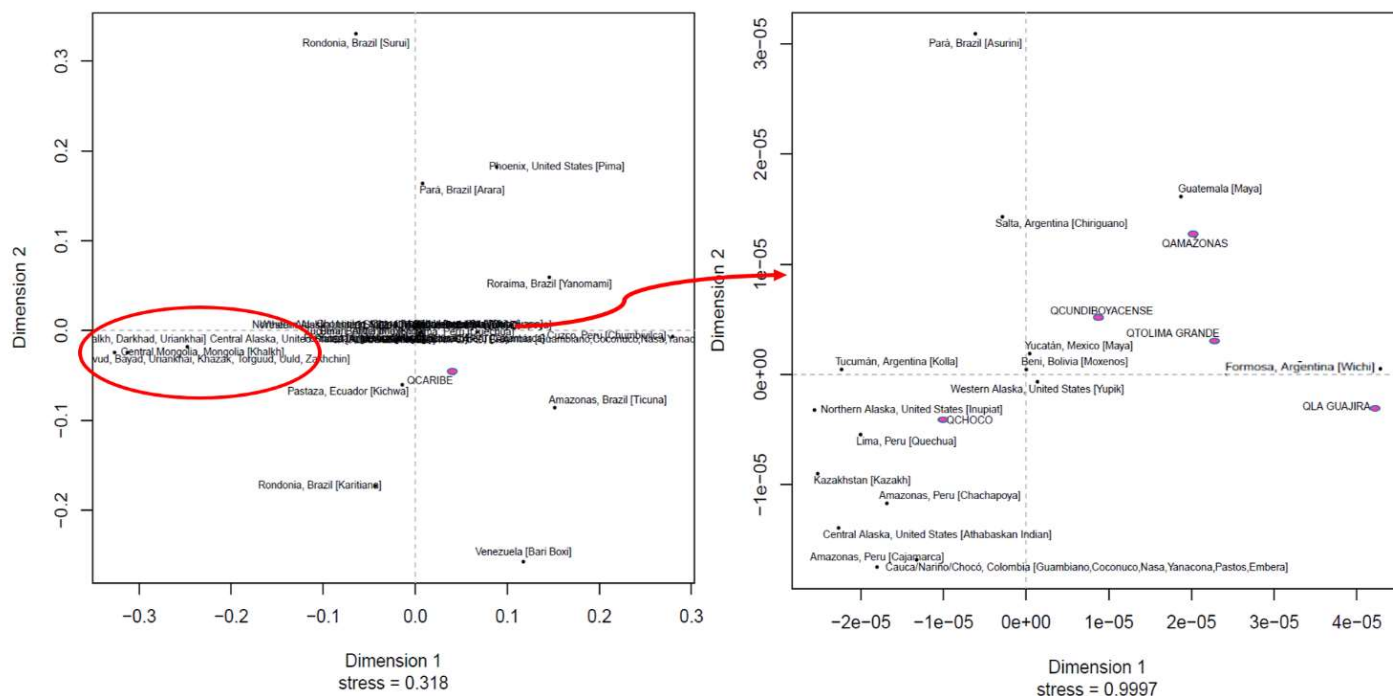
Las muestras de las poblaciones colombianas con mayor componente nativoamericano, fueron comparadas mediante MDS con otras poblaciones mezcladas latinoamericanas. Las muestras de La Guajira no se agrupan con otras poblaciones y las de Amazonas se acercan a las de Perú y Bolivia, que han sido descritas como portadoras de una alta proporción de ancestría nativoamericana (Salzano & Sanz, 2014) **Figura 4-9a**. Al considerar como referencia un conjunto de poblaciones indígenas de todo el continente Americano, La Guajira y Amazonas se encuentran dentro del cluster conformado por la mayoría de las poblaciones de análisis y solo hay separación de Surui, Karitiana, Arara y Ticuna de Brasil, Atabaskan de Alaska, Chumbivilca de Cuzco, Perú y Bari de Venezuela **Figura 4-9b**.

**Figura 4-9.** MDS para las poblaciones de Amazonas y La Guajira **a.** Respecto a otras poblaciones mezcladas Latinoamericanas. **b.** Respecto a grupos de referencia Natioamericanos.



Si se analizan solo las muestras clasificadas como haplogrupo Q por el predictor de haplogrupos, junto con las muestras de poblaciones nativoamericanas de referencia, las muestras de la región Caribe no se agrupan con las demás muestras colombianas, se encuentran más cercanamente ubicadas a las de Kichwas de Ecuador. Las muestras de las etnias Surui, Arara, Yanomami, Ticuna y Karitiana de Brasil, de Pimas de Estados Unidos y de Bari de Venezuela, se muestran dispersas y alejadas del cluster conformado por las otras etnias. Más cercanamente se observa un cluster formado por las muestras de poblaciones de Mongolia, y otro por las demás muestras de referencia y colombianas. En este último, las muestras de La Guajira, Tolima Grande, Amazonas y Cundiboyacense se agrupan cercanamente, mientras las de Chocó se ubican en el mismo agrupamiento de Inupiat en Alaska y Quechuas de Lima **Figura 4-10.**

**Figura 4-10.** MDS para las muestras del haplogrupo Q Colombianas respecto a grupos de referencia Nativoamericanos.



Los linajes Q disminuyen en frecuencia también hacia el interior. De acuerdo con los datos del censo nacional de población 2005 (DANE, 2010c), La Guajira y Amazonas forman parte de los departamentos del país con mayor proporción de población autorreconocida como indígena, con alrededor del 40%, Chocó y Nariño alrededor de 10%, Cauca 21%, Zona Cafetera, Cundiboyacense y Tolima Grande 5% o menos. Estos datos pueden estar influenciados por los criterios que las personas tienen en cuenta para considerarse pertenecientes a una etnia indígena, en muchos casos acompañado del desarraigo cultural.

Los resultados presentados en este estudio, encuentran mayor consistencia con informes provenientes de poblaciones Mexicanas, cuya proporción varía entre el 12% y el 49% (Martinez, et.al., 2012) y con los de otros trabajos de investigadores en Colombia como los de Acosta et.al., 2009, Rojas, et.al. 2010, Criollo, 2012, Noguera, et.al., 2014 y Xavier, et.al., 2015, con valores que oscilan entre el 2% y 60%; contrastan sin embargo con los datos obtenidos desde poblaciones mezcladas de Chile y Argentina (Toscanini, 2016 y Corach, 2010) donde la proporción de linajes Q se estimó entre el 0 y el 13%, o lo reportado

en Adhikari, 2017 en una publicación sobre la diversidad genética de América, que indica que aproximadamente el 90% de los linajes masculinos en Antioquia son Europeos.

De acuerdo con los resultados obtenidos en este estudio, la conservación de los linajes nativoamericanos alcanza en algunas regiones hasta el 50% de la población, apoyando la idea de que aunque en efecto la población indígena sufrió en Colombia una depresión demográfica significativa, un número importante de linajes de cromosoma Y sobrevivió y perduró.

En el momento de llegada de los españoles, las grandes culturas habitantes del territorio colombiano incluían, Wayuu, Urabaes, Paeces, Quillacingas, Panches, Muzos, Tumacos, Calimas, Pastos, Emberas, Awas, Taironas, Zenues, Muiscas y Quimbaya entre otras, donde la mayor densidad poblacional se concentraba a nivel de las sociedades más complejas y jerárquicas por ejemplo las cuatro últimas mencionadas (Zambrano, 2000), (Jaramillo, 1964). La reducción del censo poblacional se debió a enfrentamientos bélicos, al régimen de trabajo impuesto por los europeos, a las enfermedades traídas por los conquistadores, y como consecuencia de la destrucción de la organización económica y de la cultura que causó disolución de familias y tribus, el bajo interés por la reproducción y la esterilidad voluntaria (Jaramillo, 1964). Las estrategias de las etnias indígenas frente a la invasión de sus territorios fueron variadas, algunos grupos se pusieron voluntariamente al servicio de los recién llegados como los Zenues y en otras ocasiones establecieron relaciones matrimoniales como en algunos pueblos Muiscas, para conservar la paz, tratando de evitar el exterminio. Otros, prefirieron enfrentar a los conquistadores, directamente como los Tayronas y otras etnias llamadas Caribes por los cronistas por su agresividad, sufriendo numerosas bajas en la población. Los sobrevivientes y aquellos que viéndose superados abandonaron sus poblados, tuvieron que resguardarse en zonas de difícil acceso por altura o por condiciones de la selva tropical (Zambrano, 2000) (Uribe, 1993), que se han convertido en territorios ancestrales en los cuales aún habitan hoy sus descendientes, a quienes se les considera menos mezclados que otras etnias. Entre estos grupos se cuentan los Koguis, Arhuacos, Wiwas, Yukos, (cuyas zonas de resguardo se ubican en la Sierra Nevada de Santa Marta y en la Serranía del Perijá con influencia de su población en los departamentos de la Costa Caribe Colombiana), Emberas (con zonas de resguardo ubicadas en el departamento de Córdoba, la Zona Cafetera y en el Chocó

principalmente), y la mayoría de las tribus Amazónicas y de la Orinoquía donde varios reportes muestran una proporción de linajes Q de superior a 84% (Bortolini et.al., 2003), (Rojas, 2010), (Bisso-Machado et.al., 2012), (Xavier, 2015), (Franco, 2017).

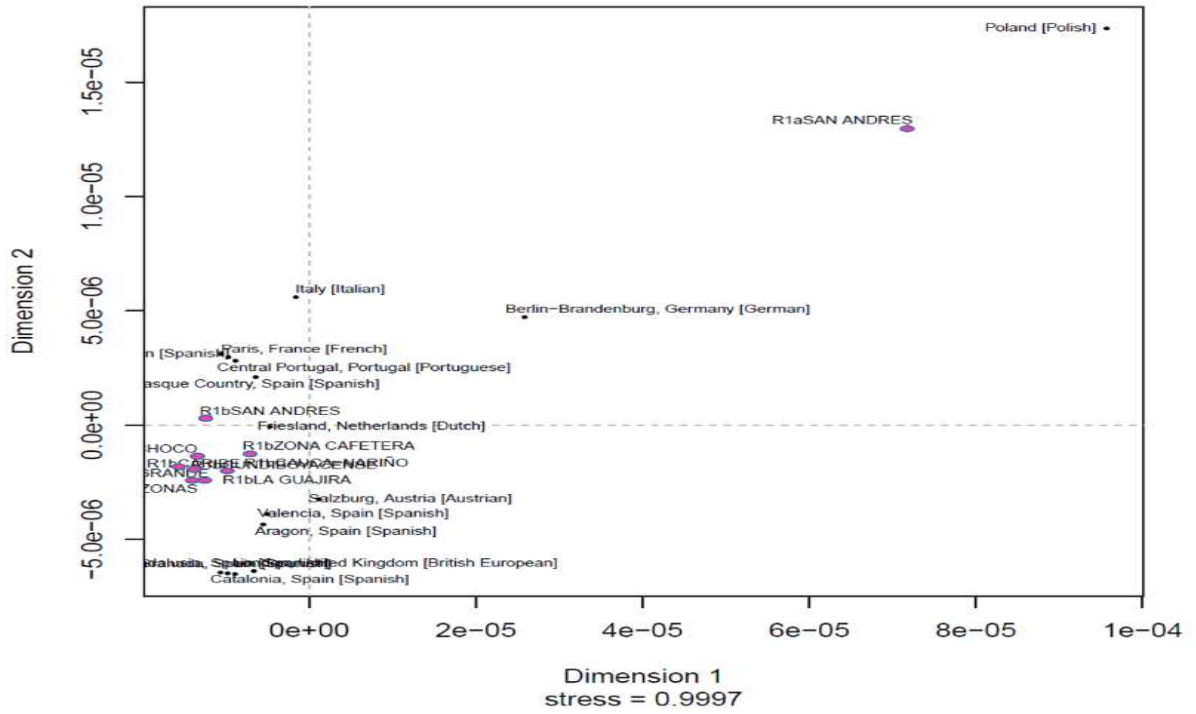
Las mayores proporciones de linajes Q alrededor del 50% encontrados en nuestro estudio en las muestras de La Guajira y del Amazonas, y en segundo lugar alrededor del 20% en la zona cafetera y del Caribe Continental Colombiano, concuerdan con los antecedentes históricos, sin desconocer que acontecimientos más recientes han golpeado con fuerza a estas comunidades, obligándolas a desplazarse a la zona urbana, en especial a las asociadas a sus zonas de influencia, creando movimientos migratorios de residencia permanente o provisionales relacionados con motivos laborales o de consecución de servicios, que han favorecido los procesos de mezcla. A nivel del centro del país, una mayor facilidad de comunicación entre las regiones y de acceso a los servicios, facilitan mayores eventos de mezcla y la disgregación de los cromosomas Y nativoamericanos.

Los individuos clasificados dentro de los haplogrupos R1b se compararon con poblaciones de referencia de Europa Occidental y de Europa Oriental. Las muestras de la región Cundiboyacense, Caribe Continental, Tolima grande, Cauca-Nariño, Zona cafetera, Amazonas, La Guajira y Chocó se agrupan muy cercanamente, mientras las de San Andrés lo hacen con las de población holandesa. Se destaca que los R1a de San Andrés se alejan de estas agrupaciones, acercándose a las muestras de Polonia. **Figura 4-11.**

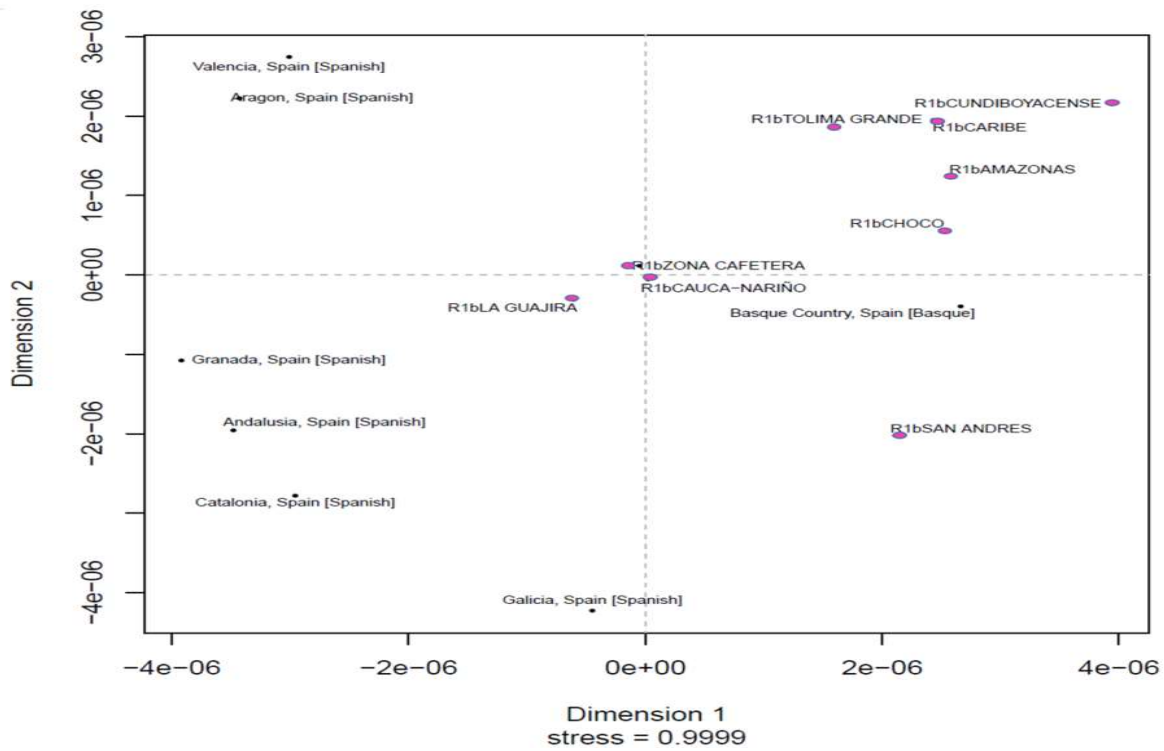
El resultado de un MDS considerando solamente poblaciones de diversas regiones de España para hacer la comparación, ubica las muestras Colombianas de San Andrés sin agruparse, las de Cauca-Nariño, Zona Cafetera y La Guajira en un agrupamiento y las restantes en otro cluster, pero en ningún caso acercándose espacialmente a ningunas de las poblaciones españolas de referencia. **Figura 4-12.**



**Figura 4-11.** MDS para las muestras del haplogrupo R1a y R1b Colombianas respecto a grupos de referencia Europeos.



**Figura 4-12.** MDS para las muestras del haplogrupo R1b Colombianas respecto a grupos de referencia de diferentes regiones de España.



La presencia de los haplogrupos R especialmente R1b, se ha relacionado con el aporte de linajes masculinos europeos posterior a la conquista. Aunque la proporción de esta contribución es variable, suele ser mayoritaria con respecto de los demás haplogrupos, en casi todas las poblaciones mezcladas americanas. Contando un promedio de 20 años por generación, al territorio americano las contribuciones más antiguas tendrán aproximadamente 26 generaciones. Las primeras formas de mezcla avaladas por la corona española, para el caso de América Latina, implicaban las uniones de hombres españoles y mujeres indígenas, además considerando que en principio la minoría de población extranjera era femenina (Jaramillo, 1964). La subsiguiente mezcla, la sensible baja de la población masculina nativa, establecimiento de las castas y ligado a ello los beneficios diferenciales así como la estratificación social, combinada con la necesidad de mejorar el estatus de los hijos y sus condiciones de vida facilitaron la expansión de estos linajes. (Jaramillo, 1964).

Las primeras exploraciones de los conquistadores europeos en Colombia sucedieron inicialmente en la Costa Caribe Colombiana y la región del Darién y el actual Urabá Antioqueño, donde alrededor de 1510 fracasaron los intentos de establecer asentamientos españoles en San Sebastián de Urabá y Santa María la Antigua del Darién. Posteriores expediciones hacia el interior del país fueron conducidas siguiendo el cauce del río Magdalena, bordeando la Costa Pacífica y luego por la zona montañosa sur occidental, a través de los Andes centrales y Orientales y desde Venezuela por la región del Orinoco. (Gamboa, 2013) (Pita, 2013). Las difíciles condiciones climáticas y de cultivo en las zonas tropicales del Chocó, el Amazonas y el Orinoco, no hacían estas tierras atractivas para los colonizadores, por lo que las primeras poblaciones fundadas por los conquistadores Españoles entre 1525 y 1580 se ubicaron en la Costa Caribe (Santa Marta y San Juan Nepomuceno (1525), Cartagena (1533), Sincelejo y Tolú (1535), Mompox (1537), Ciénaga (1538), Riohacha (1545), Valledupar (1550)), los Santanderes (Vélez (1539), Málaga (1542), Pamplona (1549) y Ocaña (1572)), Valle, Cauca y Nariño (Cali y Popayán(1536), Buenaventura y Pasto (1539), Cartago (1540), Santander de Quilichao (1543), Buga (1555), Ipiales (1568), y Roldanillo (1576)), Cundinamarca, Boyacá, Huila y Tolima (Timaná y Bogotá (1538), Tunja y Neiva (1539), Ibagué (1550), Honda (1560), Facatativá (1561), Villa de Leyva y Ortega (1572) y Fusagasugá (1580)), y en menor proporción en la Zona Cafetera (Marmato (1525) y Santafé de Antioquia (1546)). En las regiones de más difícil acceso y

subsistencia, la mayor parte de las fundaciones se llevaron a cabo en los siglos XIX y XX. Tal es el caso del Choco con el 60%, Meta con el 86% y Amazonas con el 100% de municipios fundados en siglo XIX y XX, y en Putumayo el 77% y en Caquetá, Vaupés, Guainía, Guaviare, Vichada con el 100% de fundaciones en el siglo XX. (Agustín, 2001).

La presencia de los demás haplogrupos en América se atribuye con mayor frecuencia a eventos migratorios recientes, provenientes de Europa, Asia y Oriente Medio, donde de acuerdo con la literatura, los diferentes haplogrupos se distribuyen con mayor frecuencias así: Haplogrupo G en la región del Cáucaso, el Mediterráneo y Oriente Medio, el Haplogrupo I en Europa, el haplogrupo J en Oriente Medio, el haplogrupo L en India, Pakistán, Iran y Afganistán y el haplogrupo T encontrado en bajas frecuencias en Europa, Africa y Medio Oriente (Semino, et.al., 2004), (Blanco, 2008). Sin embargo, no se puede descartar la posibilidad de que estos haplogrupos J provengan de la población española, como consecuencia de las invasiones de pueblos musulmanes de oriente medio en el siglo VIII (Capelli, et.al., 2009). De igual forma, haplogrupos H, I, J y R1 han sido encontrados en población ROM ibérica (Gusmao, et.al., 2008). En Colombia, de acuerdo con el censo poblacional del DANE, las comunidades ROM o gitanas se han establecido con mayor volumen en la zona caribe colombiana, aunque existen Kumpanias en varias regiones del país y su cultura suele proteger las uniones con las personas de su propio pueblo y forman comunidades relativamente aisladas (DANE, 2007). Durante finales del siglo XIX, pero con mayor grado en la primera mitad del siglo XX, varias migraciones provenientes de Europa, especialmente migraciones sirio-libanesas, llegaron a la zona de la costa atlántica donde se establecieron comercialmente. Algunos migraron al interior del país y se radicaron en la zona cafetera y en el Tolima, aunque en menor número (Hincapié, 1995). De igual forma, se recibieron en Colombia migrantes de Europa Oriental desplazados por las guerras mundiales, muchos de los cuales llegaban en barcos, a los puertos del caribe donde se establecieron.

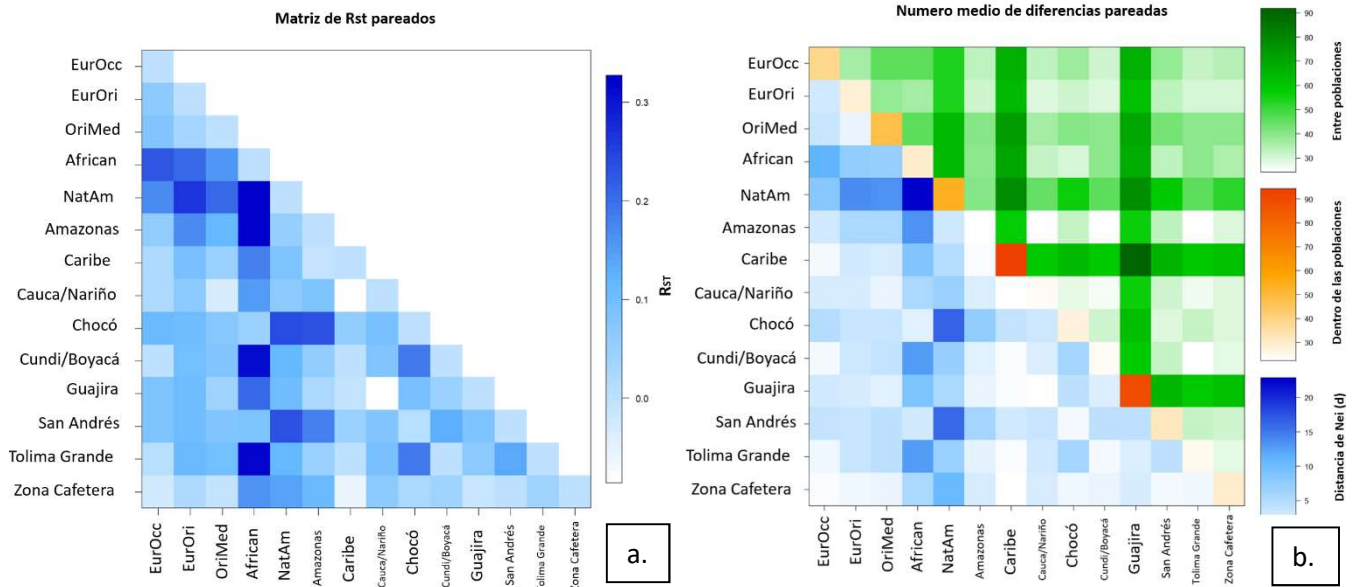
Patrones diferenciales de colonización europea se dieron como consecuencia de las dificultades para ingresar a los territorios de las selvas Amazónicas, territorios de la Orinoquía y la selva Chocoana, por su difícil acceso y por las condiciones climáticas de difícil adaptación, así como las enfermedades tropicales (Machado, 2007). En territorios de la costa caribe o la región central, siendo las zonas más tardías de colonización las regiones

montañosas del sur de Antioquia y la zona cafetera, los patrones de colonización fueron diferentes en incluso en algunos casos, medió la cultura predominante de evitar la mezcla, que fue favorecida por cierto grado de aislamiento geográfico durante un largo periodo de tiempo (Hincapié, 1995). Estas particulares condiciones históricas en Colombia, han tenido como consecuencia la distribución diferencial de las frecuencias de haplogrupos a nivel de diferentes regiones del territorio, similar a lo reportado para países como Panamá (Grugni, et. al., 2015), y en contraste con otras regiones más uniformes donde aparentemente hay menor complejidad (Corach, et.al., 2010; Resque, et.al., 2016; Martínez, et.al., 2012).

### **4.3.1 Estructura Poblacional y Relaciones Genéticas**

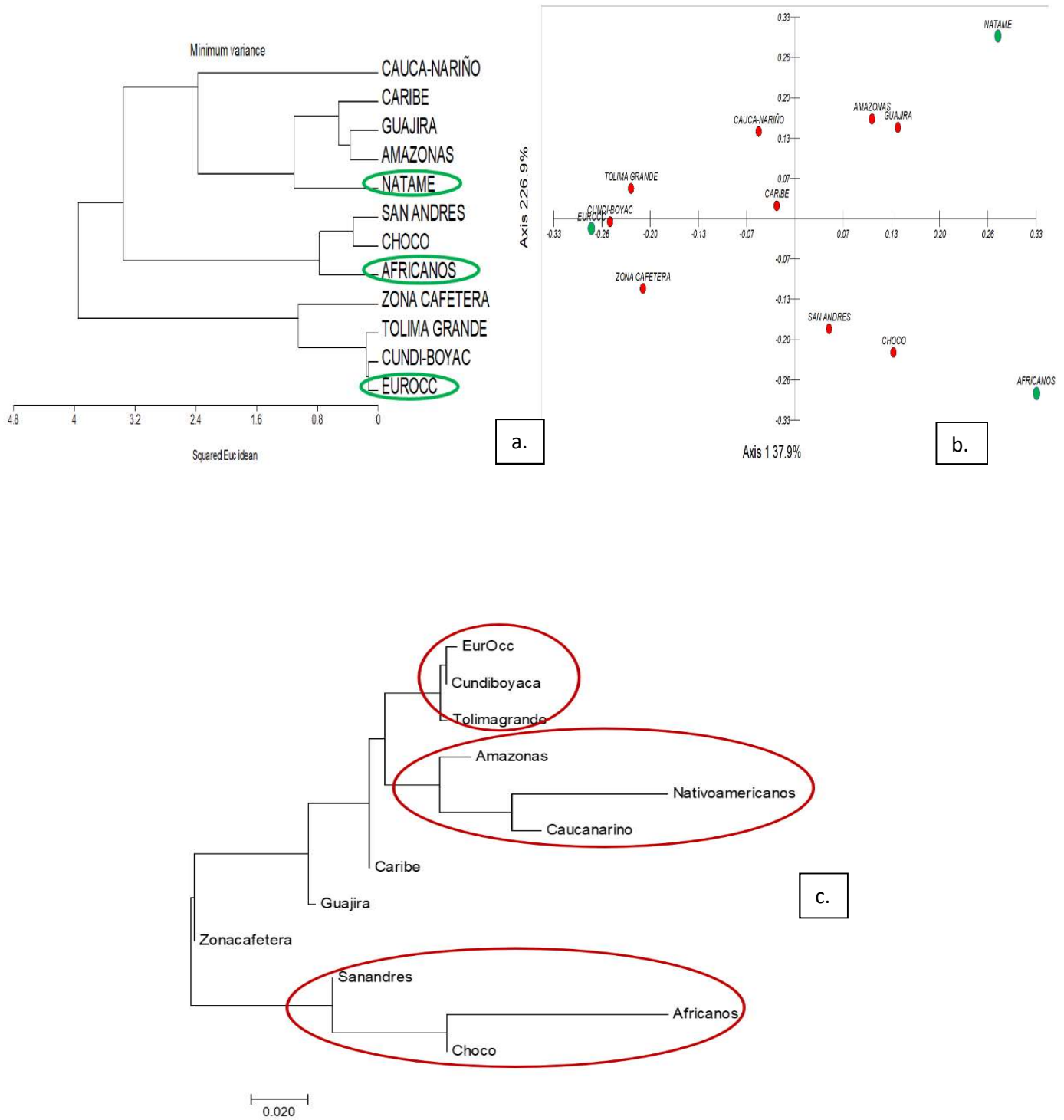
Para el análisis de la estructura de las poblaciones colombianas, se predeterminaron 5 grupos: Amazonas, Caribe-Guajira, Cauca/Nariño, Chocó-San Andrés y Cundi/Boyacá-Tolima Grande-Zona Cafetera. El AMOVA indicó que la variación más alta corresponde a las diferencias dentro de las poblaciones ( $F_{ST} = 0.07357$ ,  $p$  valor= 0.0000). A una escala mayor, todas las poblaciones colombianas fueron agrupadas para compararse con poblaciones de referencia Europeas (Europa Occidental y Europa Oriental), de Oriente Medio, Africanas y Nativoamericanas. Los resultados del AMOVA fueron los mismos que en el caso anterior ( $F_{ST} = 0.16121$ ,  $p$  valor= 0.0000). El análisis de Rst pareados muestra que la población del caribe continental tiene diferencias significativas con todas las poblaciones de referencia y con Chocó y San Andrés. Estas dos últimas no se diferencian entre sí, ni con Cauca-Nariño o con Zona Cafetera. La población de Amazonas tiene diferencias significativas con todas las demás poblaciones, excepto con Caribe y Cauca-Nariño. Esta última tiene diferencias significativas solo con las poblaciones de referencia Africanas, sin embargo, este resultado debe ser interpretado con precaución, pues puede estar influenciado por el pequeño tamaño de muestra. En general, estos resultados son consistentes con lo indicado en el número medio de diferencias pareadas, donde además se encuentra que la mayor variación tanto dentro como entre las poblaciones, se observa en la región Caribe y de La Guajira. (**Figura 4-13 a y b**)

**Figura 4-13.** Matriz de Rst pareados (a) y Número Medio de diferencias pareadas (b) incluyendo tanto las poblaciones colombianas como las 5 poblaciones de referencia Europa Occidental, Europa Oriental, Oriente Medio, Africanos y Nativoamericanos.



El árbol Neighbor Joining construido a partir de la matriz de distancias Rst, y tanto el Análisis de Componentes Principales como el Análisis de clusters a partir de las frecuencias alélicas de los STRs de Cromosoma Y, muestran cercanía entre las poblaciones Cundiboyacense y Tolima Grande con las poblaciones de referencia de Europa Occidental, las muestras de Amazonas y Cauca-Nariño con las poblaciones de referencia Nativoamericanas, y un cluster separado de los demás que agrupa Chocó y San Andrés con las de referencia Africanas. **Figura 4-14.** La diferencia en el agrupamiento para las muestras de Cauca Nariño y de Zona Cafetera en los dos dendogramas, puede estar relacionado con el tamaño de la muestra para estas poblaciones.

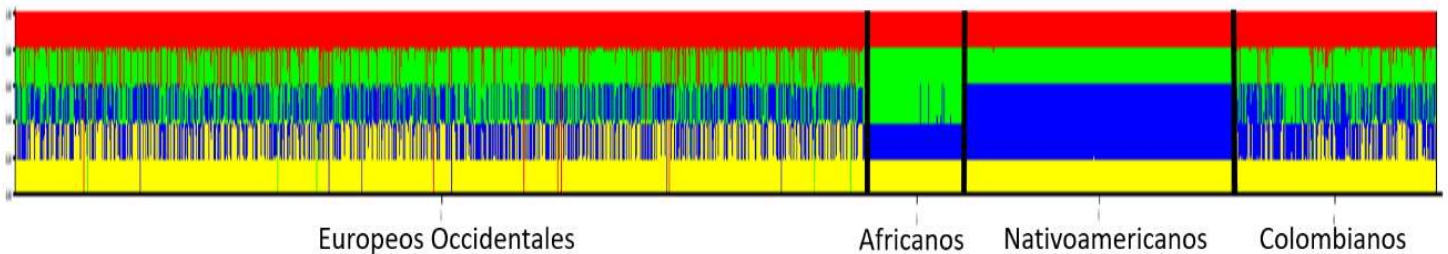
**Figura 4-14.** a. Análisis de clusters por mínima varianza y b. Análisis de Componentes Principales a partir de frecuencias alélicas. c. Arbol Neighbor joining construido con Mega 7 a partir de matriz Rst.



### 4.3.3 Exploración de datos de posible ancestría

Las muestras correspondientes a las poblaciones de referencia fueron inicialmente analizadas con el software Structure probando K=2 a K=8 y el análisis con Structure Haverster determinó el mejor como K=4, aunque en el caso de la población de Europa Occidental, se observan tres patrones de mezcla. **Figura 4-15.** Los resultados mostraron que para las muestras Natioamericanas el componente mayoritario fue el 3, mientras para las muestras africanas fue el 2. Al incluir 389 muestras de las poblaciones colombianas para estudio, se estimó la proporción de cada componente por individuo y se asignó la ancestría más probable siguiendo el patrón obtenido para las muestras de referencia en cuanto al aporte mayoritario. Solo para 8 de las muestras la clasificación fue indeterminada, 161 se asignaron como de ancestría africana, 144 europea y 76 nativoamericana.

**Figura 4-15.** Bar plot obtenido con Structure para k=4



La evaluación analítica de los resultados se llevó a cabo como se describe en el capítulo 2, considerando la asignación hecha previamente con el predictor de haplogrupos como test de referencia y la asignación por Structure de acuerdo con el patrón de las muestras de referencia. Para este caso solo se tuvieron en cuenta las 363 muestras cuyos resultados de probabilidad de asignación con el predictor fue superior al 0.8. Los resultados se muestran en la **Tabla 4-6.**

De acuerdo con estos resultados, la mayor eficiencia en el análisis se obtiene para la determinación de la ancestría nativoamericana Vs no ancestría nativoamericana en la evaluación general, y en la evaluación pareada, especialmente para distinguir entre ancestría Europea y Natioamericana. Para el caso de la ancestría europea se obtiene una alta especificidad pero una baja sensibilidad, lo que indica que los marcadores empleados

son más útiles para determinar a los individuos con mayor probabilidad de ancestría no europea, que para detectar a los individuos con ancestría europea y generando mayores proporciones de falsos negativos o individuos que teniendo un linaje masculino europeo no son detectados como tales por el análisis con Structure. La situación inversa se observa en la evaluación general para las poblaciones de origen africano, donde la sensibilidad es mayor, pero el menor valor de especificidad incrementa la probabilidad de encontrar falsos positivos, es decir individuos a los que se les considere pertenecientes a linajes masculinos de origen africano que realmente no tengan esa pertenencia étnica. Como consecuencia de estas circunstancias, la evaluación analítica pareada entre europeos y africanos, no resulta del todo eficiente.

**Tabla 4-6.** Evaluación analítica comparando la predicción de haplogrupo (Predictor de Haplogrupos de Athey) y la asignación de ancestría por Structure 2.3.4

INDICADOR	EVALUACION GENERAL			EVALUACION PAREADA		
	Eur Vs No Eur	Afr Vs No Afr	Nam Vs No Nam	Eur Vs Afr	Eur Vs Nam	Afr Vs Nam
Sensibilidad	0.697	0.941	0.971	0.700	0.993	0.941
Especificidad	0.955	0.770	0.979	1.000	1.000	0.971
Cociente Falso Positivo	0.000	0.230	0.021	0.000	0.000	0.029
Cociente Falso Negativo	0.303	0.059	0.029	0.300	0.007	0.059
VPP	1.000	0.563	0.918	1.000	1.000	0.976
VPN	0.684	0.977	0.993	0.571	0.985	0.931
Prevalencia real	0.566	0.239	0.194	0.714	0.678	0.552
Prevalencia aparente	0.394	0.225	0.189	0.500	0.673	0.519

La revisión más cercana de las muestras de referencia de Europa occidental revela al menos 4 patrones diferentes de mezcla. Individuos con componente mayoritario 1, con componente mayoritario 4 o con componentes 1 y 4 altos y 2 y 3 bajos. Sin embargo el 34% de las muestras presentan un componente mayoritario 2, igual que el patrón de las muestras africanas. Sin embargo al realizar la predicción del haplogrupo, las muestras europeas se asignan con altos valores de probabilidad en los haplogrupos I, J, G y las muestras africanas el E1b1a y E1b1b. Esta característica, explica los resultados obtenidos en la evaluación analítica para estas poblaciones. Si se considera la evaluación analítica para los individuos de estudio asignados como R1b por el predictor de haplogrupos, los resultados son mucho más eficientes. **Tabla 4-7**



**Tabla 4-7.** Evaluación Analítica considerando solo individuos de estudio asignados como R1b

SOLO R1b	EVALUACION GENERAL			EVALUACION PAREADA		
INDICADOR	Eur Vs No Eur	Afr Vs No Afr	Nam Vs No Nam	Eur Vs Afr	Eur Vs Nam	Afr Vs Nam
Sensibilidad	0.986	0.941	0.971	0.986	1.000	0.941
Especificidad	1.000	0.981	0.978	1.000	1.000	0.971
Cociente Falso Positivo	0.000	0.019	0.022	0.000	0.000	0.029
Cociente Falso Negativo	0.014	0.059	0.029	0.014	0.000	0.059
VPP	1.000	0.952	0.931	1.000	1.000	0.976
VPN	0.987	0.976	0.991	0.976	1.000	0.931
Prevalencia real	0.480	0.287	0.233	0.640	0.676	0.552
Prevalencia aparente	0.473	0.270	0.226	0.631	0.676	0.519

Los resultados obtenidos para cada población se muestran en la **Tabla 4-8**. En comparación con la proporción de haplogrupos por región presentada en la sección 4.3.1.3, se observa una sobreestimación de la ancestría africana y una desestimación de la ancestría europea, relacionada con los datos analizados arriba.

Adicionalmente se incluyó el valor *kappa* estimado para evaluar la concordancia de la clasificación realizada por Structure y por la asignación hecha por el predictor de haplogrupos. En todas las regiones la concordancia más alta se da para la asignación de ancestría nativoamericana que varía entre 0.82 y 1. De acuerdo con los resultados de la evaluación analítica, la asignación de ancestría europea y africana tiene menores valores de concordancia, la primera entre 0.44 y 0.82, y la segunda entre 0.21 y 0.82. Los valores de concordancia 0 observados para la población de Cauca-Nariño y Chocó, indican que la concordancia observada no puede ser atribuida a factores diferentes al azar, sin embargo debe tenerse en cuenta que el pequeño tamaño de muestra puede estar influenciado este resultado. Los valores de concordancia superiores a 0.7 se consideran óptimos, por encima de 0.5 buenos y aceptables si son mayores a 0.4 (Santos, J., 2011), lo cual significa que solo el 18.5% de las concordancias evaluadas no son aceptables, es decir la ancestría estimada para Cauca-Nariño, la ancestría nativoamericana en la población del Chocó y la ancestría Africana en las poblaciones de Tolima Grande y Cundiboyacense. En comparación con la proporción de haplogrupos por región presentada en la sección 4.3.1.3, se observa una sobreestimación de la ancestría africana y una desestimación de la ancestría europea.

