



ZTF-FCT
Zientzia eta Teknologia Fakultatea
Facultad de Ciencia y Tecnología

PROCESOS MICROEVOLUTIVOS EN POBLACIONES DE SUDAMÉRICA

Memoria de Tesis Doctoral dirigida por:

Dr. José Ángel Peña García

Dr. Miguel Ángel Alfonso Sánchez

para la obtención del grado de 'Doctor en Biología' de:

Luis Gómez Pérez

Departamento de Genética, Antropología Física y Fisiología Animal

Leioa, 2015

En biología nada tiene sentido, si no es a la luz de la evolución

Nothing in Biology Makes Sense, except in the Light of Evolution

Theodosius Dobzhansky (1900-1970) genetista ruso.

Este trabajo ha sido realizado gracias a la UPV/EHU, por varios motivos; en primer lugar por la posibilidad de cursar la carrera de Biología, así como el programa de Doctorado "Avances en Genética Molecular"; pero, en mayor medida por la concesión de una Beca de investigación predoctoral, sin la cual no hubiera sido posible realizar dicho proyecto. Por ello, me gustaría darles mi sincero agradecimiento a esta institución, así como a los rectores, decanos y demás personal que la ha regido durante este tiempo.

Igualmente, debo destacar el papel del Departamento de Genética, Antropología Física y Fisiología Animal, y de sus directores de Departamento, bajo el cual se ha llevado a cabo esta Tesis Doctoral.

A su vez, debo expresar mi gratitud a las instituciones que han colaborado para la realización de esta Tesis Doctoral, de manera directa o indirecta, mediante financiación económica o mediante recursos humanos o materiales; como la financiación de grupos SAIOTEK, Subvención para Unidades de Formación e Investigación ELDUNANOTEK (UPV/EHU UFI11/32), Subvención para Actividades de Grupos de Investigación Consolidados BIOMICS (Gobierno Vasco IT-833-13 y Gobierno Vasco IT-424-07), así como el Instituto de la Altura de Jujuy, la Universidad de Zaragoza y la Universidad Nacional de Colombia.

AGRADECIMIENTOS

Es para mí un placer utilizar este espacio para mostrar mi agradecimiento a aquellas personas que me han ayudado y apoyado durante el largo proceso que ha sido realizar la Tesis Doctoral.

El primer lugar, debo agradecer de manera especial y sincera a mis directores de Tesis Doctoral, el Dr. **José Ángel** Peña García y el Dr. **Miguel** Ángel Alfonso Sánchez, por varios motivos: por aceptarme para realizar esta Tesis Doctoral bajo su dirección; por su disponibilidad y paciencia para colaborar conmigo; por sus consejos, apoyo y confianza en mi trabajo; por su capacidad para guiar mis ideas, tarea que ha sido imprescindible para llevar a buen puerto este trabajo; por el tiempo que han invertido en mi formación como investigador. Debo agradecerles también el haberme facilitado siempre los medios y la financiación necesarios para el desarrollo de esta Tesis Doctoral. Pero sobre todo por el afecto y trato que han tenido para conmigo. Durante el tiempo de ocio, hemos pasado buenos momentos en forma de charlas, aportes personales y muy variadas y pintorescas reflexiones.

Quiero expresar también mi más sincero agradecimiento a **Susana** García Obregón por su ayuda e iniciación en los análisis experimentales de laboratorio, así como sus consejos en los momentos difíciles. Por supuesto, debo destacar que su colaboración en el desarrollo de esta Tesis ha sido muy importante, sin la cual hubiera sido imposible realizarla.

Del mismo modo debo agradecer a **Ibone** Espinosa Carcabilla. Su apoyo, participación activa en multitud de experimentos de laboratorio, aporte de ideas, su inagotable imaginación y su inestimable colaboración han sido de enorme valor. Hemos compartido buenos y malos momentos. Seguramente todos ellos han tenido su importancia para poder llevar a buen puerto este proyecto. Espero haberla aportado tanto como ella a mí y haberla orientado en la medida de lo posible en su Tesis Doctoral, para devolverla su ayuda, aunque sea tan solo una parte.

También debo destacar los aportes de experiencia, consejos, apoyo y orientación de la Dra. **Ana** M^a Pérez Miranda. Sus amplios conocimientos en materia de laboratorio han sido, sin duda, una ayuda impagable.

