ESTIMACIÓN DEL NIVEL DE SUB-ESTRUCTURA EN 6 POBLACIONES COLOMBIANAS MEDIANTE EL ANALISIS DE ANCESTRIA CON UN PANEL DE 46 MARCADORES INDELS-AIMS

SARA ZEA MONTOYA

Universidad Nacional de Colombia
Facultad de Medicina, Departamento de Morfología,
Bogotá, Colombia
2017.

ESTIMACIÓN DEL NIVEL DE SUB-ESTRUCTURA EN 6 POBLACIONES COLOMBIANAS MEDIANTE EL ANALISIS DE ANCESTRIA CON UN PANEL DE 46 MARCADORES INDELS-AIMS

SARA ZEA MONTOYA

Tesis o trabajo de investigación presentada como requisito parcial para optar al título de:
Magister en Genética Humana

Director

WILLIAM USAQUÉN MARTÍNEZ. PHD.
PROFESOR ASOCIADO
UNIVERSIDAD NACIONAL DE COLOMBIA

Universidad Nacional de Colombia
Facultad de Medicina, Departamento de Morfología,
Maestría en Genética Humana
Bogotá, Colombia
2017.

Dedicado a…

A mis padres y a Alejo, quienes siempre me han dado todo su amor y su apoyo incondicional aun desde la distancia.

AGRADECIMIENTOS

A mi director William Usaquén, por su colaboración, sus enseñanzas y su comprensión.

A la doctora Leonor Gusmão y las personas del laboratorio de Diagnósticos por DNA de la Universidade do Estado do Rio de Janeiro por recibirme con tanta hospitalidad y su apoyo en la parte experimental de este trabajo.

Un especial agradecimiento a mis compañeros del grupo de Genética de Poblaciones e identificación por su paciencia y colaboración.

A los profesores y compañeros de la maestría, en especial a Luz, Mauricio y Jenny por ser una grata compañía en este camino.

Finalmente quiero agradecer a mi amiga Isabel por sus consejos, a mis padres y a Alejo, mi compañero de vida, quienes fueron mi razón para esforzarme un poco más cada día y culminar esta etapa de mi vida y siempre estuvieron presentes acompañándome aun en los momentos difíciles.

RESUMEN

Actualmente es de gran interés lograr un acercamiento a la estructura genética del país debido a su diversidad étnica y cultural, hecho que se encuentra ligado a su ubicación geográfica y a los diferentes procesos de mezcla entre los distintos grupos poblacionales a lo largo del tiempo. Para abordar este enfoque, el presente estudio analiza un panel de 46 marcadores informativos de ancestría INDELS-AIMS en un total de 340 individuos de 6 muestras de población de Colombia, que se pueden entender como propias de 3 regiones biogeográficas colombianas de las 5 establecidas, y cuentan con su respectiva información genealógica y demográfica. Adicionalmente, se calcularon diferentes parámetros de genética descriptiva, componentes ancestrales y sub-estructura genética. Los resultados muestran una agrupación de las 6 muestras de población en lo que se plantea como 3 modelos tipo de poblaciones americanas, que concuerdan con diferentes datos analizados previamente por diferentes grupos que estudian el poblamiento suramericano.

ABSTRACT

Nowadays, it is of great interest to achieve an approach to the genetic structure of the country, due to its ethnic and cultural diversity, which is linked to its geographical location and the different processes of mixing between different population groups over time. To address this approach in this study we analyze a panel of 46 INDELS-AIMS ancestry information markers in a total of 340 individuals from 6 Colombian population samples, which can be understood as belonging to 3 Colombian biogeographical regions of the 5 established, and have their respective genealogical and demographic information. Different parameters of descriptive genetics, ancestral components and genetic sub-structure were analyzed. The results show a grouping of the 6 population samples in what is considered as 3 type models of American populations, which agree with different data previously analyzed by different groups that study the South American population.

CONTENIDO

INTRODUCCIÓN 9

1. EL ESTUDIO DE LA SUBESTRUCTURA GENETICA EN LAS POBLACIONES HUMANAS 10

RESUMEN 10

1.1 Perspectiva histórica del estudio de la subestructura genética 11

1.2 Marcadores moleculares utilizados en el análisis de estructura genética 12

1.3 Importancia del estudio de la estructura poblacional en Colombia 13

1.4 La subestructura genética abordada desde el estudio de ancestría en poblaciones humanas 17

CONCLUSIÓN 18

Bibliografía 19

2. COMPOSICIÓN ANCESTRAL DE 3 MODELOS TIPO DE POBLACIÓN COLOMBIANA 21

RESUMEN 21

INTRODUCCIÓN 22

MATERIALES Y MÉTODOS 23

Selección de muestras 23

Extracción de ADN genómico 24

Amplificación 25

Electroforesis capilar 25

Análisis de datos 25

RESULTADOS 26

Genotipificación 26

Análisis de diversidad y genética descriptiva 26

Análisis de sub-estructura genética 33

Estimación de componentes ancestrales 34

Análisis de componentes principales 36

Bibliografía 45

3. MARCADORES INDELS PARA ANLISIS DE ANCESTRIA VS IDENTIFICACION 47

INTRODUCCIÓN 47

3.1 Marcadores Indels (Insertion-Deletion) 48
3.2 Estudios Previos 49
3.3 Perspectivas 50
3.4 Conclusiones 52
Bibliografía 53

INTRODUCCIÓN

Es de gran interés lograr un acercamiento a la estructura genética del país dada la notable diversidad étnica y cultural de la población colombiana. Este hecho se puede atribuir tanto a su ubicación geográfica como al amplio grado de mezcla entre las poblaciones nativas, europeas y africanas que se ha presentado desde el periodo de la colonización. Teniendo en cuenta además que las causas de lo anterior son eventos como el comercio de esclavos provenientes de África y la reducción considerable de las poblaciones nativas, al igual que otras regiones en Latinoamérica (Salzano & Bortolini, 2002).

Así mismo, en la época actual se pueden observar actividades de desplazamiento por parte de múltiples poblaciones hacia otras regiones del país debido a diferentes factores, tales como la búsqueda de mejores oportunidades laborales o la evasión del conflicto armado que ha perdurado por más de 5 décadas, obligando al desarrollo conceptual de modelos migratorios diferentes al tradicional rural-urbano.

En el escenario poblacional y forense, la mayoría de estudios se fundamentan en otros trabajos preliminares o pilotos para la exploración inicial de la población y para la obtención de un primer set de frecuencias alélicas de diversos marcadores moleculares. Los marcadores STR son los más usados actualmente, sin embargo, en la práctica forense y poblacional se han desarrollado otros marcadores (ej. SNPs, INDELs), que pueden ser herramientas de gran utilidad dadas sus características como marcadores moleculares (Pereira et al., 2012).

En el presente estudio se parte de un conjunto de muestras que ya cuentan con datos adicionales como análisis de isonimia, genealogía y estudios previos con diferentes marcadores genéticos (STRs autosómicos, Cromosoma Y, etc.) y se pretende hacer un riguroso análisis de los componentes ancestrales de diferentes grupos de población colombiana.

1. EL ESTUDIO DE LA SUBESTRUCTURA GENETICA EN LAS POBLACIONES HUMANAS

RESUMEN

La población colombiana al ser analizada mediante diferentes marcadores moleculares puede presentar diferentes modelos de subestructura, los cuales son útiles para determinar diferentes linajes ancestrales, aspecto de gran importancia en Colombia debido a su complejo modelo histórico. En este sentido, es importante aclarar que la presencia de diferentes linajes no necesariamente es sinónimo de una población subestructurada, teniendo en cuenta los fenómenos migratorios y de mezcla entre individuos que se presentan en gran parte de las poblaciones actuales. La estimación de subestructura en las poblaciones humanas puede ser determinante en la práctica poblacional y forense para evaluar la pertinencia del uso de diferentes frecuencias alélicas en casos específicos entendidos desde su contexto poblacional. Por ejemplo, en un caso de filiación en Colombia generalmente se usa la misma distribución de frecuencias para individuos de diferentes procedencias, y en caso de presentarse subestructura, sería necesario emplear diferentes frecuencias para cada sub-población.

1.1 Perspectiva histórica del estudio de la subestructura genética

Sewall Green Wright (1889-1988) fue pionero en el estudio de la estructura en las poblaciones y sus implicaciones en la evolución. En su trabajo se muestra una estrecha relación entre la estructura demográfica y genética de una población, destacando en sus métodos los índices de fijación F, entre otros. (Slatkin, 1994).

Con los estadísticos F, Sewall Wright proporciona una relación completa de los procesos que conducen a la diferenciación genética dentro y entre las poblaciones, lo cual ha generado que en los últimos 50 años estos se conviertan en los estadísticos descriptivos más utilizados de la genética evolutiva y de poblaciones (Holsinger & Weir, 2009). Sin embargo, estos estimadores dependen mucho en su interpretación de los marcadores empleados, además no siempre guardan una relación directa con la condición de equilibrio de Hardy – Weinberg, es decir, encontrar desequilibrio no siempre implica subestructura y viceversa.

Los índices de fijación (F) describen la diferenciación genética en una población mediante discrepancias en las frecuencias alélicas dentro de esta. En una población se pueden analizar diferentes niveles de subestructura: Población total (t) respecto a una subpoblación (F_{st}), Población total respecto a una región (F_{rt}) y una región respecto a una subpoblación (F_{sr}) (Hartl & Grant, 2007a).

Actualmente se puede afirmar que una población presenta subestructura cuando posee una diferenciación genética en cuanto a las frecuencias alélicas entre las subpoblaciones. La diferenciación genética de una población, puede ser resultado tanto de procesos de selección natural, como de procesos aleatorios de transmisión de alelos de una generación a la siguiente o de diferente probabilidad en la frecuencia alélica entre los fundadores iniciales de la subpoblación (Hartl & Grant, 2007b).

Si una población presenta algún nivel de subestructura, se podrá evidenciar una reducción en la proporción de heterocigotos con respecto a la esperada, además de que la distribución de frecuencias alélicas y genotipos varía considerablemente entre subpoblaciones.

1.2 Marcadores moleculares utilizados en el análisis de estructura genética

Los estudios previos al uso de marcadores moleculares se han realizado mediante grupos sanguíneos y su distribución poblacional, dado su patrón de herencia mendeliana mediante sus 3 formas alélicas, A, B y O (Ramón & Manilla, n.d.). Sin embargo, con los avances de la tecnología se hizo notoria la predilección por el uso de marcadores con una mayor resolución, que, por sus diferentes características y mecanismos de herencia, ofrecen diferentes ventajas según el tipo de estudio que se desea llevar a cabo. Algunos de estos marcadores son los STRs, ADN mitocondrial, SNPs, INDELs, entre otros.

Actualmente, es común el análisis de marcadores uniparentales (mtDNA y Cromosoma Y), los cuales son útiles para obtener información de la ancestría materna y paterna por separado, pero esto solo representa una pequeña porción de todo el mapa genéticopoblacional (Saloum de Neves Manta et al., 2013).

Por otro lado, tanto la identificación humana como la genética de poblaciones se basa principalmente en el estudio de los STRs o SNPs en función de las características particulares de la investigación. No obstante, otros tipos de variación genética tales como polimorfismos de inserción/deleción (Indel) tienen un potencial considerable en el campo de la identificación, ya que pueden combinar las características deseables de ambos STRs y SNPs (Pereira et al., 2009).

Los polimorfismos de tipo Indel presentan características interesantes como marcadores genéticos; (1) se extienden por todo el genoma (autosomas y cromosomas sexuales), (2) el polimorfismo se deriva de un único evento mutacional (inserción o deleción), (3) muchos de ellos muestran diferencias significativas en las frecuencias alélicas entre grupos poblacionales geográficamente separados, de tal forma que pueden ser utilizados como marcadores informativos de ancestría (AIMS), (4) pueden ser analizados en amplicones cortos, lo cual supone una ventaja para la amplificación de muestras degradadas o con cierto grado de dificultad. (Da Costa Francez et al., 2012)

Los marcadores de tipo Indel son muy frecuentes en el genoma humano y presentan diversas ventajas para estudios de poblaciones y forenses, tales como bajas tasas de mutación respecto a los STRs, fácil interpretación, amplicones pequeños y la posibilidad de utilizar PCR multiplex. La gran capacidad de adaptación de los marcadores Indel para la amplificación de un bajo número de copias o de ADN degradado permite su uso como una interesante plataforma, tanto en casos forenses, como en el escenario poblacional. (Da Costa Francez et al., 2012).

Es importante tener en cuenta que los diversos tipos de marcadores genéticos nos brindan información distinta de las poblaciones en cuanto a su estructura genética, razón por la que los estadísticos informativos de diferenciación poblacional deben ser interpretados en función de estas características. Un ejemplo de lo anterior es que los marcadores STRs no necesariamente se consideran marcadores diferenciadores de poblaciones en el ámbito forense, y en los estudios realizados con estos marcadores se puede observar una distribución de frecuencias alélicas en campana.

En contraste, hablando de los marcadores INDEL-AIMs, encontramos una distribución diferente donde algunos alelos específicos se fijan en una población continental determinada y en otras no, o se fijan otros alelos diferentes, y gracias a esto podemos diferenciar entre poblaciones consideradas ancestrales o representativas de regiones continentales específicas (Pereira et al., 2012). En la mayoría de estudios realizados en Latinoamérica estas poblaciones ancestrales se refieren a las propias de Europa, África, América, y no tan frecuentemente del este de Asia.

1.3 Importancia del estudio de la estructura poblacional en Colombia

Colombia cuenta con 5 grandes regiones biogeográficas, que a su vez se subdividen en subregiones; la región Andina, Caribe, Costa Pacífica, Orinoquia y Amazonia (Rangel-ch, 2005), las cuales son definidas por diferentes factores climáticos y de suelo, entre otros.

Además, posee una población multicultural ampliamente diversa a nivel regional y étnico, teniendo en cuenta la existencia de 87 etnias indígenas, 4 grupos diferenciados de población afrocolombiana, los que se ubican en el corredor del pacífico colombiano, los raizales del Archipiélago de San Andrés Providencia y Santa Catalina, la comunidad de San Basilio de Palenque y la población que reside en las cabeceras municipales o en las grandes ciudades. También se hablan 64 lenguas amerindias, además de la lengua de los raizales del Archipiélago de San Andrés, Providencia y Santa Catalina, y la lengua criolla de las comunidades de San Basilio de Palenque (Departamento Administrativo Nacional de Estadistica (DANE), 2007).

Los primeros seres humanos que llegaron al territorio de la actual Colombia datan de aproximadamente 10.000 y 15.000 años. En el siglo XV existían tres grandes familias que poblaban Colombia; Los pueblos pertenecientes a la familia lingüística Caribe se ubicaban en la costa del mar Caribe, los Arawak en los ríos Caquetá, Amazonas y Putumayo, y los Chibchas en la Sierra Nevada de Santa Marta y en los altiplanos del centro del país. Esta última presentaba más pobladores y un significativo desarrollo en la agricultura, el uso de calendario, los jeroglíficos, y los rituales religiosos (Suárez & Quintin, 2004).

Diversos procesos que van desde la conquista española, hasta las migraciones actuales ligadas a la situación socio-política, han contribuido al desarrollo de la gran diversidad poblacional y cultural actual del país. Debido a esto, se cree que algunas poblaciones colombianas, a pesar de tener diferentes orígenes étnicos y geográficos, pueden presentar un nivel de subestructura genética considerablemente bajo. Además, algunos grupos aparentemente aislados en realidad presentan gradientes de mezcla con otros grupos étnicos presentes en el territorio colombiano, razón por la cual fueron de gran interés los 6 grupos poblacionales seleccionados en este estudio.