**Tabla 4-8.** Resultados de la asignación de ancestría a partir de los datos del software Structure por regiones y valores de concordancia con la predicción de haplogrupos (Athey)

REGION	N	ANCESTRÍA	INDIVIDUOS	%	CONCORDANCIA ( <i>kappa</i> )
AMAZONAS	47	EUROPEA	9	19.1	0.76
		AFRICANA	7	14.9	0.49
		NATIVOAMERICANA	29	61.7	0.82
		INDETERMINADO	2	4.3	
CARIBE	44	EUROPEA	16	36.4	0.65
		AFRICANA	15	34.1	0.44
		NATIVOAMERICANA	13	29.5	0.89
		INDETERMINADO	0	0.0	
CAUCA-NARIÑO	5	EUROPEA	2	40.0	0.62
		AFRICANA	3	60.0	0.62
		NATIVOAMERICANA	0	0.0	
		INDETERMINADO	0	0.0	
CHOCO	47	EUROPEA	12	25.5	0.71
		AFRICANA	34	72.3	0.68
		NATIVOAMERICANA	1	2.1	0.66
		INDETERMINADO	0	0.0	
CUNDI-BOYACA	47	EUROPEA	28	59.6	0.44
		AFRICANA	13	27.7	0.21
		NATIVOAMERICANA	5	10.6	1.00
		INDETERMINADO	1	2.1	
GUAJIRA	21	EUROPEA	3	14.3	0.58
		AFRICANA	6	28.6	0.58
		NATIVOAMERICANA	11	52.4	1.00
		INDETERMINADO	1	4.8	
SAN ANDRES	50	EUROPEA	17	34.0	0.79
		AFRICANA	31	62.0	0.79
		NATIVOAMERICANA	0	0.0	0.00
		INDETERMINADO	2	4.0	
TOLIMA GRANDE	92	EUROPEA	51	55.4	0.54
		AFRICANA	27	29.3	0.33
		NATIVOAMERICANA	12	13.0	0.95
		INDETERMINADO	2	2.2	
VALLE	2	EUROPEA	0	0.0	0.00
		AFRICANA	1	50.0	1.00
		NATIVOAMERICANA	1	50.0	1.00
		INDETERMINADO	0	0.0	
ZONA CAFETERA	7	EUROPEA	2	28.6	0.28
		AFRICANA	4	57.1	0.22
		NATIVOAMERICANA	1	14.3	1.00
		INDETERMINADO	0	0.0	0.00

Dado que el manual de software Structure indica que los análisis realizados con marcadores de este tipo deben ser cuidadosamente interpretados, pues los algoritmos podrían generar

resultados sesgados o inexactos, y por tanto no dar una representación real de la estructura poblacional en sí, en consulta personal realizada con el doctor Pritchard, ha indicado que el uso de esta herramienta puede ser útil para visualizar los datos, pero que el modelo no es válido debido a que no hay recombinación en el cromosoma Y.

## 4.4 Conclusiones

La reconstrucción de la historia genética de las poblaciones humanas, como producto de la interacción de las fuerzas de cambio evolutivo y los patrones culturales, las formas de relación con el ambiente, las estrategias económicas comerciales y de dominio y colonización, así como de las barreras geográficas y grandes cambios climáticos y otros fenómenos de la naturaleza, puede dilucidarse a partir del análisis de la estructura genética de las poblaciones humanas actuales, tanto conservadas como mezcladas, empleando diferentes niveles de profundidad y tipos de marcadores.

Esta historia en América, tierra de mezcla triétnica, ha sido descrita desde enfoques variados con base en distintas clases de evidencia, en diferentes partes del territorio. El análisis genético poblacional de los linajes masculinos permitió relacionar haplogrupos de cromosoma Y específicos para cada uno de los tres componentes, siendo los linajes Q los más ancestrales, llegados con los primeros pobladores. Las frecuencias de linajes masculinos en el noroccidente de Suramérica se relacionan con los patrones históricos del origen de las poblaciones, como se observa por ejemplo en los menores valores de diversidad de haplogrupos en la región de San Andrés, lo cual se relaciona con un menor número de poblaciones originales (mínimo aporte de cromosomas Y nativoamericanos), siendo más significativa la presencia de esclavos africanos y hombres europeos, así como un menor flujo genético posterior dada la condición geográfica de la isla y la dificultad para la población indígena de desplazarse y establecerse en esa región, junto con la migración efectiva de cromosomas Y Europeos, más frecuentes en la población continental colombiana actual. De igual forma, la baja diversidad haplotípica para esta región puede relacionarse con pequeños tamaños efectivos de población originales, poco tiempo de divergencia, y poco flujo genético efectivo. Por otro lado, aunque el tamaño de muestra es bajo para la zona cafetera, se encuentran valores altos de diversidad genética y haplotípica,

lo cual puede relacionarse con la posición geográfica en donde interactuaron las poblaciones de origen indígena, africano y europeo, pero adicionalmente, por la dinámica comercial y migratoria de la región gracias a la cual se ha generado una migración efectiva de población mezclada con linajes de cromosoma Y de diferentes orígenes, de forma constante desde la colonización europea, aunque con mayor énfasis durante el siglo XIX (Agustín, 2001).

Los resultados del AMOVA muestran la existencia de subestructura genética en la población Colombiana, percibiéndose las mayores diferencias entre las poblaciones de Chocó y San Andrés con las de Amazonas, Tolima Grande y Cundinamarca-Boyacá, conclusión apoyada también en el análisis de relaciones genéticas mediante el árbol neighbor joining y en el análisis de componentes principales, donde se observa la cercanía de muestras colombianas de diferentes subregiones a muestras de referencia nativoamericanas, europeas y africanas. De igual forma, el análisis de frecuencias de haplogrupos refleja 3 modelos de estructura poblacional, cada uno de los cuales se caracteriza por la presencia mayoritaria de un haplogrupo representativo de las poblaciones europeas, africanas o nativoamericanas.

Aunque claramente se observa que las muestras de San Andrés y Chocó comparten un modelo similar, los resultados de esta investigación muestran que la diversidad de haplotipos en las dos regiones es diferente, siendo los de San Andrés más cercanos a los de la población de Angola y los de Chocó más relacionados a los de la región Afrocaribeña. Al considerar sin embargo solamente las muestras clasificadas como pertenecientes al haplogrupo E1b1a en ambas localidades, estas son más cercanas a la población Yoruba de Africa, por lo cual se puede considerar que las diferencias observadas en las dos poblaciones, provienen más probablemente de la mezcla efectiva posterior a la conquista, que ha variaciones en la población esclava original que se concentró en estas áreas, de acuerdo con los datos históricos.

De otro lado, la presencia del haplogrupo E1b1b, de origen africano, se atribuye a la población europea que portaba estos linajes de cromosoma Y, como consecuencia de las invasiones musulmana del siglo VIII, lo cual también es respaldado por la estimación de la

frecuencia de este haplogrupo aparentemente inversa a la del haplogrupo E1b1a, en las diferentes regiones Colombianas analizadas.

Cabe destacar también, que aunque las muestras de La Guajira corresponden a personas autorreconocidas como pertenecientes a la etnia Wayuu, se observa introgresión de población masculina de origen europeo y africano. Sin embargo, estas muestras son diferentes de otras poblaciones mezcladas con alto componente indígena y los haplotipos de sus linajes Q difieren también de los de otras etnias nativoamericanas. Por otro lado, a pesar de que datos históricos relacionan la pertenencia de grupos de colonizadores a regiones específicas de España, con los 17 marcadores analizados, no se encontró una relación aparente con ninguna de las subregiones colombianas estudiadas.

En general, el legado histórico de los diferentes linajes masculinos tanto ancestrales como de la ocupación europea y del arribo de la población africana, son encontrados en las poblaciones mezcladas actuales, además del rastro de los posteriores aportes realizados por eventos migratorios portadores de otros linajes, creando una compleja historia de mezcla, altamente favorecida por la posición geográfica colombiana, que la hizo paso obligado tanto de las migraciones paleolíticas de los primeros pobladores, así como lugar de fundación de los primeros asentamientos europeos y puerto de arribo de los grandes barcos de comercio de esclavos, y de forma más reciente, punto de entrada por vía marítima de migrantes europeos y de medio oriente.

Finalmente, la exploración de datos para un posible análisis de ancestría de cromosoma Y de las muestras colombianas usando el software Structure, empleando datos de poblaciones de referencia de los tres orígenes étnicos, obtuvo como el mejor  $K=4$ . Los resultados de la evaluación analítica y el análisis de concordancia, son aparentemente favorables para la discriminación de la ancestría Nativoamericana, pero no tan precisas para la africana y europea, aunque para esta última, se obtienen mejores resultados si solo se consideran haplogrupos R1b. Aumentar el número de marcadores usados o las muestras de las poblaciones de referencia africanas o una mejor selección de las muestras de referencia europeas podría mejorar los resultados de la asignación de ancestría de linajes de cromosoma Y en poblaciones de origen triétnico. Sin embargo, dado el antecedente histórico con relación al legado de linajes provenientes de las invasiones

musulmanas en Europa, esta tarea podría ser más compleja. Debe tenerse en cuenta también, que estas conclusiones deben ser cuidadosamente consideradas, considerando por un lado el uso de predictor de haplogrupos para estas comparaciones, por lo que sería entonces pertinente, validar estos resultados realizando directamente la asignación de haplogrupos mediante el uso de SNPs, y por otro tener en cuenta las apreciaciones del doctor Pritchard con relación a que no es posible considerar como válido un modelo de estructura poblacional en sí, generado a partir de datos de un marcador uniparental.

# **5. Capítulo 5. Linajes de ADN Mitocondrial y de Cromosoma Y nativoamericanos, revelan la historia de poblamiento del norte de Suramérica**

## **5.1 Introducción**

La historia del poblamiento americano ha sido una cuestión abordada desde perspectivas diferentes en la antropología, el lenguaje y la genética entre otras, y que ha reunido esfuerzos multidisciplinarios para encontrar escenarios que integren las hipótesis y consigan el mayor grado de “certeza” posible con respecto a las rutas, tiempos y número de individuos en las poblaciones fundadoras (Cavalli-Sforza, et. al., 1992; Cavalli-Sforza, 1997; Ramachandran, et.al., 2005; Pitblado, 2011).

Tras los hallazgos de linajes masculinos y femeninos propios de las poblaciones nativoamericanas (Torróni et.al., 1993; Santos et.al., 1999; Lell et.al., 2002; Bortolini et.al., 2003; Shurr, 2004; Pitblado, 2011; Dulik, et.al., 2012; Regueiro, et.al., 2013), el uso de los marcadores uniparentales se implementó como herramienta para dilucidar los aspectos relacionados con las teorías de poblamiento americano y la relación entre la diversidad genética, la distancia geográfica y la variación del lenguaje (Roewer et.al., 2013; Keyeux, et.al., 2002; Bisso-Machado et.al., 2012), pasando del nivel continental a los estudios subcontinentales o enfocados en regiones específicas y empleando metodologías cada vez más robustas, considerando marcadores tipo RFLPs, STRs y llegando hasta los SNPs y

genomas completos (Cavalli-Sforza et.al., 1994; Torroni, 1994; Battaglia et.al., 2013; Tarazona-Santos et.al., 2001; Bodner, et.al., 2012; Keyeux, et.al., 2002; Melton, et.al., 2007)

En el escenario Suramericano se ha considerado que la posición geográfica de Colombia como lugar de entrada de las primeras migraciones humanas, juega un papel importante en el análisis de las rutas de dispersión y que por su característico paisaje, que cuenta con el extremo terminal de la cordillera de los Andes, dos regiones costeras separada y selvas húmedas tropicales, es posible analizar el concepto de aislamiento geográfico y su relación con la diversidad genética.

En este contexto, el estudio de la diversidad de las poblaciones indígenas colombianas, en busca de los linajes ancestrales nativoamericanos, con base en muestreos dirigidos con soporte de encuestas que permitan seleccionar las muestras por sus genealogías, además de conocer los parámetros culturales asociados a la selección de parejas y comportamientos migratorios, permite explorar con mayor robustez las relaciones genéticas entre las etnias colombianas y esto visto en el marco de las teorías de poblamiento americano.

Este trabajo presenta los resultados obtenidos del análisis de los linajes Q de cromosoma Y y de los linajes A, B, C y D de ADN mitocondrial, de muestras tomadas en poblaciones mezcladas del Caribe continental y el Amazonas biogeográfico colombiano con historias de poblamiento por comunidades indígenas, cuyas genealogías apoyaran su autodeterminación como pertenecientes a una etnia en particular.

Los resultados obtenidos se analizaron desde la perspectiva genética, lingüística y geográfica, que consideramos dan su apoyo a la teoría de la separación temprana de los primeros pobladores a su llegada a Suramérica, discutida entre otros autores por Tarazona-Santos et.al. , Wang et.al. y Keyeux et.al., destacando la presencia de una muestra de Amazonas con un linaje D1f, propuesto por Kummar et.al. 2011 como un subhaplogrupo panamericano de contemporáneo a los haplogrupos fundadores.



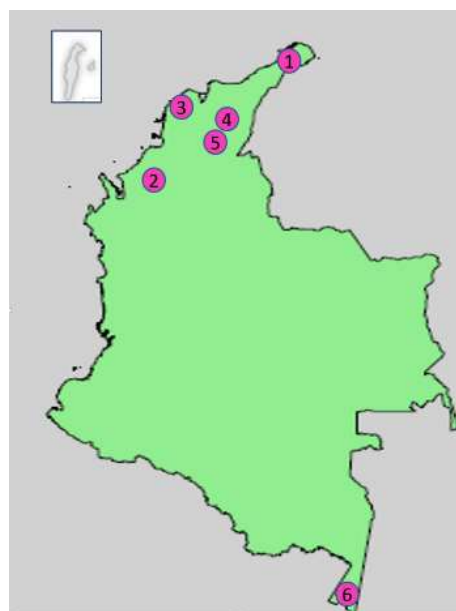
## 5.2 Materiales y Métodos

### 5.2.1 Poblaciones y encuestas

Las muestras correspondientes a las poblaciones colombianas empleadas en este estudio, han sido recolectadas por los investigadores de Grupo de Genética de Poblaciones e Identificación, del Instituto de Genética de Universidad Nacional de Colombia. **Figura 5-1.** En todos los casos se aplicó una encuesta para obtener información adicional de tipo genealógico y demográfico. 51 muestras de sangre total y/o de raspado de células epiteliales de cavidad oral, fueron colectadas previa firma del consentimiento informado, de personas mayores de edad residentes al menos por cinco años en la localidad de muestreo. Las muestras de Guajira corresponden a población Wayuu por genealogía.

**Figura 5-1:** Localización de los puntos de muestreo.

1. La Guajira, 2. Tuchín (Córdoba), 3. Tubará y Baranoa (Atlántico), 4. San Juan del Cesar (La Guajira), 5. Valledupar y Pueblo Bello (Cesar), 6. Leticia (Amazonas).



### 5.2.2 Análisis Molecular

El DNA genómico fue obtenido a partir de las muestras de sangre o de células epiteliales empleando metodologías convencionales de Salting-out, Chelex y/o por el método comercial de extracción FTA de Whatman. La amplificación de las regiones HVI y HVII del ADN Mitocondrial (ADNmt) se realizó empleando los primers:

<i>L15996</i>	<i>CTCCACCATTAGCACCCAAAGC</i>
<i>H16401</i>	<i>TGATTTACGGAGGATGGTG</i>
<i>L29</i>	<i>GGTCTATCACCTATTAACCAC</i>
<i>H408</i>	<i>CTGTAAAAGTGCATACCGCCA</i>

La electroforesis capilar se llevó a cabo en un analizador genético ABI 3500 y el análisis de las secuencias se realizó con la versión Demo de Sequencher 4.10.1.

La amplificación de 17 sistemas STRs de cromosoma Y (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 (Y GATA C4) y GATA H4), fueron realizadas empleando los kits comerciales AmpFISTR® Yfiler y Yfiler™Plus de Applied Biosystems, siguiendo las instrucciones del fabricante.

### 5.2.3 Análisis genético poblacional y estadístico

Las frecuencias haplotípicas, alélicas y de haplogrupos fueron estimadas por el método de conteo simple, y la diversidad en los tres casos, de acuerdo con Nei (1987) usando la ecuación  $DG = (n * (1 - \sum p_i^2)) / (n - 1)$ . Para calcular las frecuencias de haplogrupos de Cromosoma Y se descartaron 27 muestras porque la probabilidad estaba debajo de 80%. El software Arlequín 3.5.2.2 (Excoffier, et.al. 2010) se usó para realizar el Análisis Molecular de Varianza (AMOVA) y para construir las matrices de Rst pareados a partir del análisis de los haplotipos, y de distancias de Goldstein (1995) y para elaborar las matrices de Fst pareados a partir de los datos de frecuencia de haplogrupo de las poblaciones colombianas y de las poblaciones de referencia (**Tabla5-1**) empleadas para las comparaciones. El análisis de Escalamiento Multidimensional (Multidimensional scaling MDS) se realizó con la herramienta en línea en la plataforma YHRD para AMOVA y MDS (<https://yhrd.org/amova>) basado en Rst.

La determinación de haplogrupos de Cromosoma Y se realizó utilizando el Predictor de Haplogrupos (Athey, 2006). Solo se consideraron para los análisis aquellos haplotipos cuya probabilidad de asignación del haplogrupo señalado por el software, fue mayor o igual a 0.8. La asignación de haplogrupos (Hgr) y subhaplogrupos (subHgr) para ADNmt se realizó a través de Haplogrep 2.0 version 2.1.19 (<https://haplogrep.uibk.ac.at/>) y se confirmó con EMPOP (<https://empop.online/>). Los árboles filogenéticos de los haplogrupos fueron elaborados con la herramienta disponible en Haplogrep 2.0

**Tabla 5-1.** Poblaciones de referencia de YHRD empleadas en las comparaciones

POBLACION	N (Haplotipos)	Referencia	Datos
Tucumán, Argentina [Kolla]	14	Toscanini U., Gusmão L., Berardi G., Amorim A., Carracedo A., Salas A., Raimondi E. (2008), 'Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data.', <i>Forensic Sci Int Genet</i> 2(4), 274-80	Nativoamericanos
Río Negro, Argentina [Mapuche]	11	Roewer L., Nothnagel M., Gusmão L., Gomes V., González M., Corach D., Sala A., Alechine E., Palha T., Santos N., Ribeiro Dos Santos A., Geppert M., Willuweit S., Nagy M., Zweynert S., Baeta M., Núñez C., Martínez Jarreta B., González Andrade F., De F., Silva D., Builes JJ., Turbón D., Parra L., Arroyo Pardo E., Toscanini U., Borjas L., Barletta C., Ewart E., Santos S., Krawczak M. (2013), 'Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans.', <i>PLoS Genet</i> 9(4), e1003460	Nativoamericanos
Chubut, Argentina [Tehuelche]	10	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Chaco, Argentina [Toba]	46	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Formosa, Argentina [Wichi]	33	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Beni, Bolivia [Moxenos]	10	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Pará, Brazil [Arara]	20	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Pará, Brazil [Parakanã]	38	Palha T.J., Rodrigues EM., Santos D. (2010), 'Y-STR haplotypes of Native American populations from the Brazilian Amazon region.', <i>Forensic Sci Int Genet</i> 4(5), e121-3	Nativoamericanos
Mato Grosso do Sul, Brazil [Terena]	49	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Amazonas, Brazil [Ticuna]	23	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Amapá, Brazil [Tiriyó]	35	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Antioquia, Colombia [Emberá-Chamí]	24	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Pastaza, Ecuador [Kichwa]	42	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Puno, Peru [Aymara]	3	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Amazonas, Peru [Chachapoya]	106	Guevara EK., Palo JU., Guillén S., Sajantila A. (2016), 'MtDNA and Y-chromosomal diversity in the Chachapoya, a population from the northeast Peruvian Andes-Amazon divide', <i>Am J Hum Biol.</i> 28(6), 857-67	Nativoamericanos
Cuzco, Peru [Chumbivilca]	10	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Venezuela [Bari Boxí]	16	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos
Venezuela [Wayuu]	17	Obtenidos desde bases de datos de poblaciones YHRD	Nativoamericanos

El análisis de componentes principales (ACP) se realizó con el software MVSP 3.22 a partir de las frecuencias alélicas para Cromosoma Y usando los haplotipos previamente publicados en (Franco & Barreto, 2017; Grugni et.al., 2015; Purps, et.al., 2014 y Xavier,

et.al., 2015), y de frecuencias de haplogrupos para ADNmt, empleando las referencias citadas más adelante.

## **5.3 Resultados y Discusión**

### **5.3.1 Cromosoma Y**

Estudios que describen la variación genética de las poblaciones mezcladas americanas, han mostrado la existencia de mayor proporción de haplogrupos maternos de origen nativoamericano y de haplogrupos masculinos europeos (Salzano & Sanz, 2014; Carvajal-Carmona, et.al., 2000; Bedoya et.al., 2006). Para el caso específico del Cromosoma Y, la proporción de linajes Q oscila entre 0.05 y 0.5, en poblaciones con diferentes grados de mezcla (Lovo-Gómez, et.al., 2007; Acosta et.al., 2009; Corach, 2010; Grugni et.al. 2015; Martinez, et.al., 2012; Criollo et.al., 2018; Rojas et.al., 2010), lo que es concordante con la frecuencia de 0.2245 y 0.5417 descrita en el capítulo anterior para las regiones biogeográficas Colombianas de Caribe Continental y Amazonas respectivamente. De igual forma, los valores de diversidad de haplogrupos de 0.8044 y 0.6622 y de diversidad haplotípica de 0.9970 y 0.9978 respectivamente, se encuentran dentro del rango descrito por varias publicaciones entre 0.4 y 0.8 para el primer caso (Lovo-Gómez, et.al., 2007; Criollo et.al., 2018; Rojas et.al., 2010) y entre 0.9849 y 0.9999 para el segundo (Purps, et.al., 2014; Rangel et.al., 2007; Usme, et.al., 2013; Lovo-Gomez, et.al. 2007; Terreros 2010). Como se discutió en el capítulo anterior, esta diversidad está relacionada también con la historia de patrones de colonización posteriores a la llegada de los europeos al continente y a las migraciones recientes producto de los cambios en las oportunidades laborales por cuenta de las diferentes economías de subsistencia, explotación de los suelos para bonanzas agrícolas e industrias e incluso el narcotráfico.

Con el fin de estudiar los linajes nativoamericanos a partir de muestras obtenidas principalmente en poblaciones mezcladas colombianas, se seleccionaron muestras de personas de ascendencia indígena a partir de la información genealógica obtenida a través de las encuestas aplicadas durante la fase de muestreo, analizándose finalmente para la

región Caribe colombiana 31 individuos para linajes de Cromosoma Y y 40 individuos para ADN mitocondrial y para la región de Amazonas 20 y 11 individuos respectivamente.

De acuerdo con los resultados obtenidos a partir de marcadores STRs de cromosoma Y, el 25% de los linajes masculinos de Amazonas (15% Europeos y 10% Africanos) y el 48.4% de Caribe Continental (35% Europeos y el restante Africanos) no son de origen nativoamericano, tal y como se ha informado sobre varias poblaciones indígenas americanas, en rangos que varían entre 0% para poblaciones colombianas o mexicanas y hasta el 90% en algunas poblaciones norteamericanas (Blanco-Verea, 2008; Corach 2010; Franco & Barreto, 2017; Grugni et.al., 2015; Xavier et.al., 2015; Bortolini et.al., 2003; Sandoval et.al., 2012; Bisso-Machado, et.al., 2012; Bolnick, et.al., 2006; Criollo, et.al., 2018; Rojas et.al., 2010). La diversidad de haplogrupos de 0.7054 estimada para Caribe, es similar a reportada para Kolla, Diaguitas, Mapuches (Blanco Verea, et.al., 2008) y Nativoamericanos de Norteamérica (Bolnick, et.al., 2006), y un poco menor que la de Emberas de Antioquia y poblaciones indígenas del Cauca informadas por Xavier y colaboradores en 2015 (Xavier,et.al., 2015), mientras la de Amazonas (0.4368) es concordante con la de Indígenas Mexicanos (Sandoval, et.al., 2012), poblaciones Nasa del Tolima (Criollo, et.al., 2018) y otras comunidades indígenas colombianas (Rojas,et.al., 2010). La diversidad de haplotipos, 0.9914 y 0.9941 para las dos regiones biogeográficas colombianas, es similar a la reportada por Blanco Verea y su equipo (Blanco Verea, et.al., 2008) para poblaciones indígenas Chilenas Colla y Diaguitas, y mayor que la de Mapuches (Blanco Verea, et.al., 2008) y algunas poblaciones indígenas mexicanas (Rangel, et.al., 2007).

Para las etnias Kankuamo y Mokaná del Caribe Continental Colombiano, solo se encontró un representante, el primero con haplogrupo de origen europeo y el segundo de origen africano, mientras que para las etnias del Amazonas Tanimuca y Miraña, que también cuentan solo con un individuo, y Yagua y Huitoto, con dos cada una, en todos los casos el haplogrupo predicho es Q. Dado el N para estos grupos, los valores de diversidad deben ser considerados con precaución. **Tabla 5-2.** Para el caso de los Arhuacos, donde el haplogrupo predicho para un individuo no es de origen nativoamericano, y para Ticuna, con dos individuos con haplogrupo de origen africano, aunque los tamaños de muestra son pequeños, contrastan con los resultados de Rojas et.al., 2010, que analizó 19 y 26 individuos de estas dos etnias, todos ellos pertenecientes al haplogrupo Q. En el caso de

la etnia Zenú, las muestras tomadas en este estudio provienen de la zona de Tuchín en el departamento de Córdoba y la proporción de haplogrupos de origen europeo fue de 55% y 36% de origen nativoamericano, y un 1% de origen africano, lo que resulta similar al reporte de indígenas de esta misma etnia de la zona de Sahagun en el mismo departamento, con la diferencia de no haber encontrado haplogrupos de origen africano (Criollo, 2012). Sin embargo, lo anterior difiere de las publicaciones de Rojas 2010 y Bortolini 2003, quienes reportaron un 77% y 81% de linajes Q respectivamente, sin linajes de origen Africano, en 22 y 30 individuos estudiados. Adicionalmente, el 67% de las muestras de la comunidad Wayuu analizadas en este estudio, fueron clasificadas como haplogrupo Q, similar al reporte de Bortolini 2003, y del restante, 13% corresponden a haplogrupos de origen Africano. Para el caso de los indígenas Bora, se analizaron 3 individuos, dos de ellos pertenecientes a linajes Q.

**Tabla 5-2.** Diversidad Genética en las regiones biogeográficas colombianas del Caribe y Amazonas a partir de marcadores STRs de Cromosoma Y

ETNIA	CROMOSOMA Y					
	N	No. Htp dif.	Diversidad Htp	Diversidad Hgr	% Hgr Natam	DivGenProm/Loci
ARHUACO	3	2	0,6667	0,6667	67	0,36
KANKUAMO	1	1	NA	NA	0	NA
MOKANA	1	1	NA	NA	0	NA
WAYUU	15	11	0,9809	0,5619	67	0,63
WIWA	0	0	0	0	0	0
ZENÚ	11	9	0,9818	0,8182	36	0,71
<b>TOTAL CARIBE</b>	<b>31</b>	<b>23</b>	<b>0,9914</b>	<b>0,7054</b>	<b>52</b>	
COCAMA	5	5	1	0,7	60	0,69
TANIMUCA	1	1	1	1	100	0
TICUNA	6	6	1	0,533	67	0,44
HUITOTO	2	2	1	1	100	0,47
BORA	3	2	0,6667	0,6667	67	0,4
MIRAÑA	1	1	1	1	100	0
YAGUA	2	2	1	0	100	0,47
MACUNA	0	0	0	0	0	0
YUCUNA	0	0	0	0	0	0
<b>TOTAL AMAZONAS</b>	<b>20</b>	<b>19</b>	<b>0,9941</b>	<b>0,4368</b>	<b>75</b>	

Al estudiar la variación genética dentro de las poblaciones, se observa que la diversidad de haplotipos por etnia es mayor en Amazonas que en Caribe, mientras la diversidad de haplogrupos es mayor en Caribe que en Amazonas (**Tabla 5-2**). Comparando los datos

para las etnias Arhuaco, Ticuna y Zenú con la literatura, en este estudio se reportan mayores valores de diversidad de haplogrupos, aunque los tamaños de muestra son menores. (Rojas et.al., 2010, Criollo, 2012, Tarazona-Santos et.al., 2001).

Con el fin de ver en mayor detalle el panorama del legado nativoamericano, para los siguientes análisis se consideraron solo las muestras cuya predicción de haplogrupo las ubicó como linajes Q. Al comparar la diversidad genética promedio por loci estimada con Arlequin 3.5.2 y considerando poblaciones de referencia tomadas de las publicaciones de Grugni, et.al., 2015, Xavier, et.al., 2015 y Roewer et.al., 2013, Franco y Barreto, 2017, se observa que la etnia Parakaná la variación es mínima, (0.03433) y los valores más bajos corresponden a las muestras de la etnia panameña Kuna, la etnia brasileña Terena, la etnia venezolana Bari y a la etnia argentina Toba. En contraste, los valores más altos se presentan en las etnias Colla de Argentina y Pastos del suroccidente colombiano. Los resultados obtenidos al calcular valores promedio agrupando las poblaciones por su ubicación, de acuerdo con los criterios expuestos por Tarazona-Santos et.al., 2001, son concordantes con las conclusiones de dicho autor, al ver que las comunidades Amazónicas (orientales) cuentan con menores grados de diversidad que las Andinas (occidentales).

**Tabla 5-3.** Si se consideran categorías adicionales y se obtienen los datos de promedio para este indicador, puede notarse como las etnias de Suramérica Noroccidental (Kuna, Ngabe, Arhuaco, Zenú, Wayuu Colombiano, Wayuu Venezolano y Bari) y de la Región del Chaco Argentino (Toba), tienen menores valores de diversidad, tal como sucede con las poblaciones de referencia y las de este estudio de la región del Amazonas (Terena, Arara, Tiriyó, Parakaná, Cocama, Huitoto, Ticuna y Yagua); mientras las etnias de la Región Andina Colombiana (Coyaimas y Embera Chamí), Región suroccidental colombiana (Coconuco, Yanacona, Guambiano, Nasa, Pastos, Awa), y Chocó Colombiano (Embera Dumá), presentan valores más altos similares a los reportados por las etnias de los Andes en Suramérica (Quechua, Colla, Tehuelche y Mapuche). Esta similitud puede explicarse por la relación prehispánica creada por la expansión del Imperio Inca desde Perú hasta el Suroccidente Colombiano y a su vez, por las relaciones por cercanía geográfica en el ámbito comercial y de ocupación de territorios entre los pueblos Nasa y la Confederación Pijao en la región del Huila y del Tolima Colombiano. También se ha documentado que los Pijao habitaron originalmente el área del Chocó Colombiano, actual zona de habitación de las etnias Embera y Waunana, quienes pudieron ser responsables del desplazamiento de estos pueblos hacia las tierras de la nación Paez, tras la disputa por los territorios. (Cubillos,

1946). Estas observaciones apoyan las conclusiones de Tarazona Santos al indicar que hay mayor diversidad intrapoblacional en las comunidades Andinas que en las del Oriente del continente.

**Tabla 5-3.** Diversidad genética promedio por loci. \* Este estudio.

Población	Etnia	N	N° Loci Polimorficos	N° medio dif pareadas	Diver.Gen.Prom	s.d.	Div.Gen.Promedio
1	Kuna	19	12	71,16	0,25	0,16	<b>0,40</b>
2	Ngabe	5	12	8,00	0,39	0,24	
14	Bari	16	8	8,13	0,26	0,28	
16	WayuuVenz	18	14	61,57	0,52	0,23	
30	Arhuaco*	2	0	0,00	0,00	0,00	
31	Zenú*	4	11	21,50	0,48	0,33	
32	WayuuCol*	10	13	19,91	0,47	0,28	
22	Toba	49	15	65,09	0,33	0,23	<b>0,33</b>
10	Embera-Duma	9	11	103,11	0,44	0,32	<b>0,44</b>
15	Terena	32	14	8,71	0,26	0,16	<b>0,38</b>
17	Arara	20	12	13,33	0,39	0,24	
18	Tiriyo	35	15	23,19	0,51	0,20	
21	Parakana	38	7	0,67	0,03	0,05	
23	Bora*	2	0	0,00	0,00	0,00	
24	Cocama*	3	10	14,67	0,44	0,33	
25	Huitoto*	2	7	13,00	0,47	0,52	
26	Miraña*	1	0	0,00	0,00	0,00	
27	Tanimuca*	1	0	0,00	0,00	0,00	
28	Ticuna*	4	12	9,83	0,44	0,25	
29	Yagua*	2	7	7,00	0,47	0,52	
3	Embera-Chami	24	14	16,27	0,52	0,18	<b>0,50</b>
11	Coyaima	15	13	133,01	0,48	0,25	<b>0,49</b>
4	Coconuco	7	13	15,24	0,51	0,26	
5	Yanacona	2	8	30,00	0,53	0,52	
6	Guambiano	16	14	20,31	0,41	0,24	
7	Nasa	60	15	19,72	0,46	0,22	
8	Pastos	8	15	223,75	0,55	0,19	
9	Awa	8	13	15,93	0,50	0,25	
12	Quechua	31	15	13,94	0,48	0,17	<b>0,51</b>
13	Colla	14	15	23,42	0,57	0,17	
19	Tehuelche	10	15	17,33	0,51	0,17	
20	Mapuche	11	15	17,24	0,49	0,21	



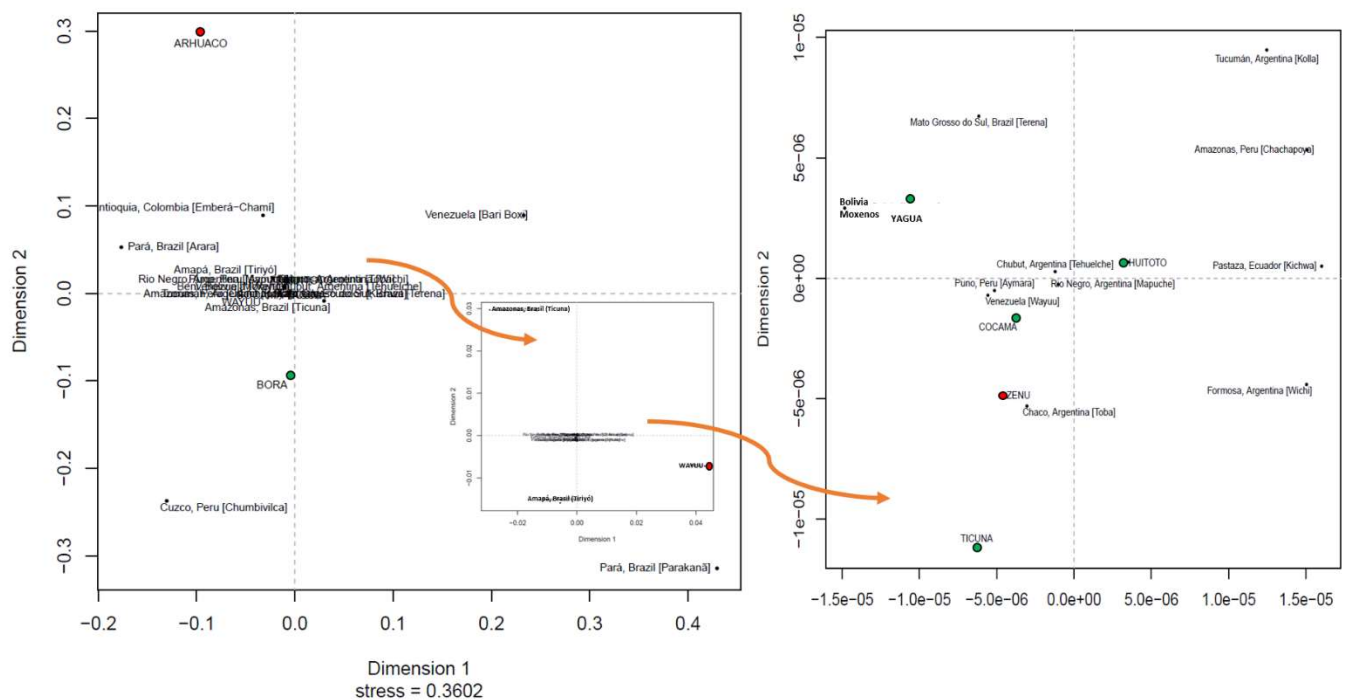
En el análisis de las diferencias entre las poblaciones de la Tabla 5-3, el AMOVA realizado mostró valores significativos de  $p$  para el FST 0.21588 ( $p=0.0000$ ) y para el FSC 0.21995 ( $p=0.0000$ ), indicando que existe diferenciación dentro de las poblaciones y entre poblaciones dentro de los grupos. Para evaluar este hallazgo entre cada una de las etnias, se realizó el análisis de Rst pareados, cuyo resultado no muestra valores significativos entre las poblaciones Colombianas de Caribe y Amazonas, aunque estos datos deben considerarse con precaución, ya que algunos tamaños de muestra son muy bajos (1 o 2 individuos). Al comparar con las muestras de referencia, se observa que las poblaciones de este estudio se diferencian poco de otras etnias colombianas, aunque se encuentren en diferentes regiones biogeográficas, pero hay mayores diferencias con poblaciones de otras áreas. En Ticuna por ejemplo, hay más diferencias que en las otras comunidades del Amazonas Colombiano, especialmente con etnias del tronco lingüístico Chibcha-Paez y Lenguas Andinas y no tiene diferencia con Wayuu de Venezuela. La muestra de Arhuacos tiene diferencias con la mayoría de los pueblos del Amazonas de Brasil, Mapuche y Bari, pero no con las otras comunidades de lenguas Andinas o Chibcha-Paez incluyendo Nasa y Awa, o con las demás poblaciones colombianas. La etnia Zenú solo muestra valores significativos de Rst con las poblaciones Terena, Arara y Paracaná todas ellas de Brasil. Por otro lado, la muestra de Wayuu de Colombia no se diferencia significativamente de Wayuu Venezuela, pero es la que más diferencias tiene incluso con etnias colombianas como Embera Chami, Guambiano, Nasa y Pastos, también de todos los pueblos Andinos y Amazónicos de Brasil y Bari de Venezuela.

La similitud entre el conjunto de individuos con haplogrupo Q de las regiones colombianas de Amazonas, La Guajira y Caribe Colombiano, sin distinguir las etnias, y poblaciones nativoamericanas disponibles en la base de datos YHRD, se mostró en el capítulo anterior mediante el MDS realizado con base en distancias de Rst pareados entre los haplotipos, donde se observa que las muestras de Caribe continental se ubican más cercanas a Quechuas del Ecuador fuera del grupo de las demás muestras colombianas, mientras los individuos de Amazonas y La Guajira están más cerca de Mayas de Guatemala y Wichi de Argentina respectivamente.