No puedo olvidar mencionar a los alumnos internos y estudiantes de máster que han pasado por el laboratorio a lo largo de estos años (Miguel, Ángela, Isa, Leyre, Ane, Ainara, Lorena, Iñigo, Tamara...) que colaboraron de manera importante o me dieron su apoyo.

Entre los profesores e investigadores miembros de nuestro grupo de investigación debo agradecer, en un papel destacado, a la Doctora **Marian** Martínez de Pancorbo y al Doctor **Sergio** Cardoso. Ellos han sido las cabezas visibles de todo su equipo en Vitoria, al que extiendo mi agradecimiento por la colaboración, así como la impartición de varios cursos, que han resultado básicos para mejorar mis conocimientos y aptitudes científicas y por el suministro de muestras proveniente de sus instalaciones.

Además, debo reconocer los aportes del doctor José E. Dipierri, la Dra. Dora Sánchez, la Dra. María Luisa Bravo y la Dra. Maite Zarrabeitia, y demás miembros de sus equipos de investigación, que han colaborado suministrando varias de las muestras empleadas en los análisis de la Tesis.

No podría dejar de lado a mi familia, que ha sido importantísima para mí durante estos años, encabezada por mis padres, **Luis** y **Ásun**, que me apoyaron en todos los aspectos para que realizara la Tesis Doctoral, aunque no sabían dónde se metían. Ellos me han apremiado para que la acabara lo antes posible, ejerciendo de una imprescindible, aunque no siempre efectiva, presión. También a mi hermana, **Laura**, mi imagen especular, especialista en el *brainstorming* y oyente (que no "escuchante") de pruebas y presentaciones. Ha sido, y seguirá siendo, un gran y visceral apoyo emocional. A mis **abuelos**; a todos mis **tíos** y **tías**; a mi primos (**Natalia, Valeria y Leo**) y **demás familia**,... todos ellos han participado de este proyecto y llevan esperando éste momento con mucha expectación; espero no defraudarles.

Por otro lado, agradecer a la gente que me ha apoyado y animado: ya sean amigos (Vali, Germán, Álvaro, Zazo, Alex, Sergio, Héctor, Dani, More, Vadó,...), con sus pesadas y reiterativas preguntas acerca del tema o de otras cuestiones no adecuadas para este espacio; o amigas (Bea, Celia, Natalia, Montse, Paula, Cristina, Sandra,...) siempre más correctas y constructivas; compañeros de la carrera, como Raquel, Aline, Elda, María, Lucía... que se han interesado por mis "avances"; A **Rosa e Iván** por sus imprescindibles aportes en materia de estilo; También a los muchos y buenos compañeros de piso que he tenido en estos años.

A **Andoni**; *conspiranoico* antagonista, catalizador de ideas, *partner* de proyectos imposibles e incansable discutiador; que siempre me ha hecho repensar las ideas preestablecidas y darle una vuelta de tuerca más.

A los vecinos del laboratorio (Lara, Alaitz, Saioa, Montse, Maru, Iratxe, Imanol, Arrate) siempre dispuestas a ayudar, prestar o colaborar en lo necesario. También a la secretaria del departamento Marifé y a María Jesús.

No olvido a **Nagore**, que fue una de las personas más importantes a la hora de tomar la decisión de realizar la Tesis, porque fue la primera en estimularme y en animarme a decidirme por este proyecto; no la guardo rencor por ello.

Pero sin duda, a **Ángela**, *mi alma melliza*. No sé si esta Tesis habría sido posible sin ella, pero seguro que habría sido muy diferente. Ella es la que más tiempo lleva aguantando este largo proceso, ha sobrellevado más que nadie los malos momentos (mas suyos que míos), las dificultades y los retrasos. Y ha presionado, sin duda, para que la terminase; ha presionado mucho... por evidentes motivos. Pues ¡ya está! Así que ¡Muchas Gracias, Amor!

Por último, no puedo olvidarme de agradecer, en general, a todos los que se han interesado por mí y mi Tesis durante estos años.