En el caso específico de la península de La Guajira, ubicada en el extremo norte de América del Sur, durante miles de años, las comunidades Wayuu se han establecido en esta área. Hoy en día, según DANE (2005), la comunidad Wayuu representa el 44,9% de la población en esta región y son el mayor grupo amerindio en Colombia (más de 144000 personas). Sin

embargo, a pesar de que esta población ha definido una identidad cultural, existen múltiples factores que han llevado a la convivencia entre los guajiros y migrantes que comparten el mismo territorio (Rojas, Alonso, Sarmiento, Eljach, & Usaquén, 2013).

También es importante el Archipiélago de San Andrés y Providencia, el cual se encuentra a 180 km de la costa de América Central, 400 km de Jamaica y 480km de la costa de Colombia y está formado por las islas de San Andrés, Providencia y Santa Catalina. Históricamente, este espacio ha sido poblado por diferentes grupos culturales a través del tiempo, tales como comunidades amerindias de centro América, colonos ingleses y esclavos africanos, lo cual da paso al surgimiento de la población raizal, que es el resultado del mestizaje de estos grupos poblacionales. Después de la independencia en 1821, las islas se convirtieron en parte de la nación colombiana (Alonso & Usaquén, 2013).

De igual forma se debe tener en cuenta la población del departamento del Chocó, en donde se encuentra ubicada la población más representativa de individuos afrodescendientes en Colombia. El departamento está localizado en el noroeste del país, en la región del Pacífico colombiano. Comprende las selvas del Darién y las cuencas de los ríos Atrato y San Juan. Es el único departamento de Colombia con costas en los océanos Pacífico y Atlántico. Durante la época previa a la conquista española el territorio estaba habitado por los pueblos Kuna o Tule en el Golfo de Urabá y Bajo Atrato, los Uaunana o Noanamaes en San Juan, y los Emberá, Baudoes o Citararaes del Alto Atrato y el Baudó (Sitio oficial de la Gobernación del departamento del Chocó, 2016).

En cuanto a las poblaciones consideradas como mestizas (o mezcladas), Tolima, se encuentra situado en el centro-occidente del país. El territorio estuvo habitado durante muchos años, antes de la llegada de Cristóbal Colón al continente, por razas de origen Caribe, llamadas por los españoles "Los Pijaos", que comprendían las tribus Putimaes y Yalcones, Pantágoras, Guarinoes y Marquetones, entre las cuales sobresalían los Panches. Sin embargo la zona comprendida entre Anserma, Toche, Cocora e Ibagué, estaba poblada por descendientes de la raza Quimbaya, diferentes a los Pijaos, los cuales presentaban predominantes rasgos

asiáticos y se entrenaban en el manejo de las armas desde la adolescencia (Sitio oficial de la Gobernación del departamento del Tolima, 2016).

Por otro lado, la historia del Huila está sostenida en el legado de la cultura San Agustín, entre los años 1000 AC y 1650 DC. La importancia de esta cultura ha sido reconocida por la UNESCO, confiriendo a los parques arqueológicos de San Agustín, Alto de Los Ídolos y Alto de Las Piedras la categoría de Patrimonio Cultural de la Humanidad. A la llegada de los conquistadores españoles en 1538, la región estaba poblada por numerosas comunidades indígenas que inicialmente se opusieron a la llegada de los colonos; en el norte habitaban los Pijaos, en el sur los Andaquíes y Yalcones, en el occidente los Paeces y en el oriente los Tamas (Sitio oficial de la Gobernación del departamento del Huila, 2016).

Finalmente, se puede hablar de Bogotá como una población típica mestiza, si bien sus primeros pobladores fueron los Muiscas, pertenecientes a la familia lingüística Chibcha. Cuando arribaron los conquistadores se calcula que había medio millón de indígenas de este grupo. Aunque Bogotá careció de un flujo importante de inmigrantes extranjeros, según los censos llevados a cabo en el siglo XIX, la población tuvo un crecimiento regular; en 1832 tenía 36.465 habitantes; en 1881, 84.723 habitantes y hacia finales del siglo casi 100.000. El crecimiento de la población se debió en parte a las reformas que ampliaron las fuentes de trabajo, gracias a que Bogotá ofrecía la posibilidad de trabajar en el comercio o de desempeñar diversas funciones. Este aumento trajo una considerable ampliación física de la ciudad (Sitio oficial Portal Bogotá, 2016).

En la actualidad, Bogotá se considera una ciudad altamente diversa poblacional y culturalmente, esto debido a un creciente número de oportunidades laborales y de mejor calidad de vida, razón por la cual una gran cantidad de individuos de otras regiones se establecen allí.

Estudios previos sugieren una estrecha relación entre la distancia geográfica y la diferenciación genética, así mismo, puede ser posible identificar algún grado de sub-

estructura en una población que se relaciona más con factores culturales que con la distancia entre las poblaciones que coexisten en una región (Rojas et al., 2013).

1.4 La subestructura genética abordada desde el estudio de ancestría en poblaciones humanas

En la actualidad es de gran interés científico el conocimiento de la composición ancestral de las poblaciones colombianas, dado que a pesar de que se han realizado diversos estudios en esta área, es poco lo que se conoce sobre la estructura genética y componentes ancestrales de estas poblaciones. En algunas ocasiones los resultados de los estudios pueden ser discordantes, debido al uso de diferente cantidad y tipo de marcadores y estrategias de muestreo (Saloum de Neves Manta et al., 2013).

El estudio de la variación genética mediante el uso de polimorfismos de ADN distribuidos a lo largo del genoma, ha permitido una mejor comprensión de la historia de la diversidad de poblaciones humanas (Pereira et al., 2009). El uso de estos marcadores también permite una mayor aproximación a la estimación de la ancestría, información muy útil para los estudios poblacionales en virtud de los datos relevantes que pueden proporcionar. Esta información esclarecería las dinámicas de la migración entre diferentes poblaciones, patrones de mezcla y la influencia de fuerzas de cambio evolutivo en la composición genética de las comunidades en cuestión (Saloum de Neves Manta et al., 2013).

Por lo tanto, se ha visto un creciente aumento en el interés por estudiar y comprender la ancestría de las poblaciones y usar esta información como una herramienta útil para estudios de asociación, geneticopoblacionales y forenses. El resultado esperado es la disminución de los posibles efectos negativos o sesgados que puedan generarse debido a una estratificación genética en la población que no ha sido considerada o ha sido subestimada.

CONCLUSIÓN

Es indiscutible la importancia de la determinación de subestructura genética en las poblaciones humanas, sin embargo, dados los diferentes puntos de vista y metodología de los grupos de trabajo y de los investigadores referente al estudio de la subestructura, en ocasiones se pueden encontrar opiniones polarizadas al respecto. No obstante, se cree que con este tipo de estudio se puede contribuir a la unificación de parámetros en cuanto al establecimiento de la importancia y las implicaciones de la estimación de los niveles de subestructura en las poblaciones, al brindar igual valor a los factores de la biología humana que se deben tener en cuenta para este oficio, tales como los aspectos biogeográficos y culturales de las poblaciones, su composición ancestral y la información que pueden suministrar los diferentes sets de marcadores utilizados en casos específicos. Así, el objetivo de este trabajo es tener la capacidad de crear una sinergia entre los elementos mencionados para comprender que estamos hablando de poblaciones dinámicas y complejas, sin olvidar que todos estos factores influyen en el desarrollo y comportamiento de las comunidades a través del tiempo.

Bibliografía

Alonso, L. A., & Usaquén, W. (2013). Y-chromosome and surname analysis of the native islanders of San Andrés and Providencia (Colombia). *HOMO- Journal of Comparative Human Biology*, *64*(1), 71–84. https://doi.org/10.1016/j.jchb.2012.11.006

Da Costa Francez, P. A., Rodrigues, E. M. R., De Velasco, A. M., & Dos Santos, S. E. B. (2012). Insertion-deletion polymorphisms-utilization on forensic analysis. *International Journal of Legal Medicine*, *126*(4), 491–496. https://doi.org/10.1007/s00414-011-0588-z

Departamento Administrativo Nacional de Estadistica (DANE). (2007). Colombia una nación multicultural. *Colombia Una Nación Multicultural, Su Diversidad Étnica*.https://doi.org/10.1371/journal.pgen.1004023

Hartl, D. L., & Grant, A. (2007a). Principles of Population Genetics. 4th Edition. *Sinauer Associates*, *14*(4), 544–545. https://doi.org/10.2980/1195-6860(2007)14[544b:POPG]2.0.CO;2

Hartl, D. L., & Grant, A. (2007b). *Principles of Population Genetics. 4th Edition*. *Sinauer Associates* (Vol. 14). https://doi.org/10.2980/1195-6860(2007)14[544b:POPG]2.0.CO;2

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature Reviews. Genetics*, *10*(September), 639–650. https://doi.org/10.1038/nrg2611

Ministerio de Cultura. (2011). Los árabes en colombia, 168.

Pereira, R., Phillips, C., Alves, C., Amorim, A., Carracedo, Á., & Gusmao, L. (2009). A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*, *30*(21), 3682–3690. https://doi.org/10.1002/elps.200900274

Pereira, R., Phillips, C., Pinto, N., Santos, C., dos Santos, S. E. B., Amorim, A., … Gusmão, L. (2012). Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS ONE*, *7*(1). https://doi.org/10.1371/journal.pone.0029684

Ramón, A., & Manilla, C. R. (n.d.). Estudio de mestizaje genético con marcadores ABO-Rh en 2 poblaciones de Veracruz ; antesala a un estudio genético de salud.

Rangel-ch, J. O. (2005). La biodiversidad de Colombia. *Palimpsestos*, *5*, 292–304. https://doi.org/10.1073/pnas.262413599

Rojas, M. Y., Alonso, L. a, Sarmiento, V. a, Eljach, L. Y., & Usaquén, W. (2013). Structure analysis of the La Guajira-Colombia population: a genetic, demographic and genealogical overview. *Annals of Human Biology*, *40*(2), 119–31. https://doi.org/10.3109/03014460.2012.748093

Saloum de Neves Manta, F., Pereira, R., Vianna, R., Rodolfo Beuttenmüller de Araújo, A., Leite Góes Gitaí, D., Aparecida da Silva, D., … Gusmão, L. (2013). Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. *PLoS ONE*, *8*(9), 1–11. https://doi.org/10.1371/journal.pone.0075145

Salzano, F. M., & Bortolini, M. C. (2002). The Evolution and Genetics of Latin American Populations, *1*, 532. https://doi.org/10.1017/CBO9780511666100

Sitio oficial de la Gobernación del departamento del Chocó. (2016). Consultado el 17 de Julio, 2016 de http://www.choco.gov.co/informacion_general.shtml#historia.

Sitio oficial de la Gobernación del departamento del Huila. (2016). Consultado el 17 de Julio, 2016 de http://www.huila.gov.co/conoce-el-huila/resena-historica.html.

Sitio oficial de la Gobernación del departamento del Tolima. (2016). Consultado el 17 de Julio, 2016 de http://www.tolima.gov.co/publicaciones.php?id=1306.

Sitio oficial Portal Bogotá. (2016). Consultado el 17 de Julio, 2016 de www.bogota.gov.co/ciudad/historia.

Slatkin, M. (1994). Flujo Génico y Estructura de las Poblaciones, 1–11.

Suárez, C. A., & Quintin, C. I. (2004). Suárez. In *"Colombia Andina". Colombia: Norma. pp. 3-4. ISBN 9580468060*.

2. COMPOSICIÓN ANCESTRAL DE 3 MODELOS TIPO DE POBLACIÓN COLOMBIANA

RESUMEN

En este trabajo se pretende analizar la estructura genética y composición ancestral de 6 muestras de población (San Andrés, Choco, Bogotá, La Guajira, Tolima y Huila) representativas de 3 regiones biogeográficas, con el fin de construir un modelo de la población colombiana basado en la hipótesis de la conformación de 3 grupos o modelos tipo de población suramericana, y complementar la información que se tiene hasta el momento de estas poblaciones con los datos obtenidos en este estudio. Las muestras de los 6 grupos estudiados han sido previamente analizadas mediante marcadores STR o marcadores de cromosoma Y; los resultados de estos análisis pueden representar una guía importante en el desarrollo de la hipótesis planteada.

INTRODUCCIÓN

Colombia se encuentra situado al noroccidente en América del Sur, tiene costas en el océano Pacífico y en el Atlántico, en el mar Caribe se ubica el archipiélago de San Andrés, Providencia y Santa Catalina (Oficina de Información Diplomatica del Departamento de Relaciones Exteriores, 2015).

Puede afirmarse que la mayor parte de su población se deriva de la mezcla poblacional entre los grupos indígenas nativos, colonos europeos y esclavos africanos, aunque también se encuentran presentes grupos afro descendientes e indígenas en algunas regiones, y es importante no desconocer la presencia de linajes de origen asiático en algunas regiones del país, por ejemplo, los grupos árabes, principalmente sirios, libaneses y palestinos que migraron a Colombia a finales del siglo XIX (Ministerio de Cultura, 2011).

La conquista de América trajo como consecuencia la ocupación del territorio de algunos pueblos que habitaban la zona, la imposición de algunas características de la cultura europea, así como el desplazamiento y una reducción considerable de las poblaciones nativas, tanto física como culturalmente (Departamento Administrativo Nacional de Estadistica (DANE), 2007).

Estos acontecimientos históricos han contribuido al desarrollo de los grupos poblacionales actuales en Colombia, dando como resultado una población ampliamente diversa, tanto étnica como culturalmente. Del mismo modo, se pueden observar diferentes actividades migratorias recientes dentro del territorio nacional.

Los resultados de este estudio pueden proporcionar información valiosa sobre su estructura genética actual, para la comprensión tanto de patrones de mezcla dentro de la población, como eventos de migración y colonización en el territorio colombiano.

MATERIALES Y MÉTODOS

Selección de muestras

Las muestras fueron seleccionadas del banco de muestras del grupo de Genética de Poblaciones e Identificación del Instituto de Genética de la Universidad Nacional de Colombia, las cuales fueron previamente procesadas y tipificadas para STRs y Cromosoma Y. Cabe recalcar que cada muestra cuenta con el consentimiento informado correspondiente y con un análisis de genealogía, realizado mediante una encuesta destinada a verificar diferentes parámetros como: tiempo de residencia, información genealógica y demográfica.

Se procesaron 50 individuos de Bogotá pertenecientes a casos de paternidad de rutina del Grupo de Genética de Poblaciones e Identificación, 73 de Choco, 37 de Tolima, 39 de Huila, 50 de San Andrés, y 91 muestras de la población de La Guajira. Estos últimos fueron clasificados específicamente como Wayuu mediante la información genealógica y demográfica recopilada en la encuesta. En total se analizaron 340 individuos no relacionados.

Los grupos poblacionales seleccionados se consideran como representativos de los diferentes posibles modelos de población que podemos encontrar en Colombia, tomando los individuos de San Andrés y Chocó como representativos de la población afrodescendiente; los individuos de Bogotá, Tolima y Huila se consideran representativos de poblaciones típicas mestizas o mezcladas; y los individuos de La Guajira, específicamente clasificados como Wayuu a partir de la información recolectada en la encuesta, serán considerados representativos de la población indígena.