El mismo tipo de análisis realizado diferenciando por etnias, teniendo en cuenta solo aquellas con dos o más individuos, muestra Huitoto y Cocama más cercanamente relacionadas con Wayuu de Venezuela, Aymara, Mapuche y Tehuelche, a Zenú con Toba,

y a Yagua con Moxenos, compartiendo el cuadrante superior derecho con Terena, sin observarse un patrón aparente de relación por filiación lingüística o por cercanía geográfica, mientras, Arhuaco, Bora, Wayuu y Ticuna de Colombia no tienen ninguna cercanía con otras poblaciones de referencia, incluso con Wayuu de Venezuela o Ticuna de Brasil, para los dos últimos casos. **Figura 5-2.**

**Figura 5-2.** MDS para muestras asignadas como haplogrupo Q de Amazonas y Caribe Colombianos, comparadas con poblaciones de referencia nativoamericanas de la base de datos YHRD

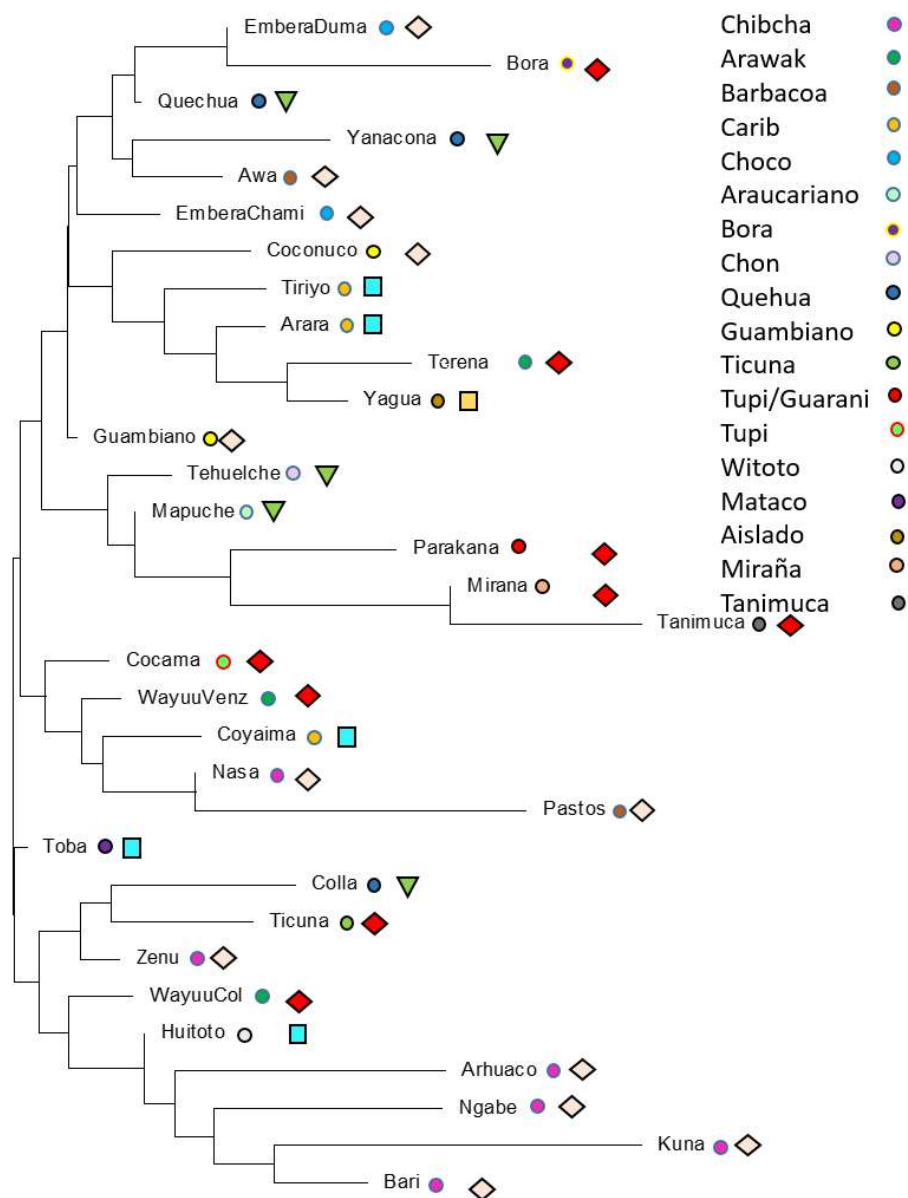


Las relaciones genéticas entre los grupos estudiados, se analizaron a través de la construcción de un árbol Neighbor joining desde la matriz de distancias pareadas de Goldstein 1995 ( $\delta\mu^2$ ) obtenida con Arlequin, a partir de número promedio de diferencias en tamaño alélico dentro de las poblaciones. **Figura 5-3.**

Se observan en el árbol 3 grandes clusters. El primero que cuenta con representantes de las 4 ramas de los lenguajes amerindios de acuerdo con la clasificación de Greenberg y Ruhlen, destacándose la ausencia de la subrama Chibcha. El subgrupo (1a) relaciona etnias de filiación lingüística Chibcha-Paez del Chocó, Antioquia y sur occidente colombiano

con etnias andinas de este último sector y del norte de Ecuador, con comunidades de Amazonas colombiano (Bora y Yagua) y Tukano ecuatorial y genopanocarib del noroccidente de Brasil. El cluster conformado por Tiriyo, Arara, Terena y Yagua, se sostiene en las topologías obtenidas por el método UPGMA con las mismas distancias y en el árbol Neighbor-Joining construido a partir de distancias Rst. El Subgrupo (1b) solo incluye etnias de filiación lingüística Andina y Tucano ecuatorial, debe sin embargo cuidarse la interpretación, dado que Miraña y Tanimuca cuentan solo con un individuo.

**Figura 5-3.** Arbol Neighbor-Joining con base en distancias de Goldstein (1995) relacionando muestras de poblaciones nativoamericanas de Suramérica.



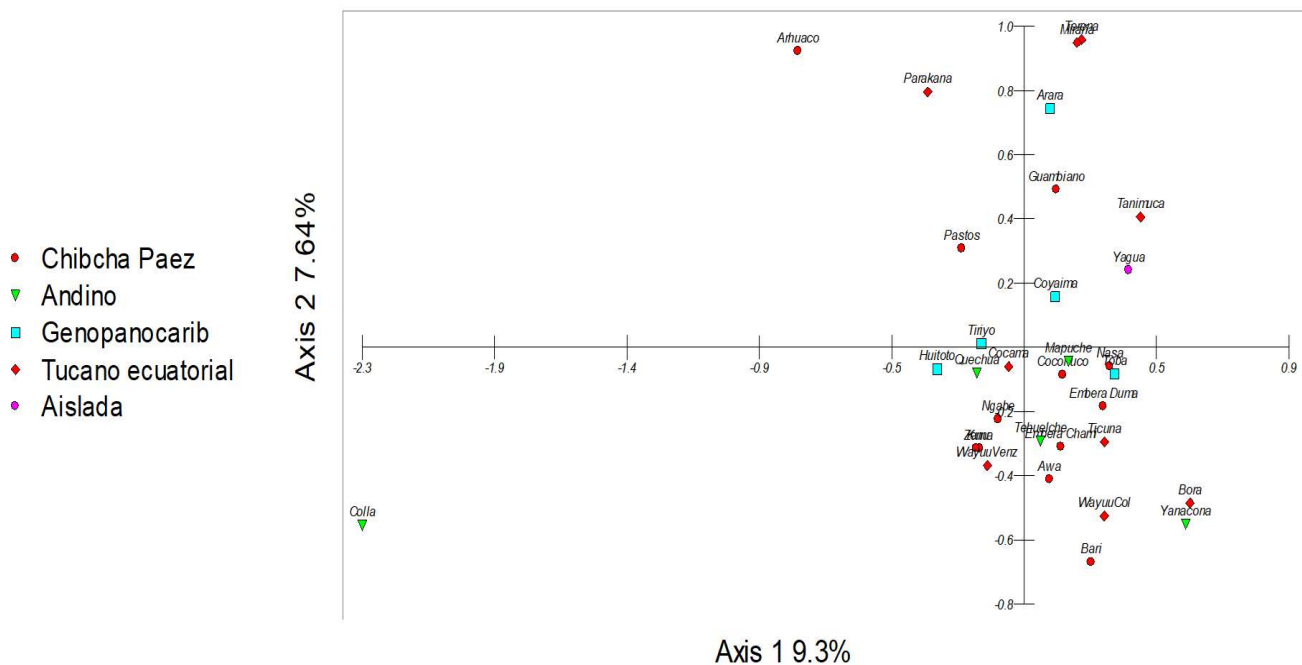
El cluster conformado por Tehuelche y Mapuche se sostiene también en los árboles alternativos descritos en el análisis de (1a). Para el grupo 1 no parece observarse una relación entre etnias dependiente exclusivamente de la pertenencia a una rama del lenguaje, sino más probablemente relacionada con la cercanía geográfica, independientemente de las barreras geográficas o ecológicas que pudiesen evitar el contacto entre estas comunidades. Cabe resaltar que a esta hipótesis escapan las etnias Tehuelche y Mapuche ubicadas en Argentina.

El grupo 2 no incluye etnias del tronco del lenguaje Andino y relaciona Cocama del Amazonas Colombiano y Wayuu de Venezuela, ambas de filiación lingüística Tucano Ecuatorial y un cluster conformado por Coyaima (Genopanocarib) y las poblaciones de lengua Chibcha-Paez Nasa y Pastos, siendo esta última la de mayor grado de divergencia. En el primer caso, esta relación puede analizarse a la luz de las hipótesis antropológicas del origen de los pueblos Arawak de la región caribe continental de Suramérica como producto de migraciones reversas provenientes de pueblos de la mismo subtronco lingüístico establecidos en la región de Amazonas y Orinoco (Bolnik, et.al., 2016; Moreno et.al., 2013). En el segundo caso, Pastos y Nasas comparten territorios, y para esta última comunidad, actualmente se conocen grupos cuyos ancestros originalmente se ubicaron en la región suroccidental de Colombia, pero que se desplazaron a la región sur del Huila y Tolima, territorios ocupados por la confederación Pijao, de la cual hacía parte los indígenas Coyaimas, lo que podría haber favorecido la mezcla entre las etnias. Esta observación también fue hecha por Franco y Barreto, 2017.

En el grupo se observa el cluster conformado por poblaciones de lengua Chibcha, que se sostiene en los árboles alternativos y relaciona las etnias Wayuu y Zenú de Colombia, lo que es concordante con el árbol UPGMA con distancias de Goldstein y se ha evidenciado en publicaciones realizadas con análisis de marcadores autosómicos tanto STRs como SNPs (Wang et.al., 2007, Hunley & Healy, 2011, He 2012, y Reich et.al., 2012). Cabe anotar que Wayuu de Colombia y de Venezuela no conforman el mismo cluster en ninguna de las topologías analizadas.

En el análisis de componentes principales realizado a partir de frecuencias alélicas, las dos primeras componentes principales solo explican el 17% de la variación, por lo que una inferencia al respecto podría no ser pertinente. **Figura 5-4.**

**Figura 5-4.** ACP a partir de frecuencias alélicas de 15 sistemas STR de cromosoma Y



### 5.3.2. ADN Mitocondrial

Respecto al análisis de la presencia de haplogrupos de origen nativoamericano a partir del estudio de las regiones HVI y HVII del ADNmt, la proporción encontrada en las poblaciones mezcladas fue de 0.875 en el Caribe Continental Colombiano y para Amazonas una región con mayor grado de aislamiento geográfico, es de 1, lo cual concuerda con reportes realizados en la literatura (Corach, 2010; Criollo et.al., 2018; Terreros, 2010). Los linajes maternos en la región del Caribe continental colombiano muestran un 10% de haplogrupos L relacionados con ancestría africana y una muestra de indígena Zenú fue asignada como perteneciente al haplogrupo ancestral M, que requiere confirmación ampliando la revisión a otras regiones del genoma mitocondrial. **Tabla 5-4.** Este es un valor comparativamente alto respecto al reportado para haplogrupos no nativoamericanos en grupos indígenas en Colombianos (1-6% en Tolima Criollo et.al. 2018; 0-5% Colombia Rojas, et.al., 2010). La introgresión de haplogrupos de origen africano en poblaciones indígenas Colombianas que se observa en este estudio en Arhuacos, Kankuamos y Wayuu, ya había sido reportada para 2% Pijaos del Tolima (Criollo, et.al., 2018), 7% Nasa del Cauca (Xavier, et.al., 2015), 6% Zenú (Criollo, 2012), pero no en Arhuacos. En el caso de Wayuu, como en varias otras etnias, los reportes indican la presencia de haplogrupos no nativoamericanos, pero a excepción de Melton et.al., 2007, no especifican cuales son, por lo que podrían encontrarse

allí incluidos los de origen africano (Bisso-Machado et.al., 2012, Usme, et.al., 2013). Sobre las poblaciones Kankuamo, Mokaná, Cocama, Yucuna y Macuna, no se encontraron reportes previos, sin embargo es importante ampliar el muestreo sobre estas etnias para obtener resultados más robustos.

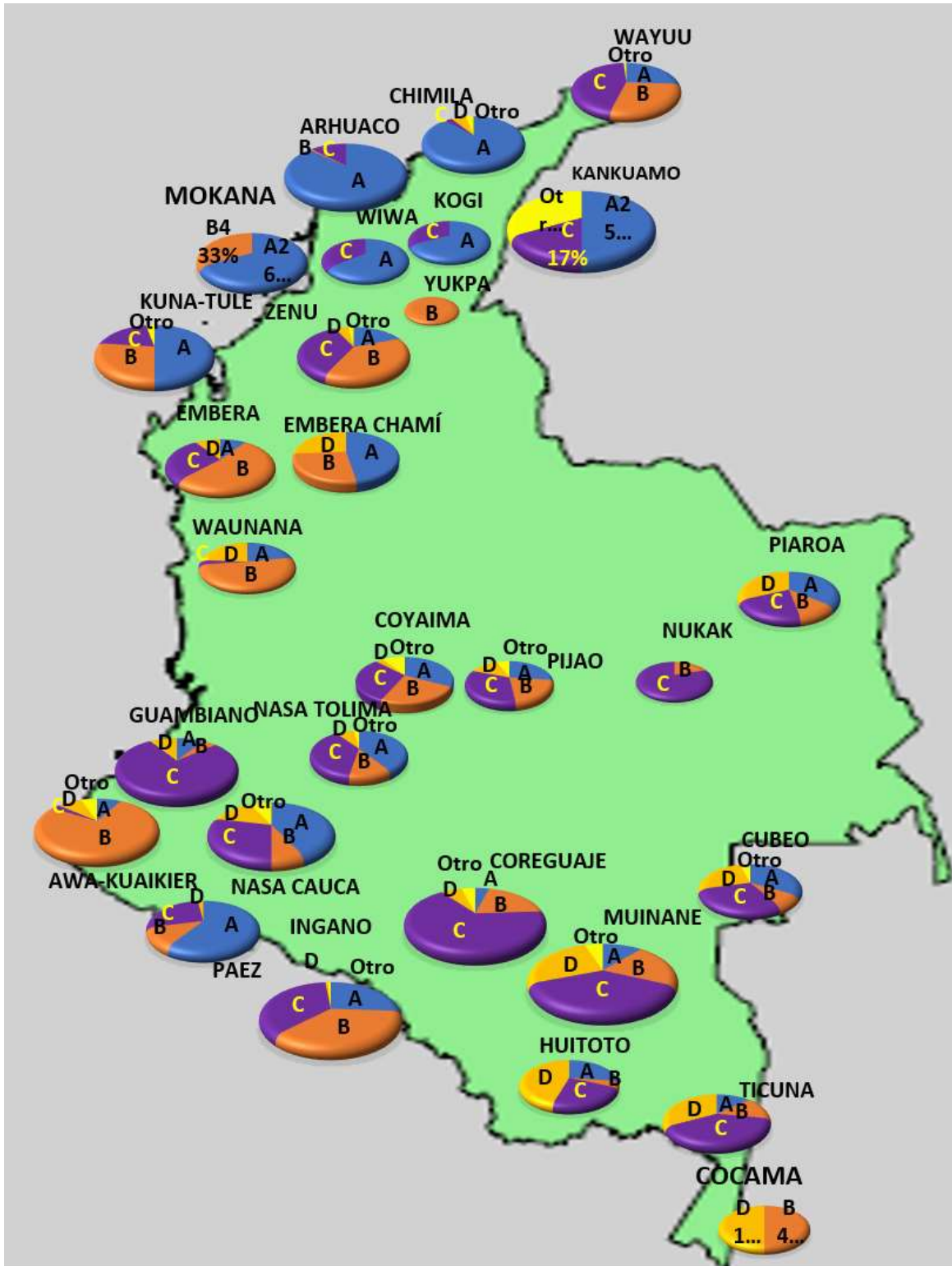
**Tabla 5-4.** Frecuencias de Haplogrupos para Caribe y Amazonas por etnias

POBLACION	N	A2	B4	C	M	D1	L
ARHUACO	10	0.700	0.000	0.200	0.000	0.000	0.100
KANKUAMO	6	0.500	0.000	0.167	0.000	0.000	0.333
MOKANA	12	0.667	0.333	0.000	0.000	0.000	0.000
WAYUU	3	0.333	0.000	0.333	0.000	0.000	0.333
WIWA	1	0.000	0.000	1.000	0.000	0.000	0.000
ZENÚ	8	0.250	0.125	0.375	0.125	0.125	0.000
<b>CARIBE TOTAL</b>	<b>40</b>	<b>0.525</b>	<b>0.125</b>	<b>0.200</b>	<b>0.025</b>	<b>0.025</b>	<b>0.100</b>
COCAMA	2	0.000	0.500	0.000	0.000	0.500	0.000
YUCUNA	1	1.000	0.000	0.000	0.000	0.000	0.000
TICUNA	4	0.750	0.000	0.250	0.000	0.000	0.000
HUITOTO	2	0.500	0.000	0.500	0.000	0.000	0.000
MACUNA	1	0.000	1.000	0.000	0.000	0.000	0.000
BORA	1	0.000	1.000	0.000	0.000	0.000	0.000
<b>AMAZONAS TOTAL</b>	<b>11</b>	<b>0.455</b>	<b>0.273</b>	<b>0.182</b>	<b>0.000</b>	<b>0.091</b>	<b>0.000</b>

A pesar de los datos reportados para Ticuna, Huitoto y Wayuu en este trabajo, publicaciones con mayores tamaños de muestra han mostrado la presencia de haplogrupos B y D en proporciones de 15% y 33% para Ticuna (Mesa et.al., 2000) y 4.5% y 45.5% para Huitoto (Keyeux, et.al., 2002) y del haplogrupo B en más del 15% para Wayuu (Mesa, et.al., 2000 y Keyeux, et.al., 2002).

La **Figura 5-5** muestra las frecuencias de haplogrupos para poblaciones indígenas colombianas. Para un análisis más conservador, en la figura se incluyen las frecuencias Zenú, Wayuu, Arhuaco, Ticuna y Huitoto tomadas de poblaciones de referencia con mayor número de muestras (Bisso-Machado et.al., 2012, Mesa et.al., 2000, Keyeux, et.al., 2002). Se observa que las poblaciones del norte cuentan con mayores proporciones de haplogrupo A y menores de haplogrupo C y haplogrupo D casi ausente, en contraste con las poblaciones del sur con predominio del haplogrupo C y presencia del haplogrupo D equivalente o mayor que la de C en Ticuna y Huitoto. La frecuencia del haplogrupo B tiende a disminuir de Occidente a Oriente.

**Figura 5-5.** Frecuencias de haplogrupos de ADN mitocondrial en poblaciones indígenas colombianas. Zenú, Wayuu, Arhuaco, Ticuna y Huitoto corresponden a poblaciones de referencia.



En la región Caribe se observan en esencia dos patrones, uno correspondiente a las comunidades de la Confederación Tayrona (Arhuaco, Kogi, Wiwa y Kankuamo) y Chimila, esta última incluyendo el haplogrupo D y con una alta presencia de haplogrupos L en Kankuamo. El segundo corresponde a las poblaciones Kuna, Zenú, Mokaná y Wayuu que tienen alta presencia de haplogrupo B. El caso del pueblo Yukpa, asentado en la Serranía del Perijá en el departamento del Cesar, ha sido discutido por Keyeux et.al. 2002, refiriendo un efecto de cuello de botella favorecido por aislamiento geográfico, que también se detecta a través del análisis de marcadores de grupos sanguíneos (Keyeux, et.al., 2002), sin embargo, un reporte más reciente (Noguera, et.al., 2015) muestra un 57% de Hgr A y el restante de Hgr B en esta comunidad.

Las etnias Zenú y Mokana no han sido claramente establecidas como descendientes de pueblos amerindios de una rama lingüística específica, ya que no se cuenta con suficiente evidencia para su asignación. Al respecto Mesa et.al., 2000 comenta la dualidad entre la propuesta de considerar la lengua nativa de esta comunidad como un dialecto Carib o Chibcha según Voegelin y Greenberg (Mesa, et.al., 2000). La estructura observada de acuerdo con la frecuencia de haplogrupos para esta etnia es sin embargo más parecida a la del pueblo Wayuu, sin que esto necesariamente implique una relación filogenética, aunque cabe destacar como se mencionó en la sección referente al cromosoma Y, que estas dos etnias se asocian en el mismo cluster para este tipo de marcadores y también con autosómicos (Wang et.al., 2007, Hunley & Healy, 2011, He, 2012, y Reich et.al., 2012). Con respecto a Mokaná, se ha propuesto que se relaciona ancestralmente con el tronco lingüístico Arawak, en especial con el pueblo Wayuu, sin embargo, en la muestra analizada no se encontró el haplogrupo C, presente también en pueblos Arawak de Suramérica con frecuencia mayores que 20% en varias etnias.

Se observa también como la distribución de las frecuencias es diferente para las muestras de Embera-Chamí de Antioquia (Xavier, et.al., 2015) y de Emberas del Chocó (Usme, et.al., 2013), de la misma forma en que se reportan en otros muestreos (Keyeux et.al., 2002, Mesa et.al., 2000, Criollo, 2012), sugiriendo explorar una posible subestructura en esta etnia que cuenta con varias comunidades asentadas en diferentes localidades, algunas de ellas separadas por accidentes geográficos que podrían tener el efecto de barreras (Franco & Barreto, 2017).



En la región sur occidental de Colombia la etnia Awa muestra un patrón único, posiblemente relacionado con flujo genético que pudo darse en razón a sus constantes relaciones comerciales con poblaciones ubicadas hacia el sur en Ecuador y de Peru, con mayor influencia del imperio Inca (Ministerio del Interior). Paez y Nasa del Cauca muestran coincidencia con la distribución de las frecuencias en las etnias del sur del Tolima Nasa, Pijao y Coyaima, mientras el de Guambiano es más compatible con el de los pueblos Amazónicos y de la Orinoquia (relaciones también observadas en el análisis de cromosoma Y), a excepción de Nukak que tiene pérdida aparente de los haplogrupos A y C, lo que es coherente con su aislamiento geográfico y cultural casi hasta el comienzo del siglo XXI.

La diversidad de haplogrupos para las regiones fue de 0.6744 en el Caribe y de 0.7455 en el Amazonas, lo que está por encima de los reportes previos para otras poblaciones Colombianas del Caribe y de la zona Sur realizados por Rojas et.al., 2010 (0.21-0.72). Comparativamente con los reportes de otras investigaciones, la diversidad de haplogrupos para Arhuaco, Wayuu y Zenú encontradas en este estudio son mayores (Melton et.al., 2007, Bisso-Machado et.al., 2012, Usme et.al., 2013, Criollo, 2012). Para Kankuamos este valor también supera el de las comunidades Wiwa, Kogi, Arhuaco y Chimila que no es mayor en ninguna publicación del 0.5, con las que comparte territorios ancestrales y filiación lingüística (Melton, et.al., 2007, Bisso-Machado, et.al., 2012, Usme, et.al., 2013). Para el caso de Mokaná, la diversidad de haplogrupos es menor que la de las etnias Arawak de referencia (Bisso-Machado et.al., 2012 y Melton et.al., 2007). **Tabla 5-5 y Tabla 5-6.**

**Tabla 5-5.** Diversidad Genética en Caribe y Amazonas Colombiano por etnias

ETNIA	ADN MITOCONDRIAL				
	N	No. Htp dif.	Diversidad Htp	Diversidad Hgr	% Hgr Natam
ARHUACO	10	4	0.9778	0.5111	90
KANKUAMO	6	5	0.9333	0.7333	67
MOKANA	12	8	0.9697	0.4848	100
WAYUU	3	3	1	1	67
WIWA	1	1	NA	NA	100
ZENÚ	8	7	0.9643	0.8571	87.5
<b>TOTAL CARIBE</b>	<b>40</b>	<b>29</b>	<b>0.9763</b>	<b>0.6744</b>	<b>87.5</b>
COCAMA	2	2	1	1	100
TANIMUCA	0	0	0	0	0
TICUNA	4	4	1	0.8333	100
HUITOTO	2	2	1	1	100
BORA	1	1	NA	NA	100
MIRANA	0	0	0	0	0
YAGUA	0	0	0	0	0
MACUNA	1	1	NA	NA	100
YUCUNA	1	1	NA	NA	100
<b>TOTAL AMAZONAS</b>	<b>11</b>	<b>11</b>	<b>1</b>	<b>0.7454</b>	<b>100</b>

**Tabla 5-6.** Frecuencias y Diversidad (D. Hgr) de haplogrupos en poblaciones de referencia de América y Colombia

Etnia	N	A	B	C	D	otro	D. Hgr	Referencia
<b>Aché</b>	63	0.1000	0.9000	0.0000	0.0000	0.0000	0.1829	Bisso-Machado et.al. 2012
<b>Ancash</b>	33	0.0900	0.5200	0.1800	0.2100	0.0000	0.6652	Bisso-Machado et.al. 2012
<b>Aónikenk Tehuelches</b>	15	0.0000	0.0000	0.2700	0.7300	0.0000	0.4224	Bisso-Machado et.al. 2012
<b>Apalaí</b>	120	0.3700	0.0100	0.3000	0.3200	0.0000	0.6762	Bisso-Machado et.al. 2012
<b>Arara</b>	70	0.5400	0.2000	0.2600	0.0000	0.0000	0.6095	Bisso-Machado et.al. 2012
<b>Araweté</b>	18	0.3900	0.0000	0.5000	0.1100	0.0000	0.6203	Bisso-Machado et.al. 2012
<b>Arequipa</b>	22	0.0900	0.6800	0.1400	0.0900	0.0000	0.5257	Bisso-Machado et.al. 2012
<b>Arhuaco</b>	21	0.9040	0.0480	0.0480	0.0000	0.0000	0.1871	Usme, et.al., 2013
<b>Arhuaco (Ijka)</b>	134	0.8700	0.0100	0.1200	0.0000	0.0000	0.2303	Bisso-Machado et.al. 2012
<b>Ijka</b>	40	0.9200	0.0300	0.0500	0.0000	0.0000	0.1541	Melton, et.al. 2007
<b>Arsario - Wiwa</b>	76	0.6300	0.0000	0.3700	0.0000	0.0000	0.4724	Bisso-Machado et.al. 2012
<b>Arsario</b>	50	0.6400	0.0000	0.3600	0.0000	0.0000	0.4702	Melton, et.al. 2007
<b>Asurini</b>	24	0.0400	0.5400	0.1700	0.2100	0.0400	0.6597	Bisso-Machado et.al. 2012
<b>Atacameño</b>	79	0.1300	0.7300	0.1000	0.0400	0.0000	0.4442	Bisso-Machado et.al. 2012
<b>Awa-Kuaikier</b>	37	0.0810	0.7570	0.0270	0.0810	0.0540	0.4216	Rondon, 2009 En Arias, 2012
<b>Awa-Guajá</b>	53	0.1300	0.8700	0.0000	0.0000	0.0000	0.2306	Bisso-Machado et.al. 2012
<b>Awa-Juriti</b>	18	0.0000	0.7200	0.1100	0.0000	0.1700	0.4665	Bisso-Machado et.al. 2012
<b>Aymara</b>	411	0.0400	0.7600	0.0800	0.1100	0.0100	0.4032	Bisso-Machado et.al. 2012
<b>Ayoreo</b>	91	0.0000	0.0000	0.8300	0.1700	0.0000	0.2853	Bisso-Machado et.al. 2012
<b>Barasano</b>	5	0.8000	0.0000	0.0000	0.2000	0.0000	0.4000	Usme, 2013
<b>Boruca</b>	14	0.2143	0.7143	0.0000	0.0714	0.0000	0.4725	Torrioni et al. 1993
<b>Bribri/Cabecar</b>	24	0.5417	0.4583	0.0000	0.0000	0.0000	0.5181	Torrioni et al. 1993
<b>Cayapa</b>	120	0.2900	0.4000	0.0900	0.2200	0.0000	0.7053	Bisso-Machado et.al. 2012
<b>Chimila</b>	35	0.8800	0.0000	0.0300	0.0600	0.0300	0.2267	Keyeux et.al., 2002
<b>Chimila</b>	21	0.9520	0.0480	0.0000	0.0000	0.0000	0.0960	Usme, et.al. 2013

<b>Chorote</b>	34	0.1500	0.4400	0.2300	0.1800	0.0000	0.7198	Bisso-Machado et.al. 2012
<b>Ciboney</b>	15	0.0700	0.0000	0.6000	0.3300	0.0000	0.5638	Melton, et.al. 2007
<b>Cinta-Larga</b>	45	0.2500	0.0000	0.2000	0.5300	0.0200	0.6302	Bisso-Machado et.al. 2012
<b>Coconuco</b>	11	0.0900	0.0900	0.7270	0.0900	0.0000	0.4919	Xavier, et.al. 2015
<b>Coreguaje</b>	69	0.0400	0.2000	0.6600	0.0600	0.0400	0.5252	Bisso-Machado et.al. 2012
<b>Coya</b>	60	0.1300	0.5700	0.2300	0.0500	0.0200	0.6126	Bisso-Machado et.al. 2012
<b>Coyaima</b>	67	0.3280	0.2390	0.3130	0.0300	0.0900	0.7394	Rondon, 2009 En Arias, 2012
<b>Cubeo</b>	22	0.2700	0.1800	0.5000	0.0500	0.0000	0.6728	Bisso-Machado et.al. 2012
<b>Cubeo</b>	24	0.3330	0.0830	0.2920	0.2500	0.0420	0.7646	Usme, et.al., 2013
<b>Curripaco</b>	22	0.4100	0.3600	0.2300	0.0000	0.0000	0.6803	Bisso-Machado et.al. 2012
<b>Embera (Cauca)</b>	21	0.3340	0.4780	0.0470	0.0940	0.0000	0.6814	Keyeux et.al., 2002
<b>Embera (Dabeiba)</b>	22	0.7300	0.2300	0.0000	0.0000	0.0500	0.4313	Mesa, et.al., 2000
<b>Embera Chocó</b>	21	0.0950	0.5240	0.2860	0.0950	0.0000	0.6569	Usme, et.al. 2013
<b>Embera Chamí (Segovia-Antioquia)</b>	38	0.4740	0.2630	0.0000	0.2630	0.0000	0.6542	Xavier,et.al., 2015
<b>Embera Córdoba</b>	29	0.3448	0.5862	0.0000	0.0000	0.0690	0.5517	Criollo, 2012
<b>Embera Panamá</b>	44	0.2300	0.5200	0.2500	0.0000	0.0000	0.6285	Melton, et.al. 2007
<b>Guambiano</b>	23	0.0400	0.0400	0.7900	0.1300	0.0000	0.3720	Keyeux et.al., 2002
<b>Guambiano</b>	24	0.0420	0.1250	0.6660	0.1670	0.0000	0.5334	Usme, et.al., 2013
<b>Guambiano</b>	22	0.0900	0.0450	0.7730	0.0900	0.0000	0.4025	Xavier, et.al., 2015
<b>Ingano</b>	27	0.1500	0.4400	0.3700	0.0000	0.0400	0.6702	Mesa, et.al., 2000
<b>Ingano</b>	21	0.3800	0.0950	0.5250	0.0000	0.0000	0.5995	Keyeux et.al., 2002
<b>Ingano</b>	43	0.1200	0.4600	0.4200	0.0000	0.0000	0.7777	Torres, et.al. 2006
<b>Ingano</b>	48	0.3960	0.3540	0.2290	0.0210	0.0000	0.6791	Usme, et.al. 2013
<b>Krahó</b>	14	0.2900	0.5700	0.1400	0.0000	0.0000	0.6154	Bisso-Machado et.al. 2012
<b>Kogi</b>	153	0.6700	0.0000	0.3300	0.0000	0.0000	0.4451	Bisso-Machado et.al. 2012
<b>Kogui</b>	32	0.7190	0.0000	0.2810	0.0000	0.0000	0.4171	Usme, et. al. 2013

<b>Kuben-Kran-Kegn</b>	19	0.5800	0.2600	0.0600	0.1000	0.0000	0.6148	Bisso-Machado et.al. 2012
<b>Kuna</b>	16	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	Torrioni et al. 1993
<b>Kuna</b>	63	0.7100	0.2900	0.0000	0.0000	0.0000	0.4184	Batista, 1995 en Melton et.al. 2007
<b>Kuna-Tule</b>	30	0.5000	0.2700	0.2000	0.0000	0.0300	0.6581	Keyeux et.al., 2002
<b>Macushi</b>	10	0.1000	0.2000	0.3000	0.4000	0.0000	0.7778	Bisso-Machado et.al. 2012
<b>Makiritare</b>	10	0.2000	0.0000	0.7000	0.1000	0.0000	0.5111	Bisso-Machado et.al. 2012
<b>Mapuche</b>	314	0.0500	0.2300	0.3200	0.3600	0.0400	0.7133	Bisso-Machado et.al. 2012
<b>Marubo</b>	10	0.1000	0.0000	0.6000	0.3000	0.0000	0.6000	Bisso-Machado et.al. 2012
<b>Matsiguenga</b>	38	0.0500	0.9200	0.0000	0.0300	0.0000	0.1543	Bisso-Machado et.al. 2012
<b>Maya</b>	27	0.5200	0.2200	0.1500	0.0700	0.0400	0.6773	Schurr 1990 en Melton et.al. 2007
<b>Maya</b>	30	0.6300	0.3300	0.0300	0.0000	0.0000	0.5103	Boles 1995 en Melton et.al. 2007
<b>Mixe5</b>	16	0.6300	0.3100	0.0600	0.0000	0.0000	0.5370	Melton et.al. 2007
<b>Mixtec (Alta)</b>	15	0.7300	0.1300	0.1300	0.0000	0.0000	0.4643	Melton et.al. 2007
<b>Mixtec (Baja)</b>	14	0.9300	0.0700	0.0000	0.0000	0.0000	0.1402	Melton et.al. 2007
<b>Mocoví</b>	5	0.8000	0.0000	0.0000	0.2000	0.0000	0.4000	Bisso-Machado et.al. 2012
<b>Moseten-Chimane</b>	71	0.3900	0.5400	0.0300	0.0000	0.0400	0.5617	Bisso-Machado et.al. 2012
<b>Movima</b>	22	0.0900	0.0900	0.6400	0.1800	0.0000	0.5676	Melton et.al. 2007
<b>Muinane</b>	19	0.1100	0.2100	0.3700	0.2600	0.0500	0.7777	Keyeux et.al., 2002
<b>Munduruku</b>	92	0.1200	0.1700	0.0900	0.5800	0.0400	0.6173	Bisso-Machado et.al. 2012
<b>Nasa Cauca</b>	14	0.4290	0.0710	0.2860	0.1430	0.0710	0.7578	Xavier et.al., 2015
<b>Nasa Tolima</b>	93	0.4086	0.1183	0.3763	0.0753	0.0215	0.6786	Criollo, 2012
<b>Navajo</b>	48	0.5833	0.3750	0.0000	0.0000	0.0417	0.5284	Torrioni et al. 1992 En Torrioni et al. 1993
<b>Ngobe</b>	46	0.6700	0.3300	0.0000	0.0000	0.0000	0.4520	Melton et.al. 2007
<b>Ngobe-Guaymi</b>	16	0.6875	0.3125	0.0000	0.0000	0.0000	0.4583	Torrioni et al. 1993
<b>Nukak</b>	20	0.0000	0.2000	0.8000	0.0000	0.0000	0.3368	Keyeux et.al., 2002

<b>Ojibwa</b>	28	0.6429	0.0357	0.0714	0.0000	0.2500	0.5370	Torroni et al. 1993
<b>Pacaás Novos (Wari)</b>	30	0.4000	0.3000	0.2700	0.0300	0.0000	0.6995	Bisso-Machado et.al. 2012
<b>Páez</b>	51	0.5900	0.1200	0.2700	0.0200	0.0000	0.5755	Bisso-Machado et.al. 2012
<b>Paez</b>	36	0.2780	0.0830	0.2780	0.3330	0.0280	0.7476	Usme, et.al. 2013
<b>Palikúr</b>	64	0.0100	0.4700	0.0400	0.4700	0.0100	0.5652	Bisso-Machado et.al. 2012
<b>Parakanã</b>	31	0.0600	0.3900	0.3200	0.2300	0.0000	0.7120	Bisso-Machado et.al. 2012
<b>Pasto</b>	9	0.6700	0.3300	0.0000	0.0000	0.0000	0.4975	Keyeux et.al., 2002
<b>Pehuenche</b>	205	0.0200	0.0900	0.4000	0.4900	0.0000	0.5943	Bisso-Machado et.al. 2012
<b>Piapoco</b>	39	0.1800	0.0300	0.1500	0.0500	0.5900	0.6092	Bisso-Machado et.al. 2012
<b>Piapoco</b>	8	0.1250	0.2500	0.2500	0.1250	0.2500	0.8929	Usme, et.al. 2013
<b>Piaroa</b>	28	0.3600	0.1100	0.2100	0.3200	0.0000	0.7382	Bisso-Machado et.al. 2012
<b>Piaroa Vichada</b>	18	0.277	0.1690	0.277	0.277	0.0000	0.7849	Keyeux et.al., 2002
<b>Pijao</b>	52	0.2692	0.2115	0.3269	0.1346	0.0577	0.7693	Criollo, 2012
<b>Pijao</b>	35	0.2290	0.2570	0.4290	0.0570	0.0290	0.7138	Rondon, 2009 En Arias, 2012
<b>Pilagá</b>	41	0.0500	0.3700	0.2700	0.2900	0.0200	0.7208	Bisso-Machado et.al. 2012
<b>Pilaga</b>	26	0.2700	0.3500	0.0400	0.3500	0.0000	0.7077	Melton et.al. 2007
<b>Pima</b>	30	0.0667	0.5000	0.4333	0.0000	0.0000	0.5770	Wallace et al. 1985; Schurr et al. 1990; Torroni et al. 1992 En Torroni et al. 1993
<b>Piratapuyo</b>	8	0.1250	0.1250	0.5000	0.1250	0.1250	0.7857	Usme, et. al. 2013
<b>Poturujara-Zoe Aislado</b>	23	0.4400	0.0000	0.2600	0.3000	0.0000	0.6783	Bisso-Machado et.al. 2012
<b>Puinave</b>	19	0.0500	0.1600	0.5800	0.1600	0.0500	0.6411	Bisso-Machado et.al. 2012
<b>Puinave</b>	61	0.0820	0.5080	0.3280	0.0660	0.0160	0.6334	Usme, et. al., 2013
<b>Quechua</b>	232	0.1400	0.6200	0.1500	0.0900	0.0000	0.5678	Bisso-Machado et.al. 2012
<b>Quechua</b>	23	0.0900	0.5700	0.3000	0.0400	0.0000	0.6016	Wang 2007 en Casas, 2017
<b>Sáliba</b>	13	0.1500	0.0000	0.5500	0.1500	0.1500	0.6825	Bisso-Machado et.al. 2012
<b>Selknam (Ona)</b>	16	0.0000	0.0000	0.5600	0.3800	0.0600	0.5743	Bisso-Machado et.al. 2012
<b>Siona</b>	12	0.7500	0.1700	0.0800	0.0000	0.0000	0.4388	Keyeux et.al., 2002