ÍNDICE

RESUMEN	1
INTRODUCCIÓN.....	5
Procesos de Colonización del Continente Americano	15
Poblamiento inicial del continente	16
Poblamiento de origen europeo y africano	26
Poblaciones estudiadas	31
Provincia de Antioquia (Colombia)	32
Población mestiza	34
Afrocolombianos	36
Jujuy (Argentina)	39
Etnia Huaorani (Ecuador)	46
Marcadores Genéticos	51
Los elementos <i>Alu</i>	51
Origen y estructura de los elementos <i>Alu</i>	54
Retrotransposición	56
Clasificación de las subfamilias <i>Alu</i>	58
Inserciones <i>Alu</i> en estudios Genético-Evolutivos.....	61
Elementos <i>Alu</i> autosómicos analizados	63
El Complejo Mayor de Histocompatibilidad (MHC)	70
Inserciones MHC- <i>Alu</i> analizadas	72
MHC-STRs analizados.....	75
HIPÓTESIS DE TRABAJO.....	81
OBJETIVOS DE LA TESIS DOCTORAL.....	85
MATERIAL Y MÉTODOS.....	91
Materiales	93

Muestras Poblacionales	93
Mestizos colombianos.....	93
Afrocolombianos	93
Jujeños de Argentina	94
Huaoranis	97
Métodos	98
Extracción del ADN.....	98
Método Fenol-Cloroformo-Isoamil Alcohol.....	98
Extracción mediante kit comercial en columna.....	100
Marcadores genéticos	102
Inserciones <i>Alu</i>	102
Microsatélites.....	105
Análisis estadísticos.....	106
Frecuencias Alélicas.....	106
Frecuencias Haplotípicas	107
Error Estándar de la Media	107
Diversidad alélica.....	107
Diversidad haplotípica.....	107
Prueba de Kolmogórov-Smirnov	108
Test del desequilibrio de ligamiento	109
Índice Garza-Williamson.....	110
Test de Neutralidad de Ewens-Watterson.....	110
<i>Networks</i>	111
AMOVA.....	111
Método del centroide.....	112
Estimas del mestizaje.....	113

Clinas	114
Coefficientes de distancia genética	115
Diferenciación poblacional	116
Análisis de escalamiento multidimensional	117
Dendrogramas	118
Test de Mantel	120
Bootstrap	121
Análisis discriminante	122
RESULTADOS.....	123
Manuscrito I. Utilidad de las inserciones <i>Alu</i> en los estudios de mestizaje	127
Manuscrito II. Genetic admixture estimates by <i>Alu</i> elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia	143
Manuscrito III. Microevolutionary Processes due to Landscape Features in the Province of Jujuy (Argentina).....	163
Manuscrito IV. <i>Alu</i> Polymorphisms in the Waorani Tribe from the Ecuadorian Amazon Reflect the Effects of Isolation and Genetic Drift	179
Manuscrito V. Young <i>Alu</i> insertions within the MHC class I region in Native American populations: insights into the origin of the MHC- <i>Alu</i> repeats	193
Manuscrito VI. Jujuy province (NW Argentina): STR loci unveil genetic footprints of microevolutionary responses to a steep orography.....	211
DISCUSIÓN	237
CONCLUSIONES.....	249
REFERENCIAS.....	255
MATERIAL ANEXO.....	285

RESUMEN

La serie de trabajos aquí presentados se han realizado con el fin de valorar los efectos del flujo génico y la deriva génica sobre el patrimonio genético de poblaciones que presentan características diferentes, tanto desde un punto de vista geográfico y ambiental, como cultural e histórico. Para ello, se han analizado 21 marcadores genéticos de origen autosómico (17 inserciones *Alu* y 4 microsatélites) en poblaciones sudamericanas: afro-descendientes y mestizos de Colombia, amerindios de la provincia de Jujuy, en Argentina, y nativos americanos de la etnia Huaorani de Ecuador. Los marcadores genéticos analizados se han mostrado eficientes para los objetivos propuestos. Con el fin de aumentar el poder de resolución de los análisis, los marcadores utilizados presentan diferentes características, además, varios marcadores, localizados en el MHC y con desequilibrio de ligamiento, se han utilizado en forma de haplotipos.

Se ha encontrado que en la población mestiza de Colombia existe un componente europeo mayoritario; a su vez muestra un importante aporte de las poblaciones amerindia y africana, con un rango de valores similares a los obtenidos en otras poblaciones próximas. Por el contrario, los individuos colombianos de ascendencia africana han mostrado muy poco mestizaje con otras etnias; ello posiblemente se deba a un aislamiento de origen social, principalmente debido a circunstancias asociadas a la esclavitud, condición de la que participaban los fundadores de sus poblaciones.