El procesamiento de las muestras fue realizado en el Laboratorio de Diagnósticos por DNA (LDD) del Instituto de Biología de la Universidad del Estado de Rio de Janeiro y en el laboratorio del Grupo de Genética de Poblaciones e Identificación del Instituto de Genética de la Universidad Nacional de Colombia (GPI).

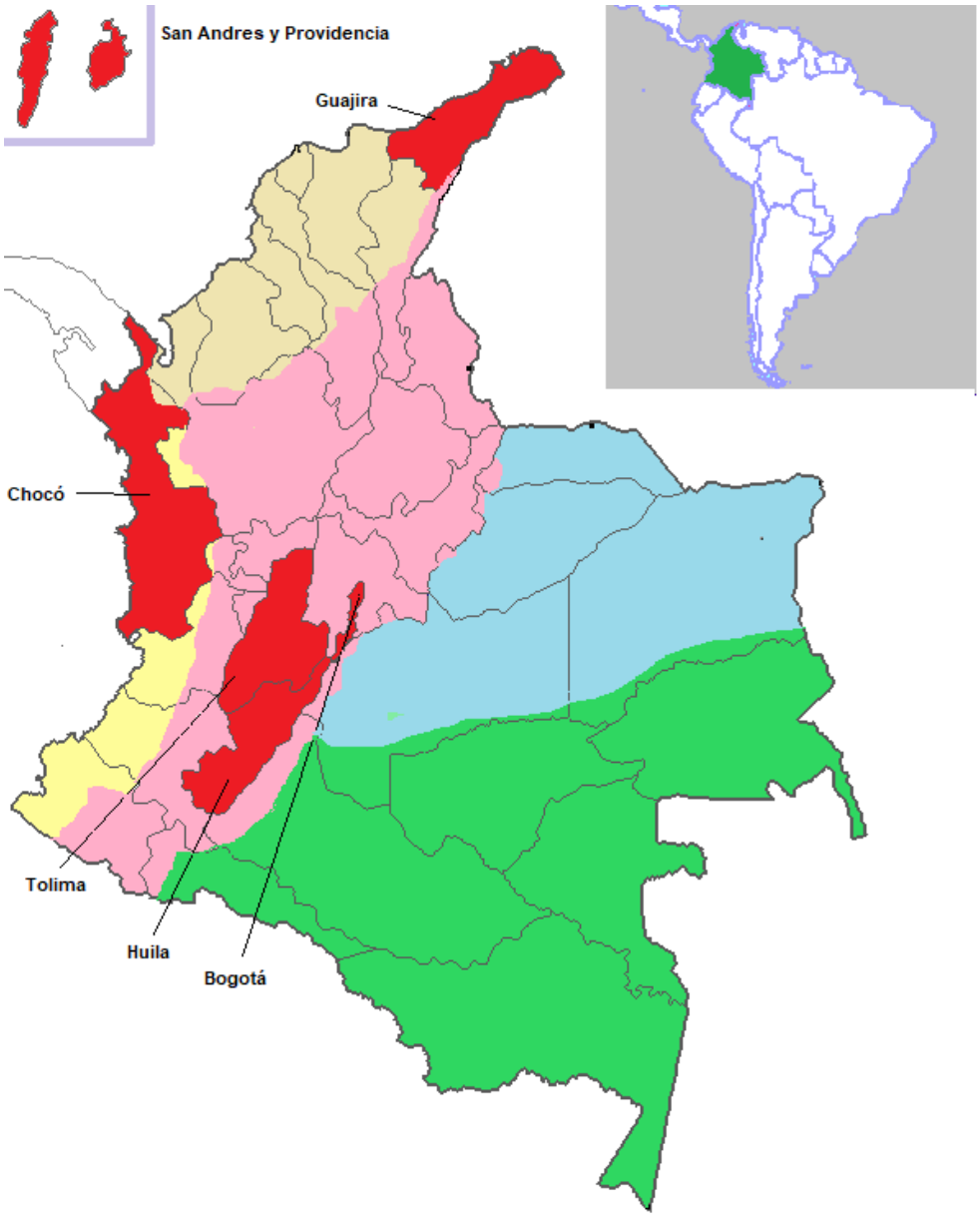

Figura 1. Ubicación biogeográfica del origen de las muestras analizadas.

Extracción de ADN genómico

Las muestras procesadas son muestras de sangre periférica y células epiteliales de mucosa bucal recolectadas en tarjetas FTA™ WHATMAN®. La extracción de ADN se realizó mediante protocolos de Chelex y buffer FTA estandarizados en cada laboratorio.

Amplificación

Se realizó una PCR multiplex con un panel de 46 marcadores INDEL descritos en el trabajo de Pereira et al., 2012, con las siguientes condiciones experimentales:

Reactivo	Volumen
Qiagen Multiplex Master Mix	2.5µL
Primer Set 46 Primers	0.5µL
Primer Set Refuerzo Primers	0.5µL
Agua	0.5µL
ADN (Chelex)/FTA	1.0µL/1 punch
Total	5.0µL

Tabla 1. Volúmenes utilizados para la amplificación de 46 marcadores INDEL.

Electroforesis capilar

Este proceso se realizó en un analizador genético ABI 3500 (Life Technologies) con protocolo estandarizado en el LDD. La recolección y análisis de datos se hizo mediante el software Genemapper v.4.1 (Life Technologies). En el GPI se realizó en un analizador genético ABI Prism 310 y software Genemapper ID v.3.2 (Life Technologies).

Análisis de datos

Se calcularon frecuencias alélicas, equilibrio de Hardy Weinberg, heterocigocidad observada y esperada y Fst mediante análisis en el software Arlequín v. 3.5.2.2 (Excoffier, L. and Lischer, 2010) y Genepop v4.6 (Rousset, 2017).

Se estimaron las proporciones de componentes ancestrales poblacionales e individuales y posibles patrones de mezcla en las poblaciones analizadas con STRUCTURE v.2.3.4 (Pritchard, Wen, & Falush, 2010), (Porras-Hurtado et al., 2013) incluyendo datos de HGDP-CEPH (Human Genome Diversity Cell Line Panel) de muestras de referencia de poblaciones africanas, europeas, del este de Asia y Nativo americanas (Pereira et al., 2012), para facilitar

la interpretación de estos resultados se realizó un análisis de componentes principales con el software MVSP v.3.22 (Kovach Computing Services).

RESULTADOS

Genotipificación

Se obtuvieron perfiles completos para la mayoría de las muestras, excepto 3 de ellas, lo cual puede deberse a la calidad y antigüedad de las mismas, ya que son muestras de células epiteliales de mucosa bucal tomadas en el año 2009, una de estas 3 muestras fue excluida de los análisis pues no se lograron obtener 17 de los 46 marcadores.

Análisis de diversidad y genética descriptiva

Se calcularon frecuencias alélicas para los 46 marcadores en cada uno de los seis grupos poblacionales analizados mediante Genepop v4.6. Los indels son marcadores bi-alelicos, sin embargo, en algunas investigaciones previas se ha descrito un tercer alelo en algunos locus de los 46 que componen el panel (Pereira et al., 2012). En el presente estudio también se detecta el alelo 3 en algunas de las poblaciones en los marcadores 2264, 159 y 360, los resultados se relacionan en la tabla 2.

Los análisis de equilibrio de Hardy Weinberg, heterocigocidad observada y esperada y estadisticos Fst mediante análisis en el software Arlequín v. 3.5.2.2

	Chocó			Bogotá			Huila			San Andres			Tolima			La Guajira		
Locus	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1470	0.226	0.774	-	0.270	0.730	-	0.316	0.684	-	0.240	0.760	-	0.368	0.632	-	0.077	0.923	-
777	0.418	0.582	-	0.570	0.430	-	0.579	0.421	-	0.460	0.540	-	0.539	0.461	-	0.824	0.176	-
196	0.541	0.459	-	0.270	0.730	-	0.316	0.684	-	0.540	0.460	-	0.395	0.605	-	0.050	0.950	-
881	0.534	0.466	-	0.900	0.100	-	0.895	0.105	-	0.540	0.460	-	0.868	0.132	-	0.984	0.016	-
3122	0.753	0.247	-	0.970	0.030	-	0.947	0.053	-	0.700	0.300	-	0.974	0.026	-	0.984	0.016	-
548	0.384	0.616	-	0.160	0.840	-	0.197	0.803	-	0.300	0.700	-	0.158	0.842	-	0.154	0.846	-
659	0.308	0.692	-	0.200	0.800	-	0.145	0.855	-	0.250	0.750	-	0.211	0.789	-	0.172	0.828	-
2011	0.432	0.568	-	0.830	0.170	-	0.789	0.211	-	0.490	0.510	-	0.776	0.224	-	0.874	0.126	-
2929	0.815	0.185	-	0.830	0.170	-	0.776	0.224	-	0.830	0.170	-	0.776	0.224	-	0.984	0.016	-
593	0.548	0.452	-	0.010	0.990	-	0.079	0.921	-	0.440	0.560	-	0.039	0.961	-	0.017	0.983	-
798	0.267	0.733	-	0.440	0.560	-	0.368	0.632	-	0.310	0.690	-	0.355	0.645	-	0.278	0.722	-
1193	0.151	0.849	-	0.180	0.820	-	0.079	0.921	-	0.210	0.790	-	0.171	0.829	-	0.129	0.871	-
1871	0.123	0.877	-	0.360	0.640	-	0.250	0.750	-	0.170	0.830	-	0.395	0.605	-	0.396	0.604	-
17	0.740	0.260	-	0.610	0.390	-	0.539	0.461	-	0.670	0.330	-	0.671	0.329	-	0.889	0.111	-
2538	0.055	0.945	-	0.260	0.740	-	0.329	0.671	-	0.190	0.810	-	0.250	0.750	-	0.061	0.939	-
1644	0.329	0.671	-	0.580	0.420	-	0.592	0.408	-	0.510	0.490	-	0.684	0.316	-	0.333	0.667	-
3854	0.473	0.527	-	0.050	0.950	-	0.118	0.882	-	0.420	0.580	-	0.171	0.829	-	0.022	0.978	-
2275	0.370	0.630	-	0.160	0.840	-	0.250	0.750	-	0.380	0.620	-	0.276	0.724	-	0.352	0.648	-
94	0.137	0.863	-	0.400	0.600	-	0.434	0.566	-	0.180	0.820	-	0.487	0.513	-	0.610	0.390	-
3072	0.315	0.685	-	0.910	0.090	-	0.961	0.039	-	0.470	0.530	-	0.934	0.066	-	0.973	0.027	-
772	0.952	0.048	-	0.85	0.15	-	0.882	0.118	-	0.94	0.06	-	0.868	0.132	-	0.806	0.194	-
2313	0.315	0.685	-	0.49	0.51	-	0.5	0.5	-	0.17	0.83	-	0.5	0.5	-	0.703	0.297	-
397	0.664	0.336	-	0.68	0.32	-	0.579	0.421	-	0.73	0.27	-	0.658	0.342	-	0.456	0.544	-
1636	0.404	0.596	-	0.85	0.15	-	0.829	0.171	-	0.52	0.48	-	0.789	0.211	-	0.983	0.017	-
51	0.247	0.753	-	0.63	0.37	-	0.724	0.276	-	0.31	0.69	-	0.776	0.224	-	0.856	0.144	-
2431	0.055	0.945	-	0.19	0.81	-	0.145	0.855	-	0.07	0.93	-	0.132	0.868	-	0.17	0.83	-
2264	0.226	0.753	0.021	0.28	0.67	0.05	0.303	0.684	0.013	0.38	0.62	-	0.395	0.592	0.013	0.101	0.899	-
2256	0.096	0.904	-	0.46	0.54	-	0.461	0.539	-	0.18	0.82	-	0.382	0.618	-	0.769	0.231	-
128	0.068	0.932	-	0.35	0.65	-	0.303	0.697	-	0.13	0.87	-	0.289	0.711	-	0.159	0.841	-
15	0.363	0.637	-	0.59	0.41	-	0.632	0.368	-	0.43	0.57	-	0.697	0.303	-	0.874	0.126	-
2241	0.616	0.384	-	0.2	0.8	-	0.171	0.829	-	0.58	0.42	-	0.092	0.908	-	0.06	0.94	-
419	0.932	0.068	-	0.78	0.22	-	0.645	0.355	-	0.82	0.18	-	0.789	0.211	-	0.544	0.456	-
943	0.432	0.568	-	0.53	0.47	-	0.526	0.474	-	0.39	0.61	-	0.592	0.408	-	0.302	0.698	-
159	0.678	0.315	0.007	0.59	0.4	0.01	0.579	0.421	-	0.73	0.27	-	0.579	0.421	-	0.578	0.25	0.172
2005	0.199	0.801	-	0.44	0.56	-	0.355	0.645	-	0.29	0.71	-	0.539	0.461	-	0.412	0.588	-
250	0.651	0.349	-	0.54	0.46	-	0.526	0.474	-	0.71	0.29	-	0.553	0.447	-	0.388	0.612	-
1802	0.356	0.644	-	0.03	0.97	-	0.053	0.947	-	0.35	0.65	-	0.092	0.908	-	0.135	0.865	-
1607	0.452	0.548	-	0.35	0.65	-	0.395	0.605	-	0.4	0.6	-	0.461	0.539	-	0.545	0.455	-
1734	0.671	0.329	-	0.53	0.47	-	0.592	0.408	-	0.7	0.3	-	0.671	0.329	-	0.577	0.423	-
406	0.288	0.712	-	0.67	0.33	-	0.724	0.276	-	0.25	0.75	-	0.789	0.211	-	0.714	0.286	-
1386	0.233	0.767	-	0.35	0.65	-	0.553	0.447	-	0.18	0.82	-	0.447	0.553	-	0.604	0.396	-
1726	0.459	0.541	-	0.68	0.32	-	0.592	0.408	-	0.39	0.61	-	0.763	0.237	-	0.802	0.198	-
3626	0.171	0.829	-	0.53	0.47	-	0.461	0.539	-	0.3	0.7	-	0.395	0.605	-	0.099	0.901	-
360	0.568	0.377	0.055	0.72	0.28	-	0.684	0.316	-	0.52	0.42	0.06	0.724	0.276	-	0.566	0.429	0.005
1603	0.219	0.781	-	0.62	0.38	-	0.487	0.513	-	0.25	0.75	-	0.684	0.316	-	0.803	0.197	-
2719	0.466	0.534	-	0.44	0.56	-	0.447	0.553	-	0.37	0.63	-	0.553	0.447	-	0.528	0.472	-

Tabla 2. Frecuencias alélicas para 46 marcadores en 6 muestras de población.

Los valores más bajos de heterocigocidad fueron encontrados en la muestra de población Wayuu, lo cual se esperaba dado que es el grupo que se considera más conservado, teniendo en cuenta la información de la encuesta para todos los individuos de la misma, en la cual en su totalidad tienen ascendencia Wayuu en las 3 generaciones anteriores, y esto reafirma la hipótesis de que a pesar de que no son una comunidad completamente aislada, han sido menos expuestos a mezcla con otros grupos respecto a las otras 5 poblaciones.

En el análisis de equilibrio de Hardy-Weinberg no se encontraron diferencias estadísticamente significativas en la mayoría de los 46 locus en las 6 poblaciones. En aquellos marcadores en los cuales sí se encontró un valor p significativo, pueden influir diferentes factores, como el tamaño de la muestra o al poco número de heterocigotos en algunos marcadores. Estos resultados se presentan en tabla 3, en la cual se muestran en amarillo los locus con p valor significativo

Este hecho puede deberse a su vez al mecanismo propio de los marcadores Indels como marcadores informativos de ancestría, en los cuales algunos alelos deben estar fijados en ciertas poblaciones ancestrales para; a partir de estas diferencias en las frecuencias alélicas, finalmente inferir una ancestría en grupos de población diferentes (Da Costa Francez et al., 2012).