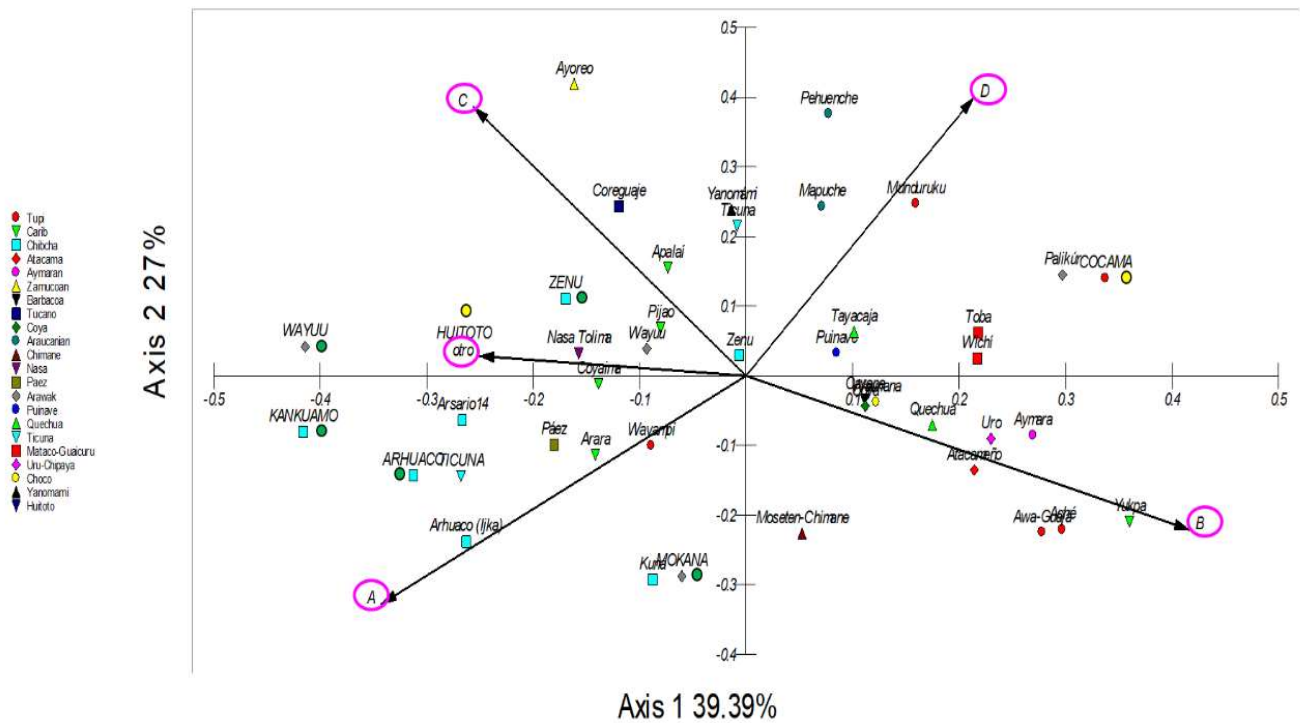
<b>Suruí</b>	44	0.0700	0.0400	0.0000	0.8900	0.0000	0.2061	Bisso-Machado et.al. 2012
<b>Tainos</b>	24	0.0000	0.0000	0.7500	0.2500	0.0000	0.3913	Melton et.al. 2007
<b>Tatuyo</b>	10	0.4000	0.0000	0.5000	0.1000	0.0000	0.6444	Usme, et.al. 2013
<b>Tayacaja</b>	61	0.2100	0.3300	0.1300	0.3000	0.0300	0.7515	Bisso-Machado et.al. 2012
<b>Tehuelche</b>	29	0.0000	0.2000	0.2400	0.5600	0.0000	0.6098	Bisso-Machado et.al. 2012
<b>Teribe</b>	20	0.8000	0.2000	0.0000	0.0000	0.0000	0.3368	Melton, et.al. 2007
<b>Ticuna</b>	371	0.2000	0.1100	0.3500	0.3300	0.0100	0.7183	Bisso-Machado et.al. 2012
<b>Tiriyó</b>	32	0.0900	0.1900	0.2200	0.4700	0.0300	0.7077	Bisso-Machado et.al. 2012
<b>Toba</b>	80	0.1500	0.4300	0.0500	0.3700	0.0000	0.6615	Bisso-Machado et.al. 2012
<b>Trinitarion</b>	35	0.1400	0.4000	0.3700	0.0300	0.0600	0.6990	Melton, et.al. 2007
<b>Tucano</b>	17	0.0000	0.1800	0.4700	0.3500	0.0000	0.6632	Keyeux et.al., 2002
<b>Tucano</b>	14	0.0710	0.2140	0.0710	0.5710	0.0710	0.6602	Usme, et.al., 2013
<b>Tucanos Guaviare</b>	24	0.2500	0.0830	0.3750	0.2920	0.0000	0.7354	Arias, 2012
<b>Tucanos Vaupés</b>	41	0.5370	0.0730	0.2930	0.0980	0.0000	0.6261	Arias, 2012
<b>Tupe</b>	16	0.0000	0.6900	0.3100	0.0000	0.0000	0.4563	Bisso-Machado et.al. 2012
<b>Tuyuca</b>	6	0.1700	0.5000	0.0000	0.1700	0.1600	0.7999	Usme, et.al. 2013
<b>Uro</b>	64	0.1100	0.6900	0.0900	0.1100	0.0000	0.4994	Bisso-Machado et.al. 2012
<b>Urubu Kaapor</b>	42	0.2100	0.3100	0.1400	0.2900	0.0500	0.7720	Bisso-Machado et.al. 2012
<b>Wai-wai</b>	26	0.1500	0.1500	0.4300	0.2700	0.0000	0.7251	Bisso-Machado et.al. 2012
<b>Wapishana</b>	12	0.0000	0.2500	0.0800	0.6700	0.0000	0.5260	Bisso-Machado et.al. 2012
<b>Waunana</b>	30	0.0000	0.6330	0.3000	0.0670	0.0000	0.5222	Keyeux et.al., 2002
<b>Waunana</b>	74	0.2200	0.5100	0.0300	0.2400	0.0000	0.6417	Rojas, et.al., 2010
<b>Wounan</b>	31	0.2900	0.1900	0.4800	0.0300	0.0000	0.6701	Melton, et.al. 2007
<b>Wayampi</b>	99	0.6200	0.1100	0.0800	0.1900	0.0000	0.5667	Bisso-Machado et.al. 2012
<b>Wayuu</b>	89	0.2600	0.2800	0.4500	0.0000	0.0100	0.6588	Bisso-Machado et.al. 2012
<b>Wayuu</b>	17	0.2940	0.1760	0.4710	0.0000	0.0590	0.6983	Usme, et.al., 2013
<b>Wichí</b>	199	0.1200	0.5100	0.0700	0.2900	0.0100	0.6396	Bisso-Machado et.al. 2012

<b>Xavante</b>	25	0.1600	0.8400	0.0000	0.0000	0.0000	0.2800	Melton, et.al. 2007
<b>Yagua</b>	12	0.2500	0.0000	0.6700	0.0800	0.0000	0.5260	Bisso-Machado et.al. 2012
<b>Yámana(Yaghan)</b>	32	0.0000	0.0000	0.6300	0.3700	0.0000	0.4812	Bisso-Machado et.al. 2012
<b>Yanacona</b>	6	0.1670	0.0000	0.8330	0.0000	0.0000	0.3339	Xavier, et.al., 2015
<b>Yanomámi</b>	491	0.0200	0.2500	0.5000	0.1900	0.0400	0.6507	Bisso-Machado et.al. 2012
<b>Yeral</b>	8	0.1250	0.1250	0.5000	0.2500	0.0000	0.7500	Usme, et.al., 2013
<b>Yukpa</b>	88	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	Bisso-Machado et.al. 2012
<b>Yungay</b>	38	0.0500	0.4500	0.3400	0.1600	0.0000	0.6715	Bisso-Machado et.al. 2012
<b>Yuracare</b>	28	0.3900	0.3200	0.2100	0.0400	0.0400	0.7241	Melton, et.al. 2007
<b>Zapotec</b>	15	0.3300	0.3300	0.3300	0.0000	0.0000	0.7214	Melton, et.al. 2007
<b>Zenú</b>	107	0.1900	0.3800	0.3600	0.0500	0.0200	0.6935	Bisso-Machado et.al. 2012
<b>Zenú Sahagun</b>	48	0.2917	0.1250	0.2083	0.1042	0.2708	0.7881	Criollo, 2012
<b>Zoro</b>	30	0.2000	0.0700	0.1300	0.6000	0.0000	0.5981	Melton, et.al. 2007

Tanto el AMOVA realizado con poblaciones de referencia americanas y colombianas a partir de las frecuencias de haplogrupos, como a partir de las secuencias, mostró valores significativos de  $F_{ST}=0.2504$  ( $p=0.000$ ) y  $0.1776$  ( $p=0.000$ ) y de  $F_{SC}=0.2031$  y  $0.1663$  ( $p=0.000$ ), respectivamente, otorgando el mayor porcentaje de variabilidad a las diferencias entre poblaciones dentro de los grupos.

En el Análisis de Componentes Principales considerando poblaciones de referencia Americanas con tamaño de muestra superior a 50, los dos primeros componentes reúnen el 66.3% de la variación. Las poblaciones del subtronco del lenguaje Chibcha se agrupan bajo la influencia del haplogrupo A, mientras las Araucarianas lo hacen por el haplogrupo D. Las poblaciones de lengua Carib, exceptuando Yukpa son separadas las por el segundo componente principal de las de lengua Quechua, Mataco-Guaicurú y de las de lengua Tupi, a excepción de Wayampi, lo que sugiere una posible relación entre frecuencias de haplogrupos y subgrupos del lenguaje. **Figura 5-6.**

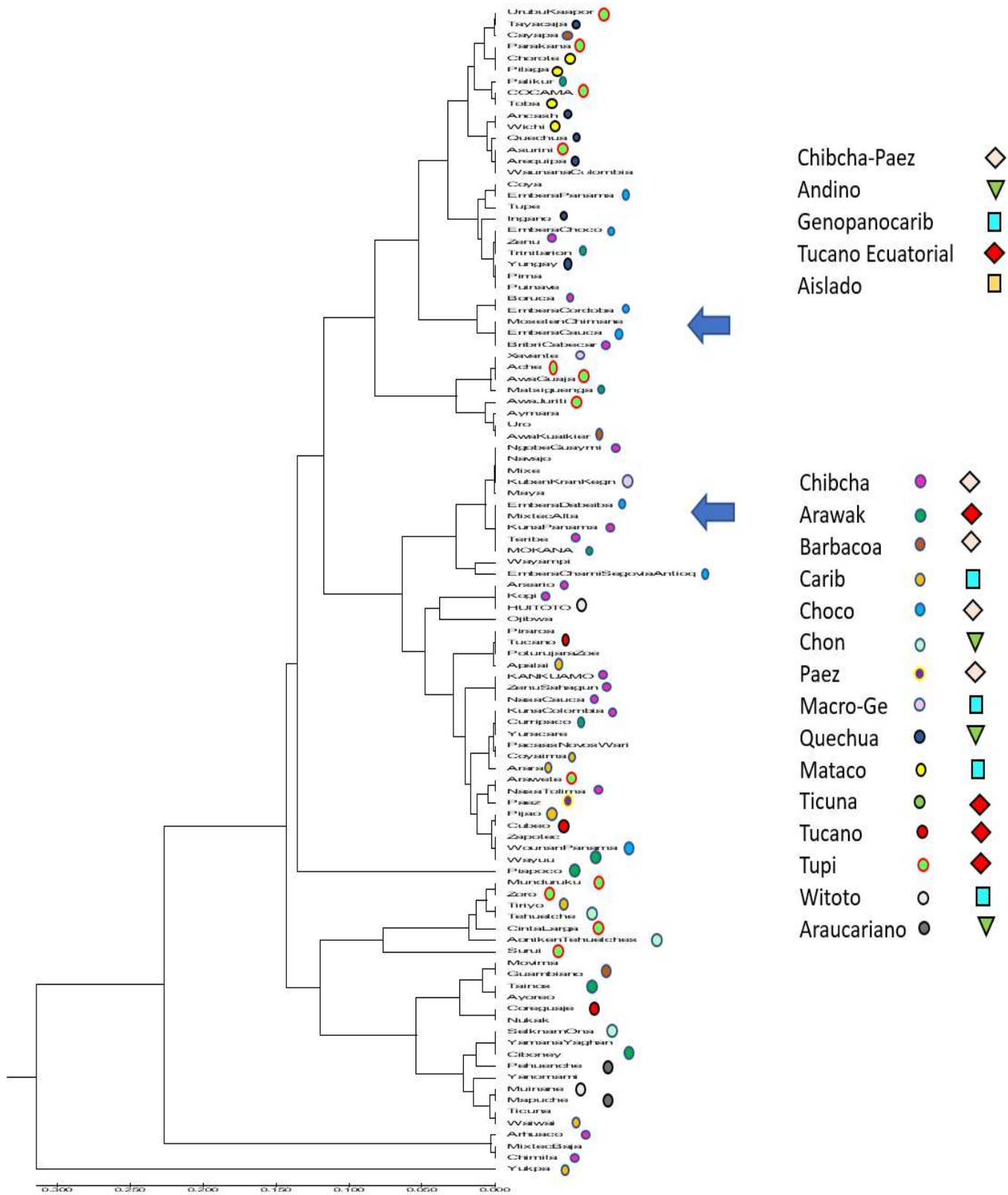
**Figura 5-6.** Análisis de Componentes Principales con base en frecuencias de haplogrupos para etnias Americanas con N>50.



Para explorar esta posibilidad, se construyó un árbol Neighbor-Joining a partir de las frecuencias de los haplogrupos. **Figura 5-7.** Un total de 102 poblaciones de America fueron incluidas y corresponden a este trabajo solo las de Kankuamo, Mokaná, Cocama y Huitoto. Se puede observar como Yukpa se separa de todas las demás poblaciones, lo cual es concordante con la presencia únicamente del Hgr B en la muestra y además un cluster aislado que relaciona Arhuaco y Chimila, habitantes del Caribe biogeográfico colombiano con lengua Chibcha, con la población Mixteca perteneciente al tronco lingüístico Amerindio Central. Las demás etnias se organizan en dos grupos. El primero de ellos, conformado por el mayor número de poblaciones, alberga las demás comunidades de lengua Chibcha, Paez, Chocó y Barbacoa, a excepción de Guambiano, las de lengua Arawak, excepto las del caribe insular Tainos y Ciboney, todas las de lengua Quechua y Mataco, el 57% de las etnias de lengua Carib, 66% de lengua Tucano y el 69% de lengua Tupi.



**Figura 5-7.** Arbol UPGMA construido a partir de la matriz Fst pareados con base en frecuencias de haplogrupos



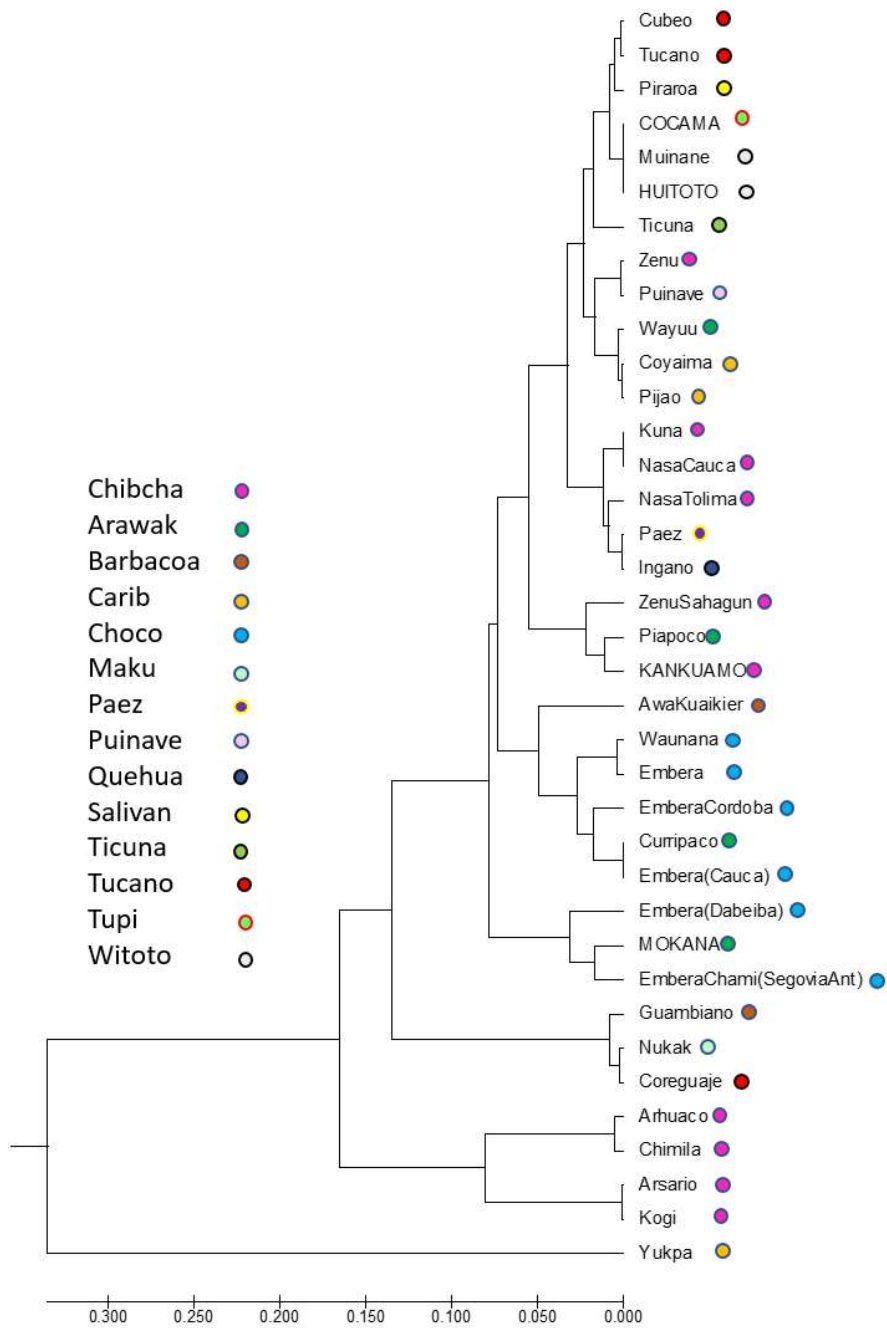
Cabe destacar también, que en este cluster se encuentran las etnias de lenguas Amerindias del Norte y Centrales, más cercanamente relacionadas con las poblaciones de lengua Chibcha mesoamericanas y de Panamá y con la población Embera de Dabeiba Antioquia y a las muestras de la etnia Mokaná en Colombia, así como un agrupamiento que reúne etnias chibcha de Mesoamerica y otras etnias Embera de lengua Choco de diferentes zonas del territorio Colombiano, lo cual permite cuestionarse si esta aparente relación genética aporta a la resolución de conflicto lingüístico sobre la pertenencia o no del lenguaje Choco al tronco Chibcha-Paez propuesto por Greenberg. Finalmente en el segundo agrupamiento, solo se encuentran de manera exclusiva las poblaciones de lengua Araucariana Pehuenche y Mapuche y de lengua Chon Tehuelches y Selknam.

Esto puede interpretarse, en concordancia con lo encontrado en el ACP, como la posible evidencia de varios grados de diferenciación en razón de la dispersión geográfica o de diferentes formas de aislamiento entre grupos étnicos a los que se les atribuye un posible origen ancestral común, por compartir el subtronco lingüístico.

Analizando solamente las etnias colombianas, de nuevo la etnia Yukpa se encuentra separada de las demás. Se observa que las poblaciones chibcha de la Sierra Nevada de Santa Marta se agrupan separadas de las demás, al igual que la mayoría de las tribus amazónicas, a excepción de las de lengua Puinave que se relacionan con etnias de la costa Caribe (Zenú y Wayuu) y del Tolima (Coyaima y Pijao), y las etnias Arawak, Piapoco, Mokaná y Curripaco, que hacen parte de un cluster con Zenú y Kankuamo de la Costa Caribe en el primer caso, y del mismo grupo con los Embera para las dos últimas. **Figura 5-8.**

En ambos árboles, aunque de manera más importante en el de las etnias colombianas, aparentemente hay una influencia de la cercanía geográfica en la forma como se relacionan genéticamente las poblaciones.

**Figura 5-8.** Arbol UPGMA a partir de distancias  $F_{st}$  con base en frecuencias de haplogrupos para las etnias colombianas



**Tabla 5-7.** Frecuencias de Subhaplogrupos para muestras de las etnias del Caribe y del Amazonas biogeográficos en Colombia

POBLACION	N	A2+(64)	A2+(64) +16129	A2+(64) +16189	A2ac	A2al	A2q	B4	B4c1c	C	C1	C1c3	M	D1	D1f	L2a1+16189 (16192)	L2a1c2a	L2a1i	L3e1b2
ARHUACO	10	0.000	0.700	0.000	0.000	0.000	0.000	0.000	0.000	0.200	0.000	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000
KANKUAMO	6	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.167	0.000	0.000	0.000	0.000	0.000	0.167	0.167
MOKANA	12	0.417	0.000	0.000	0.000	0.250	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
WAYUU	3	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.333	0.000	0.000	0.000
WIWA	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ZENÚ	8	0.125	0.000	0.000	0.000	0.125	0.000	0.125	0.000	0.000	0.000	0.375	0.125	0.125	0.000	0.000	0.000	0.000	0.000
COCAMA	2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000
YUCUNA	1	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TICUNA	4	0.000	0.000	0.000	0.500	0.000	0.250	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HUITOTO	2	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MACUNA	1	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BORA	1	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

### 5.3.2.1 Análisis de subhaplogrupos y secuencias

Para realizar el análisis de las secuencias se utilizaron como muestras de referencia las poblaciones de la publicación de Tamm et.al., 2007 que contaran con las regiones hipervariables HVI y HVII. **Tabla 5-8.**

**Tabla 5-8.** Poblaciones empleadas en el análisis de secuencias. \* Este estudio

ETNIA	N	UBICACIÓN	DivGenProm
Arhuaco*	9	Caribe Sierra Nevada de Santa Marta Colombia	0.556
Kankuamo*	4	Caribe Sierra Nevada de Santa Marta Colombia	0.500
Mokana*	12	Caribe Colombia	0.970
Wayuu*	2	Caribe Colombia	1.000
Wiwa*	1	Caribe Sierra Nevada de Santa Marta Colombia	0.000
Zenú*	7	Caribe Colombia	0.952
Cocama*	2	Amazonas Colombia	1.000
Yucuna*	1	Amazonas Colombia	0.000
Ticuna*	4	Amazonas Colombia	1.000
Huitoto*	2	Amazonas Colombia	1.000
Macuna*	1	Amazonas Colombia	0.000
Bora*	1	Amazonas Colombia	0.000
Apache	9	Norteamérica	0.694
WiwaRef	47	Caribe Sierra Nevada de Santa Marta Colombia	0.801
Cofan	2	Amazonas Colombia	1.000
Catamarca	10	Argentina	1.000
Cayapa	16	Amazonas Ecuador	0.900
Coreguaje	26	Amazonas Colombia	0.874
Dogrib	3	Norteamérica	1.000
Embera	18	Choco Colombia	0.941
ArhuacoRef	29	Caribe Sierra Nevada de Santa Marta Colombia	0.466
Kuna	9	Panamá	0.722
Kogi	48	Caribe Sierra Nevada de Santa Marta Colombia	0.668
Mocovi	2	Argentina	1.000
Ngobe	44	Panamá	0.828
PaiuteNorte	3	Centroamérica	1.000
Salta	6	Argentina	0.933
Secoya	7	Amazonas Colombia	0.667
Siona	1	Amazonas Colombia	0.000
Tucuman	7	Argentina	1.000
Vaupés	21	Amazonas Colombia	0.986
Waunana	44	Choco Colombia	0.975

<b>WayuuRef</b>	41	Caribe Colombia	0.810
<b>Zuni</b>	2	Centroamérica	1.000

El AMOVA realizado para este conjunto de poblaciones muestra un valor significativo de  $F_{ST}=0.1769$  y de  $F_{SC}=0.1705$ , ambos con p valor =0.0000, evidenciando subestructura entre las poblaciones dentro de los grupos.

**Tabla 5-9.** Polimorfismos encontrados en las muestras de Caribe y Amazonas Colombiano

PID	POLIMORFISMO HVI	POLIMORFISMO HVII	Hgr	CLASIF. GENEALOGICA
GUSJ 01023	C16193T G16213A C16223T C16239T C16278T C16294T A16309G G16390A	A73G G143A T146C T152C T195C A263G 309.1C 315.1C	L2a1c2a	ARHUACO
CEVA 01013	C16111T G16129A C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16129	ARHUACO
CEPB 01001	C16223T C16245T T16298C C16327T	A73G A249del A263G 315.1C	C	ARHUACO
CEPB 01007	C16111T G16129A C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 309.2C 315.1C	A2+(64) +16129	ARHUACO
CEPB 01008	C16111T G16129A C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16129	ARHUACO
CEPB 01021	C16223T C16245T T16298C C16327T	A73G A249del A263G 315.1C	C	ARHUACO
CEPB 01022	C16111T G16129A C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16129	ARHUACO
CEPB 01024	C16111T G16129A C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16129	ARHUACO
CEPB 01032	C16111T G16129A C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16129	ARHUACO
CEPB 01034	C16111T G16129A C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16129	ARHUACO
91001 01167	A16183C T16189C 16193.1C T16217C (LHP)	A73G A263G 309.1C 309.2C C311T 315.1C 315.2C	B4	BORA

91001 01067	C16142T C16147T C16223T T16325C	A73G T195C T196C A263G 315.1C	D1f	COCAMA
91001 01141	A16051G A16183C T16189C 16193.1C T16217C	A73G C150T A263G 309.1C 315.1C	B4c1c	COCAMA
91001 01131	C16111T T16172C G16213A C16223T C16290T G16319A T16362C	C64T A73G T125C T127C C128T T146C T152C A153G T195C A235G A263G 309.1C 315.1C	A2ac	HUITOTO
91001 01240	C16223T T16298C T16325C C16327T	A73G A188G A249del A263G A290del A291del 309.1C 315.1C	C1	HUITOTO
GUSJ 01036	C16223T A16265G T16298C G16319A T16325C C16327T	A73G A249del A263G A290del A291del 309.1C 309.2C 315.1C	C1c3	KANKUAMO
CEPB 01006	T16093C T16189C C16192T C16223T C16278T C16294T T16362C G16390A (LHP)	A73G G143A T146C T152C T195C A263G 315.1C	L2a1i	KANKUAMO
CEPB 01027	C16223T C16239T T16325del C16327T	A73G C150T G185A A189G A263G 309.1C 315.1C	L3e1b2	KANKUAMO
CEPB 01033	C16111T T16189C C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16189	KANKUAMO
CEPB 01038	C16111T T16189C C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16189	KANKUAMO
91001 01137	A16051G A16183C T16189C 16193.1C T16217C C16360T	T146C T152C A263G 309.1C 315.1C	B4	MACUNA
ATTU 01001	A16183C T16189C 16193.1C T16217C (LHP)	A73G A263G 309.1C 315.1C	B4	MOKANA
ATTU 01002	T16092C C16111T C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 309.2C 315.1C	A2+(64)	MOKANA
ATTU 01003	T16092C C16111T C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64)	MOKANA
ATTU 01004	T16092C C16111T C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64)	MOKANA
ATTU 01006	A16183C T16189C 16193.1C T16217C (LHP)	A73G A263G 309.1C 309.2C 315.1C	B4	MOKANA
ATTU 01010	T16092C C16111T C16214T C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 315.1C	A2+(64)	MOKANA
ATTU 01012	C16111T T16126C C16223T C16290T A16305G T16311C G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2al	MOKANA
ATTU 01014	A16183C T16189C 16193.1C T16217C (LHP)	A73G A263G 309.1C 315.1C	B4	MOKANA
ATTU 01016	C16111T T16126C C16223T C16245T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 315.1C	A2al	MOKANA

ATBA 01001	T16189C T16217C	A73G A263G 315.1C	B4	MOKANA
ATBA 01016	C16111T T16126C T16209C C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2al	MOKANA
ATBA 01029	T16092C C16111T A16180G C16223T C16290T A16293T G16319A T16362C	C64T A73G T146C A153G A235G A263G 315.1C	A2+(64)	MOKANA
91001 01158	C16111T T16209C C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 309.2C 315.1C	A2q	TICUNA
91001 01172	C16111T G16213A C16223T A16241T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2ac	TICUNA
91001 01195	G16129A T16178C C16223T T16298C A16318G T16325C C16327T	A73G A249del A263G A290del A291del 309.1C 315.1C	C1	TICUNA
91001 01235	C16111T G16213A C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G A291G 309.1C 309.2C 315.1C	A2ac	TICUNA
GUSJ 01009	C16223T T16298C T16325C C16327T	A73G A249del A263G A290del A291del 309.1C 315.1C	C1	WAYUU
GUSJ 01033	C16111T G16129A C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16129	WAYUU
CEVA 01007	C16111T T16189C C16223T C16290T G16319A T16362C (LHP)	C64T A73G T146C A153G A235G A263G 309.1C 315.1C	A2+(64) +16189	WAYUU
GUSJ 01038	C16071T T16189C C16192T C16223T C16278T C16294T A16309G G16390A	A73G T146C T152C T195C A263G 315.1C	L2a1 +16189 (16192)	WAYUU
GUSJ 01006	C16223T A16265G T16298C G16319A T16325C C16327T	A73G A249del A263G A290del A291del 309.1C 315.1C	C1c3	WIWA
91001 01081	C16111T G16213A C16223T C16290T G16319A T16362C	A73G T125C T127C T146C A153G A235G A263G 309.1C 315.1C	A2ac	YUCUNA
COTU 03002	T16189C T16217C	A73G A263G 309.1C 315.1C	B4	ZENÚ
COTU 03009	C16111T T16126C C16223T C16290T G16319A T16362C	C64T A73G T146C A153G G185A A235G A263G 315.1C	A2al	ZENÚ
COTU 02012	C16223T A16265G T16298C G16319A T16325C C16327T	A73G A249del A263G A290del A291del 315.1C	C1c3	ZENÚ
COTU 02019	C16223T A16265G T16298C G16319A T16325C C16327T T16362C	A73G A249del A263G A290del A291del 315.1C	C1c3	ZENÚ
COTU 02020	C16223T A16265G T16298C G16319A T16325C C16327T T16362C	A73G A249del A263G A290del A291del 315.1C	C1c3	ZENÚ



<b>COTU 02029</b>	T16189C 16193.1C C16223T T16325C T16362C (LHP)	A73G A263G 315.1C	D1	ZENÚ
<b>COTU 02045</b>	C16223T T16362C G16391A	A73G G143A A188G A263G 309.1C 315.1C	M	ZENÚ
<b>COTU 02050</b>	C16111T C16223T C16290T G16319A T16362C	C64T A73G T146C A153G A235G A263G T310C	A2+(64)	ZENÚ

Las muestras analizadas del Caribe Continental y de la Amazonia Colombianas, fueron clasificadas de acuerdo con EMPOP y Haplogrep hasta subhaplogrupos en los niveles posibles, teniendo en cuenta que no se realizó genotipificación de la región hipervariable III ni de regiones codificantes. Los resultados obtenidos se muestran en la **Tabla 5-7**. Los linajes panamericanos A2, C1 y D1 fueron detectados en las muestras analizadas, así como el subHgr D1f en una muestra de Amazonas, que ha sido propuesto como un antiguo linaje que estaba presente en las poblaciones paleoamericanas por Kummar et.al. (2011). La resolución alcanzada para el Hgr B, no permite determinar la presencia del linaje panamericano B2 en este estudio.

### Haplogrupo A2.

La **figura 5-9a** muestra la distribución global de acuerdo con EMPOP de los haplogrupos A2q, A2al y A2ac encontrados en las muestras analizadas en este trabajo. Todas las muestras de las etnias de la región Caribe tienen el SbHgr A2+64 que no se encuentra en ninguna de las muestras de la Amazonia. Todos los individuos A2 de Wayuu tienen además la mutación 16129, que comparten con el 70% de los Arhuacos de este trabajo, y como se observa en el árbol filogenético del haplogrupo, en la **Figura 5-9b**, se encuentra también en las muestras de referencia de Arhuacos y Wiwas (Tamm, et.al., 2007).

La mutación 16189 está presente solamente en Kankuamos, pero ha sido reportada por Tamm et.al. 2007 en Kogis y Wiwas con más frecuencia y menos representadas en Cayapa, Dogrib, Emberas y Waunana, y también en EMPOP con mayor frecuencia en Centroamérica. Zenú y Mokaaná cuentan con individuos A2+64 (reportado en Wiwas, Arhuacos, Kogis y Ngabes con mayor frecuencia y en menor proporción en Cayapa, Embera, Dogrib, Waunana y Wayuu entre otras por Tamm et.al., 2007) y A2al. Sobre este SubHgr se ha encontrado poca información, solo un reporte de 31 haplotipos en el Sur del

Brasil (EMPOP) sin que se especifique su pertenencia a una etnia indígena, y en 3 individuos de las muestras de referencia de las etnias Embera, Waunana y de la zona del Vaupés, sin que ninguno de ellos corresponda al mismo haplotipo aquí reportado. Cabe resaltar sin embargo que la zona indicada como de origen de las muestras referidas en este subclado en EMPOP, fue habitada por pueblos de lengua Tukano ecuatorial y Genopanocarib.

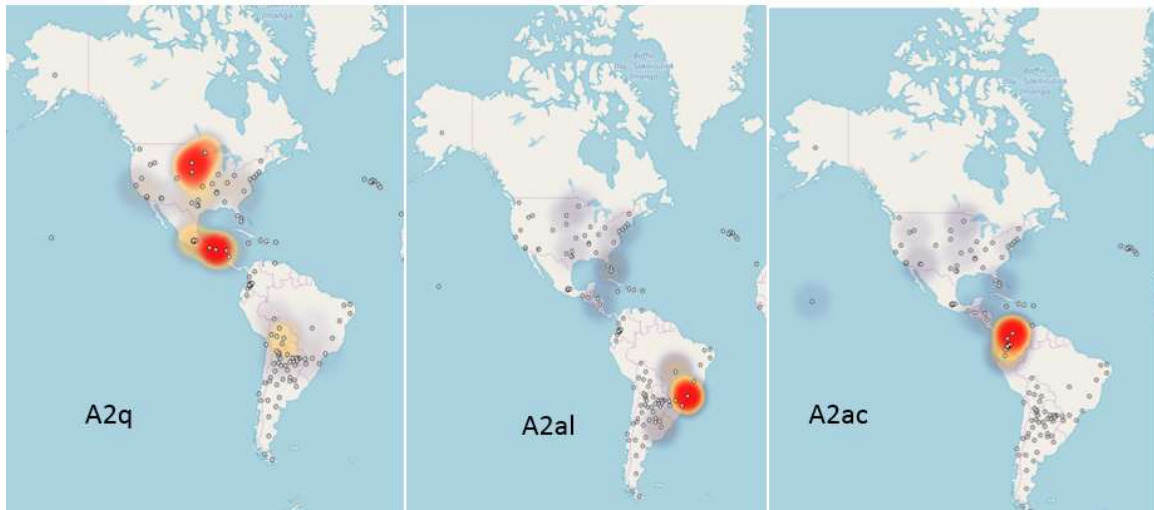
Para el caso de Amazonas, el subhaplogrupo más frecuente es A2ac (Yucuna, Huitoto y Ticuna), reportado en EMPOP para 40 muestras en Antioquia (Colombia), en Embera Chamí y Guambianos (Xavier, et.al., 2015) y en las muestras de referencia del departamento del Vaupés y en muestras de la etnia Cayapo, sin haplotipos coincidentes (Tamm, et.al., 2007), además de 7 muestras colectadas en población mezclada de Ecuador (Brandini, et.al., 2017) y 2 muestras de Ticuna de Colombia (Rieux, et.al., 2014). Brandini et.al., 2017 reporta la edad de este subHgr entre 12300 a 16900 años, e indica que se ha encontrado previamente en un hispano de USA, por lo que difiriendo de su apreciación, debería considerarse en principio y hasta obtener más evidencia, que este subHgr es propio de Suramérica, y que la presencia de un individuo cubano tipificado como A2ac1 (Behar, et.al., 2008), un subHgr con un origen aproximado hace 4000 a 6000 años (Brandini, et.al., 2017; Casas, 2017), podría estar relacionado con la hipótesis del origen de los Tainos a partir de migraciones reversas de poblaciones amazónicas (Shroeder, et.al., 2018), sin embargo, se requiere verificar la genealogía del individuo estudiado y ampliar el muestreo para confirmar esta posibilidad. El subHgr A2ac ha sido reportado por Casas 2017 en muestras antiguas de Muisca Siglo X-XVI y Formativo 1826+40 a.P con un tiempo estimado al Ancestro Común Más Reciente de 11480, y con base en la distribución de sublinajes como A2ac1, A2ac2a y A2y, propone continuidad genética con expansiones a varias regiones del territorio colombiano, y a otros países vecinos incluida la Amazonía Ecuatoriana.

De acuerdo con Kummar 2011, el linaje A2q fue reportado en población norteamericana por VanOven et.al., 2009. El presente trabajo reporta un individuo Ticuna A2q que está representado en 97 haplotipos de Centro y Norteamérica, con 8 coincidencias completas en nativoamericanos de Guatemala en una base de 3374 perfiles de esta metapoblación según EMPOP. Desafortunadamente no se cuenta con más información ni publicaciones al respecto, aunque podría tratarse de etnias Mayas que corresponden al mayor porcentaje

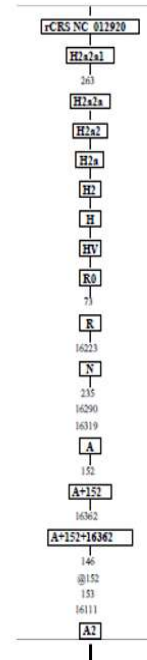
de grupos indígenas en Guatemala. Dado que Behar et.al., 2008 ha propuesto el origen de este subHgr en 13000 años aproximadamente, podría considerarse esta como una evidencia de un linaje que llegó con los primeros pobladores y que posteriormente sufrió deriva genética, ya que no se ha reportado en otras poblaciones suramericanas.

El árbol filogenético para el haplogrupo A, construido usando las muestras de referencia de poblaciones nativoamericanas publicadas por Tamm et.al. 2007 (**Figura 5-9b**), muestra las posibles mutaciones reversas o perdidas, las mutaciones privadas locales, es decir aquellas que se detectan en la muestra pero que no están relacionadas con la definición del SubHgr, y las mutaciones privadas globales, correspondientes a las no reportadas en PhyloTree, dos de estas últimas encontradas en Ticuna y Mokaneá.

**Figura 5-9a.** Representación de subHgr A2 mitocondriales en EMPOP. Imagen Tomada de EMPOP.



- Apache ●
- WiwaRef ●
- Cayapa ●
- Dogrib ●
- ArhuacoRef ●
- Mocovi ●
- Ngobe ●
- Paiute Norte ●
- Salta ●
- Secoya ●
- Siona ●
- Zuni ○
- Tucuman ●
- Ticuna ●
- Vaupes ●
- Waunana ●
- WayuuRef ●
- Arhuaco ●
- Kankuamo ●
- Mokaná ●
- Wayú ●
- Wiwa ●
- Zenú ●
- Huitoto ●
- Embera ●



**KEY**

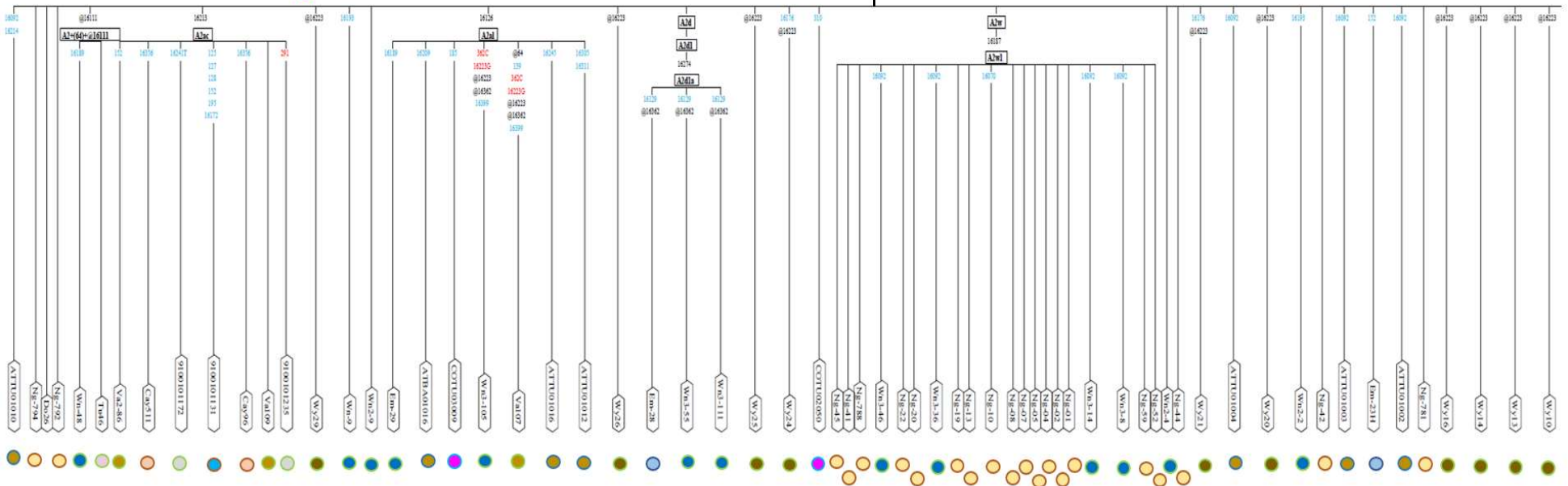
Local private mutation ●

Global private mutation ●

@ = assumed back mutation or missing mutation

Heteroplasmic mutation ●

Figura 5-9b. Arbol filogenético del Haplogrupo A elaborado por Haplogrep 2.0, a partir de las secuencias de las regiones HVI y HVII del ADNmt.