En Jujuy, el impacto de las poblaciones colonizadoras europeas sobre las nativas se ha revelado como poco influyente. Los resultados han mostrado que esta población tiene un acervo genético mayoritariamente amerindio. Por otro lado, esta población ha mostrado una variabilidad en su grado de mestizaje en relación con la altitud. A tenor de los resultados obtenidos en Jujuy, el medio que habita esta población se ha revelado como un factor capaz de condicionar el flujo génico, lo que combinado con los efectos de la deriva génica, debido a sus pequeños tamaños poblacionales, ha potenciado los procesos de microdiferenciación entre las regiones de Jujuy.

En el caso de los Huaorani se ha observado que el aislamiento reproductivo que han sufrido, producto de algunas prácticas socioculturales asociadas a conductas violentas, con los consiguientes altos niveles de endogamia, ha desembocado en una población con una diversidad genética muy reducida. Además, presentan una especial susceptibilidad a la acción de la deriva genética debido a su escaso tamaño poblacional. Debido a ello, se ha revelado como una población con un alto grado de diferenciación con otras poblaciones amerindias.

Así, la orografía en Jujuy, el devenir histórico de los afrocolombianos y las tradiciones de los Huaorani han condicionado en gran medida el balance flujo/deriva. En el caso que la incidencia de estos factores sobre las poblaciones persistiera el tiempo suficiente terminaría moldeando su patrimonio genético. De esta forma se ha llegado a la conclusión de que los procesos microevolutivos pueden verse afectados por diferentes factores, tanto intrínsecos como externos a las poblaciones.

INTRODUCCIÓN

La variabilidad genética humana es una valiosa fuente de información que utilizan varias disciplinas, como la Antropología, la Biología Humana, la Antropogenética, la Genética de Poblaciones, la Genética Médica o la Genética Forense. Paradójicamente, la especie humana puede considerarse relativamente homogénea, lo que dificulta la tarea de caracterizar dicha variabilidad. Ya durante la segunda mitad del siglo XX, a partir del análisis de marcadores de determinación indirecta (serológicos) se encontraron indicios de que la diversidad genética de los seres humanos era baja en relación a otras especies. Posteriormente, estos datos han sido confirmados por el análisis de marcadores de determinación directa (genéticos) (Li y Sadler, 1991; Zhivotovsky *et al.* 2003; Barbujani y Colonna, 2010).

Esta característica ha sido interpretada como el resultado de un corto periodo de evolución, de modo que el ser humano pertenecería a una especie relativamente joven y muy posiblemente con una población fundadora de pequeño tamaño, de no más de 10.000 individuos (Hawks *et al.* 2000) o incluso menos (Fagundes *et al.* 2007). Por ello, las poblaciones habrían tenido relativamente poco tiempo para divergir unas de otras (Harpending y Rogers, 2000; Zhivotovsky *et al.* 2003). Así, dos personas tomadas al azar difieren en alrededor de 1 de cada 1.000 pares de nucleótidos (0,1%), mientras que dos chimpancés difieren en alrededor de 1 de cada 500 pares de nucleótidos (0,2%) (Fischer *et al.* 2004). En datos absolutos, dos seres humanos diferirán en promedio en alrededor de 6,4 millones de nucleótidos.

El análisis de F_{ST} (Wright, 1950, 1951, 1965), una medida de variabilidad interpoblacional, muestra un panorama similar. Los valores promedio de F_{ST} para la especie humana oscilan entre 0,05 y 0,13, dependiendo del polimorfismo genético analizado, lo que da una idea de la homogeneidad de nuestra especie (Barbujani y Colonna, 2010). Nuestra diversidad equivale aproximadamente a una tercera parte de la mostrada por los gorilas (F_{ST} de 0,38) (Thalmann *et al.* 2007). Otro ejemplo lo expone Zhivotovsky *et al.* (2003) donde explica que la variabilidad del mtDNA de nuestra especie es similar a la de un pequeño grupo social de chimpancés.

Por otra parte, la mayor parte de la variabilidad que se puede encontrar entre individuos se debe a polimorfismos neutros o prácticamente neutros, lo que implica que probablemente sólo una fracción de los polimorfismos sean efectivos y por lo tanto responsables de las diferencias fenotípicas observables entre individuos (Fay *et al.* 2002).