Para observar esto detalladamente, se presentan las figuras 2.A y 2.B. En estas se plantean dos ejemplos de la distribución de frecuencias alélicas, el primero (2.A) para uno de los marcadores (593), en este se evidencia claramente el comportamiento de las frecuencias para las diferentes poblaciones; y el segundo (2.B), contemplando las frecuencias los 46 marcadores para una sola población, en este caso Bogotá. Respecto a la segunda gráfica (2B), se realizó el mismo ejercicio para las 6 poblaciones y los resultados fueron similares.

	CHO			BOG			HUI			SAN			TOL			GUA		
Locus	Obs	Exp	P-val	Obs	Exp	P-val	Obs	Exp	P-val	Obs	Exp	P-val	Obs	Exp	P-val	Obs	Exp	P-val
1470	0.315	0.352	0.500	0.340	0.398	0.303	0.474	0.438	0.718	0.400	0.368	0.705	0.474	0.472	1.000	0.132	0.143	0.418
777	0.397	0.490	0.149	0.460	0.495	0.773	0.368	0.494	0.180	0.400	0.502	0.165	0.500	0.504	1.000	0.330	0.291	0.288
196	0.507	0.500	1.000	0.380	0.398	0.730	0.474	0.438	0.717	0.440	0.502	0.407	0.474	0.484	1.000	0.100	0.096	1.000
881	0.438	0.501	0.348	0.200	0.182	1.000	0.158	0.191	0.335	0.560	0.502	0.568	0.263	0.232	1.000	0.033	0.033	1.000
3122	0.411	0.374	0.532	0.060	0.059	1.000	0.105	0.101	1.000	0.480	0.424	0.501	0.053	0.052	1.000	0.033	0.033	1.000
548	0.438	0.476	0.621	0.280	0.272	1.000	0.342	0.321	1.000	0.400	0.424	0.741	0.316	0.269	0.563	0.286	0.262	0.684
659	0.342	0.429	0.100	0.320	0.323	1.000	0.184	0.251	0.147	0.300	0.379	0.149	0.316	0.337	0.649	0.300	0.287	1.000
2011	0.562	0.494	0.339	0.260	0.285	0.612	0.368	0.337	1.000	0.420	0.505	0.265	0.395	0.352	0.652	0.209	0.222	0.628
2929	0.315	0.304	1.000	0.300	0.285	1.000	0.237	0.352	0.059	0.300	0.285	1.000	0.289	0.352	0.347	0.033	0.033	1.000
593	0.329	0.499	0.004	0.020	0.020	1.000	0.158	0.147	1.000	0.400	0.498	0.248	0.079	0.077	1.000	0.033	0.033	1.000
798	0.425	0.394	0.564	0.560	0.498	0.402	0.579	0.472	0.180	0.420	0.432	1.000	0.605	0.464	0.078	0.467	0.403	0.187
1193	0.274	0.258	1.000	0.240	0.298	0.173	0.158	0.147	1.000	0.300	0.335	0.424	0.237	0.287	0.275	0.258	0.226	0.346
1871	0.247	0.218	0.586	0.360	0.465	0.130	0.395	0.380	1.000	0.300	0.285	1.000	0.316	0.484	0.043	0.418	0.481	0.271
17	0.438	0.388	0.363	0.460	0.481	0.774	0.447	0.504	0.526	0.460	0.447	1.000	0.289	0.447	0.060	0.156	0.199	0.069
2538	0.082	0.104	0.183	0.440	0.389	0.469	0.553	0.447	0.265	0.300	0.311	1.000	0.500	0.380	0.078	0.122	0.115	1.000
1644	0.329	0.444	0.033	0.360	0.492	0.082	0.447	0.489	0.737	0.420	0.505	0.267	0.421	0.438	1.000	0.444	0.447	1.000
3854	0.562	0.502	0.352	0.100	0.096	1.000	0.237	0.212	1.000	0.440	0.492	0.563	0.237	0.287	0.273	0.022	0.043	0.033
2275	0.438	0.469	0.619	0.240	0.272	0.589	0.447	0.380	0.399	0.480	0.476	1.000	0.447	0.405	0.691	0.330	0.459	0.010
94	0.247	0.238	1.000	0.400	0.485	0.248	0.395	0.498	0.320	0.320	0.298	1.000	0.447	0.506	0.528	0.407	0.478	0.186
3072	0.384	0.435	0.414	0.180	0.165	1.000	0.079	0.077	1.000	0.580	0.503	0.392	0.079	0.125	0.131	0.055	0.054	1.000
772	0.096	0.092	1.000	0.260	0.258	1.000	0.237	0.212	1.000	0.120	0.114	1.000	0.263	0.232	1.000	0.300	0.315	0.735
2313	0.438	0.435	1.000	0.500	0.505	1.000	0.579	0.507	0.515	0.340	0.285	0.322	0.474	0.507	0.751	0.484	0.420	0.209
397	0.425	0.449	0.793	0.440	0.440	1.000	0.474	0.494	1.000	0.300	0.398	0.143	0.316	0.456	0.073	0.451	0.499	0.400
1636	0.370	0.485	0.053	0.260	0.258	1.000	0.289	0.287	1.000	0.560	0.504	0.571	0.368	0.337	1.000	0.034	0.033	1.000
51	0.329	0.374	0.344	0.500	0.471	0.765	0.447	0.405	0.691	0.380	0.432	0.506	0.289	0.352	0.346	0.222	0.249	0.381
2431	0.110	0.104	1.000	0.380	0.311	0.175	0.289	0.251	1.000	0.140	0.132	1.000	0.263	0.232	1.000	0.297	0.284	1.000
2264	0.397	0.383	0.782	0.480	0.475	1.000	0.447	0.446	0.389	0.480	0.476	1.000	0.421	0.500	0.309	0.202	0.183	0.592
2256	0.164	0.175	0.497	0.520	0.502	1.000	0.605	0.504	0.327	0.320	0.298	1.000	0.395	0.478	0.316	0.352	0.357	1.000
128	0.137	0.128	1.000	0.500	0.460	0.755	0.447	0.428	1.000	0.260	0.228	0.581	0.474	0.417	0.460	0.297	0.269	0.454
15	0.479	0.466	1.000	0.540	0.489	0.561	0.579	0.472	0.183	0.460	0.495	0.772	0.395	0.428	0.706	0.231	0.222	1.000
2241	0.411	0.476	0.322	0.280	0.323	0.380	0.237	0.287	0.274	0.520	0.492	0.774	0.132	0.169	0.262	0.121	0.114	1.000
419	0.110	0.128	0.281	0.400	0.347	0.417	0.447	0.464	1.000	0.240	0.298	0.172	0.368	0.337	1.000	0.560	0.499	0.288
943	0.425	0.494	0.244	0.500	0.503	1.000	0.421	0.505	0.342	0.500	0.481	1.000	0.447	0.489	0.738	0.407	0.424	0.804
159	0.425	0.444	0.854	0.620	0.497	0.103	0.526	0.494	0.746	0.380	0.398	0.732	0.474	0.494	1.000	0.589	0.577	0.000
2005	0.315	0.321	1.000	0.440	0.498	0.565	0.553	0.464	0.298	0.380	0.416	0.730	0.658	0.504	0.099	0.560	0.487	0.193
250	0.589	0.458	0.020	0.440	0.502	0.407	0.263	0.505	0.004	0.420	0.416	1.000	0.632	0.501	0.186	0.438	0.477	0.502
1802	0.411	0.462	0.444	0.060	0.059	1.000	0.105	0.101	1.000	0.420	0.460	0.549	0.184	0.169	1.000	0.247	0.235	1.000
1607	0.411	0.499	0.156	0.460	0.460	1.000	0.526	0.484	0.736	0.520	0.485	0.768	0.500	0.504	1.000	0.506	0.499	1.000
1734	0.411	0.444	0.596	0.540	0.503	0.776	0.395	0.489	0.314	0.360	0.424	0.319	0.395	0.447	0.482	0.648	0.491	0.003
406	0.356	0.413	0.263	0.460	0.447	1.000	0.395	0.405	1.000	0.380	0.379	1.000	0.316	0.337	0.648	0.352	0.410	0.198
1386	0.329	0.360	0.512	0.460	0.460	1.000	0.421	0.501	0.346	0.280	0.298	0.642	0.632	0.501	0.187	0.484	0.481	1.000
1726	0.534	0.500	0.640	0.440	0.440	1.000	0.500	0.489	1.000	0.500	0.481	1.000	0.368	0.366	1.000	0.374	0.319	0.178
3626	0.233	0.286	0.202	0.500	0.503	1.000	0.395	0.504	0.206	0.320	0.424	0.098	0.421	0.484	0.501	0.176	0.179	1.000
360	0.493	0.536	0.308	0.400	0.407	1.000	0.421	0.438	1.000	0.540	0.555	0.144	0.395	0.405	1.000	0.396	0.499	0.041
1603	0.274	0.345	0.092	0.440	0.476	0.763	0.605	0.506	0.328	0.460	0.379	0.251	0.474	0.438	0.717	0.326	0.318	1.000
2719	0.384	0.501	0.059	0.560	0.498	0.401	0.421	0.501	0.344	0.580	0.471	0.131	0.526	0.501	1.000	0.522	0.501	0.832
Total	**0.359**	**0.383**	-	**0.377**	**0.379**	-	**0.383**	**0.387**	-	**0.397**	**0.406**	-	**0.376**	**0.381**	-	**0.299**	**0.297**	-

Tabla 3. Heterocigocidad observada y esperada con respectivo p-valor para los 46 locus.

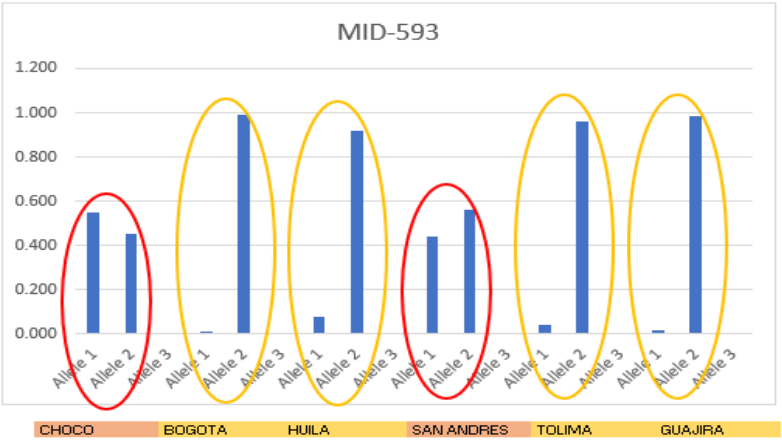

Fig 2.A. Distribución de frecuencias alélicas en el marcador 593 para los 6 grupos analizados.

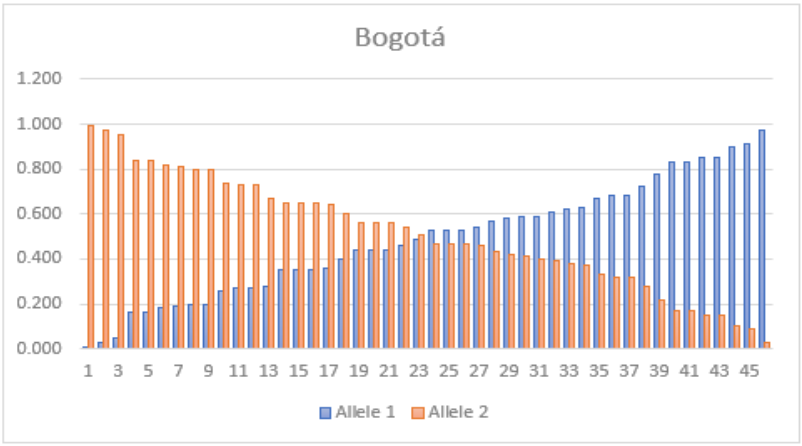

Fig 2.B Distribución de frecuencias alélicas en los 46 marcadores analizados para la muestra de población de Bogotá.

Se presenta además en la figura 5 una forma gráfica de la posible fijación de alelos en las poblaciones para diferenciar ancestría. Si tomamos como ejemplo el marcador MID-94, se observa que este indica una frecuencia entre 81% y 100% del alelo 2 para Chocó y San Andrés, mientras que, para La Guajira, la frecuencia para este alelo en el mismo marcador esta entre 21 y 40%.

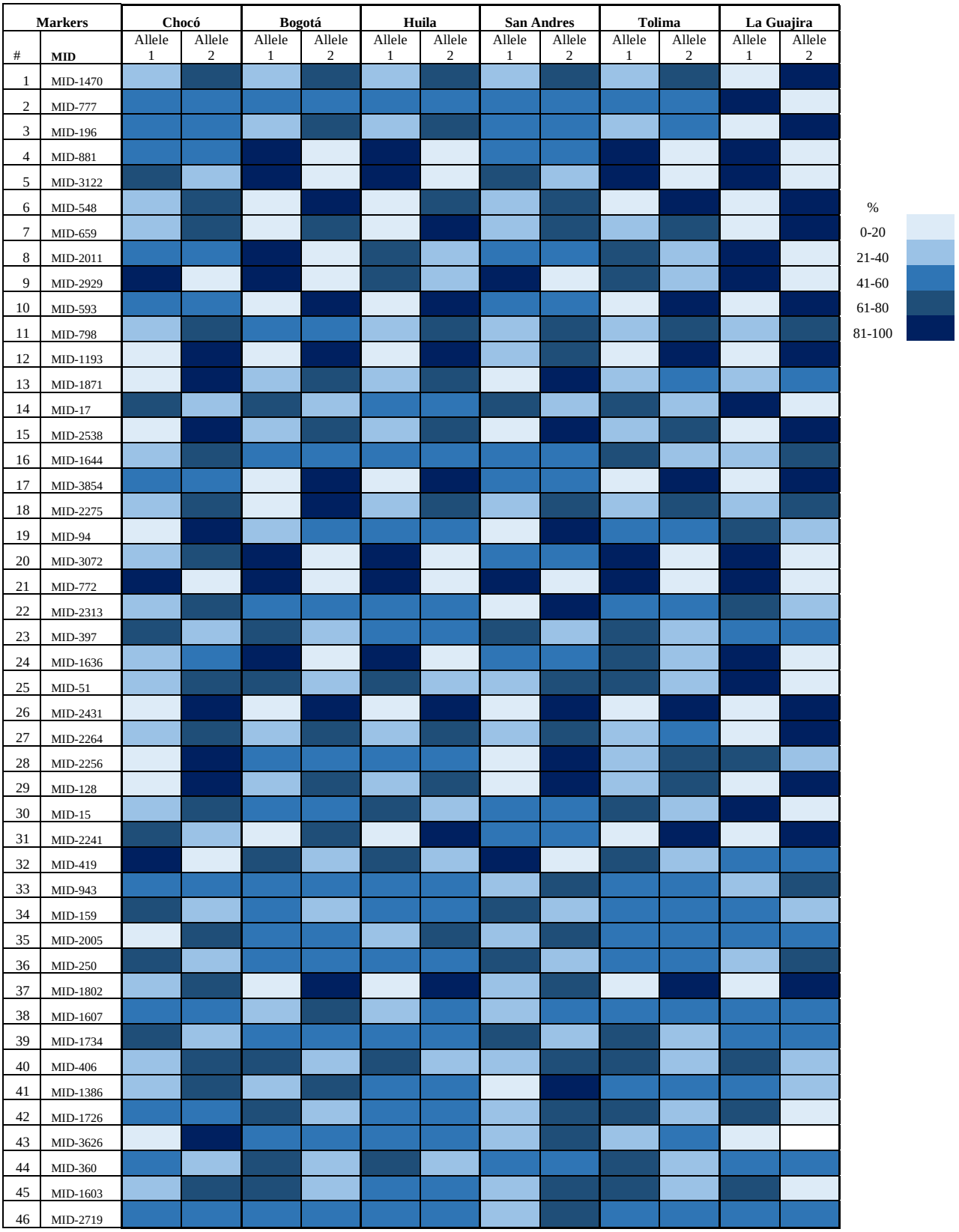

Fig 5. Gradiente de frecuencias alélicas en 46 marcadores indels para 6 grupos analizados.