- |              |   |          |   |
|--------------|---|----------|---|
| Apache       | ● | Huitoto  | ● |
| WiwaRef      | ● | Embera   | ● |
| Cayapa       | ● | Kankuamo | ● |
| Dogrib       | ● | Kogi     | ● |
| ArhuacoRef   | ● | Zenú     | ● |
| Mocovi       | ● | Kuna     | ● |
| Ngobe        | ● |          |   |
| Paiute Norte | ● |          |   |
| Salta        | ● |          |   |
| Secoya       | ● |          |   |
| Siona        | ● |          |   |
| Zuni         | ● |          |   |
| Tucuman      | ● |          |   |
| Ticuna       | ● |          |   |
| Vaupés       | ● |          |   |
| Waunana      | ● |          |   |
| WayuuRef     | ● |          |   |
| Arhuaco      | ● |          |   |
| Kankuamo     | ● |          |   |
| Mokaná       | ● |          |   |
| Wayúú        | ● |          |   |
| Wiwa         | ● |          |   |

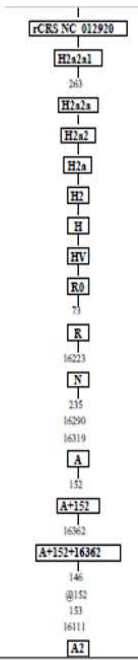
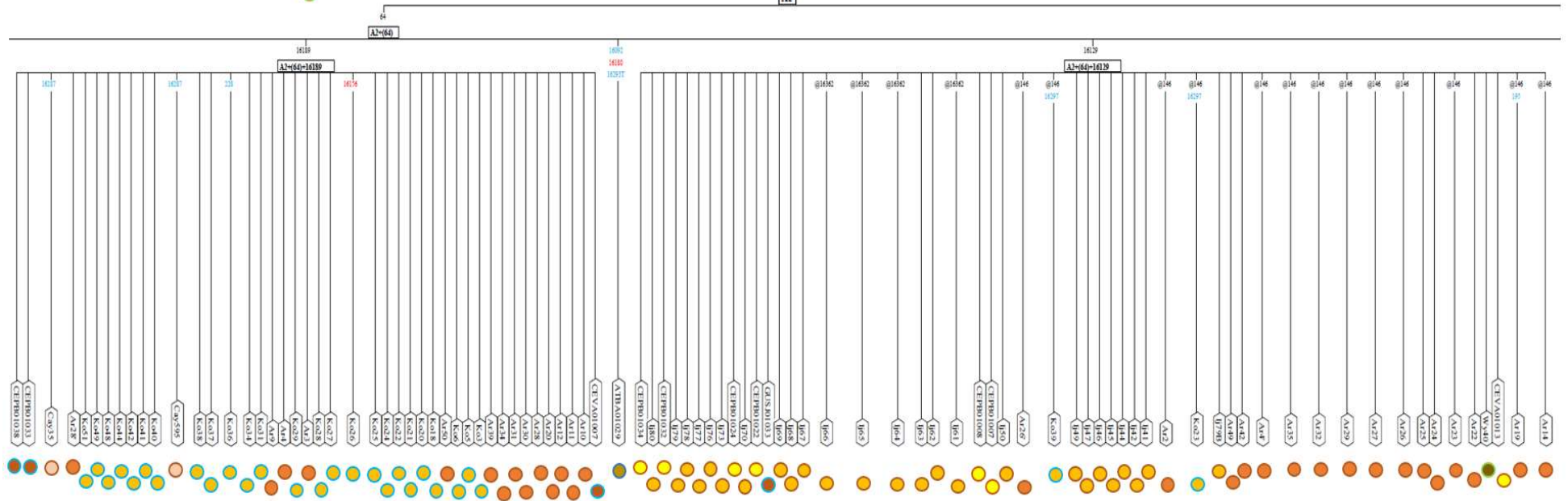


Figura 5-9b Continuación. Arbol filogenético del Haplogrupo A elaborado por Haplogrep 2.0, a partir de las secuencias de las regiones HVI y HVII del ADNmt.



- Apache ●
- WiwaRef ●
- Cayapa ●
- Dogrib ●
- ArhuacoRef ●
- Mocovi ●
- Ngobe ●
- Paiute Norte ●
- Salta ●
- Secoya ●
- Siona ●
- Zuni ●
- Tucuman ●
- Ticuna ●
- Vaupes ●
- Waunana ●
- WayuuRef ●
- Arhuaco ●
- Kankuamo ●
- Mokaná ●
- Wayuú ●
- Wiwa ●
- Zenú ●

- Huitoto ●
- Embera ●
- Kankuamo ●
- Kogi ●
- Kuna ●
- Catamarca ●
- Yucuna ●

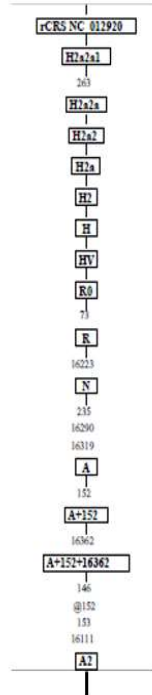
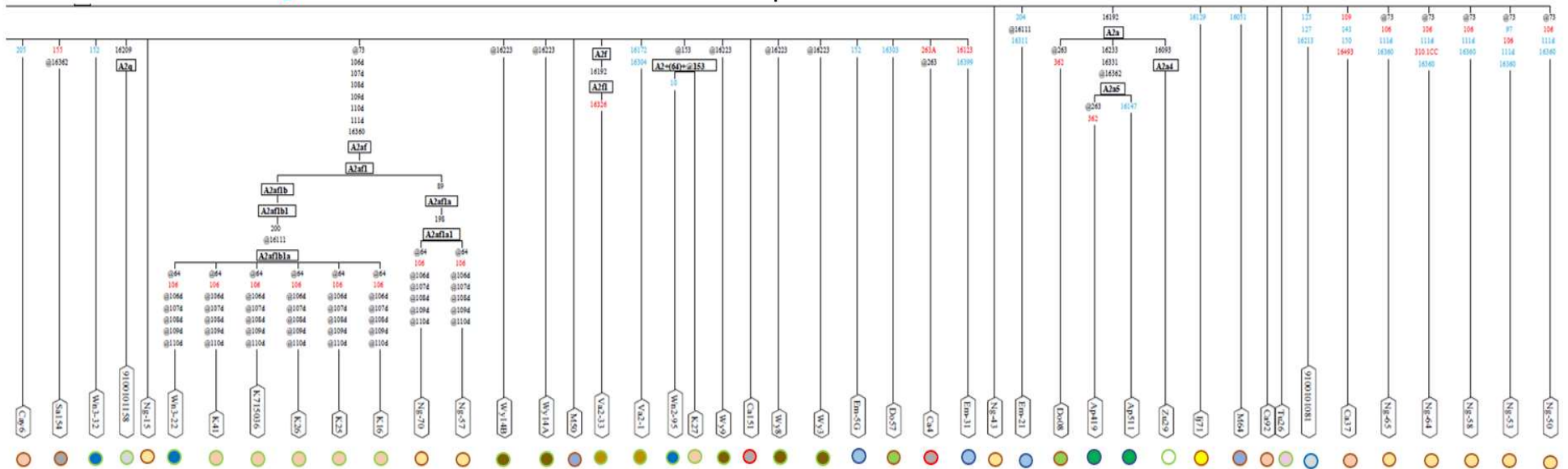


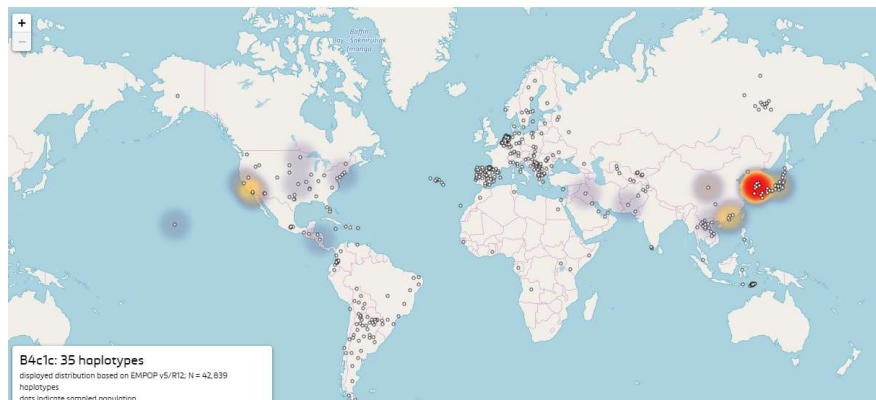
Figura 5-9b Continuación. Arbol filogenético del Haplogrupo A elaborado por Haplogrep 2.0, a partir de las secuencias de las regiones HVI y HVII del ADNmt.



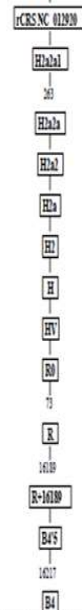
## Haplogrupo B.

Dada la resolución técnica en este trabajo, no fue posible diagnosticar este haplogrupo mas allá de B4, sin embargo el análisis realizado con las herramientas EMPOP y Haplogrep, se clasificó un individuo Cocama como B4c1c. Este SubHgr se encuentra con mayor frecuencia en poblaciones de Asia Oriental, y con menor representación en muestras del sur de Estados Unidos, como se reporta en la base de datos EMPOP. **Figura 5-10a.** Aunque no hay haplotipos coincidentes si se hallan vecinos por uno o dos cambios mutacionales en población nativoamericana de Mexico, Honduras, Estados Unidos y Guatemala. Behar 2008 reporta una edad aproximada de 9000 años para este subclado, tiempo después de la llegada de los primeros pobladores. Como se observa en el árbol filogenético del haplogrupo B2 (**Figura 5-10b**), solo se encontró relación con un individuo Ngobe de las muestras de referencia (Tamm, et.al., 2007) clasificado con Haplogrep 2.0 como B4c1a1a. Las demás muestras, una de Macuna, una de Cocama, una de Bora, una de Zenú y 4 de Mokaná, solo se clasificaron como B4, sin embargo, como una circunstancia particular, las tres muestras tomadas en la localidad de Tubará, Atlántico, sitio de asentamiento Mokaná, presenta una posible heteroplasmia de longitud causada por 16189C.

**Figura 5-10a.** Frecuencia del subhaplogrupo B4c1c de acuerdo con la base de datos EMPOP. Figura tomada de EMPOP.



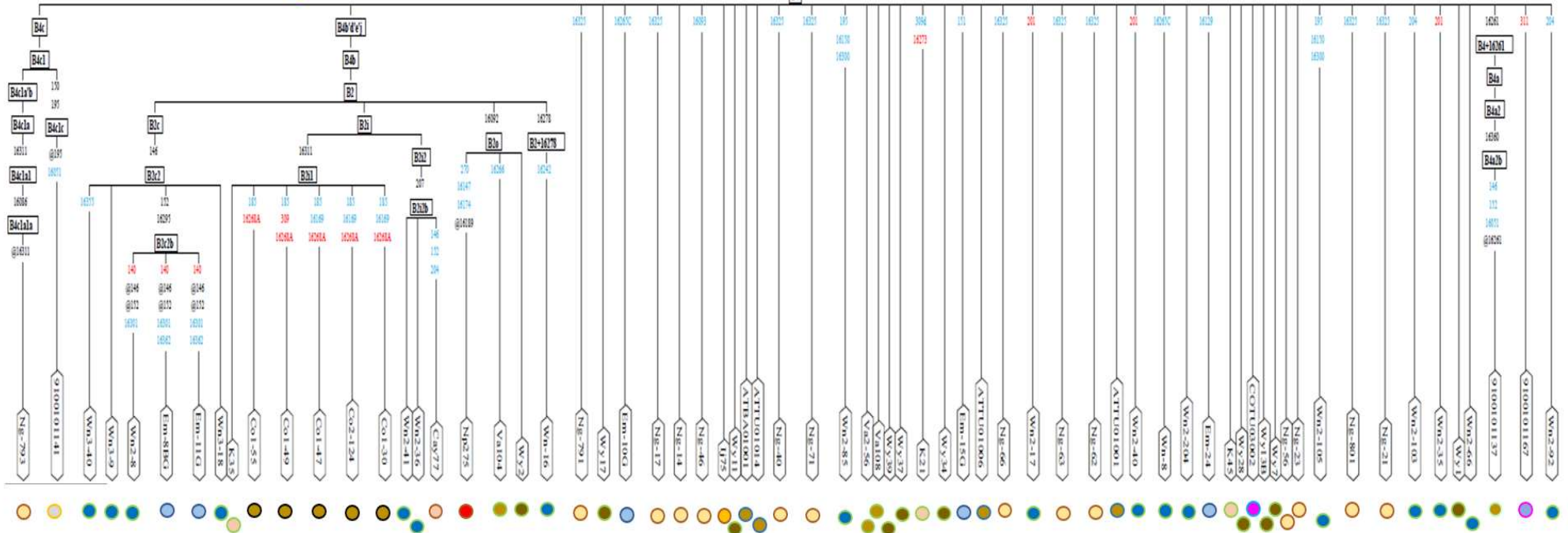
- |              |   |           |   |
|--------------|---|-----------|---|
| Apache       | ● | Huitoto   | ● |
| WiwaRef      | ● | Embera    | ● |
| Cayapa       | ● | Kankuamo  | ● |
| Dogrib       | ● | Kogi      | ● |
| ArhuacoRef   | ● | Kuna      | ● |
| Mocovi       | ● | Catamarca | ● |
| Ngobe        | ● | Yucuna    | ● |
| Paiute Norte | ● | Wayuú     | ● |
| Salta        | ● | Wiwa      | ● |
| Secoya       | ● | Zenú      | ● |
| Siona        | ● | Cocama    | ● |
| Zuni         | ● | Coreguaje | ● |
| Tucuman      | ● | Macuna    | ● |
| Ticuna       | ● | Bora      | ● |
| Vaupes       | ● |           |   |
| Waunana      | ● |           |   |
| WayuuRef     | ● |           |   |
| Arhuaco      | ● |           |   |
| Kankuamo     | ● |           |   |
| Mokaná       | ● |           |   |



**KEY**

- Local private mutation
- Global private mutation
- @ = assumed back mutation or missing mutation
- Heteroplasmic mutation

Figura 5-10b. Arbol filogenético del Haplogrupo B elaborado por Haplogrep 2.0, a partir de las secuencias de las regiones HVI y HVII del ADNmt.





### **Haplogrupo C.**

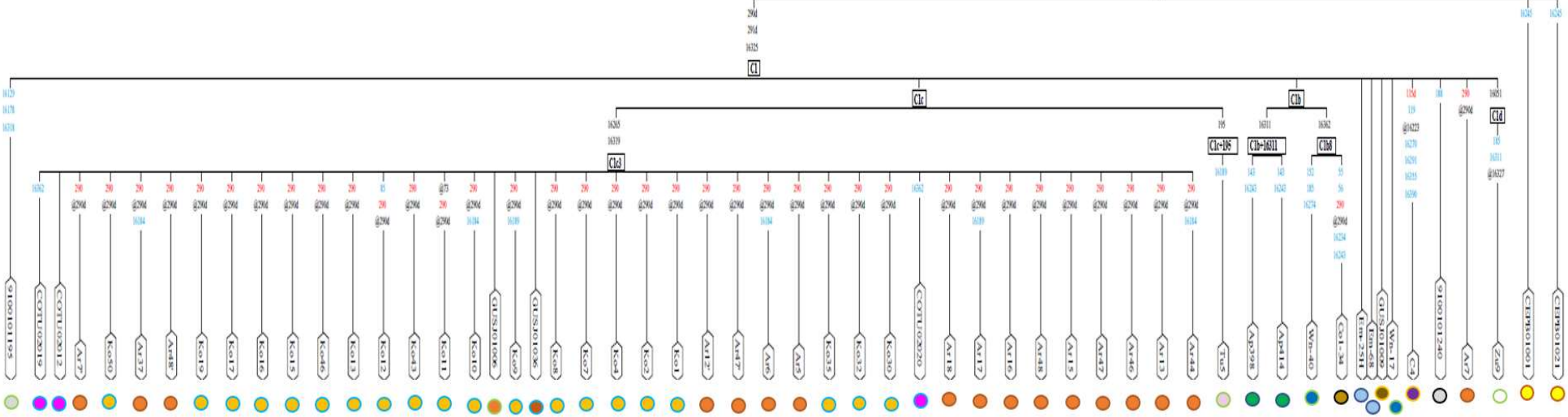
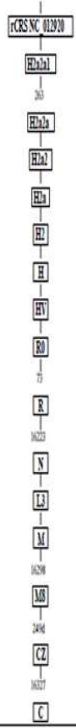
En la región Caribe, 3 muestras de Zenú, 1 de Kankuamo y la única de Wiwa, fueron catalogadas como C1c3, un subhgr de origen aproximado 2700 años (Behar et.al., 2008). Este subclado ha sido previamente reportado en 2 muestras Zenú y una de Kogi (Rieux, et.al. 2014). De igual forma se encuentra en las muestras de referencia de Kogis y Wiwas (Tamm, et.al., 2007) **Figura 5-11**. Dados los escasos reportes para este subHgr, podría sugerirse como particular a las poblaciones de la Sierra Nevada de Santa Marta Kogi y Wiwa, dada su ausencia en muestras de Arhuacos tanto de este estudio como de las muestras de referencia, y la transferencia a la etnia Zenú puede ser consecuencia de flujo genético reciente. Tamm et.al., 2007 muestra a C1c como un subclado específico de América con una amplia distribución que tiene un tiempo estimado de divergencia de 15800 años  $\pm$  4700. Se encontraron además dos Arhuacos con el mismo haplotipo que solo pudieron ser clasificados como C y una muestra de Ticuna, una Witoto y de Wayuu, se catalogaron como C1. Esta última con un haplotipo encontrado 25 veces en la base de datos EMPOP para la metapoblación Nativoamericana de Estados Unidos.

**Figura 5-11.** Arbol filogenético del Haplogrupo C elaborado por Haplogrep 2.0, a partir de las secuencias de las regiones HVI y HVII del ADNmt.

- |              |   |           |   |
|--------------|---|-----------|---|
| Apache       | ● | Huitoto   | ● |
| WiwaRef      | ● | Embera    | ● |
| Cayapa       | ● | Kankuamo  | ● |
| Dogrib       | ● | Kogi      | ● |
| ArhuacoRef   | ● | Kuna      | ● |
| Mocovi       | ● | Catamarca | ● |
| Ngobe        | ● | Yucuna    | ● |
| Paiute Norte | ● | Wayú      | ● |
| Salta        | ● | Wiwa      | ● |
| Secoya       | ● | Zenú      | ● |
| Siona        | ● | Cocama    | ● |
| Zuni         | ● | Coreguaje | ● |
| Tucuman      | ● | Macuna    | ● |
| Ticuna       | ● | Bora      | ● |
| Vaupés       | ● | Cofan     | ● |
| Waukana      | ● | Huitoto   | ○ |
| WayuuRef     | ● |           |   |
| Arhuaco      | ● |           |   |
| Kankuamo     | ● |           |   |
| Mokaná       | ● |           |   |

**KEY**

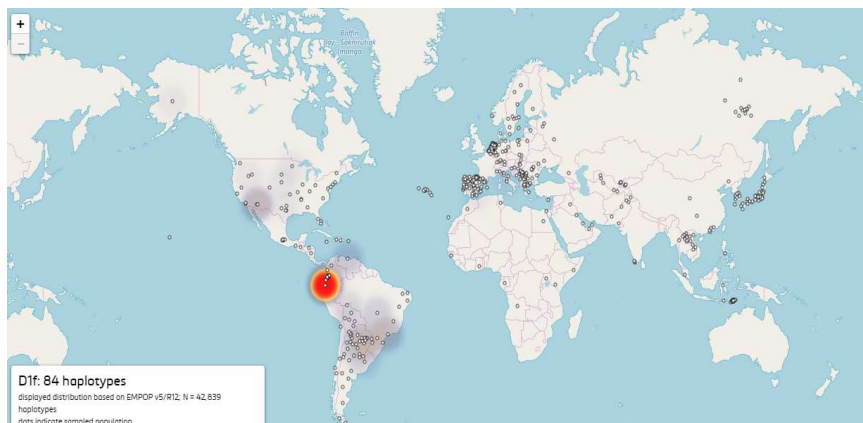
- Local private mutation
- Global private mutation
- @ = assumed back mutation or missing mutation
- Heteroplasmic mutation



## Haplogrupo D.

En la región Caribe el subhaplogrupo D1 solo fue identificado en Zenú. Sin embargo, en Amazonas se encontró una muestra asignada como D1f por EMPOP donde 84 haplotipos en Ecuador corresponden con este Subhgr (**Figura 5-12a**), pero ninguno es coincidente con el haplogrupo de la muestra de la etnia Cocama, ni se encuentran vecinos cercanos con uno o dos cambios. Xavier, et.al., 2015 lo reporta en Nasas de Cauca. También Perego et.al, 2012 reporta 2 individuos de las etnias Embera-Waunana pertenecientes al mismo SubHgr, pero al incluir estas secuencias para el análisis filogenético en Haplogrep 2.0, estos linajes fueron asignados como A23, mientras muestras provenientes de etnias del Vaupés en la Amazonía Colombiana y de los indígenas Coreguajes del Caquetá en esta misma región biogeográfica, fueron asignados como D1f. Se ha sugerido para este subclado un origen hace 18000-15000 años (Brandini, et. al., 2017; Kummar, et.al., 2011; Behar, et.al., 2008). Por su patrón de dispersión geográfica se considera como evidencia probable de una separación temprana de poblaciones a la llegada al norte de Suramérica, de tal modo que se encuentran representantes de este linaje tanto hacia el oriente por la ruta de la costa caribe, como hacia el sur por la ruta del pacífico (Brandini, et.al., 2017). Esto no es por supuesto incoherente con encontrar estos subHgr en la región Amazónica, dado que se estima que la población del Orinoco y del Amazonas llegó posteriormente a un paso por la costa caribe sur americana (Roewer et.al., 2013; Wang et.al., 2007). Sin embargo teniendo en cuenta que no se reporta este linaje en el cono sur de Suramérica y que su frecuencia es menor en las muestras de la ruta occidental, parece haberse generado un fenómeno de deriva genética hacia el sur del continente.

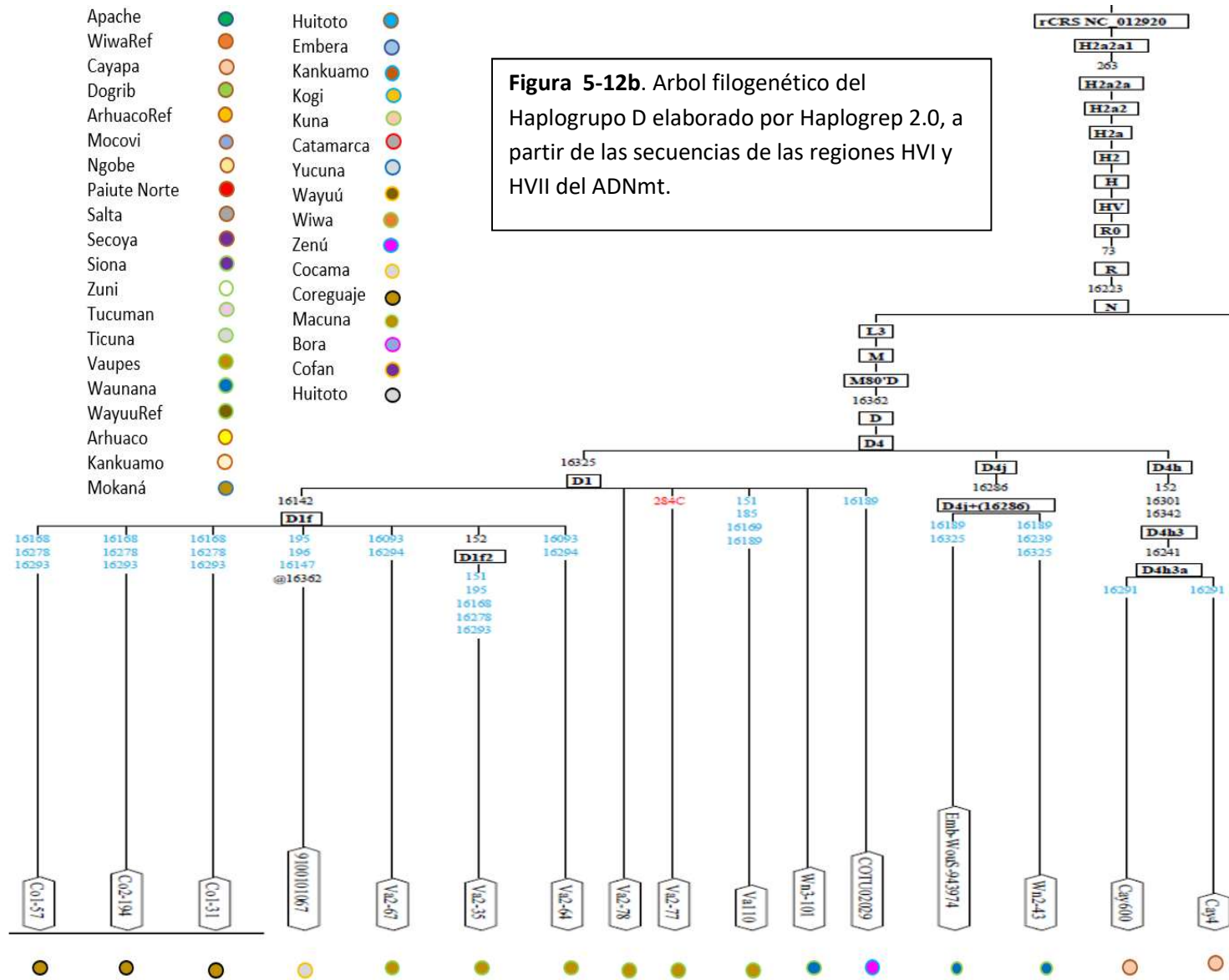
**Figura 5-12a.** Frecuencia del subhaplogrupo D1f de acuerdo con la base de datos EMPOP. Figura tomada de EMPOP.



- |              |   |           |   |
|--------------|---|-----------|---|
| Apache       | ● | Huitoto   | ● |
| WiwaRef      | ● | Embera    | ● |
| Cayapa       | ● | Kankuamo  | ● |
| Dogrib       | ● | Kogi      | ● |
| ArhuacoRef   | ● | Kuna      | ● |
| Mocovi       | ● | Catamarca | ● |
| Ngobe        | ● | Yucuna    | ● |
| Paiute Norte | ● | Wayú      | ● |
| Salta        | ● | Wiwa      | ● |
| Secoya       | ● | Zenú      | ● |
| Siona        | ● | Cocama    | ● |
| Zuni         | ● | Coreguaje | ● |
| Tucuman      | ● | Macuna    | ● |
| Ticuna       | ● | Bora      | ● |
| Vaupés       | ● | Cofan     | ● |
| Waunana      | ● | Huitoto   | ○ |
| WayuuRef     | ● |           |   |
| Arhuaco      | ● |           |   |
| Kankuamo     | ● |           |   |
| Mokaná       | ● |           |   |

**Figura 5-12b.** Arbol filogenético del Haplogrupo D elaborado por Haplogrep 2.0, a partir de las secuencias de las regiones HVI y HVII del ADNmt.

**KEY**  
 Local private mutation  
 Global private mutation  
 @ = assumed back mutation or missing mutation  
 Heteroplasmic mutation



## Otros linajes

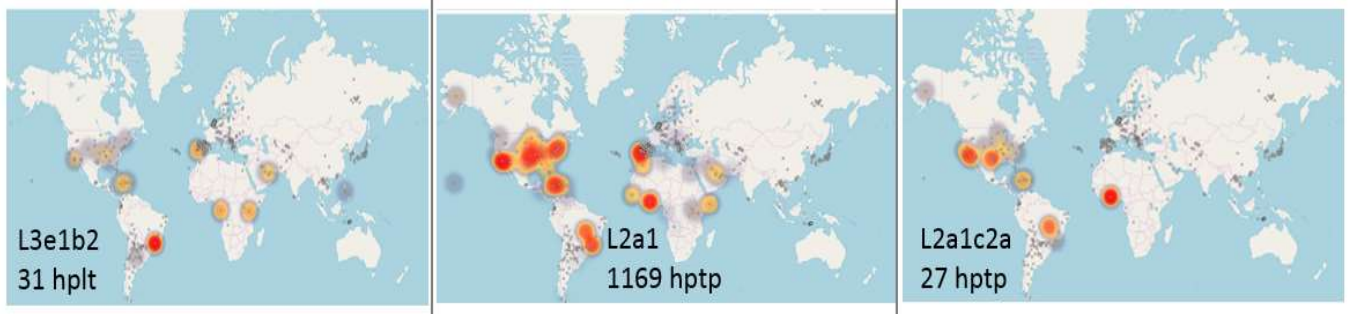
Una muestra de Zenú fue clasificada como haplogrupo M, un linaje que se considera ancestral a C y D, sin embargo, se requiere la confirmación de esta asignación a través de nuevas regiones del ADNmt, ya que este linaje no se ha reportado directamente para poblaciones colombianas pero si más de 7000 haplotipos en poblaciones del sur de Norteamérica, asiáticas orientales y en la India (EMPOP) **Figura 5-13**, ninguno de ellos coincidente con el de la muestra de Zenú, ni se reportan vecinos cercanos por uno o dos cambios.

**Figura 5-13.** Frecuencia del haplogrupo M de acuerdo con la base de datos EMPOP. Figura tomada de EMPOP.



Los SbHgr L de origen Africano, solo encontrados en las muestras de Caribe, no se comparten entre las etnias y fueron más frecuentes en Wayuu y Kankuamo, lo que es concordante con los porcentajes de ancestría africana de hasta 20% en Arhuacos y hasta del 30% en Kankuamos y Wayuu, de acuerdo con los resultados obtenidos por el análisis de SNPs de Ancestría analizados en el capítulo 3. El subHgr L2a1 es el más frecuente en la base de datos EMPOP y se ha reportado en Estados Unidos, Brasil y Caribe insular, así como en el occidente africano. **Figura 5-14**. Ninguno de los haplotipos L ha sido reportado en población nativoamericana.

**Figura 5-14.** Frecuencia de los subhaplogrupos L de acuerdo con la base de datos EMPOP. Figura tomada de EMPOP.



## 5.4 Conclusiones

El efecto fundador serial propuesto por Cavalli-Sforza como hipótesis del poblamiento inicial, soporta la idea de un continuo aislamiento por distancia y disminución de la diversidad, como consecuencia de la división de una población original con la subsecuente dispersión de las poblaciones derivadas.

En el marco del poblamiento suramericano dos teorías fundamentales se han presentado. La primera con relación a la separación de las poblaciones al momento de la llegada a Sur América, generándose una migración hacia el sur por la vía de los Andes y la otra hacia el oriente por la vía inicialmente de la costa caribe y posteriormente hacia el sur del continente. (Tarazona-Santos et.al., 2001; Keyeux, 2002; Wang et.al., 2007, Rothhammer, 2009). La segunda considera la entrada inicial al subcontinente solamente por la vía occidental, con una migración transandina posterior con flujo genético en ambas vías (Bodner et.al., 2012).

La relación entre linajes femeninos amerindios centrales y del norte con las etnias de lengua Chibcha, así como la menor diversidad genética por frecuencias de haplogrupos, encontrada tanto con marcadores de cromosoma Y como de ADN mitocondrial para estas comunidades con respecto de las etnias andinas, la presencia del subhaplogrupo panamericano de ADN mitocondrial D1f (Kummar et.al., 2011) en una muestra de Amazonas, apoyan la teoría de la separación temprana a la llegada de los primeros pobladores a Suramérica.

Considerando sin embargo a Colombia como una región de encuentros y complejidad en el flujo genético, aun desde los tiempos prehispánicos (Keyeux, et.al., 2002; Jaramillo, 1964) sin descontar fenómenos de aislamiento y cuellos de botella y aunque Hunley & Healy, 2011 propone que el efecto del flujo genético entre poblaciones por cercanía geográfica es pequeño como para impactar sobre la diversidad genética continental, la concordancia parcial entre geografía y datos lingüísticos con los datos genéticos a partir de linajes masculinos y femeninos muestran que a nivel subcontinental o regional si puede existir una influencia en la conformación de las relaciones genéticas entre las etnias.

Un caso interesante es el de la población Wayuu que se agrupa en una buena parte de los reportes de la literatura con diferentes tipos de marcadores con la etnia Zenú o con las etnias Chibchas de la Sierra Nevada de Santa Marta, poblaciones geográficamente cercanas aunque de otro tronco lingüístico, y tiene menos relación aparente con otras poblaciones de lengua Arawak.

Estos resultados también pueden estar relacionados con prácticas conservadas actualmente de mezcla de etnias, que se ve también reflejada en las encuestas tanto de la región Caribe como del Amazonas, donde se informan personas provenientes linajes de etnias diferentes. Adicionalmente los patrones ancestrales conservados aun en la actualidad de matrilocidad y sesgo en la selección de las parejas por género, contribuyen a encontrar con más frecuencia el flujo genético de linajes masculinos entre las diferentes poblaciones. Considerando los patrones menos evidentes de segregación cultural y reproductiva entre las etnias colombianas, junto con un fenómeno de mestizaje reciente, impulsado entre otras razones por la aparente mejoría en la calidad de vida y bienestar de las poblaciones que no pertenecen a pueblos indígenas, sumado al desplazamiento que sufren las comunidades nativas por violencia y usurpación de tierras por intereses económicos, es casi obligatoria la reinterpretación de la ancestría de la población colombiana vista desde diferentes profundidades, que se alcanzan con el uso de marcadores genéticos y herramientas interdisciplinarias como análisis genealógicos, lingüísticos y antropológicos que permiten obtener interpretaciones diferentes a la luz de las distintas evidencias.

Teniendo en cuenta estos argumentos cabe preguntarse incluso, si los resultados que se obtienen como producto de muestreos nuevos y que difieren de los publicados con base en muestras tomadas de un generación o dos atrás, sin considerar además criterios genealógicos que puedan contribuir al conocimiento lo más cercano posible a los linajes familiares y ancestrales de una persona que sea incluida como perteneciente a una etnia específica, esté mostrando el resultado de la mezcla de una generación a otra, haciendo que considerar solamente la variable del lugar de nacimiento, sea insuficiente para trazar la ancestría individual y generando sesgos en la interpretación a nivel de las poblaciones.



## 6. Conclusiones y Recomendaciones

### 6.1 Conclusiones

Las poblaciones americanas contemporáneas se han definido como poblaciones triétnicas, por la mezcla histórica de los pueblos originarios nativoamericanos y los linajes europeos y africanos posteriores al descubrimiento de América. En Colombia, la proporción en la que esta mezcla de linajes se dió en los años posteriores a la llegada de los europeos, depende en gran medida de las dinámicas de poblamiento, actividades económicas, supervivencia de los pueblos nativos y el desarrollo cultural y social alrededor de estas realidades. En el interés por realizar el análisis de la ancestría de la región Caribe Colombiana, originalmente habitada por pueblos de filiación lingüística Chibcha, Caribe y Arawak, se han explorado varios niveles históricos, con base en los resultados obtenidos empleando marcadores STRs de Cromosoma Y, SNPs Autosomicos y Secuencias Hipervariable I y II del ADN mitocondrial.

Inicialmente a un nivel global, los tres muestran la mezcla de los linajes nativoamericanos y africanos, en muestras de personas que fueron seleccionadas con base en su autodeterminación y confirmación de la pertenencia étnica por análisis genealógico. Los resultados con marcadores autosómicos muestran entre un 10% y 15% de esta introgresión y los marcadores de cromosoma Y y ADNmt muestran porcentajes más altos y variables, concordantes con las dinámicas de poblamiento de las diferentes localidades muestreadas, llegando incluso a superar

el 30% en los linajes maternos de las muestras de la etnia Kankuamo. En el caso de la Etnia Zenu, se observa una mayor mezcla de los componentes nativoamericanos con linajes europeos masculinos que con linajes de origen africano.

En cuanto al análisis específico de la ancestría nativoamericana, los marcadores SNPs autosómicos muestran que la variabilidad en las frecuencias alélicas es mayor para las etnias colombianas (a excepción de los Pastos), que para las etnias Mesoamericanas y Andinas, lo cual puede estar relacionado con el flujo genético reciente, producto de las dinámicas culturales y comerciales de los pueblos indígenas en Colombia. A pesar de esto, se observa un aparente patrón de disminución de la diversidad genética de Occidente a Oriente que es también observado en el análisis de ADNmt de poblaciones Andinas a poblaciones Amazónicas, y que está de acuerdo con reportes realizados por otros investigadores con marcadores de cromosoma Y, y autosómicos (Tarazona-Santos et.al., 2001; Wang, et.al., 2007).

Los resultados de la distribución de frecuencias de haplogrupos de ADNmt para las muestras de las etnias de la costa caribe colombiana muestran la presencia de al menos dos estructuras diferentes: una relacionada con los indígenas de la sierra y otra con los indígenas de las zonas bajas, lo cual es compatible con el origen histórico de las etnias Kogi, Arhuaco, Wiwa y Kankuamo, mas relacionado con los pueblos chibchas y la relación de las etnias Wayuu, Mokaná, Zenú entre ellas y con poblaciones de lengua Arawak o Genopanocarib.

De igual forma, el hallazgo de una muestra Ticuna A2q encontrada en Poblaciones Maya de Guatemala y de una muestra D1f, un linaje descrito como contemporáneo de los haplogrupos fundadores, es congruente con la teoría de una separación temprana de las poblaciones originales a su llegada a sur america, en una rama que siguió hacia el sur por la cordillera de los Andes y otra al oriente por la costa Caribe hacia Venezuela y Brasil, que luego se expandió hacia la región amazónica. De la misma forma el hallazgo de A2ac en muestras de Amazonas y su relación con otros reportes realizados en una muestra cubana, podrían apoyar la observación de una migración reversa de pueblos Arawak desde el Amazonas como teoría de poblamiento de la región insular caribe.

Las etnias de la costa caribe fueron relacionadas mediante análisis filogenético tanto para marcadores de cromosoma Y y de ADNmt con otras poblaciones de origen lingüístico Chibcha en congruencia con análisis de otros autores. Otra coincidencia corresponde al cluster Zenú Wayuu que es recurrente en diferentes tipos de análisis y se observa también en este trabajo apoyando la observación general de que en muchos casos para la etnias colombianas, las relaciones genéticas entre las poblaciones están más mediadas por la cercanía geográfica y cultural, que por las filiaciones lingüísticas. Por lo anterior, a continuación se discute el contexto histórico de las comunidades indígenas muestreadas.