La distribución de la variabilidad de los polimorfismos neutrales en la especie humana refleja su historia demográfica (Harpending y Rogers, 2000). Evidencias genéticas y arqueológicas indican que durante los últimos 100.000 años los seres humanos se han dispersado desde África para colonizar el resto del mundo, aumentando progresivamente su tamaño de población (Pickrell y Reich, 2014; Sousa *et al.* 2014). Este proceso ha afectado a la distribución geográfica de la variación genética de dos maneras importantes: *i)* las sucesivas poblaciones colonizadoras por lo general portan sólo una parte de la variación genética total de la población ancestral, debido al proceso conocido como efecto fundador. Por lo tanto, la población que no se desplaza debería contener una mayor variabilidad genética; éste es el caso de las poblaciones africanas. *ii)* ya que los grupos fundadores a menudo colonizan amplias regiones cada vez más separadas entre sí, la probabilidad de emparejamiento entre los individuos se hace cada vez más dependiente de la distancia. Con el tiempo, los emparejamientos selectivos han provocado un aumento de la heterogeneidad genética interpoblacional (Bamshad *et al.* 2004; DeGiorgio *et al.* 2011; Slatkin y Excoffier, 2012).

El análisis de la variación genética en las poblaciones ayuda a encontrar respuestas a numerosas cuestiones evolutivas. Habitualmente, la evolución se divide para su estudio en dos niveles, microevolutivo y macroevolutivo. Un proceso microevolutivo implica un cambio en las frecuencias alélicas; se encuentra restringido a poblaciones dentro de una especie, generalmente durante un pequeño número de generaciones. La macroevolución, por su parte, se relaciona con la formación de especies u otras categorías taxonómicas mayores (Carroll, 2001; Relethford, 2012). Aunque la microevolución y la macroevolución se relacionan en un sentido teórico, hay un continuo debate sobre el grado en que los acontecimientos macroevolutivos son una extrapolación

directa de tendencias microevolutivas a largo plazo (Simons, 2002), es decir si la macroevolución puede ser explicada por acumulación de cambios microevolutivos o si por el contrario existen otros mecanismos y en ese caso, dónde se encontrarían los límites entre ambas escalas (Carroll, 2001; Hendry y Kinnison, 2001). Los mecanismos o fuerzas implicadas en el proceso evolutivo son la deriva genética, el flujo génico, la mutación y la selección natural. En el caso de los procesos microevolutivos son considerados de especial interés los dos primeros, mostrando una menor incidencia la mutación y la selección natural (Relethford, 2012).

Los efectos de la deriva genética sobre la variabilidad genética son una función del tamaño poblacional efectivo (N_e), o número de individuos potencialmente reproductores de cada generación; por lo cual, grandes poblaciones efectivas pueden mantener niveles de variabilidad altos, mientras que las pequeñas están más sujetas a fluctuaciones aleatorias en las frecuencias alélicas y a pérdidas de variabilidad (Tishkoff y Verrelli, 2003). Por la acción de la deriva genética, algunas mutaciones pueden alcanzar unas frecuencias mínimas que les permita mantenerse en el patrimonio genético, incrementando el número de variantes raras en las poblaciones, o ciertas variantes alélicas neutras pueden llegar a fijarse (Halligan y Keightley, 2009; Sousa *et al.* 2014).

Por el contrario, el flujo génico tiene capacidad homogeneizadora por el hecho de intercambiar genes entre poblaciones. El flujo génico es dependiente de la tasa de migración (m) y del tamaño poblacional efectivo (N_e). La intensidad del flujo génico determina que una población se convierta en una unidad evolutiva independiente o que se mantenga evolutivamente ligada al resto de poblaciones de una especie (Slatkin, 1994b). El equilibrio entre Flujo génico y deriva genética afecta de manera importante a la distribución de la variabilidad genética en las poblaciones y a su heterogeneidad genética interpoblacional (Slatkin, 1994b; Relethford, 2012; Slatkin y Excoffier, 2012).

El desarrollo de investigaciones sobre la influencia del territorio en la genética (*the landscape genetics*) (Manel *et al.* 2003; Manel y Holderegger, 2013) ha permitido que se reconozca la

heterogeneidad espacial como un factor clave para explicar el flujo génico diferencial, así como otros procesos de microdiferenciación (McRae y Beier 2007; Duforet-Frebourg y Blum, 2014), siendo para algunos autores un factor clave para entender las diferencias entre macroevolución y microevolución (Simpson, 1944; Simpson, 1953; Arnold *et al.* 2001).