En los resultados de AMOVA arrojados por el software Arlequin se evidencia que el porcentaje de variación es considerablemente mucho mayor dentro de las poblaciones que entre ellas.

	Among populations	Within populations	Fixation indices:	
Locus	%variation	%variation	FST	P-value
1470	5.7842	94.2158	0.0578	0.0000
777	11.0289	88.9711	0.1103	0.0000
196	19.1392	80.8608	0.1914	0.0000
881	25.0811	74.9189	0.2508	0.0000
3122	15.3431	84.6569	0.1534	0.0000
548	5.1339	94.8661	0.0513	0.0000
659	1.2672	98.7328	0.0127	0.0489
2011	18.3161	81.6839	0.1832	0.0000
2929	5.3776	94.6224	0.0538	0.0000
593	36.4234	63.5766	0.3642	0.0000
798	1.0860	98.9140	0.0109	0.0196
1193	0.3580	99.6420	0.0036	0.2405
1871	7.0096	92.9904	0.0701	0.0000
17	7.1400	92.8600	0.0714	0.0000
2538	8.4563	91.5437	0.0846	0.0000
1644	7.8339	92.1662	0.0783	0.0000
3854	24.5668	75.4332	0.2457	0.0000
2275	2.4133	97.5867	0.0241	0.0020
94	16.5772	83.4229	0.1658	0.0000
3072	45.2922	54.7078	0.4529	0.0000
772	2.9705	97.0295	0.0297	0.0000
2313	15.1413	84.8587	0.1514	0.0000
397	4.3448	95.6552	0.0435	0.0000
1636	27.6539	72.3461	0.2765	0.0000
51	28.5511	71.4489	0.2855	0.0000
2431	2.1248	97.8752	0.0213	0.0029
2264	5.9215	94.0785	0.0592	0.0000
2256	30.0237	69.9763	0.3002	0.0000
128	7.0715	92.9285	0.0707	0.0000
15	17.9227	82.0773	0.1792	0.0000
2241	30.9391	69.0609	0.3094	0.0000
419	12.0987	87.9013	0.1210	0.0000
943	4.1084	95.8916	0.0411	0.0000
159	2.5272	97.4728	0.0253	0.0000
2005	5.2481	94.7519	0.0525	0.0000
250	5.4127	94.5873	0.0541	0.0000
1802	12.5948	87.4052	0.1260	0.0000
1607	1.3700	98.6300	0.0137	0.0284
1734	0.9622	99.0378	0.0096	0.0538
406	21.7598	78.2402	0.2176	0.0000
1386	12.9245	87.0755	0.1292	0.0000
1726	12.0301	87.9699	0.1203	0.0000
3626	14.4075	85.5926	0.1441	0.0000
360	1.6984	98.3016	0.0170	0.0196
1603	24.9525	75.0475	0.2495	0.0000
2719	0.7280	99.2720	0.0073	0.1124

Tabla 4. Resultados de AMOVA, porcentajes de variación dentro y entre las poblaciones para 46 INDELs.

En promedio se obtuvo un porcentaje de variación entre poblaciones de 12.7% y dentro de las poblaciones de 87.3. Este valor de variación dentro de las poblaciones puede deberse a que en todos los grupos la mayoría de individuos tienen una composición ancestral muy heterogénea, mientras que, a nivel de población, como se puede observar en los análisis tanto de ancestría como de Fst, es evidente que algunos de los grupos tienen gran similaridad, por ejemplo, Bogotá, Tolima y Huila presentan porcentajes de componentes ancestrales semejantes, igualmente sucede entre Chocó y San Andrés.

En el caso del grupo Wayuu, a pesar de que este es el que más se diferencia de los otros 5 grupos, puede compartir también cierta similaridad con los 3 grupos de población mezclada (Bogotá, Tolima, Huila) debido al porcentaje considerablemente marcado de ancestría nativa en estos últimos.

Análisis de sub-estructura genética

Con el fin de estimar el nivel de subestructura genética en los grupos poblacionales estudiados, se calcularon estadísticos Fst por pares, encontrando una mayor diferenciación entre el grupo Wayuu y los dos grupos Afro-colombianos. Por otro lado, los valores más bajos de diferenciación se evidencian entre las muestras de población de Bogotá, Tolima y Huila.

Estos resultados sugieren una posible agrupación entre las muestras de población en los siguientes tres cluster: 1. Chocó-San Andrés; 2. Bogota-Huila-Tolima; 3. La Guajira (Wayuu). Esta agrupación es confirmada con un Fst pareado entre los 3 grupos y se hace evidente en el análisis de componentes principales, siendo más cercanos entre sí los grupos 2 y 3, lo cual es coherente debido al considerable aporte de ancestría nativa en las poblaciones de Bogotá, Tolima y Huila.

	Chocó	Bogotá	Huila	San Andres	Tolima	La Guajira
Chocó	*	**0.0000**	**0.0000**	**0.00901**	**0.0000**	**0.0000**
Bogotá	0.15041	*	0.5135	**0.0000**	0.05405	**0.0000**
Huila	0.14469	**0.00028**	*	**0.0000**	0.14414	**0.0000**
San Andres	**0.00667**	0.11328	0.10830	*	**0.0000**	**0.0000**
Tolima	0.15760	**0.00454**	**0.00363**	0.12211	*	**0.0000**
La Guajira	0.25033	**0.07989**	**0.07404**	0.22983	**0.07751**	*

Tabla 5. Fst por pares para las 6 poblaciones analizadas y su respectivo p-valor.

	1	2	3
1	0		
2	0.13414	0	
3	0.23256	0.07158	0

Tabla 6. Fst por pares para los 3 cluster evidenciados. *1. Chocó-San Andrés; 2. Bogota-Huila-Tolima; 3. La Guajira (Wayuu).*

En este análisis se puede observar que el Fst entre Chocó y San Andrés es de 0.00667, lo que indica una diferenciación genética baja, según la escala de Wright (Hartl & Grant, 2007a), sin embargo, el p-valor correspondiente es de 0.00901, lo cual sugiere que es estadísticamente significativo, lo que a su vez puede deberse a las diferencias en el trasfondo histórico del poblamiento de ambas regiones.

Estimación de componentes ancestrales

A continuación, se representan gráficamente los porcentajes de ancestría obtenidas para cada grupo poblacional.

Departamento	N	EUR	EAS	AFR	NAM
Chocó	73	0.139	0.059	0.708	0.094
Bogotá	50	0.544	0.082	0.050	0.324
Huila	38	0.511	0.079	0.064	0.346
San Andrés	50	0.252	0.056	0.604	0.088
Tolima	38	0.520	0.076	0.056	0.347
La Guajira	90	0.048	0.037	0.028	0.887
Total	339				

Tabla 7. Porcentajes de composición ancestral para las 6 poblaciones analizadas. *N=Número de individuos.*

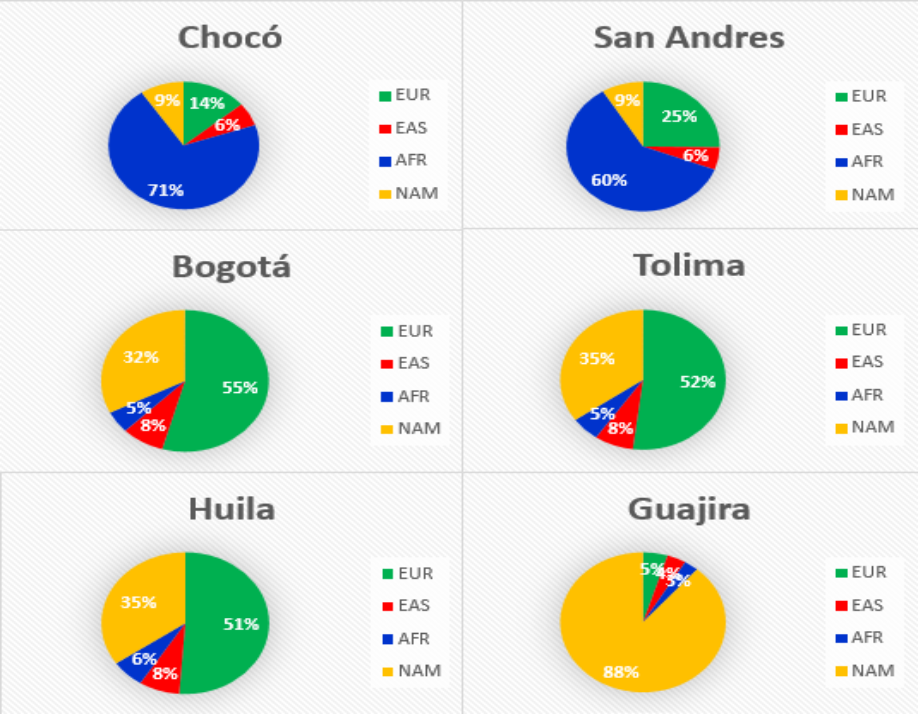

Fig 4. Porcentaje de componentes ancestrales para los 6 grupos poblacionales analizados.

En la gráfica del análisis de STRUCTURE se muestra la composición ancestral para todos los individuos de las 6 poblaciones analizadas, incluyendo las muestras de referencia de África, Europa, Este de Asia y América.

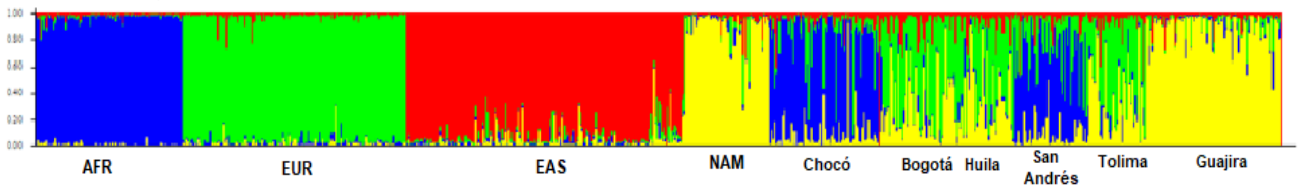

Fig 5. Estimación de ancestria usando STRUCTURE para las muestras de referencia del panel HGDP-CEPH y los individuos de los 6 grupos poblacionales de Colombia para 46 marcadores INDEL-AIMs. (AFR: Africa; EUR: Europa; EAS: Este de Asia; NAM: Nativo americano). Se asumió un K=4, Burn-in Steps= 100000, MCMC= 100000. Use admixture model.

Los porcentajes de ancestría encontrados para cada uno de los 6 grupos fueron consistentes con los resultados del análisis de diferenciación genética por Fst. Es evidente que Bogotá, Tolima y Huila presentan porcentajes de ancestría y un patrón de mezcla muy similares, Igual a lo que sucede con Chocó y San Andrés, teniendo este último una mayor proporción de ancestría europea, hecho que concuerda con los datos históricos del poblamiento del archipiélago.

De la misma forma se puede observar en la población de La Guajira (Wayuu) que a pesar de tener un fuerte componente ancestral nativo americano (88%), presentan un gradiente de mezcla con otros grupos ancestrales, lo que sugiere que deja de ser un aislado genético.

Análisis de componentes principales

En las gráficas de componentes principales se observan todos los individuos de cada grupo poblacional, la mayoría agrupados en el vector correspondiente a la ancestría inferida previamente por medio de STRUCTURE. Sin embargo, se encuentran algunos individuos dispersos que corresponden a los que tienen mayor porcentaje de mezcla, o a los que según la información de la encuesta tienen una ancestría diferente por parte de alguno de sus predecesores.

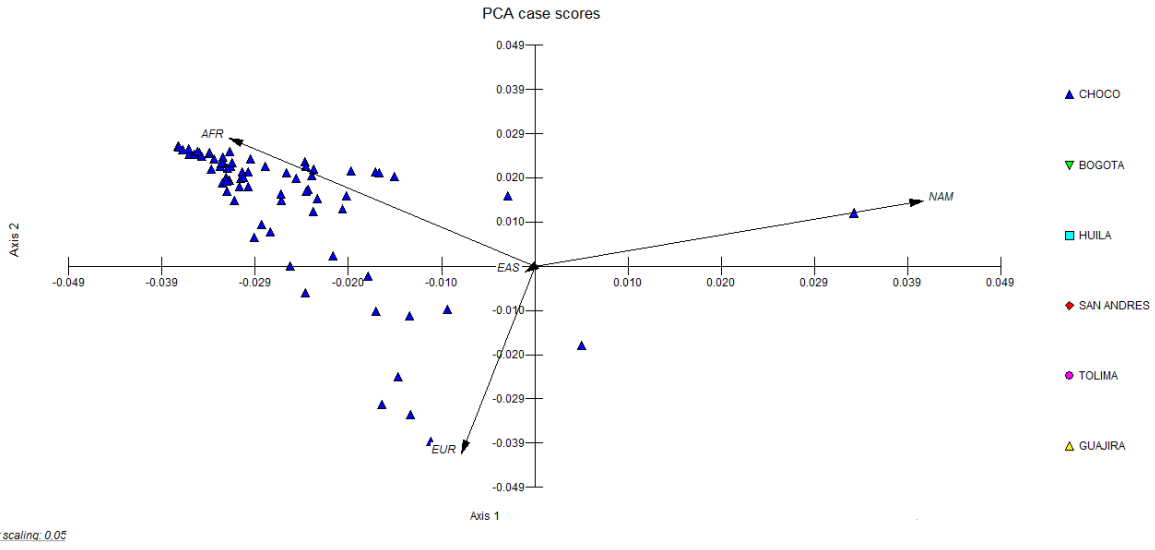

Fig 6A. Análisis de componentes principales para 73 individuos de Chocó.