Primeras evidencias de vida sedentaria en Suramérica se encuentran en la región Amazónica hace 6000 años, con la consiguiente dispersión de esta forma de vida a las zonas de la costa caribe colombiana y Pacífica del Ecuador (Zambrano, 2000). La facilidad de consecución de recursos facilitó esta nueva forma de vida y la relación con el ambiente y la adaptación generó el crecimiento de las poblaciones y la posterior expansión, migración y en ocasiones jerarquización de las mismas, con el fin de obtener resultados más eficientes a partir de la organización de las comunidades. Zambrano propone que las comunidades que tradicionalmente conocemos por su tradición lingüística como Chibcha, Caribe y Arawak, también se distinguían por sus rasgos culturales y de relación con su entorno. Así, describe las poblaciones Chibchas de la región caribe colombiana como pueblos de alta densidad poblacional con un patrón de subsistencia basado en el cultivo del maíz, donde la organización física de la comunidad fue influenciada por el uso altitudinal del suelo, de tal modo que se establecían pequeñas aldeas en las zonas montañosas alrededor de los grandes centros poblados y de mayor productividad. Mientras los Arawak tenían un patrón de subsistencia y dispersión horizontal con mayor tendencia al nomadismo y dependiente de actividades de caza y recolección, con un mayor desarrollo del intercambio para obtención de productos cultivados, provenientes de otros grupos con tierras más favorables para estas actividades. Finalmente, con respecto a los grupos Caribe, más diversos en sus costumbres, contaban con la base de su subsistencia en el cultivo de la yuca amarga, con un patrón de dispersión paralelo a las costas y riveras de los ríos, lo cual hacía compleja la subsistencia ante el crecimiento poblacional, a lo cual respondían generando migraciones de grupos a otras regiones generando incluso conflictos entre comunidades relacionadas, por lo

cual se asociaba a los caribes con comportamientos belicosos permanentes y prácticas de canibalismo. (Zambrano, 2000)

De acuerdo con Zambrano 2000, trazar los orígenes de las poblaciones indígenas a los pueblos prehispánicos puede resultar una difícil tarea considerando por ejemplo desde el punto de vista lingüístico que no existe evidencia de escritura de las lenguas originales, que el estudio profundo de esta área es relativamente reciente y por tanto no fue posible rescatar a tiempo las evidencias que permitan dar mayor seguridad en los hallazgos en esta área. Así mismo, debe considerarse que en la región norte de Suramérica, no era común limitar las áreas de influencia de las diferentes poblaciones prehispánicas, en contextos comerciales o bélicos, de tal modo que las zonas limítrofes podían estar influenciadas por varias culturas en momentos históricos diferentes (Zambrano ,2000).

Tarazona Santos (2001) sugiere que los menores valores de diversidad en las poblaciones Amazónicas son consecuencia de mayor aislamiento entre las comunidades, fundamentalmente como producto de barreras geográficas creadas por las zonas de bosque húmedo tropical. En concordancia con esta hipótesis, los recuentos de cronistas de indias de la época de la conquista y colonia, sugieren por ejemplo que los pueblos del Orinoco, que se encontraban dispersos y tenían costumbres errantes además de contar con prácticas culturales cuyas consecuencias eran la disminución de la natalidad y el aumento de la mortalidad especialmente de niños (Jaramillo, 1964).

Se describe también de forma contrastante, como los pueblos sedentarios del Oriente la Sierra Nevada de Chita (Cocuy), tenían mayores densidades de población y una mayor variedad entre poblaciones representada por su diversidad de lenguas (Jaramillo, 1964).

Entre las causas descritas de la disminución demográfica de las poblaciones indígenas tras la llegada de los Españoles, Jaramillo se incluyen la violencia del proceso de conquista, los trabajos forzados en las minas y las enfermedades traídas por los colonizadores, como la “peste” que afectó a los pueblos indígenas desde Cuzco hasta el territorio Quimbaya, las dos grandes epidemias de viruela registradas en 1566 y 1587 que arrasaron poblaciones enteras, las epidemias de

viruela y sarampión registradas en Tamalameque (Cesar) en 1579, al que según las crónicas no sobrevivió mas de la décima parte de la población indígena y la de Tifo en 1566 que afecto a 4/5 de los nativos de la Sabana. Adicionalmente, la disminución de alimentos y recursos disponibles para los pueblos nativos por cuenta del consumo de los europeos que no contribuían con la producción, la destrucción de la cultura y la organización de las comunidades indígenas, separación de familias y la apatía por la reproducción fueron factores determinantes en el cuello de botella sufrido por las etnias indígenas habitantes de la región a la llegada de los españoles (Jaramillo, 1964).

Establecer el número original de pobladores y por consiguiente la proporción de la reducción de los pueblos nativos resulta difícil por las inconsistencias o imprecisiones en los registros numéricos y por la ausencia de censos reales de población. De acuerdo con el DANE, el primer censo en Colombia se realizó en las provincias de Tunja, por Juan de Otálora, quien contabilizó entre 400.000 y 500.000 Chibchas en 1550 (Rueda, 2012).

El pueblo Mokaná, habitante de varias zonas del departamento del Atlántico, provenía de las tribus indígenas de la región caribe Venezolana, entre Maracapaná y Caracas, y según Chavez Mendoza, por navegación, de lengua Carib, con algunos modelos de fortificación de las aldeas tomados de las tribus Arawak, así como el uso de hamacas más relacionado con las culturas del Orinoco y la Amazonia, mientras la vestimenta estaba más relacionada con la de etnias carib del litoral de Venezuela y Colombia. Tenían varias relaciones comerciales con otras tribus como Tayronas y Zenúes. Tenían guerras constantes con otras tribus, en las que incluían a las mujeres incluso como guerreras. (Chaves, 1979). Controversia sobre filiación lingüística, Baquero 2010 dice que son Arawak, provenientes de Wayuu. Se permitía la poligamia y el incesto padre/hija y está documentada la influencia africana genética y cultural. (Baquero, 2010).

El pueblo Zenú en la época prehispánica contaba con 3 provincias: Finzenú ubicada en la sabana y el valle de Río Sinú (Centro religioso), Panzenú en el valle de Río San Jorge (Centro agrícola y de pesca y manufacturero) y Zenúfana en los valles del

bajo Cauca y Nechí, centro del poder político y de la explotación de oro (Banco de la República, 1996). Su ubicación estratégica les hacia paso obligado de los grupos humano que se desplazaban tanto Hacia el sur como al oriente del continente bordeando la costa caribe, con lo cual tenían una dinámica importante a nivel comercial. Las familias nucleares grandes se mantienen aun hoy y se practica la exogamia dentro de la comunidad, es decir se establecen las parejas entre personas de dos asentamientos diferentes de la misma etnia y se permite la poligamia excluyendo familiares (Ministerio de Cultura, 2013). Han sido incluso relacionados con los Embera (Ramirez, 2011).

La etnia Kankuamo, la cuarta etnia de la Sierra Nevada de Santa Marta, es un pueblo que a partir de 1993 se encuentra en proceso de recuperación de su identidad, tras el desplazamiento y disgregación de la comunidad sufrido desde finales del siglo XIX, como consecuencia de la ocupación de los territorios por colonos de otras regiones huyendo de la violencia, y de otros atraídos por terrenos “valdios” para la siembra de cultivos ilícitos y la persecución de los capuchinos en su intento por eliminar las costumbres paganas e instaurar definitivamente la fe Católica, así como la agudización del conflicto armado en Colombia en los años 80’s. Por tal motivo, hoy se consideran un pueblo con alto grado de mezcla y aculturización (Arias, 2011).

Es así como se aprecia la influencia de los sucesos históricos, económicos culturales y políticos en la forma como cambia la estructura genética de las poblaciones indígenas colombianas vistas en todos los contextos históricos. Dada la historia de mezcla de las poblaciones prehispánicas, aun desde los primeros pobladores de américa, es posible que las expectativas que surgen acerca de la posición geográfica de Colombia y de su papel en las dispersiones de los pueblos en diferentes rutas hacia el resto de Suramérica, así como de la riqueza de sus etnias en las que se espera encontrar la huella de los linajes fundadores, constituyen en realidad un panorama más complejo de lo que parece, pues en sí misma, la amplia diversidad genética y su especial geografía, la han convertido en una región de encuentros genéticos y culturales que reescriben con cada generación la historia de los nativoamericanos.

Es por esto que teniendo en cuenta que el trabajo realizado en Genética de Poblaciones Humanas ha tenido aplicación en varias áreas especialmente en los

análisis de ancestría, genética forense y en área clínica para determinación de estatificación poblacional en los estudios de casos y controles, el camino trazado en el desarrollo de cada capítulo de este documento, nos ha llevado a hacer las siguientes consideraciones frente a esta disciplina sobre los aspectos más importantes que deben tenerse en cuenta al plantear una metodología de trabajo en torno a una hipótesis especialmente en el campo del estudio de la ancestría de las poblaciones humanas:

1. La selección de los marcadores a emplear debe considerar claramente, qué nivel de profundidad que se desea alcanzar. La utilidad de dichos marcadores depende en mucho de una juiciosa evaluación que pueda hacerse sobre la aplicabilidad de los desarrollos que se han realizado en el marco de estudios continentales, subcontinentales o regionales. Un tipo de marcador que favorece en análisis en poblaciones europeas, asiáticas o africanas o que se ha probado con éxito para determinar la ancestría biogeográfica a nivel de continentes y en este caso específico, que permite discriminar la ancestría nativoamericana, puede no ser tan apropiado cuando se contextualiza un problema en el marco de las diferencias dentro del subcontinente. Es esto lo que se ha observado como parte de los resultados del capítulo 2, en la comparación del panel miniLace, que cuenta con una selección de 79 de los marcadores del panel Lace (Galanter, et.al., 2012) probado para determinar la ancestría de poblaciones Americanas, con el SNPforID 52plex, que a pesar de ser diseñado para su uso en identificación, ha sido empleado en varias oportunidades con el mismo propósito del miniLace. Se observa aquí que existe un sesgo en la asignación de las proporciones de ancestría nativoamericana, europea y africana al emplear el panel de identificación para este uso. Sin embargo, al evaluar este mismo conjunto de marcadores con una aplicación a nivel regional, en poblaciones indígenas sur y centroamericanas, se encuentra un mejor poder de resolución, permitiendo acercar los análisis a conclusiones más apropiadas en el contexto regional. Sin embargo, aun se requiere mejorar la eficiencia en tal sentido, por lo que una búsqueda de combinaciones de SNPs

autosómicos mas informativos, para el cumplimiento de estos objetivos es necesaria. Esto nos dirige a la segunda consideración.

2. Las poblaciones de referencia que se emplean con frecuencia para realizar las comparaciones y llegar a conclusiones relacionadas con las ancestrías, suelen proceder de datos publicados en artículos o bases de datos con información genealógica insuficiente, que permita seleccionar los individuos mas informativos. Aun así, es posible que en algunos casos no se realice un apropiado filtrado de los genotipos a emplear, al no verificar en lo posible, que todas muestras compartan una ancestría comun, de tal modo que las interpretaciones realizadas a la luz de los análisis desarrollado a partir de esas poblaciones de referencia, pueden ser sesgadas. Es aquí donde cobra importancia el uso de herramientas de apoyo como encuestas y contextualizaciones históricas recientes o ancestrales, o ambas, dependiendo del objeto del análisis, permitiendo mayor robustez al momento de probar las hipótesis y por supuesto lograr que las poblaciones de estudio pasen a ser poblaciones de referencia para otros investigadores con un nivel de confianza mayor. Este punto conduce a tener en cuenta en el diseño experimental, que los muestreos a conveniencia, desarrollados con criterios juiciosos de selección de las poblaciones de estudio, enmarcados en contextos genealógicos, demográficos e históricos, mejoran la posibilidad de obtener información desde el punto de vista genético poblacional mas precisa en diferentes niveles de alcance, dependiendo de los objetivos del estudio. Es así como para el caso del presente trabajo, fue posible encontrar en poblaciones mezcladas, muestras con altos porcentajes de pertenencia étnica nativoamericana con marcadores AIMs autosómicos, linajes Q de cromosoma Y en proporciones superiores a las usualmente reportadas para las poblaciones colombianas y subhaplogrupos mitocondriales no reportados en trabajos previos. Es entonces cuando puede evaluarse la idea de las “poblaciones aisladas”, como fuente haplotipos o genotipos ancestrales que conduzcan al esclarecimiento definitivo de las hipótesis de poblamiento. Este punto lleva entonces a la siguiente consideración.
3. El análisis de los datos genéticos interpretados a la luz de contextos culturales y de las dinámicas poblacionales en contextos históricos antiguos,



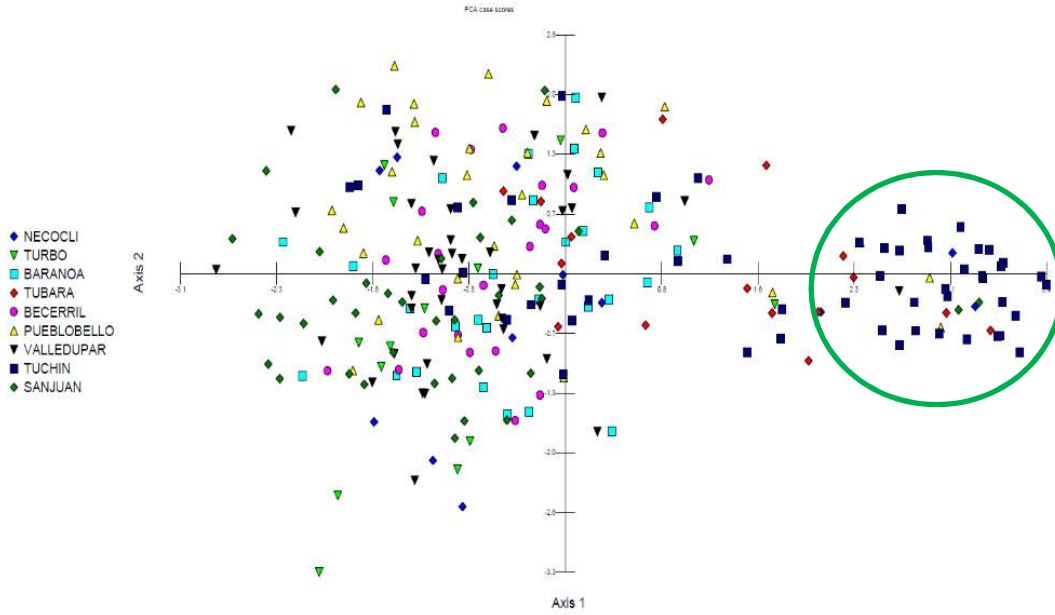
recientes y actuales, ofrece la oportunidad de dilucidar panoramas más probables sobre la ancestría de las regiones. En la búsqueda de aclarar los múltiples cuestionamientos con relación al poblamiento americano, varios trabajos se han enfocado en analizar los genotipos y su correlación con datos lingüísticos y de ubicación geográfica. Es así como se ha considerado que la posición geográfica estratégica de Colombia, en el norte de Suramérica, la hace fuente de una rica información que permita probar o desvirtuar las hipótesis acerca de este tema. Los resultados de esta investigación muestran diferentes lados de la historia, en función de los marcadores utilizados. Sin embargo, los tres muestran con claridad, que las etnias indígenas tienen una mayor diversidad de la que espera a la luz de la consideración general de que poblaciones más recientes son menos diversas, en términos ancestrales. Esto nos ha llevado a reafirmar que existe más de un tipo nativoamericano, por lo que generalizar en el contexto de la genética de poblaciones a nivel subcontinental no es posible. El panorama quizás un poco nublado por la idea de la ancestría triétnica que sesga nuestra percepción acerca de la diversidad real del componente nativoamericano y con ello las interpretaciones que hacemos de los resultados que se obtienen en juiciosos, largos y robustos trabajos de investigación. Como resultado del análisis de la ancestría de nativoamericana en el Caribe continental colombiano, los marcadores autosómicos, de cromosoma Y y de ADN mitocondrial cuentan historias diferentes y permiten evidenciar la complejidad de la dinámica poblacional en el norte de Suramérica aun desde la llegada de los primeros pobladores.

4. Dada la historia de mezcla de las poblaciones prehispánicas, aun desde los primeros pobladores de América, es posible que las expectativas que surgen acerca de la posición geográfica de Colombia y de su papel en las dispersiones de los pueblos en diferentes rutas hacia el resto de Suramérica, así como de la riqueza de sus etnias en las que se espera encontrar la huella de los linajes fundadores, constituyen en realidad un panorama más complejo de lo que parece, pues en sí misma, la amplia diversidad genética y su especial geografía, la han convertido en una región de encuentros

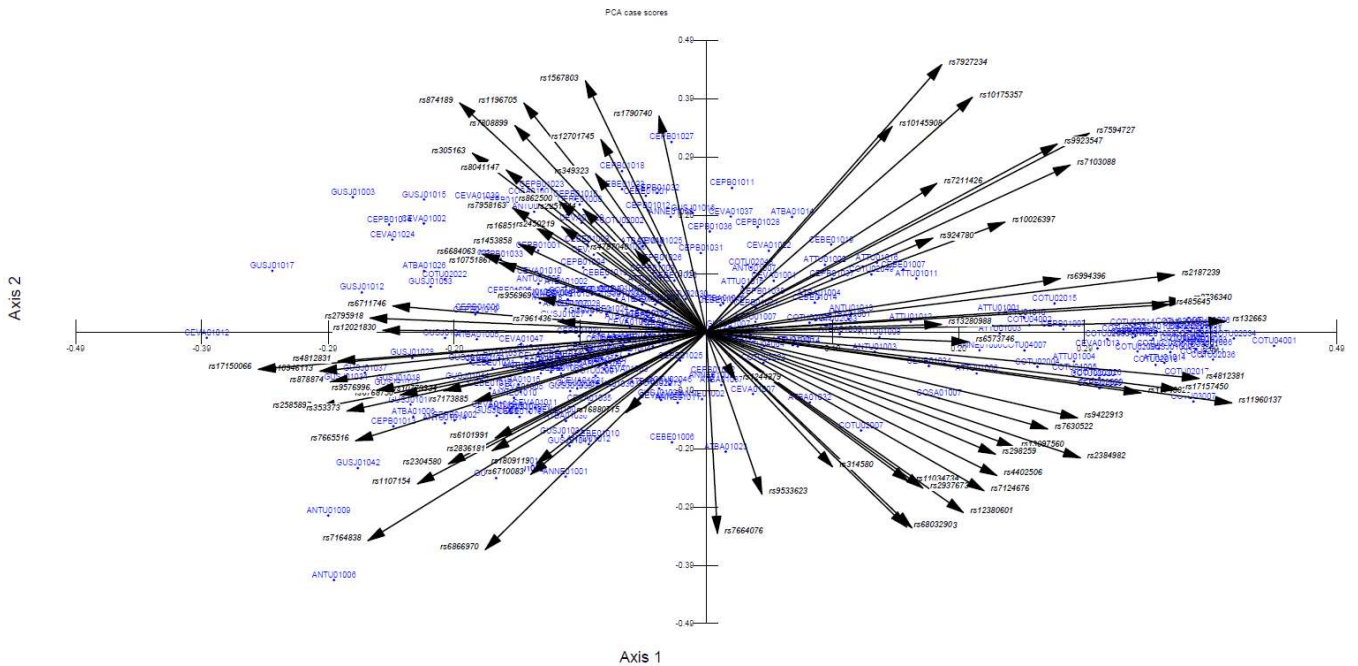
genéticos y culturales que reescriben con cada generación la historia de los nativoamericanos.

El desarrollo de este trabajo permite reflexionar sobre tres condiciones básicas que deberían tenerse en cuenta al transitar el mundo de la Genética de poblaciones Humanas: Humildad para reconocer que se requiere trabajo interdisciplinario, Rigurosidad para la selección de los marcadores, las poblaciones de referencia y Flexibilidad para realizar la interpretación de los resultados a la luz de las diferencias demográficas y genealógicas de las poblaciones muestreadas. La genética de poblaciones sin estos componentes se convierte entonces en un conjunto de datos, matemáticas y secuencias que son hilos sueltos en un panorama donde se requiere entretejer para construir.

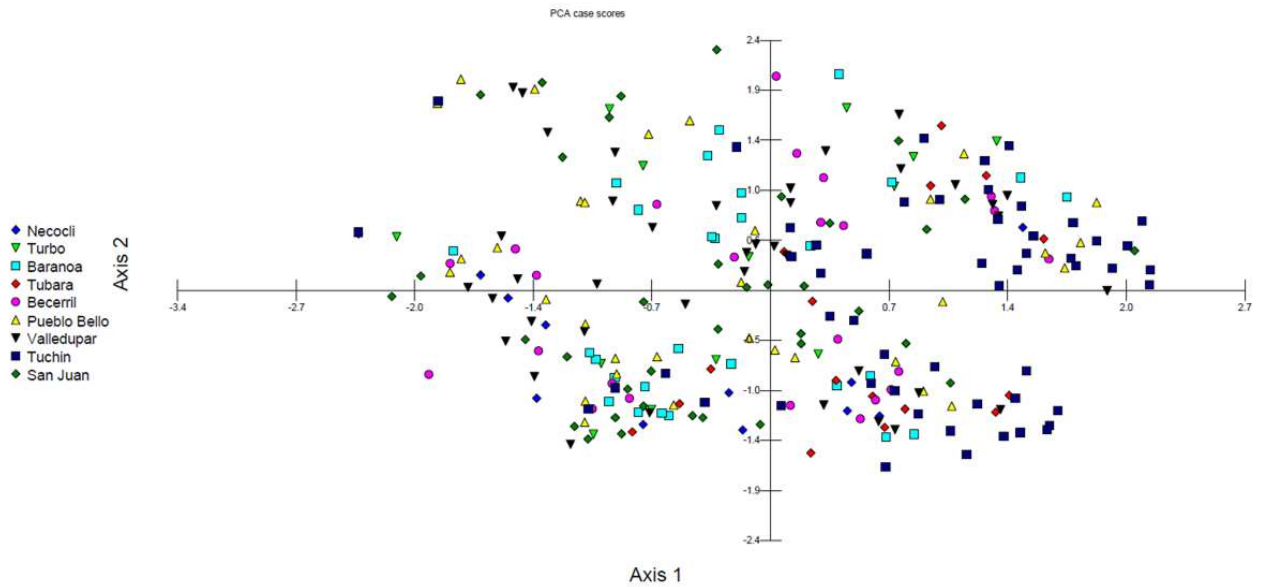
# Material Suplementario Capitulo 2



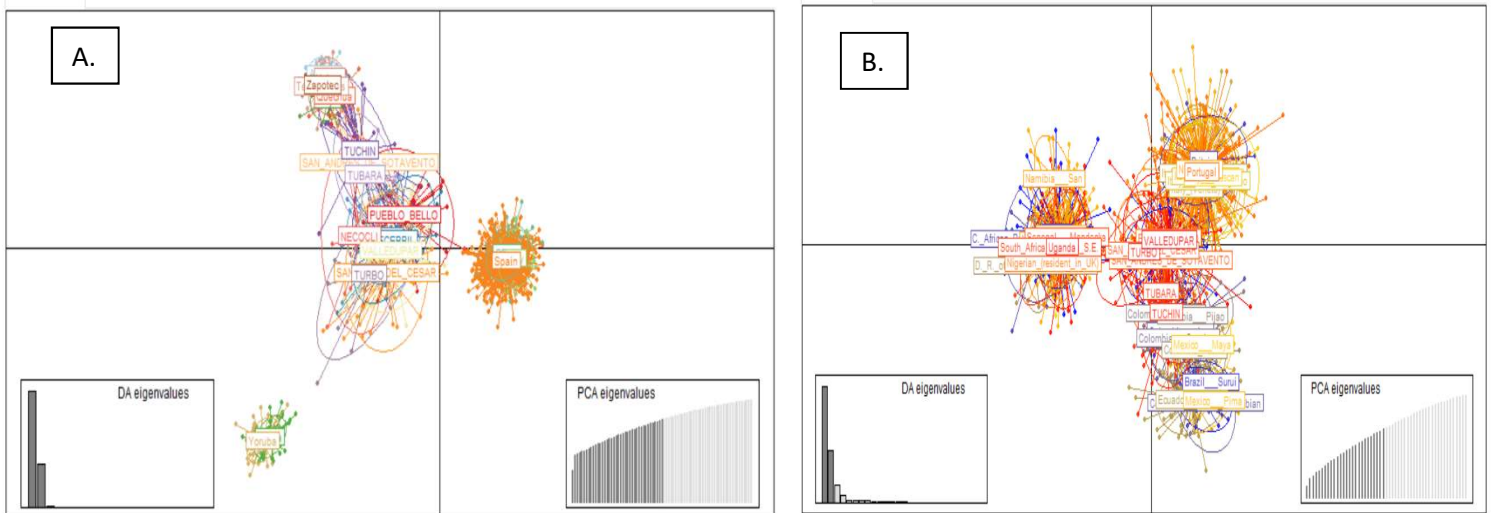
**Figura Suplementaria2- 1.** Análisis de componentes principales para los marcadores AIMs, solo para muestras de la región Caribe Colombiana con el software MVSP



**Figura Suplementaria 2-2.** Gráfica Biplot del ACP con el software MVSP para las muestras de población Colombiana de este estudio con marcadores AIMs.

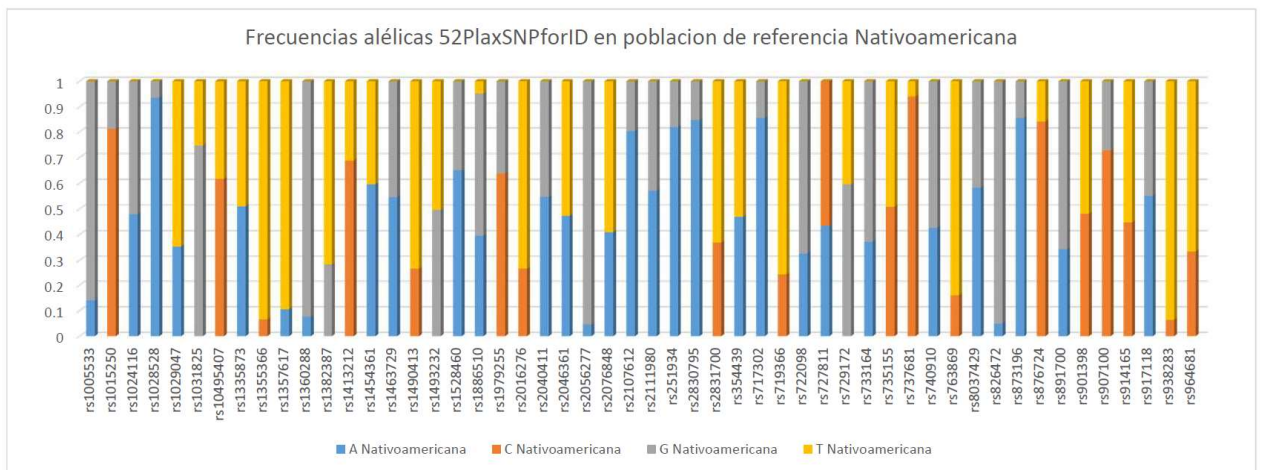
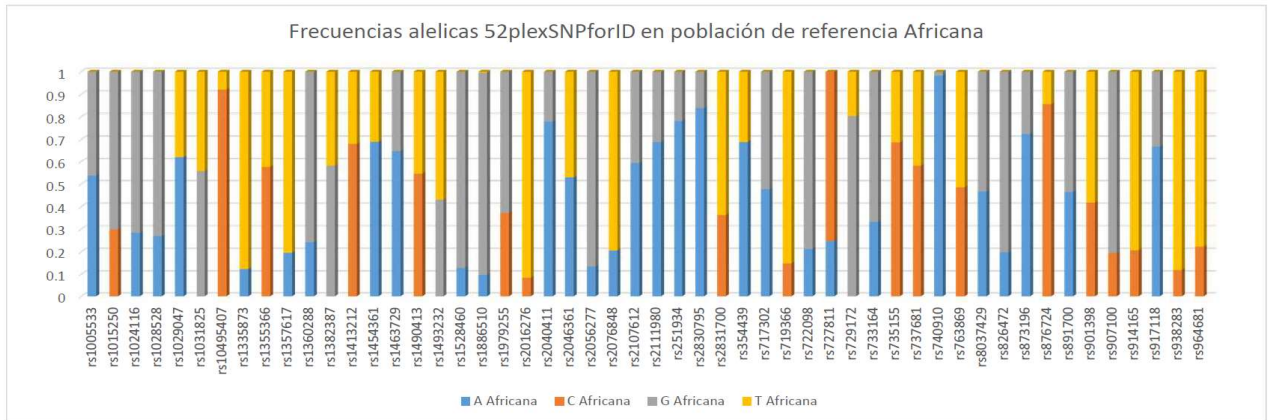
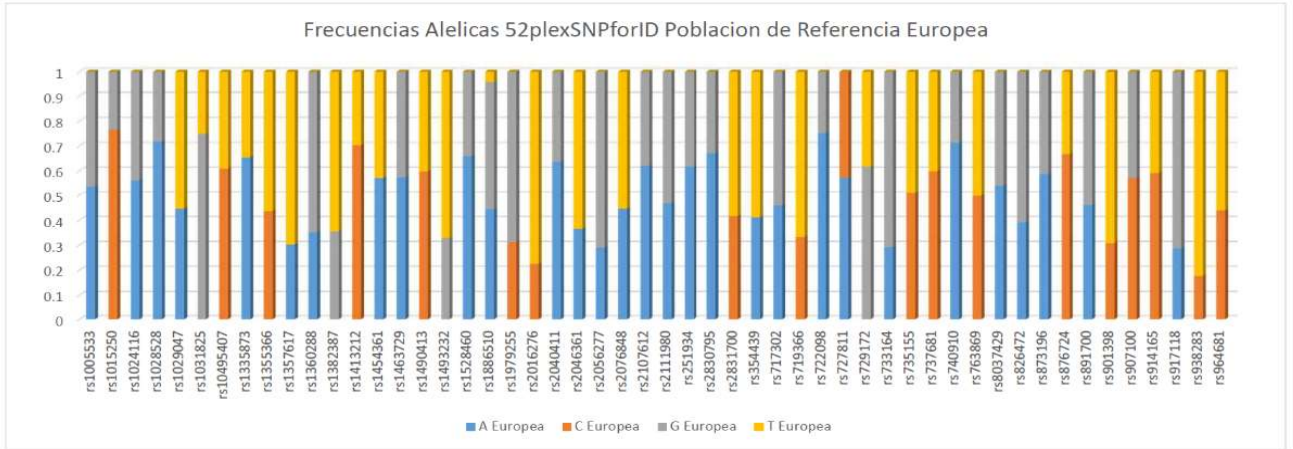


**Figura Suplementaria 2-3.** Análisis de componentes principales con MVSP para los marcadores SNPforID 52PLEX



**Figura Suplementaria 2-4.** Gráfica de DAPC usando clusters predefinidos en el nivel de municipios para las muestras colombianas y las subpoblaciones de los grupos de referencia, empleando marcadores AIMs (A) y SNPforID 52PLEX (B).

**Figura Suplementaria 2-5.** Frecuencias alélicas de los marcadores del panel SNPforID 52plex para las poblaciones de referencia. Datos obtenidos de SNPforID Browser.



**Tabla Suplementaria 2-1. Datos Geográficos y demográficos de las poblaciones de la Región Caribe de Colombia incluidas en el estudio**

Departamento	Municipio	% de pertenencia étnica por autodeterminación (Fuente DANE 2005)		Etnias asociadas a la región
		Indígena	Afrodescendiente	
ANTIOQUIA	Necoclí	3,5	58,9	Kuna
	Turbo	1,3	81,3	Kuna
ATLANTICO	Baranoa	7,4	14,3	Mokaná
	Tubará	61,5	0,5	Mokaná
CESAR	Becerril	12,4	8,1	Wiwa, Yuko
	Pueblo Bello	59,4	0,2	Arhuaco, Kankuamo, Kogui
	Valledupar	8,4	12,2	Arhuaco
CORDOBA	San Andrés de Sotavento-Tuchin	95,0	0,3	Zenú
LA GUAJIRA	San Juan del Cesar	24,3	10,0	Arhuaco, Wiwa

**Tabla Suplementaria 2-2. A. Descripción de los AIMs genotificados en el ensayo. Este conjunto de 79 SNPs, hacen parte del panel LACE de ancestría desarrollado con un total de 446 marcadores: 202 europeos, 115 Africanos y 129 Nativo americanos. (Tomado de Galanter, et.al., 2012). B. Descripción del SNPforID 52 - plex Identification Panel genotificado en el ensayo (Columnas 1 a 5 Tomado de Sanchez, et.al. 2006, columnas 6-9 tomado de SNPforID Browser )**

A. Marcadores de ancestría AIMs									
SNP rsID	Cromosoma	Posición	A1	A2	NAM_AF	EUR_AF	AFR_AF	Población	LSBL(In)
rs6684063	1	30471927	T	G	0.183	0.841	0.218	EUR	0.223
rs12021830	1	119363905	T	C	0.107	0.838	0.995	NAM	0.376
rs10779334	1	217173027	G	A	0.008	0.097	0.837	AFR	0.36
rs6711746	2	29292053	T	C	0.102	0.809	0.874	NAM	0.308
rs10175357	2	35015176	C	T	0.038	0.059	0.82	AFR	0.352
rs305163	2	36163166	G	A	0.172	0.712	0.063	EUR	0.196
rs7594727	2	96853597	A	C	0.008	0.136	0.869	AFR	0.366
rs1567803	2	100709450	T	C	0.019	0.642	0.029	EUR	0.253
rs1196705	2	149927468	G	A	0.015	0.638	0.117	EUR	0.198
rs298259	2	157033071	C	T	0.16	0.761	0.131	EUR	0.206
rs1809119	2	174234175	C	A	0.046	0.207	0.917	AFR	0.357
rs6710083	2	192480837	T	G	0.004	0.078	0.84	AFR	0.38
rs16851773	2	214897054	G	A	0.126	0.647	0.024	EUR	0.195
rs6768750	3	7215675	G	A	0.073	0.695	0.976	NAM	0.325
rs6803290	3	59667208	C	T	0.023	0.636	0.01	EUR	0.262
rs862500	3	64247689	A	T	0.134	0.821	0.117	EUR	0.268
rs13097560	3	85212019	A	G	0.046	0.658	0.068	EUR	0.221
rs7630522	3	108635778	C	T	0.058	0.787	0.262	EUR	0.208
rs2937673	3	118715077	G	A	0.019	0.553	0.044	EUR	0.19
rs7665516	4	11243690	G	A	0.023	0.637	0.888	NAM	0.333
rs10026397	4	41260644	A	G	0.046	0.194	0.932	AFR	0.381
rs924780	4	148905447	A	C	0.115	0.19	0.942	AFR	0.364
rs7664076	4	172625303	A	G	0.344	0.803	0.034	EUR	0.191
rs353373	5	40261722	G	A	0.191	0.863	1	NAM	0.315
rs349323	5	100804840	T	A	0.031	0.629	0.015	EUR	0.247
rs6866970	5	100978603	G	A	0.015	0.06	0.858	AFR	0.403
rs17157450	5	104384818	C	T	0.145	0.864	0.903	NAM	0.305
rs11960137	5	155270659	C	G	0.122	0.875	0.942	NAM	0.353
rs1107154	5	165319891	T	C	0.061	0.39	0.985	AFR	0.352
rs11743823	5	168097322	C	T	0.103	0.814	0.947	NAM	0.342

rs10946113	6	158504679	C	A	0.151	0.896	0.971	NAM	0.354
rs12701745	7	4558764	G	A	0.012	0.526	0.024	EUR	0.194
rs7808899	7	14779699	T	G	0.298	0.83	0.146	EUR	0.197
rs2384982	7	103493636	C	T	0.187	0.938	0.398	EUR	0.242
rs17150066	8	9347921	G	A	0.088	0.858	0.985	NAM	0.405
rs2736340	8	11381382	C	T	0.089	0.792	0.879	NAM	0.315
rs956969	8	62227592	T	C	0.034	0.623	0.049	EUR	0.218
rs16880715	8	111524781	T	G	0.019	0.091	0.859	AFR	0.379
rs13280988	8	112439692	A	G	0.031	0.581	0.058	EUR	0.19
rs6994396	8	114881166	G	T	0.115	0.865	0.85	NAM	0.31
rs2450219	8	119372638	T	C	0.057	0.641	0.058	EUR	0.21
rs12380601	9	73838451	A	G	0.031	0.583	0.029	EUR	0.21
rs10751867	10	2339064	T	C	0.351	0.916	0.175	EUR	0.241
rs2795918	10	55847021	T	C	0.046	0.692	0.85	NAM	0.311
rs9422913	10	127335597	C	T	0.771	0.118	0.718	EUR	0.218
rs874189	11	15838275	A	G	0.023	0.689	0.107	EUR	0.232
rs7927234	11	27392932	C	G	0.012	0.11	0.852	AFR	0.363
rs1396883	11	34714023	A	G	0.023	0.559	0.029	EUR	0.201
rs11034734	11	38427247	C	T	0.019	0.512	0.01	EUR	0.192
rs7124676	11	64069867	A	G	0.008	0.521	0.029	EUR	0.191
rs1790740	11	66886698	T	C	0.225	0.929	0.17	EUR	0.308
rs878874	11	94228113	A	G	0.107	0.843	0.985	NAM	0.374
rs485645	11	125938189	A	G	0.046	0.641	0.966	NAM	0.328
rs7103088	11	129514516	A	G	0.019	0.111	0.917	AFR	0.429
rs7961436	12	10134132	G	A	0.057	0.64	0.084	EUR	0.196
rs7958163	12	112759625	C	A	0.034	0.702	0.172	EUR	0.201
rs2585897	13	20296979	G	A	0.076	0.857	0.884	NAM	0.364
rs9576996	13	39813272	G	C	0.223	0.953	0.971	NAM	0.329
rs9533623	13	43278753	A	G	0.05	0.552	0.01	EUR	0.192
rs314580	13	88709705	A	G	0.366	0.953	0.403	EUR	0.207
rs1244379	14	33750102	C	T	0.088	0.635	0.063	EUR	0.189
rs2251244	14	61093017	C	A	0.717	0.181	0.985	EUR	0.237
rs10145908	14	62823082	C	T	0.011	0.086	0.83	AFR	0.36
rs6573746	14	66576406	A	C	0.512	0.056	0.743	EUR	0.196
rs7164838	15	32754866	G	A	0.08	0.702	0.976	NAM	0.32
rs8041147	15	35085650	T	C	0.092	0.902	0.223	EUR	0.311
rs2304580	15	38816959	T	C	0.136	0.842	0.927	NAM	0.315



rs1453858	15	46091402	T	A	0.206	0.911	0.228	EUR	0.273
rs4402506	15	91931362	C	T	0.05	0.588	0.015	EUR	0.209
rs7173885	15	93608135	G	C	0.008	0.249	0.917	AFR	0.351
rs4787040	16	7500981	T	A	0.08	0.714	0.039	EUR	0.255
rs9923547	16	12685472	A	C	0.011	0.258	0.942	AFR	0.374
rs7211426	17	51009547	A	G	0.092	0.212	0.951	AFR	0.376
rs4812381	20	37420974	A	G	0.046	0.694	0.845	NAM	0.31
rs6101991	20	38515722	G	A	0.038	0.104	0.853	AFR	0.352
rs4812831	20	42451674	G	A	0.168	0.904	0.961	NAM	0.338
rs2836181	21	38477680	T	C	0.016	0.049	0.801	AFR	0.358
rs2187239	21	42293624	A	C	0.023	0.684	0.819	NAM	0.329
rs132663	22	34892970	A	T	0.027	0.684	0.893	NAM	0.351

B. Marcadores de identificación SNPforID 52plex								
SNP rsID	Cromosoma	Posición	A1	A2	NAM_AF	EUR_AF	AFR_AF	Población
rs1005533	20	40172539	A	G	0.860	0.464	0.462	Nativoamericana
rs1015250	9	1813774	C	G	0.188	0.235	0.700	Africana
rs1024116	18	73559363	A	G	0.521	0.438	0.715	Africana
rs1028528	22	46574531	A	G	0.936	0.717	0.270	Africana
rs1029047	6	1080939	A	T	0.647	0.554	0.378	Nativoamericana
rs1031825	20	4442483	A	C	0.253	0.252	0.442	Africana
rs10495407	1	235480457	G	A	0.617	0.609	0.921	Africana
rs1335873	13	18699724	A	T	0.510	0.654	0.124	Africana
rs1355366	3	192127021	A	G	0.934	0.564	0.424	Nativoamericana
rs1357617	3	936782	A	T	0.895	0.698	0.804	Europea
rs1360288	9	124344108	C	T	0.923	0.649	0.757	Nativoamericana
rs1382387	16	79885888	G	T	0.283	0.354	0.582	Africana
rs1413212	1	239753521	A	G	0.312	0.297	0.319	Europea
rs1454361	14	23840960	A	T	0.596	0.571	0.690	Africana
rs1463729	9	122257493	A	G	0.454	0.426	0.351	Africana
rs1490413	1	4037521	A	G	0.264	0.597	0.546	Nativoamericana
rs1493232	18	1117986	C	A	0.496	0.328	0.431	Europea
rs1528460	15	52926761	C	T	0.349	0.341	0.871	Africana
rs1886510	13	20172700	C	T	0.557	0.512	0.899	Africana
rs1979255	4	191013970	C	G	0.362	0.688	0.627	Nativoamericana
rs2016276	15	22119157	A	G	0.736	0.775	0.918	Africana

rs2040411	22	46048653	A	G	0.453	0.364	0.221	Africana
rs2046361	4	10719942	A	T	0.473	0.365	0.530	Africana
rs2056277	8	139370172	C	T	0.954	0.711	0.865	Nativoamericana
rs2076848	11	134205198	A	T	0.591	0.554	0.794	Africana
rs2107612	12	741262	A	G	0.197	0.380	0.406	Nativoamericana
rs2111980	12	104830721	A	G	0.571	0.468	0.689	Europea
rs251934	5	174759601	C	T	0.819	0.617	0.782	Europea
rs2830795	21	27530034	A	G	0.846	0.670	0.839	Europea
rs2831700	21	28601558	G	A	0.631	0.583	0.637	Europea
rs354439	13	104636412	A	T	0.531	0.589	0.311	Africana
rs717302	5	2932133	A	G	0.146	0.541	0.522	Nativoamericana
rs719366	19	33155177	C	T	0.242	0.332	0.149	Africana
rs722098	21	15607469	A	G	0.326	0.751	0.213	Europea
rs727811	6	164954622	A	C	0.566	0.428	0.752	Africana
rs729172	16	5606490	A	C	0.596	0.617	0.804	Africana
rs733164	22	26141338	A	G	0.628	0.708	0.667	Europea
rs735155	10	3328178	A	G	0.508	0.510	0.688	Africana
rs737681	7	154850085	C	T	0.059	0.402	0.418	Nativoamericana
rs740910	17	5907188	A	G	0.426	0.713	0.983	Africana
rs763869	8	1363017	C	T	0.160	0.498	0.486	Nativoamericana
rs8037429	15	51332965	C	T	0.417	0.459	0.533	Africana
rs826472	10	2360631	C	T	0.049	0.390	0.198	Europea
rs873196	14	96835572	C	T	0.146	0.413	0.275	Europea
rs876724	2	104974	C	T	0.840	0.667	0.857	Europea
rs891700	1	236923075	A	G	0.343	0.461	0.465	Nativoamericana
rs901398	11	11060530	C	T	0.481	0.307	0.418	Europea
rs907100	2	239850329	C	G	0.272	0.429	0.805	Africana
rs914165	21	41336325	A	G	0.447	0.591	0.207	Africana
rs917118	7	4201341	C	T	0.450	0.713	0.331	Europea
rs938283	17	78065617	C	T	0.936	0.824	0.880	Nativoamericana
rs964681	10	132172819	C	T	0.666	0.561	0.777	Europea

**Tabla Suplementaria 2-3. Poblaciones de referencia. A. Marcadores AIMS, B. Marcadores 52plexSNPforID**

<b>A. MARCADORES AIMS</b>		
<b>ANCESTRÍA</b>	<b>LOCALIDAD</b>	<b>N</b>
<b>EUROPEA</b>	España	616
	Residentes de Utah con ascendencia de norte y occidente europeo	56
	Italia	44
	<b>Total</b>	716
<b>AFRICANA</b>	Luhya	50
	Yoruba	53
	<b>Total</b>	103
<b>NATIVOAMERICANA</b>	Aymara	25
	Maya	22
	Nahua	14
	Quechua	23
	Tepehuanos	22
	Zapoteca	21
	<b>Total</b>	127

B. MARCADORES 52plexSNPforId		
ANCESTRIA	LOCALIDAD	N
EUROPEA	Británicos	16
	Italianos	88
	España Noroccidental	125
	Portugal	167
	<b>Total</b>	396
AFRICANA	Angola	65
	C. African Republic Biaka Pygmy	20
	D. R. of Congo Mbuti Pygmy	12
	Kenya Bantu N.E	9
	Mozambique	133
	Namibia	6
	Nigeria Yoruba	19
	Nigerian_(resident_in_UK)	10
	Senegal Mandenka	19
	South Africa Bantu S.E.,	7
	Uganda	40
	<b>Total</b>	340
NATIVOAMERICANA	Brasil Karitiana	14
	Brasil Surui	6
	Colombia Awa	19
	Colombia (Piapoco-Curripaco)	6
	Colombia Coyaima	27
	Colombia Embera	4
	Colombia Pijao	8
	Colombia Pastos	33
	Ecuador Kichwa	76
	Mexico Maya	16
	Mexico Pima	10
	<b>Total</b>	219

**Tabla Suplementaria 2-4a. Frecuencias genotípicas para AIMs**

SNP rsID	N	AA	AC	AG	AT	CC	CG	CT	GG	GT	TT
rs10026397	258	0.124		0.395					0.481		
rs10145908	258					0.054		0.279			0.667
rs10175357	258					0.062		0.302			0.636
rs10751867	258					0.360		0.446			0.194
rs10779334	258	0.628		0.291					0.081		
rs10946113	258	0.236	0.430			0.333					
rs11034734	258					0.027		0.252			0.721
rs1107154	258					0.422		0.422			0.155
rs11743823	258					0.271		0.453			0.275
rs11960137	258					0.314	0.426		0.260		
rs1196705	258	0.593		0.322					0.085		
rs12021830	258					0.252		0.411			0.337
rs12380601	258	0.078		0.252					0.671		
rs1244379	258					0.205		0.357			0.438
rs12701745	258	0.694		0.271					0.035		
rs13097560	258	0.085		0.399					0.516		
rs132663	258	0.244			0.426						0.329
rs13280988	258	0.043		0.310					0.647		
rs1396883	258	0.058		0.337					0.605		
rs1453858	258	0.298			0.465						0.236
rs1567803	258					0.535		0.380			0.085
rs16851773	258	0.415		0.473					0.112		
rs16880715	257								0.623	0.311	0.066
rs17150066	254	0.264		0.378					0.358		
rs17157450	258					0.357		0.403			0.240
rs1790740	258					0.213		0.500			0.287
rs1809119	258	0.504	0.391			0.105					
rs2187239	258	0.174	0.477			0.349					
rs2251244	258	0.403	0.434			0.163					
rs2304580	258					0.155		0.504			0.341
rs2384982	258					0.202		0.504			0.295
rs2450219	253					0.585		0.352			0.063
rs2585897	258	0.267		0.407					0.326		
rs2736340	257					0.280		0.377			0.342
rs2795918	258					0.376		0.415			0.209
rs2836181	258					0.663		0.306			0.031
rs2937673	257					0.066		0.272			0.661
rs298259	258					0.097		0.395			0.508
rs305163	255	0.553		0.333					0.114		

rs314580	255	0.290		0.486				0.224		
rs349323	258	0.624			0.314					0.062
rs353373	258	0.178		0.434				0.388		
rs4402506	258					0.054		0.345		0.601
rs4787040	258	0.442			0.407					0.151
rs4812381	258	0.236		0.395				0.368		
rs4812831	258	0.244		0.345				0.411		
rs485645	258	0.236		0.407				0.357		
rs6101991	258	0.624		0.295				0.081		
rs6573746	257	0.311	0.440			0.249				
rs6684063	258							0.333	0.434	0.233
rs6710083	258							0.678	0.279	0.043
rs6711746	258					0.349		0.426		0.225
rs6768750	258	0.329		0.419				0.252		
rs6803290	258					0.054		0.310		0.636
rs6866970	258	0.620		0.295				0.085		
rs6994396	258							0.318	0.407	0.275
rs7103088	258	0.058		0.322				0.620		
rs7124676	258	0.039		0.264				0.698		
rs7164838	258	0.306		0.438				0.256		
rs7173885	258					0.609	0.310	0.081		
rs7211426	258	0.116		0.430				0.453		
rs7594727	258	0.074	0.314			0.612				
rs7630522	258					0.171		0.477		0.353
rs7664076	258	0.159		0.473				0.368		
rs7665516	258	0.364		0.419				0.217		
rs7808899	258							0.384	0.473	0.143
rs7927234	258					0.039	0.337	0.624		
rs7958163	258	0.516	0.384			0.101				
rs7961436	258	0.496		0.384				0.120		
rs8041147	253					0.383		0.435		0.182
rs862500	258	0.360			0.496					0.143
rs874189	258					0.562		0.302		0.136
rs878874	258					0.302		0.384		0.314
rs924780	258	0.105	0.376			0.519				
rs9422913	257					0.249		0.432		0.319
rs9533623	258	0.050		0.295				0.655		
rs956969	258					0.562		0.376		0.062
rs9576996	258					0.147	0.380	0.473		
rs9923547	258	0.112	0.357			0.531				

**Tabla Suplementaria 2-4b. Frecuencias genotípicas para SNPforID 52 -plex**

SNP rsID	N	AA	AC	AG	AT	CC	CG	CT	GG	GT	TT
rs1005533	256	0.172		0.414					0.414		
rs1015250	256					0.094			0.473		
rs1024116	256					0.219		0.531			0.250
rs1028528	256	0.508		0.363					0.129		
rs1029047	256	0.113			0.348						0.539
rs1031825	256	0.125	0.461			0.414					
rs10495407	256	0.086		0.414					0.500		
rs1335873	256	0.297			0.531						0.172
rs1355366	256					0.074		0.414			0.512
rs1357617	256	0.016			0.199						0.785
rs1360288	256					0.641		0.313			0.047
rs1382387	256	0.496	0.402			0.102					
rs1454361	256	0.176			0.512						0.313
rs1463729	256					0.125		0.488			0.387
rs1493232	256	0.387	0.438			0.160					
rs1528460	256					0.199		0.531			0.270
rs1886510	256	0.211		0.512					0.277		
rs1979255	256					0.215	0.492		0.293		
rs2016276	256					0.039		0.238			0.723
rs2040411	256	0.539		0.348					0.113		
rs2046361	256	0.160			0.434						0.406
rs2056277	256					0.723		0.266			0.012
rs2076848	256	0.500			0.422						0.078
rs2107612	256	0.645		0.277					0.078		
rs2111980	256					0.195		0.492			0.313
rs251934	256	0.453		0.453					0.094		
rs2830795	256	0.621		0.328					0.047		
rs2831700	256	0.313		0.473					0.215		
rs354439	256	0.438			0.391						0.172
rs717302	256	0.535		0.359					0.105		
rs719366	256	0.531		0.379					0.066		
rs722098	256	0.148		0.430					0.402		
rs727811	256								0.371	0.453	0.176
rs729172	256								0.391	0.422	0.184
rs733164	256	0.125		0.473					0.402		
rs735155	256					0.289		0.484			0.207
rs737681	256					0.559		0.367			0.074

<b>rs740910</b>	256	0.402		0.445				0.152		
<b>rs763869</b>	256	0.406		0.434				0.160		
<b>rs8037429</b>	256					0.305		0.492		0.203
<b>rs826472</b>	256					0.523		0.395		0.082
<b>rs873196</b>	256					0.047		0.320		0.633
<b>rs876724</b>	256					0.629		0.320		0.051
<b>rs891700</b>	256	0.172		0.477				0.352		
<b>rs901398</b>	256					0.133		0.445		0.422
<b>rs907100</b>	256					0.234	0.469	0.297		
<b>rs914165</b>	256	0.383		0.445				0.172		
<b>rs917118</b>	256					0.289		0.488		0.223
<b>rs938283</b>	256					0.008		0.195		0.797
<b>rs964681</b>	256					0.141		0.473		0.387
<b>rs1413212</b>	256					0.551		0.355		0.094
<b>rs1490413</b>	256	0.273		0.496				0.230		



**Tabla Suplementaria 2-5.** Matriz de P valores de Fst pareados (Método de distancias: Diferencias pareadas). Sobre la diagonal con marcadores SNPforID 52plex). Bajo la diagonal para marcadores tipo AIMs. Se resaltan los valores significativos.

		SNPforID 52plex								
		Necoclí	Turbo	Baranoa	Tubará	Becerril	Pueblo Bello	Valledupar	Tuchín	San Juan del Cesar
AIMs	Necoclí	*	0.50817+- 0.0045	0.17087+- 0.0042	0.11435+- 0.0029	0.30254+- 0.0050	0.10969+- 0.0031	0.05297+- 0.0025	0.00000+- 0.0000	0.44233+- 0.0048
	Turbo	0.11850+- 0.0032	*	0.12444+- 0.0035	0.00139+- 0.0003	0.18523+- 0.0039	0.10603+- 0.0030	0.19384+- 0.0040	0.00000+- 0.0000	0.85516+- 0.0037
	Baranoa	0.36630+- 0.0043	0.05752+- 0.0023	*	0.17078+- 0.0038	0.17691+- 0.0037	0.10454+- 0.0029	0.11979+- 0.0032	0.00000+- 0.0000	0.05772+- 0.0023
	Tubará	0.00426+- 0.0006	0.00000+- 0.0000	0.00000+- 0.0000	*	0.00069+- 0.0003	0.00624+- 0.0007	0.00000+- 0.0000	0.18612+- 0.0042	0.00010+- 0.0001
	Becerril	0.02416+- 0.0014	0.11514+- 0.0034	0.99248+- 0.0008	0.00000+- 0.0000	*	0.62469+- 0.0055	0.60380+- 0.0048	0.00000+- 0.0000	0.24374+- 0.0038
	Pueblo Bello	0.00861+- 0.0009	0.00069+- 0.0003	0.03158+- 0.0018	0.00010+- 0.0001	0.04950+- 0.0021	*	0.45629+- 0.0052	0.00000+- 0.0000	0.04336+- 0.0020
	Valledupar	0.00772+- 0.0008	0.06881+- 0.0023	0.49124+- 0.0047	0.00000+- 0.0000	0.48688+- 0.0042	0.00644+- 0.0007	*	0.00000+- 0.0000	0.19147+- 0.0041
	Tuchín	0.00040+- 0.0002	0.00000+- 0.0000	0.00000+- 0.0000	0.01940+- 0.0013	0.00000+- 0.0000	0.00000+- 0.0000	0.00000+- 0.0000	*	0.00000+- 0.0000
San Juan del Cesar	0.00257+- 0.0005	0.34967+- 0.0040	0.03554+- 0.0018	0.00000+- 0.0000	0.08504+- 0.0031	0.00020+- 0.0001	0.25225+- 0.0043	0.00000+- 0.0000	*	

**Tabla Suplementaria 2-6.** Resultados promedio de los coeficientes de membresía de cada individuo a las poblaciones ancestrales de referencia empleadas con el software STRUCTURE. La tercera columna indica el número de individuos que registran ancestría nativoamericana mayor a 0.6 con marcadores AIMs y las siguientes, los valores de ancestría promedio por población con los paneles de estudio.

POBLACION	N	Personas con Ancestría Nativoamericana >0.6 con AIMs	Ancestría Nativoamericana promedio		Ancestría Africana promedio		Ancestría Europea promedio	
			AIMs	52plex	AIMs	52plex	AIMs	52plex
<b>NECOCLI</b>	12	3 (25%)	0.4483	0.3315	0.2876	0.3158	0.2643	0.3526
<b>TURBO</b>	14	0	0.2968	0.2959	0.3701	0.3618	0.3333	0.3424
<b>BARANOA</b>	28	1 (3.57%)	0.3839	0.3345	0.2480	0.2941	0.3681	0.3715
<b>TUBARA</b>	16	10 (62.5%)	0.6231	0.4159	0.1344	0.2869	0.2424	0.2971
<b>BECERRIL</b>	24	0	0.3465	0.3219	0.2300	0.2713	0.4233	0.4068
<b>PUEBLO BELLO</b>	33	4 (12.12%)	0.3901	0.3186	0.1552	0.2759	0.4546	0.4053
<b>VALLEDUPAR</b>	40	1 (2.5%)	0.3218	0.3194	0.2612	0.2803	0.4171	0.4002
<b>TUCHIN</b>	54	35 (64.82%)	0.6990	0.4736	0.1059	0.2483	0.1951	0.2782
<b>SAN JUAN DEL CESAR</b>	37	2 (5.41%)	0.2675	0.3175	0.3171	0.3291	0.4154	0.3534
<b>TOTAL</b>	258	56 (21.71%)	0.4339	0.3582	0.2181	0.2870	0.3480	0.3547

**Tabla Suplementaria 2-7.** Datos de proporción de ancestría con marcadores autosómicos en poblaciones colombianas de diversas publicaciones. \*Población nativoamericana colombiana

LOCALIDAD	ANCESTRÍA EUROPEA	ANCESTRÍA AFRICANA	ANCESTRÍA NATIVOAMERICANA	MARCADOR	REFERENCIA
<b>Bolivar</b>	0.233	0.438	0.329	11 AIMs (Indels y SNPs)	Rojas et.al. 2010
<b>Magdalena</b>	0.5	0.282	0.218	11 AIMs (Indels y SNPs)	Rojas et.al. 2010
<b>Antioquia</b>	0.635	0.106	0.26	11 AIMs (Indels y SNPs)	Rojas et.al. 2010
<b>Antioquia</b>	0.52	0.06	0.39	325 AIMs SNPs	Galanter et.al. 2012
<b>Colombia (Medellín)</b>	0.625	0.092	0.274	SNPs Genome Wide	Homburger et.al, 2015
<b>Medellin</b>	0.65	0.1	0.25	678 STRs	Wang, et.al. 2008
<b>Córdoba</b>	0.4219	0.2248	0.3533	100 SNPs	Criollo, 2012
<b>Colombia</b>	0.612	0.096	0.292	93328 SNPs	Adhikari, et.al. 2016
<b>Awa*</b>	0.17	0.02	0.80	325 AIMs SNPs	Galanter et.al. 2012
<b>Coyaima*</b>	0.09	0.02	0.86	325 AIMs SNPs	Galanter et.al. 2012
<b>Pastos*</b>	0.16	0.02	0.83	325 AIMs SNPs	Galanter et.al. 2012
<b>Nasa*</b>	0.0344	0.0066	0.959	100 SNPs	Criollo 2012
<b>Pijao*</b>	0.1281	0.0176	0.8543	100 SNPs	Criollo 2012
<b>Embera*</b>	0.1226	0.0334	0.844	100 SNPs	Criollo 2012
<b>Zenú*</b>	0.3491	0.1721	0.4788	100 SNPs	Criollo 2012

**Tabla Suplementaria 2-8. Conceptos empleados para la evaluación analítica de la determinación del componente ancestral mayoritario**

INDICADOR	DEFINICION	EJEMPLO
<b>Sensibilidad</b>	Probabilidad de que un individuo tenga el componente ancestral mayoritario indicado por análisis del software STRUCTURE con base en la genotipificación con el panel SNPforID 52plex, dado que el mismo componente ancestral mayoritario es determinado por el análisis del panel AIMs	$63/112 = 0.563$
<b>Especificidad</b>	Probabilidad de que un individuo no tenga el componente ancestral mayoritario que se analiza, de acuerdo con el resultado de software STRUCTURE con base en la genotipificación con el panel SNPforID 52plex, dado que el mismo resultado se obtiene por el análisis del panel AIMs	$102/142 = 0.718$
<b>Falsos Positivos</b>	Número de individuos cuyo resultado por análisis con SNPforID 52plex indica la presencia del componente ancestral mayoritario analizado, que no se detecta con el test de referencia (AIMs)	40
<b>Cociente Falso Positivo</b>	Proporción de falsos positivos dados por el análisis con SNPforID 52plex, en el total de negativos determinados con el panel AIMs	$40/142 = 0.282$
<b>Falsos Negativos</b>	Número de individuos cuyo resultado por análisis con SNPforID 52plex indica la ausencia del componente ancestral mayoritario analizado, que si se detecta con el test de referencia (AIMs)	49
<b>Cociente Falso Negativo</b>	Proporción de falsos negativos dados por el análisis con SNPforID 52plex, en el total de positivos determinados con el panel AIMs	$49/112 = 0.438$
<b>VPP</b>	Probabilidad a posteriori de presentar realmente el componente ancestral mayoritario indicado dado que es el indicado por el análisis con SNPforID 52plex	$63/103 = 0.612$
<b>VPN</b>	Probabilidad a posteriori de no presentar realmente el componente ancestral mayoritario indicado dado que es lo indicado por el análisis con SNPforID 52plex	$102/151 = 0.675$
<b>Prevalencia real</b>	Proporción de individuos que tienen el componente ancestral mayoritario indicado por el test de referencia (AIMs)	$112/254 = 0.441$
<b>Prevalencia aparente</b>	Proporción de individuos que tienen el componente ancestral mayoritario indicado por el test alternativo (SNPforID 52plex)	$63/254 = 0.248$

*Tabla de referencia para los datos del ejemplo:*

		AIMs		TOTAL
		Europeo	No Europeo	
52plex	Europeo	63	40	103
	No Europeo	49	102	151
TOTAL		112	142	254

**Tabla Suplementaria 2-9.** A. Evaluación analítica de la determinación del componente ancestral mayoritario **A.** Cuando la proporción del componente ancestral mayoritario detectado con AIMs es igual o superior a 0.6. **B.** Cuando la proporción del componente ancestral mayoritario detectado con AIMs es igual o superior a 0.8, en población mezclada de la Costa Caribe Colombiana, empleando como método de referencia el uso de marcadores AIMs y como test alternativo el uso de los marcadores SNPforID 52plex. (Eur= Europeo, Afr= Africano, Nam=Nativoamericano)

INDICADOR	A.						B.		
	EVALUACION GENERAL			EVALUACION PAREADA			EVALUACION GENERAL	EVALUACION PAREADA	
	Eur Vs No Eur	Afr Vs No Afr	Nam Vs No Nam	Eur Vs Afr	Eur Vs Nam	Afr Vs Nam	Nam Vs No Nam	Eur Vs Nam	Afr Vs Nam
<b>Sensibilidad</b>	0.800	1,000	0.909	0.857	0.923	1,000	0.943	1,000	1,000
<b>Especificidad</b>	1,000	1,000	1,000	1,000	0.943	0.962	1,000	0.971	0.971
<b>Cociente Falso Positivo</b>	0.000	0.000	0.000	0.000	0.057	0.038	0.000	0.029	0.029
<b>Cociente Falso Negativo</b>	0.200	0.000	0.091	0.143	0.077	0.000	0.057	0.000	0.000
<b>VPP</b>	1,000	1,000	1,000	1,000	0.800	0.600	1,000	0.500	0.500
<b>VPN</b>	0.951	1,000	0.783	0.600	0.980	1,000	0.500	1,000	1,000
<b>Prevalencia real</b>	0.205	0.041	0.753	0.824	0.197	0.055	0.946	0.029	0.029
<b>Prevalencia aparente</b>	0.164	0.041	0.685	0.706	0.182	0.055	0.892	0.029	0.029

**Tabla Suplementaria 2-10.** Valores de divergencia (In) de cada SNP del set SNPforID 52plex. Se muestran las comparaciones de cada población de referencia Vs las demás (Ej. Europeo Vs No Europeo) y las comparaciones por pares. La última columna indica la población que se distingue el SNP según las frecuencias poblaciones. (Eur= Europeo, Afr= Africano, Nam=Nativoamericano)

SNP	Divergencias por población			Divergencias pareadas			Población
	In Eur vs No Eur	In Afr Vs No Afr	In Nam vs No Nam	In Afr Vs Eur	In Afr Vs Nam	In Eur Vs Nam	
rs1005533	0.0312	0.046	0.2615	0.0013	0.279	0.2471	Nativoamericana
rs1015250	0.1017	0.2846	0.0981	0.2699	0.315	0.0092	Africana
rs1024116	0.0489	0.0876	0.0075	0.1022	0.0647	0.0049	Africana
rs1028528	0.106	0.3157	0.2525	0.2692	0.5231	0.1214	Africana
rs1029047	0.0227	0.05	0.1141	0.0288	0.1447	0.0997	Nativoamericana
rs1031825	0.0235	0.0473	0.0105	0.0507	0.0427	0.0029	Africana
rs10495407	0.0581	0.1884	0.0328	0.1886	0.188	0	Africana
rs1335873	0.1853	0.3478	0.0253	0.4079	0.2622	0.0247	Africana
rs1355366	0.0195	0.0778	0.2789	0.0205	0.3371	0.2342	Nativoamericana
rs1357617	0.038	0.0017	0.0873	0.0147	0.0578	0.1178	Europea
rs1360288	0.0514	0.0019	0.1125	0.0201	0.0715	0.1542	Nativoamericana
rs1382387	0.0248	0.117	0.062	0.0951	0.1632	0.0123	Africana
rs1413212	0.0037	0.003	0.002	0.0047	0.0025	0.0037	Europea
rs1454361	0.0128	0.0227	0.0025	0.0266	0.0173	0.0024	Africana
rs1463729	0.0068	0.0276	0.0175	0.0222	0.0448	0.0054	Africana
rs1490413	0.0241	0.0071	0.1057	0.0017	0.0932	0.1173	Nativoamericana
rs1493232	0.0262	0.0114	0.0182	0.0245	0.0128	0.0354	Europea
rs1528460	0.1343	0.3376	0.0881	0.3468	0.325	0.001	Africana
rs1886510	0.087	0.177	0.0261	0.1827	0.1613	0.004	Africana
rs1979255	0.0428	0.0041	0.1245	0.0078	0.097	0.1527	Nativoamericana
rs2016276	0.0142	0.0739	0.029	0.0629	0.0951	0.0076	Africana
rs2040411	0.0103	0.0666	0.0371	0.0514	0.0983	0.0099	Africana
rs2046361	0.0183	0.0387	0.006	0.0435	0.0339	0.0048	Africana
rs2056277	0.0932	0.0198	0.0869	0.0663	0.0251	0.1572	Nativoamericana
rs2076848	0.0587	0.0757	0.0065	0.1043	0.0446	0.0291	Africana
rs2107612	0.004	0.0259	0.086	0.0064	0.1084	0.0683	Nativoamericana
rs2111980	0.0414	0.047	0.001	0.0683	0.0216	0.0172	Europea
rs251934	0.0528	0.0177	0.0252	0.0471	0.0035	0.0642	Europea
rs2830795	0.0726	0.0401	0.0159	0.0811	0.0029	0.062	Europea

<b>rs2831700</b>	0.0113	0.0137	0.0003	0.019	0.0067	0.0037	Europea
<b>rs354439</b>	0.0399	0.1068	0.0656	0.0977	0.1462	0.0445	Africana
<b>rs717302</b>	0.0414	0.0174	0.1993	0.0018	0.1827	0.215	Nativoamericana
<b>rs719366</b>	0.0303	0.049	0.0014	0.0608	0.0306	0.0058	Africana
<b>rs722098</b>	0.2902	0.1916	0.05	0.3462	0.0259	0.22	Europea
<b>rs727811</b>	0.0822	0.1108	0.0015	0.1489	0.0619	0.0238	Africana
<b>rs729172</b>	0.0169	0.0685	0.0186	0.0634	0.078	0.001	Africana
<b>rs733164</b>	0.0015	0.0015	0.0084	0.0003	0.0086	0.0082	Europea
<b>rs735155</b>	0.0188	0.0443	0.0122	0.0451	0.0449	0.0044	Africana
<b>rs737681</b>	0.0208	0.0339	0.2392	0.0015	0.2546	0.2263	Nativoamericana
<b>rs740910</b>	0.0191	0.2387	0.1733	0.1694	0.3996	0.0818	Africana
<b>rs763869</b>	0.0379	0.0183	0.171	0.0076	0.1564	0.1845	Nativoamericana
<b>rs8037429</b>	0.0002	0.014	0.0171	0.0071	0.031	0.0087	Africana
<b>rs826472</b>	0.1105	0.0152	0.114	0.0714	0.0389	0.2001	Europea
<b>rs873196</b>	0.0603	0.0034	0.0739	0.033	0.0318	0.1251	Europea
<b>rs876724</b>	0.0504	0.0193	0.0209	0.0474	0.0015	0.0563	Europea
<b>rs891700</b>	0.0009	0.0085	0.026	0.0018	0.0336	0.0202	Nativoamericana
<b>rs901398</b>	0.0303	0.0058	0.0191	0.0235	0.0046	0.044	Europea
<b>rs907100</b>	0.0455	0.2078	0.1291	0.1664	0.3115	0.0415	Africana
<b>rs914165</b>	0.0963	0.1326	0.0014	0.1766	0.0741	0.0261	Africana
<b>rs917118</b>	0.1417	0.1227	0.0043	0.204	0.0379	0.079	Europea
<b>rs938283</b>	0.0185	0.0035	0.0147	0.0132	0.0034	0.0294	Nativoamericana
<b>rs964681</b>	0.034	0.0333	0.0006	0.0507	0.0109	0.0152	Europea
<b>Divergencia acumulada</b>	<b>2.6132</b>	<b>4.0218</b>	<b>3.3829</b>	<b>4.3743</b>	<b>5.6117</b>	<b>3.4355</b>	

**Tabla Suplementaria 2-11.** Datos de autodeterminación y de asignación genealógica y los resultados de estimación de proporción de ancestría de acuerdo con los resultados de *Structure*, asignación por *DAPC* y clasificación bayesiana con *Snipper*. Los individuos resaltados corresponden a los que se ubican en la nube intermedia en la gráfica de ACP (Figura Suplementaria 1)

MUESTRA	DATOS DESDE LA ENCUESTA			STRUCTURE			DAPC	SNIPPER	
	Lugar Nacimiento	Autodeterminación	Clasificación genealógica	Europeo	Africano	Nativoamericano		LR	CLASIFICACION BAYESIANA
ANNE01003	APARTADO	ZENU	ZENU	0.066	0.028	0.906	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
ANNE01005	NECOCLI	ZENU	ZENU-ANTIOQUEÑO	0.083	0.07	0.847	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>ANNE01006</b>	ARBOLETES	ZENU	ANTIOQUEÑO	0.265	0.118	0.617	MEZCLADA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>ATTU01001</b>	TUBARA	MOKANA	MOKANA	0.226	0.084	0.69	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>ATTU01003</b>	TUBARA	MOKANA	MOKANA	0.16	0.088	0.752	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
ATTU01004	TUBARA	MOKANA	MOKANA	0.084	0.125	0.791	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>ATTU01006</b>	TUBARA	NS	MOKANA	0.091	0.172	0.736	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>ATTU01008</b>	TUBARA	MOKANA	MOKANA	0.166	0.159	0.675	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
ATTU01010	TUBARA	MOKANA	MOKANA	0.205	0.07	0.725	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>ATTU01012</b>	TUBARA	MOKANA	MOKANA	0.183	0.154	0.663	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
ATTU01014	JUAN MINA-BARRANQUILLA	MOKANA	MOKANA	0.058	0.048	0.894	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
CEPB01007	SIERRA NEVADA DE SANTA MARTA	NS	ARHUACO	0.107	0.075	0.818	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
CEPB01008	PUEBLO BELLO	ARHUACO	ARHUACO	0.056	0.08	0.864	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
CEPB01024	PUEBLO BELLO	ARHUACO	ARHUACO	0.154	0.05	0.797	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>CEPB01034</b>	PUEBLO BELLO	ARHUACO	ARHUACO	0.151	0.151	0.698	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA



CEVA01013	PUEBLO BELLO	ARHUACO	ARHUACO	0.061	0.057	0.882	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COSA01006	SAN ANDRES DE SOTAVENTO	ZENU	ZENU	0.06	0.042	0.898	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>COSA01007</b>	SAN ANDRES DE SOTAVENTO	INDIGENA	INDETERMINADO	0.153	0.154	0.693	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>COTU02004</b>	SAN ANDRES DE SOTAVENTO	ZENU	ZENU	0.148	0.115	0.737	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02005	SAN ANDRES DE SOTAVENTO	ZENU	ZENU	0.07	0.078	0.852	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02006	TUCHIN	ZENU	ZENU	0.059	0.034	0.907	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02011	SAN ANDRES DE SOTAVENTO	NS	CORDOBES	0.033	0.034	0.933	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02012	TUCHIN	ZENU	ZENU	0.07	0.07	0.86	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02014	SAN ANDRES DE SOTAVENTO	ZENU	ZENU	0.116	0.046	0.838	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>COTU02015</b>	TUCHIN	ZENU	ZENU	0.12	0.042	0.838	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02016	TUCHIN	NS	CORDOBES	0.051	0.023	0.925	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02018	SAN ANDRES DE SOTAVENTO	ZENU	ZENU	0.042	0.069	0.889	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02019	TUCHIN	ZENU	ZENU	0.055	0.067	0.879	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02020	SAN ANDRES DE SOTAVENTO	ZENU	ZENU	0.064	0.068	0.869	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02021	TUCHIN	NS	INDETERMINADO	0.043	0.018	0.938	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02026	TUCHIN	ZENU	ZENU	0.069	0.132	0.798	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02029	SAN ANDRES DE SOTAVENTO	ZENU	ZENU	0.085	0.059	0.856	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA

COTU02031	SAN ANDRES DE SOTAVENTO	ZENU	ZENU	0.075	0.054	0.871	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02034	TUCHIN	ZENU	ZENU	0.05	0.021	0.929	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02036	TUCHIN	ZENU	ZENU	0.06	0.031	0.91	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02045	TUCHIN	ZENU	ZENU	0.082	0.042	0.876	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU02050	LORICA	ZENU	ZENU	0.076	0.036	0.888	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU03001	PEÑITAS	NS	CORDOBES	0.036	0.023	0.941	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU03006	SAN ANDRES DE SOTAVENTO	NS	ZENU	0.064	0.027	0.91	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU03007	TUCHIN	ZENU	ZENU	0.042	0.057	0.901	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU03009	TUCHIN	ZENU	ZENU	0.031	0.029	0.94	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU03011	TUCHIN	ZENU	ZENU	0.056	0.023	0.921	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU04001	TUCHIN	ZENU	ZENU	0.018	0.016	0.966	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>COTU04002</b>	SAN ANDRES DE SOTAVENTO	INDIGENA	ZENU	0.201	0.049	0.75	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
COTU04005	TUCHIN	ZENU	ZENU	0.087	0.151	0.761	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
<b>COTU04007</b>	TUCHIN	ZENU	ZENU	0.107	0.058	0.836	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
GUSJ01006	SIERRA NEVADA DE SANTA MARTA	WIWA	WIWA	0.051	0.027	0.923	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA
GUSJ01009	SAN JUAN DEL CESAR	NS	WAYUU-GUAJIRO	0.035	0.057	0.908	NATIVOAMERICANA	>10 <sup>9</sup>	NATIVOAMERICANA

## Bibliografía

1. ACOSTA, Amparo. et.al. The genetic male component of two South-Western Colombian populations. *Forensic Sci Int Genet.* 2009;3(2):2008-2010. doi:10.1016/j.fsigen.2008.06.006
2. ADHICARI, Kaustubh. et.al. Admixture in Latin America. *Curr Opin Genet Dev.* 2016 Dec;41:106-114. doi: 10.1016/j.gde.2016.09.003.
3. ADHICARI, Kaustubh. et.al. The Genetic Diversity of the Americas. *Annu. Rev. Genom. Hum. Genet.* 2017.18:277-296.
4. AGUSTIN, Jose. FUNDACIONES COLONIALES Y REPUBLICANAS EN COLOMBIA: NORMAS, TRAZADO Y RITOS FUNDACIONALES. FUNDACIONES DE CIUDADES Y POBLACIONES. En Banco de la Republica Biblioteca Virtual. *Revista Credencial de Historia* N° 141. Septiembre, 2001 <http://www.banrepcultural.org/biblioteca-virtual/credencial-historia/numero-141/fundaciones-de-ciudades-y-poblaciones>
5. ALONSO, Angela y USAQUÉN, William Y-chromosome and surname analysis of the native islanders of San Andrés and Providencia ( Colombia ). *HOMO - Journal of Comparative Human Biology* " (2013a) Vol. 64: 71-84.
6. ALONSO, Angela. Caracterización de la población humana de los departamentos de tolima y huila. *Perspectivas: demográficas, genéticas y socioculturales.* Tesis Maestria. Universidad Nacional de Colombia. 2013b.
7. ALONSO, Angela. et.al. Paternal portrait of populations of the middle Magdalena River region (Tolima and Huila, Colombia): New insights on the peopling of Central America and northernmost South America. *PLoS One.* 2018 Nov 15;13(11):e0207130. doi: 10.1371/journal.pone.0207130
8. ANDERSON, S. et.al. Sequence and organization of the human mitochondrial genome. *Nature.* 1981 Apr 9;290(5806):457-65.
9. ARANGO, R. y SÁNCHEZ, E. *Los Pueblos Indígenas de Colombia en el Umbral del Nuevo Milenio.* Bogotá: Departamento Nacional de Planeación de Colombia, 2006.