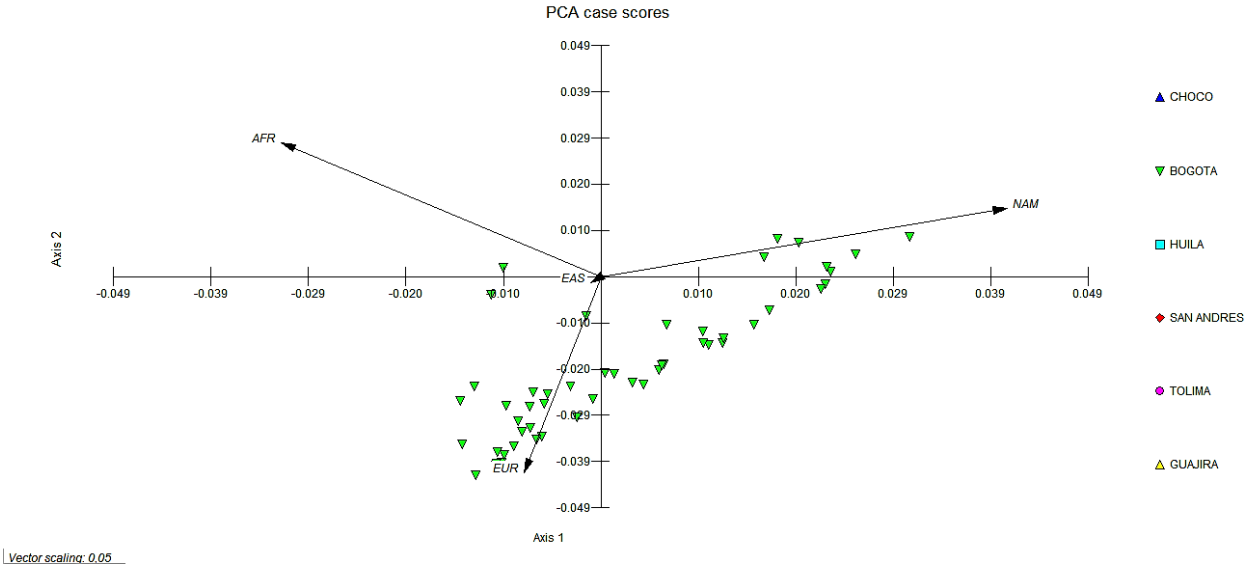

Fig 6B. Análisis de componentes principales para 50 individuos de Bogotá.

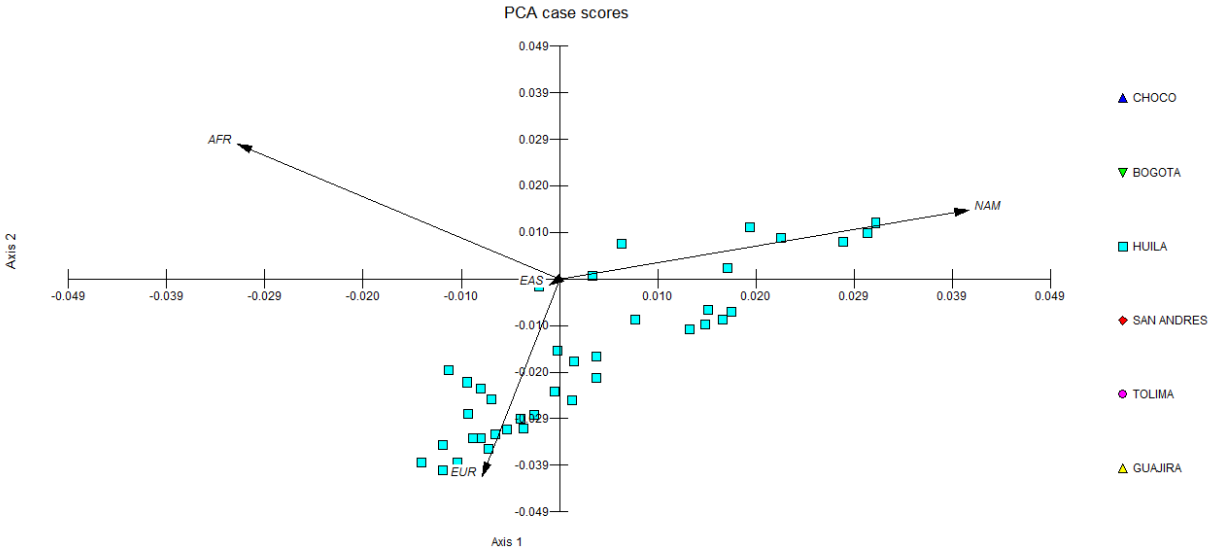

Fig 6C. Análisis de componentes principales para 38 individuos de Huila.

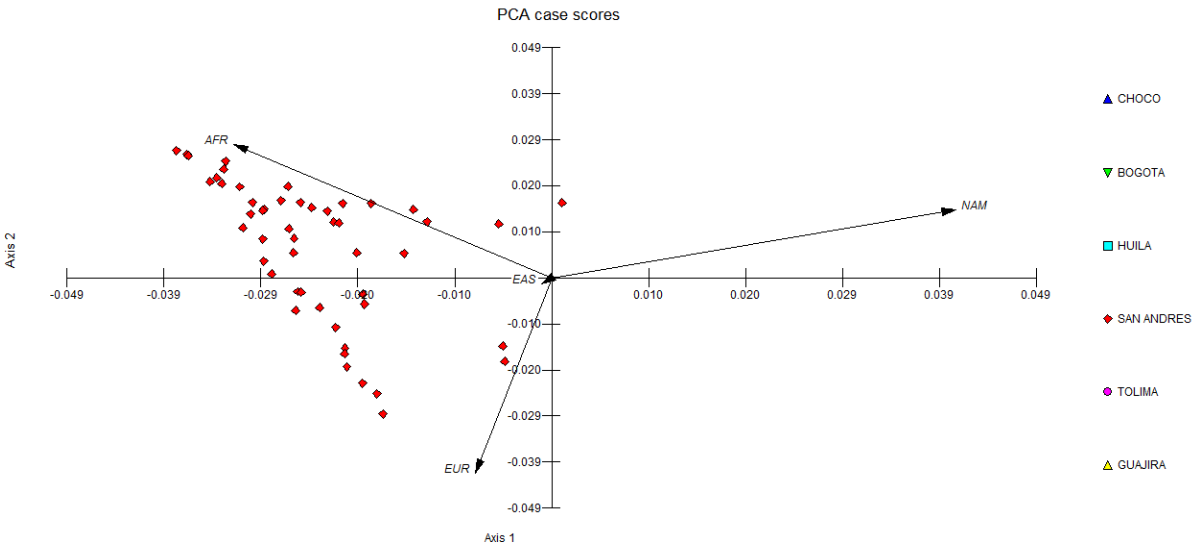

Fig 6D. Análisis de componentes principales para 50 individuos de San Andrés.

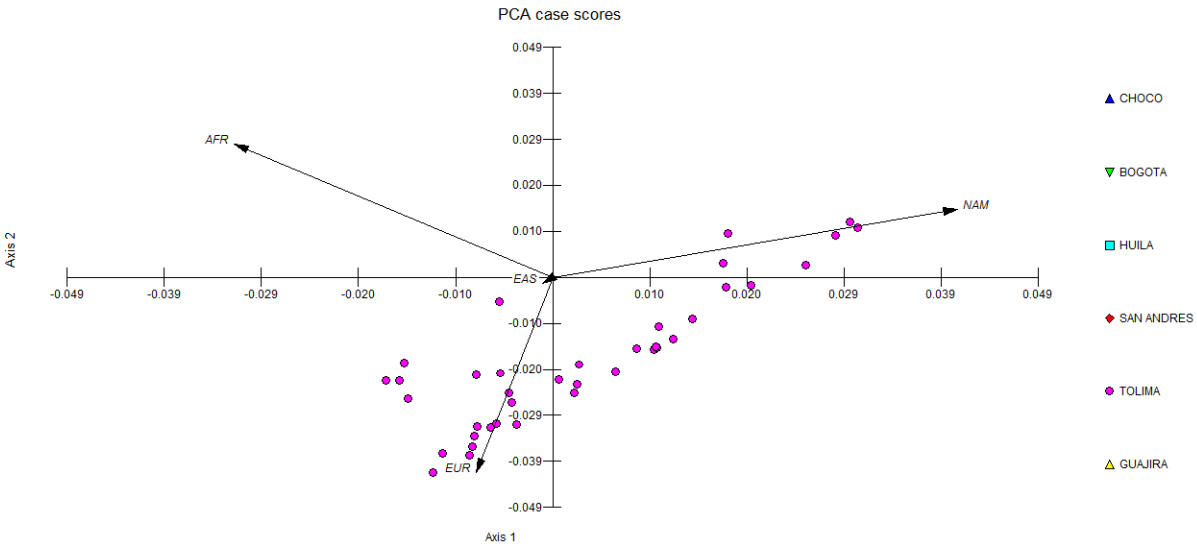

Fig 6E. Análisis de componentes principales para 38 individuos de Tolima.

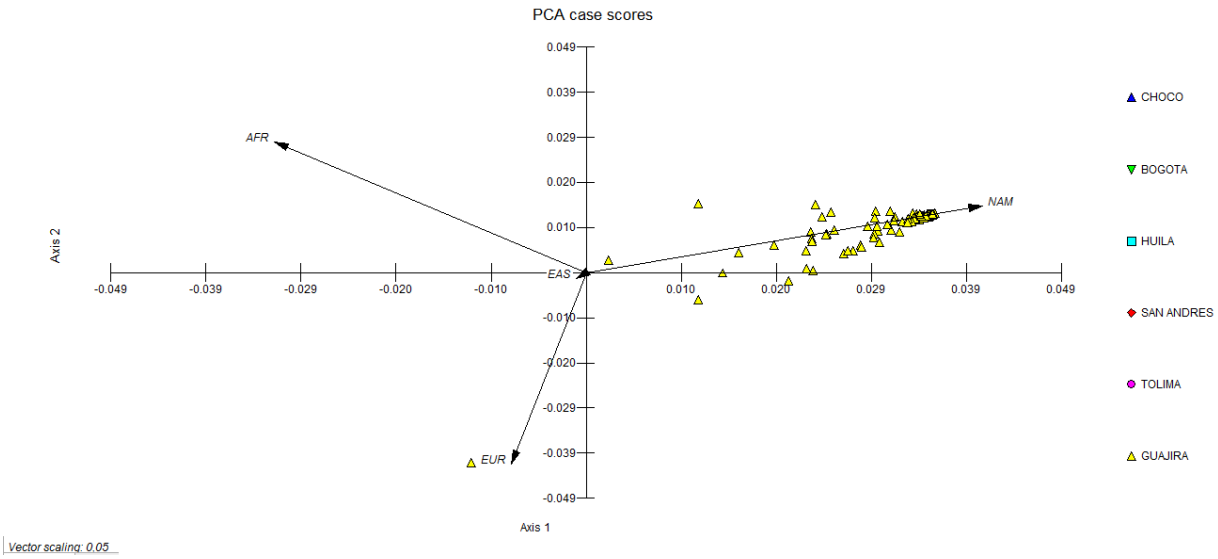

Fig 6F. Análisis de componentes principales para 90 individuos Wayuu.

Posteriormente se realizó un nuevo análisis de componentes principales asumiendo los 3 grupos descritos previamente y el resultado es concordante con lo encontrado en el análisis de Fst; La Guajira (Wayuu) se encuentra ubicada en el vector correspondiente a la ancestría nativo americana; Chocó y San Andrés se encuentran agrupados en el vector que indica ancestría africana; y Bogotá, Tolima y Huila se encuentran en un solo grupo indicando un mayor componente europeo.

Adicionalmente, respecto al último grupo se observa una ligera cercanía al vector que indica la ancestría asiática que, si bien no es necesariamente significativa, puede dar pistas de una posible asociación o similaridad entre estas dos poblaciones continentales, dada su propia historia desde la llegada de los primeros pobladores de américa que, según las hipótesis más acogidas del poblamiento americano, fueron provenientes desde Siberia (Reich et al., 2012).

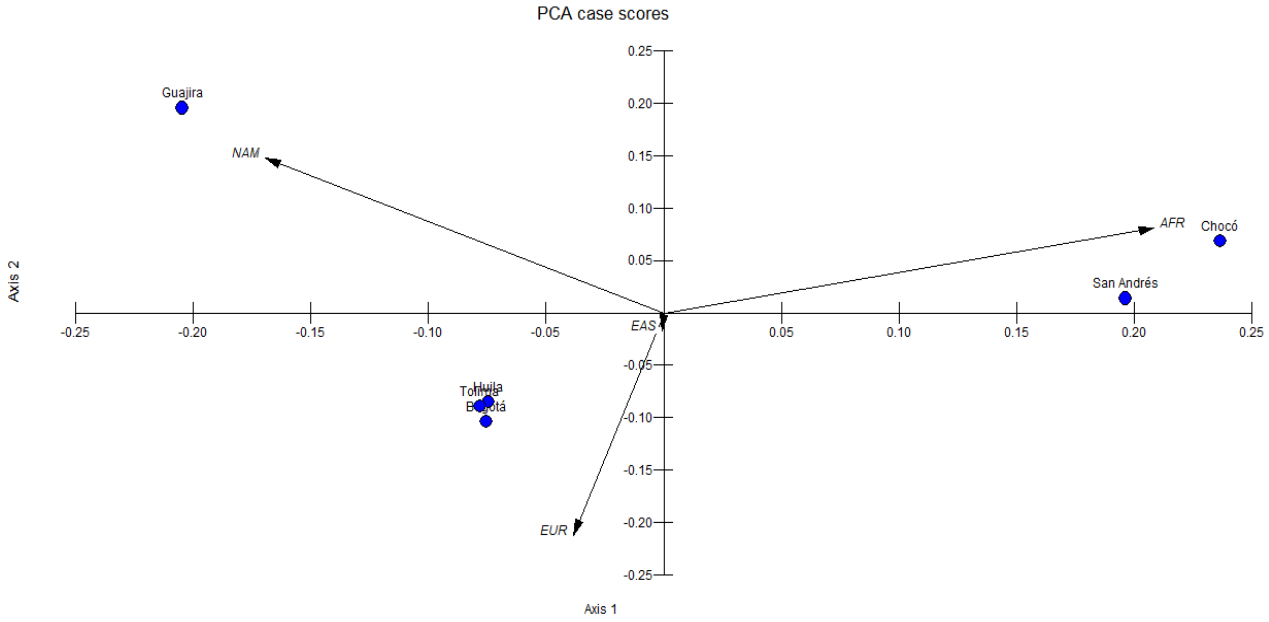

Fig 7. Análisis de componentes principales para los 6 grupos poblacionales analizados basado en los componentes ancestrales obtenidos en STRUCTURE.

DISCUSIÓN

En primer lugar, cabe resaltar que en esta investigación no se ha realizado un muestreo aleatorio o probabilístico, por lo que no se realiza un cálculo de tamaño de muestra ni de un error muestral, más la robustez del estudio es garantizada gracias a que los individuos analizados fueron seleccionados bajo criterios muy estrictos, según la información contenida en la encuesta. Algunos de estos parámetros son: la genealogía paterna y materna de 3 generaciones hacia atrás, municipio y departamento de nacimiento y de residencia (si aplica o si es relevante), pertenencia a un grupo étnico o a una casta especifica (autodeterminación), y la información demográfica sobre la pareja e hijos.

Por otro lado, a partir de los resultados de los índices Fst y los porcentajes de componentes ancestrales obtenidos con STRUCTURE se pueden inferir 3 grupos que representan los 3 posibles modelos tipo de población representativa en Colombia y sus porcentajes de mezcla promedio. Bogotá, Tolima y Huila conforman un primer grupo; Chocó y San Andrés

constituyen otro grupo representando a las poblaciones afro; y La Guajira representa a las poblaciones indígenas.

En general se ha considerado a Bogotá como una población típica mestiza, y como se puede evidenciar en este trabajo, la muestra de población bogotana analizada no representa significativamente población indígena ni afrodescendiente, al igual que lo observado en los resultados para Tolima y Huila que presentan porcentajes de ancestría similares, lo cual es concordante con el hecho de que estas 3 poblaciones se encuentran en la misma región biogeográfica. Esto consolida la idea de que la composición y diferenciación genética de un grupo poblacional no se debe limitar a las divisiones político-administrativas de un territorio.

A pesar de que en cada uno de los 3 grupos ya descritos se evidencia una fuerte diferencia en la ancestría, los valores encontrados revelan que no hay aislamiento genético en ninguno de ellos, lo cual lleva a afirmar que en la actualidad es erróneo hablar de poblaciones puras.