10. ARIAS, Hiliana. Territorio Indígena Kankuamo. Proceso de reconfiguración del resguardo desde las dimensiones socioculturales. Tesis de Maestría. Universidad Nacional de Colombia. 2011
11. ARIAS, L. Estudio de la variación genética en el ADN mitocondrial de nativos Americanos de la amazonía y restos óseos prehispánicos de los andes colombianos. Tesis de maestría. Universidad del Valle. 2012.
12. ATHEY, T. Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach. *Journal of Genetic Genealogy* 2006 2:34-39
13. AULCHENKO, Yurii. GenABEL, Tutorial. (2014). <http://www.r-project.org/>
14. BAJIC, Vladimir. et.al. Genetic structure and sex-biased gene flow in the history of southern African populations. *Am J Phys Anthropol.* 2018 Nov;167(3):656-671. doi: 10.1002/ajpa.23694.
15. BANCO DE la república. Sinú Amerindio. Montería: Editora Géminis Ltda, 1996
16. BANDELT, H. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Ann Hum Genet.* 2003 Nov;67(Pt 6):512-24.
17. BAQUERO, A. y DE LA HOZ, A. Cultura y tradición oral en el Caribe Colombiano. Barranquilla: Ediciones Uninorte, 2010.
18. BATTAGLIA, Vincenza. et.al. The First Peopling of South America: New Evidence from Y-Chromosome Haplogroup Q. *PLoS One.* 2013; 8(8): e71390.
19. BEDOYA, Gabriel. et.al. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci U S A.* 2006 May 9;103(19):7234-9.
20. BEHAR, Doron. et.al. The Dawn of Human Matrilineal Diversity. *The American Journal of Human Genetics* 2008 May; 82, 1130–1140
21. BISSO-MACHADO, Rafael; BORTOLINI, Maria Cátira y SALSANO, Francisco. Uniparental genetic markers in south amerindians. *Genetics and Molecular Biology.* 2012 Vol 35 No. 2; p. 365-387.
22. BLANCO-VEREA, Alejandro. Linajes de cromosoma Y humano: Aplicaciones genético-poblacionales y forenses. Tesis Doctoral. Universidad de Santiago de Compostela, España. 2008

23. BLANCO-VEREA, Alejandro. et.al. Y-chromosome lineages in native South American population. *Forensic Sci Int Genet.* 2010 Apr;4(3):187-93. doi: 10.1016/j.fsigen.2009.08.008.
24. BODNER, Martin. et.al. Rapid coastal spread of First Americans: novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res.* 2012 May;22(5):811-20. doi: 10.1101/gr.131722.111.
25. BOLNICK, Deborah. et.al. Asymmetric male and female genetic histories among Native Americans from Eastern North America. *Mol Biol Evol.* 2006 Nov;23(11):2161-74. Epub 2006 Aug 17.
26. BOLNICK, Deborah. et.al. Native American Genomics and Population Histories. *Annu. Rev. Anthropol.* 2016. 45:319–40
27. BORTOLINI, M. et.al. Y Chromosome evidence for differing ancient demographic histories in the Americas. *American Journal of Human Genetics.* 2003 Vol.73: 524-539.
28. BOSCH, Elena. Aplicaciones de los SNPs del cromosoma Y en genética forense y en el estudio de poblaciones. *Curso On Line de Genética Forense.* Universidad de Zaragoza. (2008). [Consultado Dic. 2008]. Disponible en <http://ebro2.unizar.es/genforense/curso>
29. BRANDINI, Stefania. et.al. The Paleo-Indian Entry into South America According to Mitogenomes. *Mol. Biol. Evol.* 2017 35(2):299–311 doi:10.1093/molbev/msx267
30. BRAVI, C. La evidencia molecular del poblamiento humano de América. *Rev del Mus La Plata.* 2013; 20(177):40-57. <http://ri.conicet.gov.ar/handle/11336/24509>.
31. BRYC, Katarzyna. et.al. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci U S A.* 2010 May 11; 107(Suppl 2): 8954–8961.
32. BULBUL, O. et.al. Inference of biogeographical ancestry across central regions of Eurasia. *Int J Legal Med.* 2016 Jan; 130(1):73-9. doi: 10.1007/s00414-015-1246-7.
33. CANN, Rebecca. Mitochondrial DNA and human evolution. *Nature.* 1987 Jan 1-7;325(6099):31-6.
34. CAPELLI, Cristian. et.al. Moors and Saracens in Europe: estimating the medieval North African male legacy in southern Europe. *European Journal of Human Genetics* (2009) 17, 848 – 852

35. CARVAJAL-CARMONA, Luis. 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 Nov; 67(5):1287-95.
36. CASAS, Andrea. Análisis microevolutivo de la población prehispánica del norte del altiplano cundiboyacense a partir de adn mitocondrial en restos óseos antiguos. Tesis Doctoral. Universidad Nacional de Colombia. 2017.
37. CAVALLI-SFORZA, L. y EDWARDS, W. Phylogenetic Analysis Models and Estimation Procedures. *Am. J. Hum. Genet.* (1967)19 (3): 233-257.
38. CAVALLI-SFORZA, L. Luca, *et.al.* "Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data." *Proceedings of the National Academy of Sciences* (1988); Vol. 85 (16): 6002-6006.
39. CAVALLI-SFORZA, L. Luca, MINCH, Eric. y MOUNTAIN, J. L. Coevolution of genes and languages revisited. *Proceedings of the National Academy of Sciences* (1992); Vol. 89 (12): 5620-5624.
40. CAVALLI-SFORZA, L. Luca; MENOZZI, Paolo y PIAZZA, Alberto. *The history and geography of human genes.* Princeton, New Jersey: Princeton University Press, 1994. 535 p.
41. CAVALLI-SFORZA, L. Luca. Genes, peoples and languages. *En: Proceedings of the National Academy of Sciences.* (Julio, 1997); Vol. 94: 7719-7724.
42. \_\_\_\_\_. Human evolution and its relevance for genetic epidemiology. *Annual Review of Genomics and Human Genetics.* [en línea] (2007); Vol. 8: 1-15. [Consultado 9 Ene. 2012] Disponible en <<http://www.annualreviews.org>>
43. CHAVES, Alvaro. Panorama prehistórico de la costa caribe colombiana. **Universitas Humanística**, [S.l.], jul. 1979, v. 10, n. 10,. ISSN 2011-2734. Disponible en: <<http://revistas.javeriana.edu.co/index.php/univhumanistica/article/view/10470>> . Fecha de acceso: 24 may. 2018
44. CHIARONI, Jacques; UNDERHILL, Peter y CAVALLI-SFORZA, L.Luca. Y chromosome diversity, human expansion, drift and cultural evolution. *Proceedings of the National Academy of Sciences.* [en línea] (Diciembre, 2009) Vol.106 (48):20174-20179. [consultado 9 Ene. 2012] Disponible en <http://www.pnas.org/cgi/doi/10.1073/pnas0910803106>
45. CORACH, Daniel. Inferring Continental Ancestry of Argentineans from Autosomal, Y-Chromosomal and Mitochondrial DNA. *Annals of Human Genetics* (2010) 74:65–76

46. CRAWFORD, Michael; RUBIICZ, Rohina y ZLOJUTRO, Mark. Origins of aleuts and genetic structure of populations of the archipelago: molecular and archaeological perspectives. *Human Biology*. (2010); Vol. 82 No. 5-6: 695-717.
47. CRIOLLO, Angel. Caracterización molecular de la variación genética en cuatro etnias indígenas (Pijao, Paez, Embera y Zenu) y dos poblaciones mestizas de Colombia (Tolima y Córdoba) mediante marcadores del mDNA, NRY Y AIMs. [Tesis de Maestría]. Colombia: Universidad del Tolima, Facultad de Ciencias, Departamento de Biología; 2012
48. CRIOLLO, Angel. et.al. Native American gene continuity to the modern admixed population from the Colombian Andes: Implication for biomedical, population and forensic studies. *Forensic Science International: Genetics* (2018) 36 e1–e7
49. CUBILLOS, Julio. Apuntes para el estudio de la cultura Pijao. ICANH. *Boletín de Arqueología*. 1946; 2 (1): 47-82.
50. DANE. Colombia una nación multicultural. Su diversidad étnica. 2007
51. DANE. Boletín Censo General de población 2005. Perfil San Andrés de Sotavento, Córdoba. 2010a.
52. DANE. Boletín Censo General de población 2005. Perfil Tubará, Atlántico. 2010b.
53. DANE. La visibilización estadística de los grupos étnicos colombianos. 2010c
54. DE LA PUENTE, Maria. et.al. The Global AIMs Nano set: A 31-plex SNaPshot assay of ancestry-informative SNPs. *Forensic Sci Int Genet*. 2016 May 22: 81-88. doi: 10.1016/j.fsigen.2016.01.015.
55. DILLEHAY, Tom. Profiles in Pleistocene History. En SILVERMAN, Helaine y ISBELL, William. *Handbook of South American Archaeology*. Springer: New York, 2008.
56. DING, Lili. et.al. Comparison of measures of marker informativeness for ancestry and admixture mapping. *BMC Genomics*. 2011 Dec 20;12:622. doi: 10.1186/1471-2164-12-622.
57. DULIK, Matthew. et.al. Mitochondrial ADN and Y chromosome variation provides evidence for a recent common ancestry between native americans and indigenous altaians. *American Journal of Human Genetics*. (2012); Vol. 90: 229-246.
58. EARL, D. y vonHOLDT, B. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Res.* (2012) 4: 359-361

59. EASTMAN, J. El Archipiélago de San Andrés y Providencia: formación histórica hasta 1822. *Revista Credencial Historia* (1992) No. 36. En <http://www.banrepcultural.org/biblioteca-virtual/credencial-historia/numero-36/el-archipelago-de-san-andres-y-providencia:-formacion-historica-hasta-1822>
60. ELHAIK, Eran. et.al. The GenoChip: A New Tool for Genetic Anthropology. *Genome Biol Evol.* 2013; 5 (5):1021-31. doi: 10.1093/gbe/evt066.
61. ELHAIK, Eran. et.al. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun.* 2014 Apr 29;5:3513. doi: 10.1038/ncomms4513.
62. EXCOFFIER, Laurent y LISCHER, HEIDI. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* (2010) 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x.
63. EYHERAMENDY, Susana. et.al. Genetic structure characterization of Chileans reflects historical immigration patterns. *Nat Commun.* 2015 Mar 17;6:6472. doi: 10.1038/ncomms7472.
64. FAGUNDES, Nelson. et.al. Mitochondrial Population Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas. *Am. Jour. Hum. Genet.* 2008 Mar 82: 583–592. doi: 10.1016/j.ajhg.2007.11.013.
65. FIEDEL, Stuart. The Peopling of the New World: Present Evidence, New Theories, and Future Directions. *Journal of Archaeological Research.* 2000, Vol. 8 (1): 39-103.
66. FONDEVILA, Manuel. et.al. Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Sci Int Genet.* 2013 Jan;7(1):63-74. doi: 10.1016/j.fsigen.2012.06.007.
67. FRANCO, Fabio y BARRETO, Guillermo. Estructura genética de poblaciones indígenas del occidente colombiano mediante el uso de marcadores ligados al cromosoma Y. *Rev. Acad. Colomb. Cienc. Ex. Fis. Nat.* (2017) Julio-Septiembre, 41 (160): 281-291
68. GALANTER, Jhosua. et.al. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet.* 2012; 8(3):e1002554.
69. GALLARDO, Milton. *Evolución: El curso de la vida.* Buenos Aires: Médica Panamericana, 2011. p. 201-226.



70. GAMBOA, J. La expedición de Gonzalo Jiménez de Quesada por el río Magdalena y el origen del Nuevo Reino de Granada (1536-1537). En Banco de la Republica Biblioteca Virtual. Revista Credencial de Historia N° 283 Julio 2013. <http://www.banrepcultural.org/biblioteca-virtual/credencial-historia/numero-283/expedicion-gonzalo-jimenez-de-quesada-por-rio-magdalena>
71. GOLDSTEIN, D. et.al. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA*. 1995 July. 92:6723-6727
72. GOLDSTEIN, David. B. y CHIKHI, Lounes. Human migrations and population structure: What we know and why it matters. *Annual review of genomics and human genetics* (2002); Vol 3: 129-152.
73. GONZALEZ, Juan. et.al. SNPAssoc: an R package to perform whole genome association studies. *Bioinformatics*. 2007 Mar 1;23(5):644-5. Epub 2007 Jan 31.
74. GREENBERG, J. 1987 *Languages in the Americas*. Stanford: Stanford University Press.
75. GRUGNI, V. et.al. Exploring the Y chromosomal ancestry of modern panamanians. *PLoS One*. 2015; 10(12):1-24. doi:10.1371/journal.pone.0144223
76. GUERRA, Carlos. et.al. A bayesian approach to genome/linguistic relationships in native South Americans. *PLoS One*. 2013 May 16; 8(5):e64099. doi: 10.1371/journal.pone.0064099.
77. GUSMAO, A. et.al. A Perspective on the History of the Iberian Gypsies Provided by Phylogeographic Analysis of Y-Chromosome Lineages. *Annals of Human Genetics* (2008) 72,215–227
78. GUTIERREZ, Omar. Conflictos sociales y violencia en el departamento del Cesar, Colombia. *Revista Colombiana de Sociología*; (2012) Vol. 35 (1): 1-16.
79. HE, Y. et.al. Genetic divergence disclosing a rapid prehistorical dispersion of natives americans in Central and South America. *Plos One* [en línea] Vol. 7 No 9 (2012); p. e44788. [consultado 8 Dic. 2012] Disponible en <http://www.plosone.org>
80. HEINZ, Tanja, et.al. Ancestry analysis reveals a predominant Native American component with moderate European admixture in Bolivians. *Forensic Sci Int Genet*. 2013 Sep;7(5):537-42. doi: 10.1016/j.fsigen.2013.05.012.
81. HENN, Brenna, CAVALLI-SFORZA, Lugi y FELDMAN, Marcus. The great human expansion. *Proc Natl Acad Sci U S A*. 2012 Oct 30;109 (44):17758-64. doi: 10.1073/pnas.1212380109.

82. HEY, Y. On the number of new world founders: a population genetic portrait of the peopling of the Americas. En Plos Biology [en línea] (2005); Vol. 3 (6): e193. [consultado 1 mar. 2013] Disponible en <http://www.plosbiology.org>
83. HINCAPIE, C. Inmigrantes extranjeros en el desarrollo del Quindío. Armenia: Editorial QuinGraficas. 1 Ed., 1995. 371p.
84. HOMBURGER, Julian. et.al. Genomic insights into the ancestry and demographic history of South America. PLoS Genet. 2015 11(12): e1005602. doi:10.1371/journal.pgen.1005602
85. HUNLEY, Keith. et.al. The impact of group fissions on genetic structure in Native South America and implications for human evolution. Am J Phys Anthropol. 2008 Feb;135(2):195-205.
86. HUNLEY, Keith. y HEALY, Meghan. The impact of founder effects, gene flow, and European admixture on native American genetic diversity. Am J Phys Anthropol. 2011 Dec;146(4):530-8. doi: 10.1002/ajpa.21506.
87. IBARRA, Adriana. et.al. Comparison of the genetic background of different Colombian populations using the SNPforID 52plex identification panel. Int J Legal Med. 2014 Jan; 128(1):19-25. doi: 10.1007/s00414-013-0858-z.
88. JARAMILLO, J. La población indígena de Colombia en el momento de la conquista y sus transformaciones posteriores. Anuario Colombiano de Historia Social y de la Cultura. 1964 No. 2, pag 239-293 ISSN electrónico 2256-5647. ISSN impreso 0120-2456. **DOI: 10.15446/achsc**  
<http://www.bdigital.unal.edu.co/30742/1/29632-106428-1-PB.pdf>
89. JARAMILLO, Jaime. Mestizaje y diferenciación social en el Nuevo Reino de Granada en la segunda mitad del siglo XVIII. Anuario Colombiano de Historia Social y de la Cultura. No. 3, 1965. Pag 21-48 ISSN electrónico 2256-5647. ISSN impreso 0120-2456. **DOI: 10.15446/achsc.**  
<https://revistas.unal.edu.co/index.php/achsc/article/view/29674>
90. JOHNSON, M. et.al. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol. 1983;19 (3-4):255-71.
91. JOHNSON, Nicholas. et.al. Ancestral components of admixed genomes in a Mexican cohort. PLoS Genet. 2011 Dec; 7 (12): e1002410. doi: 10.1371/journal.pgen.1002410.
92. JOMBART, Thibaut. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008; 24:1403-1405. doi: 10.1093/bioinformatics/btn129.

93. JOMBART, Thibaut y COLLINS, Caitlin. Analysing genome-wide SNP data using adegenet 2.0.0. 2015. <http://adegenet.r-forge.r-project.org/files/tutorial-genomics.pdf>
94. KARAFET, Tatiana, et.al. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* May 2008; Vol 18 (15): 830-8
95. KEYEUX, G. et.al. Possible migration routes into South America deduced from mitochondrial ADN studies in colombian amerindian populations. *Human Biology.* (2002) Vol. 74; p.211-233.
96. KIDD, Judith. R. et.al. Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. *Investig Genet.* 2011 Jan 5;2(1):1. doi: 10.1186/2041-2223-2-1.
97. KIDD, Kenneth. et.al. Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Sci Int Genet.* 2014 May;10:23-32. doi: 10.1016/j.fsigen.2014.01.002
98. KOSOY, Roman. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat.* 2009 Jan;30(1):69-78. DOI: 10.1002/humu.20822
99. KUMAR, Satish. et.al. Large scale mitochondrial sequencing in Mexican Americans suggests reappraisal of Native American origins. *BMC Evol Biol.* 2011 Oct 7;11:293. doi: 10.1186/1471-2148-11-293
100. KUMAR, Sudhir. et.al. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016 Jul;33(7):1870-4. doi: 10.1093/molbev/msw054.
101. LANGELLA, Oliver. 1999. <http://bioinformatics.org/populations/>
102. LELL, Jeffrey. et.al. Y chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes. *Hum Genet.* 1997 Oct;100(5-6):536-43.
103. LELL, Jeffrey. et.al. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet.* 2002 Jan;70(1):192-206
104. LEWIS, Cecile. et.al. Mitochondrial DNA and the peopling of South America. *Hum Biol.* 2007 Apr;79(2):159-78.
105. LOVO-GÓMEZ, J. et.al. The genetic male legacy from El Salvador. *Forensic Sci Int.* 2007;171(2-3):198-203. doi:10.1016/j.forsciint.2006.07.005

106. LÓPEZ, Carlos y RANERE, Anthony. Diversidad Cultural durante el Pleistoceno Tardío y el Holoceno Temprano en la Baja Centro América y en el Noroeste de Suramérica, En C. López & G. Ospina. Ecología Histórica: Interacciones Sociedad-Ambiente a Distintas Escalas Socio-Temporales. pp. 45-53. Pereira: Postergraph S.A. Universidad Tecnológica de Pereira, Universidad del Cauca, Sociedad Colombiana de Arqueología, 2008. p. 45-54.
107. LÓPEZ, Luis. Humedales de Córdoba: reconocimiento arqueológico y etnohistórico en los municipios de Pueblo Nuevo y Lorica. En RANGEL. Orlando. Colombia Diversidad Biótica IX. Bogotá D.C.: ARFO Editores e Impresores, (2007a) p 621-649.
108. LÓPEZ, Luis. Etnohistoria y Ocupaciones en la Vertiente Occidental de la Serranía de Perijá. En RANGEL. Orlando. Colombia Diversidad Biótica V. Bogotá D.C: ARFO Editores e Impresores Ltda, 2007b. p. 275-328.
109. LOPEZ, Carlos y CANO, Martha. En torno a los primeros poblamientos en el noroccidente de Sudamérica: acercamientos desde el valle interandino del Magdalena, Colombia. BOLETÍN DE ARQUEOLOGÍA PUCP N.º 15 / 2011, 43-79 / ISSN 1029-2004.
110. LLERAS, Roberto. Diferentes oleadas de poblamiento en la prehistoria tardía de los Andes Orientales. Boletín del Museo del Oro. Banco de la República No 38-39 (1995)
111. MACHADO, M. Un rastro del África Central en el Pacífico colombiano: tallas sagradas entre los indígenas Chocó y su legado africano (Congo y Angola). Afrorreparaciones: Memorias De La Esclavitud Y Justicia Reparativa Para Negros, Afrocolombianos, Y Raizales. En: Colombia ISBN: 978-958-8063-50-8 ed: Centro De Estudios Sociales Ces Universidad Nacional De Colombia, v. , p.1 - ,2007. <http://www.bdigital.unal.edu.co/1237/22/20CAPI19.pdf>
112. MANTA, Fernanda. et.al. Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. Ann Hum Biol. 2013 Jan;40 (1):94-8. doi: 10.3109/03014460.2012.742138.
113. MARCHECO-TERUEL, B. et.al., Cuba: Exploring the History of Admixture and the Genetic Basis of Pigmentation Using Autosomal and Uniparental Markers, PLoS Genet. 2014; Jul24;10(7): e1004488. doi: 0.1371/journal.pgen.1004488.
114. MARTINEZ-CORTES, G. et.al. Admixture and population structure in Mexican-Mestizos based on paternal lineages. J Hum Genet. 2012;57(9):568-574. doi:10.1038/jhg.2012.67
115. McFARLANE, A. Cimarrones y Palenques en Colombia: Siglo XVIII. Historia y Espacio No. 14, Junio 1991.

<http://bibliotecadigital.univalle.edu.co/bitstream/10893/7466/1/Cimarrones%20y%20palenques%20en%20Colombia%20-%20McFarlane%20Anthony.pdf>

116. MEISEL, A. y AGUILERA, M. Economía y medio ambiente del archipiélago de San Andrés, Providencia y Santa Catalina. Colección de Economía Regional Banco de la República. Bogotá: La imprenta editores S.A. 2016. ISBN 978-958-664-328-3. 277 p.  
[http://www.banrep.gov.co/docum/Lectura\\_finanzas/pdf/lbr\\_econ\\_med\\_amb\\_arch\\_san\\_andres\\_prov\\_san\\_cat.pdf](http://www.banrep.gov.co/docum/Lectura_finanzas/pdf/lbr_econ_med_amb_arch_san_andres_prov_san_cat.pdf)
117. MESA, Natalia. et.al. Autosomal, mtADN, and Y-chromosome diversity in amerinds: pre and post-columbian patterns of gene flow in South America. *American Journal of Human Genetics*. (2000) Vol. 67; p.1277-1286.
118. MELTON, P. et.al. Biological relationship between Central and South American chibchan speaking populations: evidence from mtADN. *En: American Journal of Physical Anthropology* [en línea]. Vol. 133 (2007); p. 753-770. [consultado 8 Dic. 2012] Disponible en <http://www.interscience.wiley.com/DOI/10.1002/ajpa.20581>
119. MINISTERIO DE CULTURA. Caracterización de los pueblos indígenas en riesgo. Bogotá D.C. 2013
120. MONCADA, Julie. Análisis de la estructura genética de una muestra de población del departamento del Amazonas, Colombia. Tesis Maestría. Universidad Nacional de Colombia. 2018.
121. MORENO-ESTRADA, Andrés. et.al. Reconstructing the Population Genetic History of the Caribbean. *PLoS Genet*. 2013 Nov;9(11):e1003925. doi: 10.1371/journal.pgen.1003925.
122. MORENO, F. et.al. SNP variation with latitude: Analysis of the SNPforID 52-plex markers in north, mid-region and south Chilean populations. *Forensic Sci Int Genet*. 2014 May;10:12-6. doi: 10.1016/j.fsigen.2013.12.009.
123. MULLIGAN, Connie. et.al. Updated three-stage model for the peopling of the Americas. *PLoS One*. 2008 Sep 17;3(9):e3199. doi: 10.1371/journal.pone.0003199.
124. MYRES, Natalie. et.al. Y-chromosome short tandem repeat intermediate variant alleles DYS385.2 delineate new phylogenetic substructure in human Y-chromosome haplogroup tree. *Croat Med J*. 2009 Jun;50(3):239-49.
125. NASSIR, Rami, KOSOY, Roman, TIAN, Chao, et.al. An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels. *BMC Genet*. 2009 Jul 24;10:39. doi: 10.1186/1471-2156-10-39.

126. NATIONAL GEOGRAPHIC e IBM. Proyecto Genográfico. En: <http://genographic.nationalgeographic.com>
127. NAVARRETE, C. Génesis y Desarrollo de la Esclavitud en Colombia Siglo XVI y XVII. Cali: Editorial Universidad del Valle, 2005.
128. NEI, Masatoshi. Genetic Distance between Populations. *The American Naturalist*. (May - Jun., 1972) Vol. 106, No. 949, pp. 283-292
129. NEI, M.; TAJIMA, F. y TATENO, Y. Accuracy of estimated phylogenetic trees from molecular data II. Gene frequency data. *J. Mol. Evol.* 1983 19:153-170.
130. NEI, Masatoshi. *Molecular evolutionary genetics*. New York: Columbia University Press; 1987.
131. NOGUERA, M.C., et.al. Colombia's racial crucible: Y chromosome evidence from six admixed communities in the department of bolivar. *Ann Hum Biol.* 2014; 41(5):453-459. doi:10.3109/03014460.2013.852244
132. NOGUERA, M. C. et.al. Mitochondrial DNA analysis suggests a Chibchan migration into Colombia. *Univ.Sci.*; 2015, Vol 20 (2):261-278.
133. OSORIO, J. Pueblos itinerantes de Urabá. La historia de las exclusiones. Retrato. Tesis Maestría en Historia Latinoamericana. Universidad Internacional de Andalucía, Sede Iberoamericana Santa María de la Rábida. 2006
134. OSSA, Humberto. et.al. Outlining the ancestry landscape of Colombian admixed population. *PLoS One.* 2016 Oct 13; 11 (10):e0164414. doi: 10.1371/journal.pone.0164414.
135. PACHON, Ximena y CAILLAVET, Chantal. *Frontera y poblamiento: estudios de historia y antropología de Colombia y Ecuador*. (Bogotá: Instituto Francés de Estudios Andinos, Instituto Amazónico de Investigaciones Científicas, Universidad de los Andes, 1996).
136. PARDO-SECO, Jacobo. et.al. Evaluating the accuracy of AIM panels at quantifying genome ancestry. *BMC Genomics.* 2014 Jun 30;15:543. doi: 10.1186/1471-2164-15-543.
137. PEREIRA, Rui. et.al. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One.* 2012;7(1):e29684. doi: 10.1371/journal.pone.0029684. Epub 2012 Jan 17.
138. PEREGO, Ugo. et.al. Distinctive Paleo-Indian Migration Routes from Beringia Marked by Two Rare mtDNA Haplogroups. *Current Biology.* 2009 January 13; 19: 1-8. DOI 10.1016/j.cub.2008.11.058

139. PEREGO, Ugo. et.al. Decrypting the Mitochondrial Gene Pool of Modern Panamanians. PLoS One. 2012;7(6):e38337. doi: 10.1371/journal.pone.0038337.
140. PITA PICO, Roger. Primeras incursiones de conquista por el rio grande de la Magdalena. En Banco de la Republica Biblioteca Virtual. Revista Credencial de Historia N° 283 Julio 2013. <http://www.banrepcultural.org/biblioteca-virtual/credencial-historia/numero-283/primeras-incursiones-de-conquista-por-el-rio-grande-de-la-magdalena>
141. PITBLADO, Bonnie. A Tale of Two Migrations: Reconciling Recent Biological and Archaeological Evidence for the Pleistocene Peopling of the Americas. Journal of Archaeological Research. (2011); Vol 19 (4): 327-375.
142. PHILLIPS, Christopher. et.al. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet. 2007 Dec;1(3-4):273-80. doi: 10.1016/j.fsigen.2007.06.008.
143. PHILLIPS, Christopher, FONDEVILA, Manuel, LAREAU, Maria Victoria, A 34-plex autosomal SNP single base extension assay for ancestry investigations. Methods Mol Biol. 2012;830:109-26. doi: 10.1007/978-1-61779-461-2\_8
144. PHILLIPS, Chistopher. et.al. Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries. Forensic Sci Int Genet. 2013a May;7(3):359-66. doi: 10.1016/j.fsigen.2013.02.010.
145. PHILLIPS, Chistopher. et.al., Development of a novel forensic STR multiplex for ancestry analysis and extended identity testing. Electrophoresis. 2013b Apr;34(8):1151-62. doi: 10.1002/elps.201200621.
146. PHILLIPS, Christopher. et.al. Building a forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set. Forensic Sci Int Genet. 2014 Jul;11:13-25. doi: 10.1016/j.fsigen.2014.02.012.
147. PHILLIPS, Christopher. et. al. Tetra-allelic SNPs: Informative forensic markers compiled from public whole-genome sequence data. Forensic Sci Int Genet. 2015a Nov;19:100-106. doi: 10.1016/j.fsigen.2015.06.011.
148. PHILLIPS, Christopher. Forensic genetic analysis of bio-geographical ancestry. Forensic Sci Int Genet. 2015b Sep;18: 49-65. doi: 10.1016/j.fsigen.2015.05.012.
149. PHILLIPS, Christopher. et. al. Inference of ancestry in forensic analysis: Autosomal ancestry-informative marker sets. En Goodwin, William. Forensic DNA typing protocols. Preston, Lancashire, UK: School of forensic and applied sciences, University of Central Lancashire, 2016. pp 233-253.

150. POLITIS, Gustavo. La estructura del debate sobre el poblamiento de América. *Boletín de Arqueología. Fundación de Investigaciones Arqueológicas Nacionales.* (1999); Vol. 14 (2): 25-51.
151. PORRAS, Liliana. et.al. Genetic variability of the SNPforID 52-plex identification-SNP panel in Central West Colombia. *Forensic Sci Int Genet.* 2009 Dec;4(1):e9-10. doi: 10.1016/j.fsigen.2008.12.003.
152. PRICE, Alkes. et.al. A genomewide admixture map for Latino populations. *Am J Hum Genet.* 2007 Jun; 80 (6):1024-36.
153. PRITCHARD, Jonathan. et.al. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*; 2000 (Junio) 155: 945–959
154. PURPS, Josephine. et.al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci Int Genet.* 2014 Sep;12:12-23. doi: 10.1016/j.fsigen.2014.04.008.
155. RAMACHANDRAN, Sohini. et.al. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci U S A.* 2005 Nov 1;102(44):15942-7
156. RAMIREZ, Bernardo. Los dos Caribes colombianos. [cepensar.blogspot.com/2011/06/el-encuentro-de-dos-caribes-por.html](http://cepensar.blogspot.com/2011/06/el-encuentro-de-dos-caribes-por.html)
157. RANGEL, Hector. et.al. Haplotipos STRs en cromosomas y amerindios de poblaciones mexicanas: evidencia genética del origen biparental de la etnia huichol. En: CIVERA, Magali y HERRERA, Martha. *Estudios de antropología biológica Vol XIII.* Instituto de Investigaciones Antropológicas. 2007 México:585-599.
158. REGUEIRO, Maria. et.al. On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists. *Am J Phys Anthropol.* 2013 Mar;150(3):333-48.
159. REICH, David. et.al. Reconstructing Native American Population History. *Nature.* 2012 Aug 16; 488(7411): 370–374. doi: 10.1038/nature11258
160. REICHEL-DOLMATOFF, Gerardo. (1997). *Arqueología de Colombia. Un texto introductorio, Publicación digital en la página web de la Biblioteca Luis Ángel Arango del Banco de la República.* , Bogotá: Presidencia de la República, 1997.
161. RESQUE, R. et.al. Male lineages in Brazil: Intercontinental admixture and stratification of the European background. *PLoS One.* 2016;11(4):1-17. doi:10.1371/journal.pone.0152573



162. RIEUX, Adrien. et.al. Improved Calibration of the Human Mitochondrial Clock Using Ancient Genomes. *Mol. Biol. Evol.* 2014 31(10):2780–2792. doi:10.1093/molbev/msu222
163. RODRIGUEZ, José Vicente. Algunos aspectos metodológicos bioantropológicos relacionados con el poblamiento de América. *Revista Maguaré*. Departamento de Antropología Universidad Nacional de Colombia. (1987); Vol. 5 p. 9-40
164. RODRIGUEZ, José Vicente. La diversidad poblacional de Colombia en el tiempo y el espacio: estudio craneométrico. *Revista Academia Colombiana de Ciencias*. Septiembre (2007); Vol. 31 (120): 321-346.
165. RODRIGUEZ, José Vicente y VARGAS, Clemencia. Evolución y tamaño dental en poblaciones humanas de Colombia. *Revista Academia Colombiana de Ciencias*. (2010); Vol. 34 (133): 423-439.
166. ROEWER, L. et.al. Continent-Wide Decoupling of Y-Chromosomal Genetic Variation from Language and Geography in Native South Americans. *PLoS Genet.* 2013; 9(4). doi:10.1371/journal.pgen.1003460
167. ROJAS, Madelene. *et. al.* Structure analysis of the La Guajira-Colombia population: a genetic, demographic and genealogical overview. *Annals of human biology*. (2013); Vol. 40 (2): 119-131.
168. ROJAS, W. et.al. Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers. *Am J Phys Anthropol.* 2010; 143(1):13-20. doi:10.1002/ajpa.21270
169. ROSENBERG, Noah. et.al. Informativeness of Genetic Markers for Inference of Ancestry. *Am J Hum Genet.* 2003 Dec; 73(6):1402-22.
170. ROTHHAMMER, Francisco y DILLEHAY, Tom. The late pleistocene colonization of South America: an interdisciplinary perspective. *Annals of Human Genetics*. (2009); Vol 73 p. 540–549
171. RUEDA, José. *Historia de los censos en Colombia*. DANE. Imprenta Nacional de Colombia: Bogotá D.C., 2012
172. RUHLEN, M. (1991). *A Guide to the World's Languages* (Stanford Univ. Press).
173. RUIZ, Y. et.al. Analysis of the SNPforID 52-plex markers in four Native American populations from Venezuela. *Forensic Sci Int Genet.* 2012 Sep; 6(5):e142-5. doi: 10.1016/j.fsigen.2012.02.007.

174. RUIZ-LINARES, Andrés. et.al. Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7.342 Individuals. *PLoS Genet.* 2014 Sep 25;10 (9):e1004572. doi: 10.1371/journal.pgen.1004572.
175. RUIZ-NARVAEZ, E. et.al. Genetic variation of the Y chromosome in Chibcha-speaking Amerindians of Costa Rica and Panama. *Hum Biol.* 2005; 77(1):71-91. doi:10.1353/hub.2005.0032
176. RUSSO, Maria. et.al. Evaluación del número mínimo de marcadores para estimar ancestría individual en una muestra de la población argentina. *Revista del museo de antropología.* 2016: 9 (1): 49-56
177. SAINT PIERRE, Michelle. et.al. Arrival of Paleo-Indians to the Southern Cone of South America: New Clues from Mitogenomes. *PLoS One.* 2012; 7(12): e51311.
178. SAITOU, N y NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987 Jul;4(4):406-25.
179. SALZANO, Francisco y SANZ, Monica. Interethnic admixture and the evolution of Latin American populations. *Genet Mol Biol.* 2014 Mar; 37(1 Suppl): 151–170. doi: 10.1590/s1415-47572014000200003.
180. SANDOVAL, Karla. et.al. Y-Chromosome Diversity in Native Mexicans Reveals Continental Transition of Genetic Structure in the Americas. *Am J Phys Anthropol.* 2012 Jul;148(3):395-405. doi: 10.1002/ajpa.22062.
181. SANCHEZ, Juan. et.al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis.* 2006 May; 27(9):1713-24. doi: 10.1002/elps.200500671.
182. SANTOS, Carla. et.al. Completion of a worldwide reference panel of samples for an ancestry informative Indel assay. *Forensic Sci Int Genet.* 2015a Jul;17:75-80. doi: 10.1016/j.fsigen.2015.03.011.
183. SANTOS, Carla. Ancestry analysis in forensic genetics: new markers and methodology. Tesis doctoral. Universidade Santiago de Compostela. Santiago de Compostela, España, 2015b.
184. SANTOS, Fabricio. et.al. The central Siberian origin for native American Y chromosomes. *Am. J. Hum.Gen* 1999 Feb; 64(2): 619-628.
185. SANTOS, J. L. *Epidemiología Genética. Principios y Métodos.* Santiago de Chile: Editorial Mediterráneo Ltda, 2011. 233p.

186. SCHURR, Theodore. The peopling of the new world: perspectives from molecular anthropology. *Annual Review of Anthropology*. (2004); Vol 33: 551-583.
187. SCHROEDER, K. et.al. A private allele ubiquitous in the Americas. *Biol Lett*. 2007 Apr 22;3(2):218-23.
188. SCHROEDER, Hannes. et.al. Origins and genetic legacies of the Caribbean Taino. *Proc Natl Acad Sci USA*. 2018 Mar 6;115(10):2341-2346. doi: 10.1073/pnas.1716839115.
189. SEMINO, Ornella. et.al. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europa and later migratory events in the Mediterranean area. *Am J Hum Genet*. 2004 May;74(5):1023-34.
190. SILVA-ZOLEZZI, Irma. et.al. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico *Proc Natl Acad Sci USA*. 2009 May 26;106(21):8611-6. doi: 10.1073/pnas.0903045106.
191. STEFFLOVA, K. et.al. Dissecting the Within-Africa ancestry of populations of African descent in the Americas. *PLoS One*. 2011;6(1). doi:10.1371/journal.pone.0014495
192. TABOADA, Patricia. et. al. The genetic legacy of the pre-colonial period in contemporary Bolivians. *PLoS One*. 2013;8(3):e58980. doi: 10.1371/journal.pone.0058980.
193. TAMM, Erika. et.al. Beringian Standstill and Spread of Native American Founders. *PLoS One*. 2007 Sep 5;2(9):e829.
194. TARAZONA-SANTOS, Eduardo. et.al. Genetic Differentiation in South Amerindians Is Related to Environmental and Cultural Diversity: Evidence from the Y Chromosome. *Am J Hum Genet*. 2001 Jun; 68(6): 1485–1496.
195. TERREROS, Grace. 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 de Maestría. Universidad Javeriana. 2010
196. TORRES, M. et.al. A revertant of the major founder native American haplogroup C common in populations from northern South America. *Am. J. Hum. Biol*. 2006. 18:59-65.
197. TORRONI, A. et.al. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet*. 1993 Sep; 53(3): 563–590.

198. TORRONI, A. et.al Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. *Proc Natl Acad Sci U S A*. 1994 Feb 1; 91(3):1158-62.
199. TOSCANINI, Ulises. Analysis of Y-chromosome STRs in Chile confirms an extensive introgression of European male lineages in urban populations. *Forensic Sci Int Genet*. 2016 Mar; 21:76-80. doi: 10.1016/j.fsigen.2015.12.005.
200. UNDERHILL, Peter. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci U S A*. 1996 Jan 9; 93(1): 196–200.
201. URIBE, C. La Etnografía de la Sierra Nevada de Santa Marta y las Tierras Bajas Adyacentes. En Instituto Colombiano de Cultura Hispánica. *Geografía Humana de Colombia. Nordeste Indígena*. Santafé de Bogotá: Giro Editores Ltda. 1993. Pag. 9-214.
202. USME, Solangy. 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
203. vanOVEN, M. et.al. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 2009 Feb; 30(2): E386-94.
204. vanOVEN, M. et.al. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Hum Mutat*. 2014 Feb;35(2):187-91
205. VASQUEZ, Patricia; GARDE, Marisa y MARTINEZ, Begoña. Amplificación y tipado de marcadores genéticomoleculares de tipo microsatélites (STRs) autosómicos y del cromosoma Y. En: Curso On Line de Genética Forense. Universidad de Zaragoza. (2008). [Consultado Dic. 2008]. Disponible en <http://ebro2.unizar.es/genforense/curso>
206. WALLACE, Douglas y TORRONI, Antonio. American Indian prehistory as written in the mitochondrial DNA: a review. *Hum Biol*. 1992 Jun; 64(3):403-16.
207. WALLACE, Douglas. Mitochondrial DNA sequence variation in human evolution and disease. *Proc Natl Acad Sci U S A*. 1994 Sep 13; 91(19): 8739–8746.
208. WANG, S. et.al. Genetic variation and population structure in native American. *Plos Genetics* [en línea] Vol 3 No 11 (2007); p e185. [consultado 3 Mar. 2013] Disponible en <http://www.plosgenetics.org>
209. WANG, Sijia. et.al. Geographic Patterns of Genome Admixture in Latin American Mestizos. *PLoS Genet*. 2008 Mar 21; 4(3):e1000037. doi: 10.1371/journal.pgen.1000037.

210. WILLUWEIT, Sascha y ROEWER, Lutz. The new Y Chromosome Haplotype Reference Database. *Forensic Sci Int Genet.* 2015 Mar; 15:43-8. doi: 10.1016/j.fsigen.2014.11.024.
211. XAVIER, Catarina. et.al. Admixture and Genetic Diversity Distribution Patterns of Non-Recombining Lineages of Native American Ancestry in Colombian Populations. *PLoS One.* 2015 Mar 16; 10(3):e0120155. doi: 10.1371/journal.pone.0120155. eCollection 2015.
212. ZAMBRANO, Fabio. (2000). Poblamiento Prehispánico. En ABELLO, Alberto y GIAIMO, Silvana. *Poblamiento y Ciudades del Caribe Colombiano.* Bogotá D.C.: Editorial Gente Nueva Ltda, 2000. p 10-25.
213. ZARANTE, Ignacio. Marcadores genéticos de las comunidades indígenas en Colombia, visitadas por la expedición humana. En: INSTITUTO COLOMBIANO DE CULTURA HISPÁNICA. *Geografía humana de Colombia.* Tomo I. Variación biológica y cultural en Colombia.