Es de resaltar que en algunos estudios que evalúan la ancestría de diferentes poblaciones en Latinoamérica se subestima o se pasa por alto la posible contribución ancestral de poblaciones asiáticas. Actualmente en Colombia puede evidenciarse su influencia, específicamente de ancestros sirio-libaneses y del medio oriente en las poblaciones del Caribe, entre otros.

Sin embargo, la estimación de ancestría asiática en américa debe ser evaluada con suma precaución ya que tiene un trasfondo bastante complejo. Para ejemplificar, hoy en día se habla de migraciones de grupos originarios de oriente medio hacia el caribe colombiano (Ministerio de Cultura, 2011) y otros países como México, Brasil y Argentina alrededor del siglo XIX. Adicionalmente se conoce que alrededor del siglo VIII ocurrió un desplazamiento de grupos árabes hacia la península ibérica, y que estos influyeron drásticamente en gran parte de la cultura española (Malpica & De Granada, 2010). En este orden de ideas puede pensarse que posteriormente haya llegado un componente ancestral asiático al nuevo mundo por medio de los colonos españoles.

Así mismo, varias hipótesis acerca del poblamiento americano indican que los primeros pobladores de América pueden ser originarios de Asía, llegando en distintas olas migratorias (Gravel et al., 2013). Para llegar al continente americano se cree que estas comunidades provenientes de Siberia cruzaron por Beringia (Battaglia et al., 2013), lo que lleva a creer que pueden existir similitudes en cuanto a la ancestría asiática y la que actualmente llamamos nativoamericana.

Por lo tanto, se considera con la posibilidad de que se esté asumiendo una diferenciación sesgada entre los componentes asiático y americano. No obstante, según Pereira et al., (2012), este set especifico de 46 marcadores indels puede lograr una clara diferenciación de los 4 grupos, además de su capacidad para asignar proporciones de ancestría en grupos poblaciones sin tener información a priori del origen geográfico de los mismos.

Partiendo de esta hipótesis se realizó un análisis complementario en STRUCTURE en el cual se incluyeron muestras de referencia del este de Asia, pero se asumió un total de 3 poblaciones ancestrales en lugar de 4 y las muestras de EAS y NAM aparecen como si fueran un solo grupo.

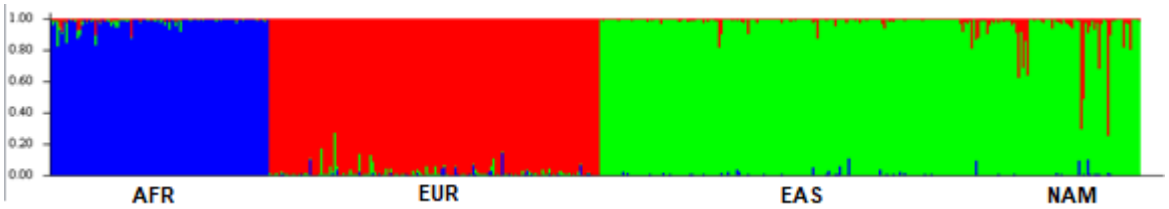

Fig 8. Estimación de ancestria usando STRUCTURE para las muestras de referencia del panel HGDP-CEPH (AFR: Africa; EUR: Europa; EAS: Este de Asia; NAM: Nativo americano) asumiendo un K=3.

Adicionalmente se realizó un análisis de componentes principales con estos mismos 4 grupos y en este sí se pueden evidenciar los núcleos de inercia esperados, diferenciando los 4 componentes ancestrales.

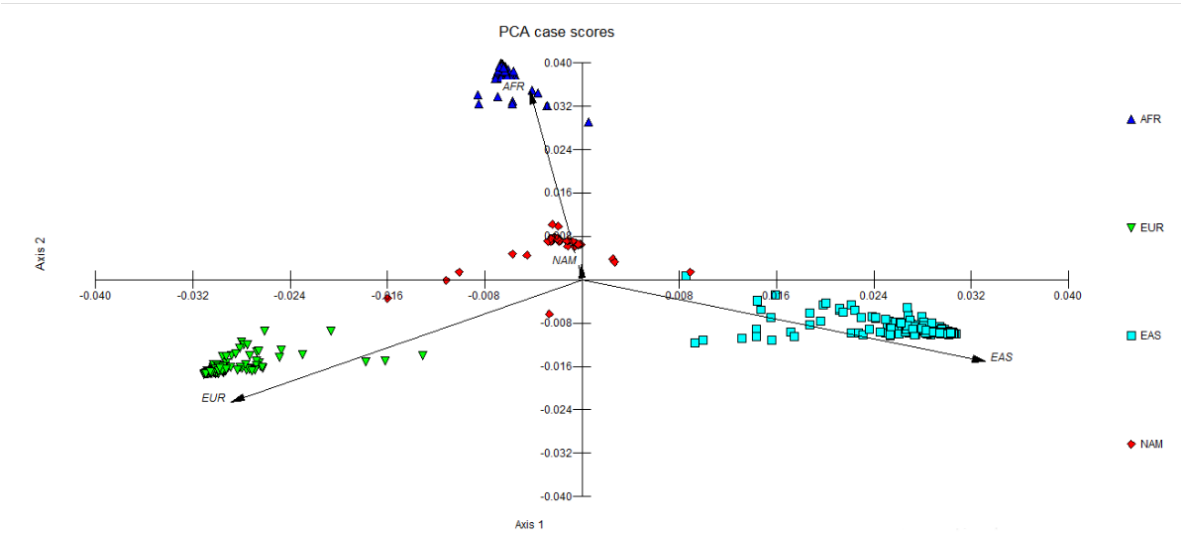

Fig 9. Analisis de componentes principales para las muestras de referencia del panel HGDP-CEPH **(**AFR: Africa; EUR: Europa; EAS: Este de Asia; NAM: Nativo americano)

No obstante, este es un análisis que aún no cuenta con los datos suficientes, pero puede considerarse como objeto de estudio a futuro. Sería necesario hacer una evaluación rigurosa de la diferenciación entre estos dos grupos continentales, para así garantizar mayor robustez en los estudios geneticopoblaciones asociados a la ancestría, respecto a la asertividad de las estimaciones con diferentes paneles de marcadores AIMs y las muestras de referencias empleadas en cada uno.

Surge además la inquietud de si se puede en realidad considerar las muestras de referencia utilizadas, adecuadas para el análisis de ancestría nativoamericana, esto debido a que se cree que en realidad al hablar de ancestría nativa, esta misma puede englobar a su vez varios tipos, tomando como punto de partida el hecho de que según las diferentes hipótesis del poblamiento americano, los primeros pobladores pudieron llegar desde Siberia en diferentes oleadas migratorias a América (Gravel et al., 2013) y por lo tanto constituir un componente ancestral común para los grupos indígenas nativos emergentes.

No obstante, los diferentes grupos indígenas establecidos en diferentes regiones del territorio americano a lo largo del tiempo han estado expuestos a diferentes condiciones ambientales y socioculturales que han podido influir en su desarrollo y diferenciación genética respecto a

otros grupos, y así actualmente se pueden encontrar poblaciones indígenas más conservadas que otras en su composición genética.

CONCLUSIÓN

Después de los análisis realizados se construye un posible modelo de la población colombiana, en cual se encuentran los 3 tipos de subpoblación que la componen, según la ancestría de cada una de ellas. Se tiene la subpoblación tipo Afroamericana representada en este caso por una muestra de individuos de San Andrés y de Chocó, el grupo de población mezclada compuesta por individuos de Bogotá, Tolima y Huila; y finalmente el grupo nativoamericano o indígena representado por individuos de ascendencia Wayuu. Es de vital importancia tener en cuenta algunas consideraciones relativas a este modelo, por ejemplo, es evidente que los grupos Afroamericano y Nativoamericano poseen algún porcentaje ancestral considerable de otras ancestrías y así mismo se observa una aproximación a sus respectivos porcentajes de mezcla, reafirmando la hipótesis de que en la actualidad es errado o subjetivo hablar de poblaciones completamente puras, en cambio se puede interpretar como grupos que permanecen más conservados que otros en cuanto a su linaje ancestral, y de esta forma se considera necesario replantar algunos conceptos como el de Caucásico-Mestizo. También es importante resaltar que esta clase de estudios deben ser útiles para comprender el sesgo que se tiene respecto a la relación que hay entre la diferenciación de las poblaciones y las divisiones político-administrativas dentro de una región geográfica, de modo que, que la composición genética de una población no está necesariamente ligada solo a estas divisiones creadas en otros ámbitos, sino que también influyen factores culturales y biogeográficos. Finalmente se plantea la posible presencia de varios tipos de ancestría amerindia y asiática y probablemente una similaridad considerable entre estos dos linajes ancestrales.

Bibliografía

Battaglia, V., Grugni, V., Perego, U. A., Angerhofer, N., Gomez-Palmieri, J. E., Woodward, S. R., … Semino, O. (2013). The First Peopling of South America: New Evidence from Y-Chromosome Haplogroup Q. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0071390

Da Costa Francez, P. A., Rodrigues, E. M. R., De Velasco, A. M., & Dos Santos, S. E. B. (2012). Insertion-deletion polymorphisms-utilization on forensic analysis. *International Journal of Legal Medicine*, *126*(4), 491–496. https://doi.org/10.1007/s00414-011-0588-z

Departamento Administrativo Nacional de Estadistica (DANE). (2007). Colombia una nación multicultural. *Colombia Una Nación Multicultural, Su Diversidad Étnica*.

Excoffier, L. and Lischer, H. E. L. (2010). An Integrated Software Package for Population Genetics Data Analysis. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodriguez-Flores, J. L., … Bustamante, C. D. (2013). Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genetics*, *9*(12). https://doi.org/10.1371/journal.pgen.1004023

Hartl, D. L., & Grant, A. (2007a). Principles of Population Genetics. 4th Edition. *Sinauer Associates*, *14*(4), 544–545. https://doi.org/10.2980/1195-6860(2007)14[544b:POPG]2.0.CO;2

Heinz, T., Álvarez-Iglesias, V., Pardo-Seco, J., Taboada-Echalar, P., Gómez-Carballa, A., Torres-Balanza, A., … Salas, A. (2013). Ancestry analysis reveals a predominant Native American component with moderate European admixture in Bolivians. *Forensic Science International: Genetics*, *7*(5), 537–542. https://doi.org/10.1016/j.fsigen.2013.05.012

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature Reviews. Genetics*, *10*(September), 639–650. https://doi.org/10.1038/nrg2611

Malpica, A., & De Granada, U. (2010). La Vida Urbana En Al-Andalus Y Su Papel En La Estructura Del Poblamiento Y En La Organización Social. *Imago TemporIs. medIum Aevum*, *IV*, 401–420. Retrieved from http://www.raco.cat/index.php/ImagoTemporis/article/viewFile/256920/343962

Ministerio de Cultura. (2011). Los árabes en colombia, 168.

Oficina de Información Diplomatica del Departamento de Relaciones Exteriores. (2015). Ficha Pais - República de Colombia.

Pereira, R., Phillips, C., Alves, C., Amorim, A., Carracedo, Á., & Gusmao, L. (2009). A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*, *30*(21), 3682–3690. https://doi.org/10.1002/elps.200900274

Pereira, R., Phillips, C., Pinto, N., Santos, C., dos Santos, S. E. B., Amorim, A., … Gusmão, L. (2012). Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS ONE*, *7*(1). https://doi.org/10.1371/journal.pone.0029684

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, ??ngel, & Lareu, M. V. (2013). An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers in Genetics*, *4*(MAY), 1–13. https://doi.org/10.3389/fgene.2013.00098

Pritchard, J. K., Wen, X., & Falush, D. (2010). Documentation for structure software : Version 2 . 3. *University of Chicago. IL*, *6*(3), 321–326. https://doi.org/10.1002/spe.4380060305

Ramón, A., & Manilla, C. R. (n.d.). Estudio de mestizaje genético con marcadores ABO-Rh en 2 poblaciones de Veracruz ; antesala a un estudio genético de salud.

Rangel-ch, J. O. (2005). La biodiversidad de Colombia. *Palimpsestos*, *5*, 292–304. https://doi.org/10.1073/pnas.262413599

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., … Ruiz-Linares, A. (2012). Reconstructing Native American population history. *Nature*, *488*(7411), 370–374. https://doi.org/10.1038/nature11258

Rojas, M. Y., Alonso, L. a, Sarmiento, V. a, Eljach, L. Y., & Usaquén, W. (2013). Structure analysis of the La Guajira-Colombia population: a genetic, demographic and genealogical overview. *Annals of Human Biology*, *40*(2), 119–31. https://doi.org/10.3109/03014460.2012.748093

Rousset, F. (2017). Genepop version 4.6.9, (March).

Saloum de Neves Manta, F., Pereira, R., Vianna, R., Rodolfo Beuttenmüller de Araújo, A., Leite Góes Gitaí, D., Aparecida da Silva, D., … Gusmão, L. (2013). Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. *PLoS ONE*, *8*(9), 1–11. https://doi.org/10.1371/journal.pone.0075145

Salzano, F. M., & Bortolini, M. C. (2002). The Evolution and Genetics of Latin American Populations, *1*, 532. https://doi.org/10.1017/CBO9780511666100

Suárez, C. A., & Quintin, C. I. (2004). Suárez. In *"Colombia Andina". Colombia: Norma. pp. 3-4. ISBN 9580468060*.

3. MARCADORES INDELS PARA ANLISIS DE ANCESTRIA VS IDENTIFICACION

INTRODUCCIÓN

En la actualidad se ha generado una gran discusión en cuanto a la diferencia entre marcadores Indels utilizados para ancestría y para identificación humana según sus características. En este capítulo de revisión se pretende abordar dicha discusión desde diferentes enfoques basados en estudios publicados previamente, en los cuales se han implementado marcadores Indels con diferentes objetivos. Entre estas finalidades se puede incluir el análisis de la composición ancestral y estructura genética de algunas poblaciones (Pereira et al., 2012), así como análisis de identificación y la utilidad de estos marcadores en casos forenses (Pereira et al., 2009).

En dichos estudios, los marcadores Indels han demostrado un gran potencial para determinar ancestría en los grupos poblacionales continentales principales (Africa, Europa, Asia y América) debido a que presentan diferencias significativas en cuanto a la frecuencia alélica entre poblaciones separadas geográficamente. Estos marcadores pueden usarse además para análisis de subestructura en poblaciones mezcladas y mezcla interétnica individual (Pereira et al., 2009).

En otros trabajos realizados con marcadores de linaje como mtDNA y cromosoma Y, se ha obtenido información valiosa sobre la ancestría materna y paterna respectivamente, aun cuando puede afirmarse que estos solo representan una pequeña proporción del genoma (Romanini et al., 2015).

3.1 Marcadores Indels (Insertion-Deletion

Los marcadores INDEL son polimorfismos de longitud que se dan por una deleción o una inserción en uno o más nucleótidos y se encuentran distribuidos en todo el genoma (autosómicos y cromosomas sexuales) (Da Costa Francez et al., 2012).
El estudio de la variación genética usando polimorfismos de ADN distribuidos a lo largo del genoma ha permitido una mejor comprensión de la historia de la diversidad de poblaciones humanas, así como también ha proporcionado también un sistema eficiente para la identificación genética de individuos (Pereira et al., 2009).

Es importante recordar las características que presentan los Indels como marcadores genéticos que los hacen interesantes para diferentes tipos de análisis; como lo son su distribución en todo el genoma (autosomas y cromosomas sexuales), la clase de polimorfismo como tal que se deriva de un único evento mutacional (inserción o deleción), en su mayoría presentan diferencias significativas en las frecuencias alélicas entre grupos poblacionales geográficamente separados, esta diferencia en la distribución de frecuencias es factor determinante para su uso como marcadores informativos de ancestría (AIMS), también generalmente son fragmentos pequeños, lo cual supone una ventaja para la amplificación de muestras degradadas o con cierto grado de dificultad (Da Costa Francez et al., 2012).

En contraste, actualmente los STR se consideran el método de elección estándar en estudios forenses ya que permiten un alto poder de discriminación, adecuado para resolver los casos de identificación humana (Pereira et al., 2009).

Sin embargo, se han descrito otras ventajas además de las características mencionadas previamente, del uso de los marcadores Indels y SNPs sobre los marcadores STR en estudios poblacionales y forenses, como una menor tasa de mutación, fácil interpretación, amplicones pequeños; lo cual mejora su capacidad para amplificar pequeñas cantidades de ADN o de muestras degradadas, que son características comunes de las muestras analizadas en el escenario forense y pueden significar inconvenientes para la correcta tipificación de las muestras (Da Costa Francez et al., 2012).

También se ha encontrado una gran utilidad de los marcadores Indels como herramienta complementaria en análisis de parentesco con STRs, específicamente en los casos donde se encuentran pocas incompatibilidades entre el presunto padre y el hijo en cuestión. En circunstancias como esta, aunque las incompatibilidades podrían explicarse como eventos mutacionales o alelos silentes, puede hallarse un margen de duda en caso de que exista relación biológica entre los presuntos padres. Para estas situaciones puede ser útil ampliar el número de marcadores, preferiblemente con baja tasa de mutación, como SNPs (2.3x10-8) o INDELs (2.3x10-9) (Da Costa Francez et al., 2012).

3.2 Estudios Previos

En la época actual se han realizado diferentes estudios en Colombia y America Latina en donde puede afirmarse la existencia de una distribución heterogénea de 3 contribuciones ancestrales principales (Nativos americanos, Europeos, Africanos) (Saloum de Neves Manta et al., 2013) El objetivo de estos análisis es entender la dinámica de algunos eventos de migración y colonización, además de determinar patrones de mezcla dentro de una población mediante el uso de marcadores genéticos uniparentales y autosómicos, incluyendo Marcadores de cromosoma Y, ADN mitocondrial, SNPs, STRs e Indels.

Particularmente, los marcadores Indels han sido empleados tanto en estudios de poblaciones como forenses, con el objetivo principal de evaluar su utilidad y eficacia en análisis que involucren muestras de diverso tipo y calidad, así como diferentes características poblacionales de los individuos.

A pesar de las diferencias en cuanto a la metodología y objetivos de las diferentes investigaciones realizadas, se podría decir que hay una conclusión en común para la gran mayoría de ellos, y es que los marcadores informativos de ancestría (AIMs) pueden ser útiles en los análisis de muestras tomadas como evidencia en casos forenses, los resultados confirman que los indels utilizados son muy informativos en 3 grupos poblacionales grandes,

por lo cual pueden ser de gran interés en los estudios en los cuales la información del origen de la muestra analizadas sea necesaria.

El trabajo realizado por Pereira et al., (2009) en el cual se estandariza una multiplex de 38 Indels para identificación humana, ha servido como soporte para muchas de las investigaciones realizadas en este tema en América Latina. La causa de esto es que en el estudio mencionado se concluye que la eficiencia del indel-plex desarrollado, hace que sea una valiosa herramienta complementaria en estudios de identificación, especialmente en los análisis de ADN altamente degradado.

Según Heinz et al., (2013) se sugiere que los análisis forenses basados en marcadores de ancestría podrían ser mejorados usando bases de datos locales, según la población analizada.

En otros estudios (Da Costa Francez et al., 2012) y (Romanini et al., 2015), se ha hecho especial énfasis en la utilidad y eficacia de estos marcadores en muestras de interés forense. Igualmente se ha encontrado que se puede obtener, de manera eficiente, una estimación de ancestría en muestras óseas degradadas y de diferente origen continental al trabajar con los problemas que se presentan usualmente en muestras forenses.

3.3 Perspectivas

Los marcadores empleados en paneles de Idels-Aims, al mostrar grandes diferencias en las frecuencias alélicas entre grupos poblacionales continentales, permiten su uso potencial en estudios de genética de poblaciones y en estudios forenses en los que el análisis de proporción de mezcla interétnica individual es necesario (Da Costa Francez et al., 2012).

Esta información puede ser útil en la perspectiva forense y en la identificación de individuos, especialmente en programas que requieran los análisis de diferentes ancestrías, como las investigaciones de migración ilegal y de desaparecidos. El conocimiento de la ancestría asociada a restos óseos humanos abre la posibilidad de redireccionar investigaciones a través de información nueva en escenarios diferentes, involucrando personas desaparecidas de la

misma región, y la recolección adecuada de material de referencia proveniente de familiares (Romanini et al., 2015).

Si bien se puede afirmar que los marcadores Indels en general presentan un importante desempeño en diferentes aplicaciones, hay que resaltar la utilidad de los diferentes paneles estandarizados tanto para análisis de ancestría como de identificación según el objetivo de la investigación. Por ejemplo, en un panel desarrollado para análisis de ancestría es ideal encontrar que se presente una diferencia significativa en las frecuencias alélicas en los grupos considerados como continentales o parentales, según (Russo et al., 2016) esta diferencia debe ser al menos del 40%.

Al aplicar entonces un panel de marcadores de ancestría a una muestra de población ya sea que se tenga a priori la información sobre su origen o no, se puede inferir una estimación de los componentes ancestrales de dicha población y la mezcla interétnica individual y a partir de estos datos se puede calcular un índice de subestructura genética. Este último puede ser de vital importancia no solo en estudios genético poblacionales, sino también en el ámbito clínico y otras áreas de la investigación genética (Wang et al., 2007), ya que la subestructura en una población puede influir de manera negativa en estudios de asociación si esta no es tenida en cuenta (Pereira et al., 2012).

Como se menciona previamente, la estimación de la ancestría en las poblaciones humanas y específicamente a nivel de individuo puede ser útil en algunos casos en el campo forense, por ejemplo, si el origen de una muestra o la mezcla interétnica de la misma es relevante para la resolución de un caso particular. Sin embargo, en general para casos de identificación de parentesco puede ser de más utilidad aplicar marcadores más neutrales en cuanto a las diferencias en las frecuencias alélicas.

Teniendo en cuenta esto, al utilizar en paralelo un panel de AIMs y uno desarrollado para identificación humana y realizar, por ejemplo, un análisis de subestructura genética, se espera que ambos arrojen resultados diferentes, dado el comportamiento mismo de los diferentes sets de marcadores genéticos.

Se considera entonces que para futuros estudios es importante tener en cuenta todas estos factores y características en cuanto a los Indels como marcadores genéticos y evitar su uso indiscriminado para cualquier tipo de estudio. Ya que, si bien no se debe negar la estrecha relación entre los ámbitos forense, poblacional y clínico, se debe también comprender las diferencias que pueden existir entre los mismos y hacer uso adecuado de las diversas herramientas desarrolladas para cada uno de ellos y así lograr un mejor acercamiento a resultados veraces y aplicables según el objetivo de cada investigación.

3.4 Conclusiones

Los resultados obtenidos a partir de los marcadores informativos de ancestría de tipo Indel (INDELS-AIMS) evidencian que el análisis de ancestría siempre tendrá una estrecha relación con la identificación humana, pues el componente poblacional no debe ser excluido en los análisis forenses.

Por otro lado, las características moleculares de los marcadores Indel son útiles tanto para análisis forenses como poblacionales, ya que los paneles aplicables para ancestría e identificación en su mayoría no solo cuentan con un alto número de marcadores, sino que cada uno de estos marcadores también son cuidadosamente seleccionados en base a las características necesarias según su futura aplicación. Esto con el fin de garantizar un alto poder de discriminación en los análisis de identificación, y lograr una estimación precisa en los análisis de ancestría en diferentes grupos poblacionales. En cuanto a esta última, como se describe en (Russo et al., 2016), la precisión de estos análisis puede aumentar a medida que se aumenta el número de marcadores.

Finalmente debe reconocerse la gran utilidad de estos marcadores para obtener tipificaciones de manera satisfactoria en muestras con niveles considerables de dificultad, debido su antigüedad, estado de degradación o presencia de inhibidores de PCR, como es el caso generalmente de las muestras de casos forenses (Restos óseos, muestras en parafina, etc).

Bibliografía

Alonso, L. A., & Usaquén, W. (2013). Y-chromosome and surname analysis of the native islanders of San Andrés and Providencia (Colombia). *HOMO- Journal of Comparative Human Biology*, *64*(1), 71–84. https://doi.org/10.1016/j.jchb.2012.11.006

Battaglia, V., Grugni, V., Perego, U. A., Angerhofer, N., Gomez-Palmieri, J. E., Woodward, S. R., … Semino, O. (2013). The First Peopling of South America: New Evidence from Y-Chromosome Haplogroup Q. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0071390

Da Costa Francez, P. A., Rodrigues, E. M. R., De Velasco, A. M., & Dos Santos, S. E. B. (2012). Insertion-deletion polymorphisms-utilization on forensic analysis. *International Journal of Legal Medicine*, *126*(4), 491–496. https://doi.org/10.1007/s00414-011-0588-z

Departamento Administrativo Nacional de Estadistica (DANE). (2007). Colombia una nación multicultural. *Colombia Una Nación Multicultural, Su Diversidad Étnica*.

Excoffier, L. and Lischer, H. E. L. (2010). An Integrated Software Package for Population Genetics Data Analysis. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodriguez-Flores, J. L., … Bustamante, C. D. (2013). Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genetics*, *9*(12). https://doi.org/10.1371/journal.pgen.1004023

Hartl, D. L., & Grant, A. (2007a). Principles of Population Genetics. 4th Edition. *Sinauer Associates*, *14*(4), 544–545. https://doi.org/10.2980/1195-6860(2007)14[544b:POPG]2.0.CO;2

Hartl, D. L., & Grant, A. (2007b). *Principles of Population Genetics. 4th Edition*. *Sinauer Associates* (Vol. 14). https://doi.org/10.2980/1195-6860(2007)14[544b:POPG]2.0.CO;2

Heinz, T., Álvarez-Iglesias, V., Pardo-Seco, J., Taboada-Echalar, P., Gómez-Carballa, A., Torres-Balanza, A., … Salas, A. (2013). Ancestry analysis reveals a predominant Native American component with moderate European admixture in Bolivians. *Forensic Science International: Genetics*, *7*(5), 537–542. https://doi.org/10.1016/j.fsigen.2013.05.012

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nature Reviews. Genetics*, *10*(September), 639–650. https://doi.org/10.1038/nrg2611

Malpica, A., & De Granada, U. (2010). La Vida Urbana En Al-Andalus Y Su Papel En La Estructura Del Poblamiento Y En La Organización Social. *Imago TemporIs. medIum*

Aevum, *IV*, 401–420. Retrieved from http://www.raco.cat/index.php/ImagoTemporis/article/viewFile/256920/343962

Ministerio de Cultura. (2011). Los árabes en colombia, 168.

Oficina de Información Diplomatica del Departamento de Relaciones Exteriores. (2015). Ficha Pais - República de Colombia.

Pereira, R., Phillips, C., Alves, C., Amorim, A., Carracedo, Á., & Gusmao, L. (2009). A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*, *30*(21), 3682–3690. https://doi.org/10.1002/elps.200900274

Pereira, R., Phillips, C., Pinto, N., Santos, C., dos Santos, S. E. B., Amorim, A., … Gusmão, L. (2012). Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS ONE*, *7*(1). https://doi.org/10.1371/journal.pone.0029684

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, ??ngel, & Lareu, M. V. (2013). An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers in Genetics*, *4*(MAY), 1–13. https://doi.org/10.3389/fgene.2013.00098

Pritchard, J. K., Wen, X., & Falush, D. (2010). Documentation for structure software : Version 2 . 3. *University of Chicago. IL*, *6*(3), 321–326. https://doi.org/10.1002/spe.4380060305

Ramón, A., & Manilla, C. R. (n.d.). Estudio de mestizaje genético con marcadores ABO-Rh en 2 poblaciones de Veracruz ; antesala a un estudio genético de salud.

Rangel-ch, J. O. (2005). La biodiversidad de Colombia. *Palimpsestos*, *5*, 292–304. https://doi.org/10.1073/pnas.262413599

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., … Ruiz-Linares, A. (2012). Reconstructing Native American population history. *Nature*, *488*(7411), 370–374. https://doi.org/10.1038/nature11258

Rojas, M. Y., Alonso, L. a, Sarmiento, V. a, Eljach, L. Y., & Usaquén, W. (2013). Structure analysis of the La Guajira-Colombia population: a genetic, demographic and genealogical overview. *Annals of Human Biology*, *40*(2), 119–31. https://doi.org/10.3109/03014460.2012.748093

Romanini, C., Romero, M., Salado Puerto, M., Catelli, L., Phillips, C., Pereira, R., … Vullo, C. (2015). Ancestry informative markers: Inference of ancestry in aged bone samples using an autosomal AIM-Indel multiplex. *Forensic Science International: Genetics*, *16*, 58–63. https://doi.org/10.1016/j.fsigen.2014.11.025

Rousset, F. (2017). Genepop version 4.6.9, (March).

Russo, M. G., Di, F., Rocca, F., Doldán, P., Cardozo, D. G., Dejean, C. B., … Avena, A. (2016). Evaluación del número mínimo de marcadores para estimar ancestría individual

en una muestra de la población argentina, *9*(1), 49–56.

Saloum de Neves Manta, F., Pereira, R., Vianna, R., Rodolfo Beuttenmüller de Araújo, A., Leite Góes Gitaí, D., Aparecida da Silva, D., … Gusmão, L. (2013). Revisiting the Genetic Ancestry of Brazilians Using Autosomal AIM-Indels. *PLoS ONE*, *8*(9), 1–11. https://doi.org/10.1371/journal.pone.0075145

Salzano, F. M., & Bortolini, M. C. (2002). The Evolution and Genetics of Latin American Populations, *1*, 532. https://doi.org/10.1017/CBO9780511666100

Sitio oficial de la Gobernación del departamento del Chocó. (2016). Consultado el 17 de Julio, 2016 de http://www.choco.gov.co/informacion_general.shtml#historia.

Sitio oficial de la Gobernación del departamento del Huila. (2016). Consultado el 17 de Julio, 2016 de http://www.huila.gov.co/conoce-el-huila/resena-historica.html.

Sitio oficial de la Gobernación del departamento del Tolima. (2016). Consultado el 17 de Julio, 2016 de http://www.tolima.gov.co/publicaciones.php?id=1306.

Sitio oficial Portal Bogotá. (2016). Consultado el 17 de Julio, 2016 de www.bogota.gov.co/ciudad/historia.

Slatkin, M. (1994). Flujo Génico y Estructura de las Poblaciones, 1–11.

Suárez, C. A., & Quintin, C. I. (2004). Suárez. In *"Colombia Andina". Colombia: Norma. pp. 3-4. ISBN 9580468060*.

Wang, S., Lewis, C. M., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., … Ruiz-Linares, A. (2007). Genetic variation and population structure in Native Americans. *PLoS Genetics*, *3*(11), 2049–2067. https://doi.org/10.1371/journal.pgen.0030185