La acumulación de diferencias genéticas entre grupos lleva a una estructuración de la diversidad genética en la forma de alelos específicos de poblaciones o, en ocasiones, diferencias significativas entre las frecuencias alélicas. Esto hace posible la caracterización de marcadores informativos de ancestralidad (AIMs, *Ancestry Informative Markers*), que pueden ser utilizados para identificar el origen de una población o un individuo. Además, los AIMs se pueden utilizar para estimar tasas de mestizaje y son muy útiles para aplicaciones forenses, clínicas y poblacionales (Shriver y Kittles, 2004; Manta *et al.* 2013; Elhaik, 2014). Cabe señalar que, para la identificación y validación de los AIMs, es necesario un conocimiento de la composición genética de las poblaciones originarias de los diferentes continentes o subcontinentes. Si estas poblaciones se encuentran mal caracterizadas, las estimaciones de las proporciones de ascendencia pueden ser inexactas. Actualmente los AIMs que se encuentran disponibles se utilizan para distinguir entre un pequeño número de poblaciones fuente o parentales claramente diferenciadas, (fundamentalmente europeos, asiáticos, amerindios y africanos). Sería muy interesante desarrollar AIMs con el fin de distinguir los aportes del flujo génico entre grupos más estrechamente relacionados (Bamshad *et al.* 2004; Shriver y Kittles, 2004; Li *et al.* 2014).

A lo largo de los trabajos realizados en esta Tesis Doctoral, se ha tratado de evaluar el impacto de los procesos de flujo génico y su interacción con la deriva genética sobre la diversidad genética de varias poblaciones sudamericanas. La diversidad biológica de América proporciona ejemplos muy interesantes de la compleja interacción entre factores biológicos y sociales en la construcción de identidades de los grupos humanos y de las dificultades en la definición de las poblaciones de nuestra especie (Ruiz-Linares, 2014). Esta interacción abarca desde sucesivos

procesos de efecto fundador y cuello de botella hasta una compleja orografía que mediante importantes barreras geográficas ha proporcionado una amplia gama de biomas (Zhivotovsky *et al.* 2003; Schurr, 2004; Roewer *et al.* 2013). Por ello, las poblaciones sudamericanas presentan una oportunidad única de analizar el efecto de la acción de los procesos microevolutivos con grados de intensidad variable sobre el patrimonio genético de las poblaciones (Watkins *et al.* 2012).

Sin embargo, un estudio acerca de la estructura de las poblaciones debe tener presente que además de los factores geográficos y de los movimientos poblaciones, existen otros factores lingüísticos y socioculturales que pueden influir sobre la estructura genética de las poblaciones humanas (Jorde, 1980; Jorde *et al.* 1982, Wood *et al.* 1982; Relethford, 1996; Manel *et al.* 2003; McRae y Beier, 2007; Capocasa *et al.* 2014); afectando a las tasas de flujo génico e indirectamente a la acción de la deriva genética, promoviendo bajos tamaños poblacionales u ocasionando reducciones drásticas en él (Relethford, 2012).

Con el fin de considerar un amplio rango de escenarios, se han seleccionado poblaciones que presentaran notables diferencias en su composición étnica, desarrollo histórico y emplazamiento geográfico y de este modo poder examinar una amplia casuística de procesos microevolutivos.

Para ello se han seleccionado una muestra de población general mestiza y otra de afroamericanos, ambas de Antioquia, en Colombia, varias muestras de amerindios procedentes de diferentes regiones de Jujuy, en el Noroeste Argentino y por último una muestra de individuos amerindios de la etnia Huaorani, procedente de Ecuador.

A su vez, los marcadores seleccionados pertenecen a dos grupos claramente diferenciados con el fin de obtener resultados complementarios:

Por un lado, se han utilizado varias inserciones *Alu*, las cuales se estima que representan alrededor del 11% del genoma humano, con un promedio de un elemento cada 4 kb (Lander *et al.* 2001, Ade *et al.* 2013). Los elementos *Alu* se caracterizan por ser marcadores bialélicos, con un

patrón codominante, identificados por la presencia o ausencia de la inserción. Estos marcadores son útiles en Antropogenética por una serie de interesantes propiedades que los caracterizan: *i*) Los elementos *Alu* son considerados idénticos por descendencia, debido a que la probabilidad de que dos elementos se inserten de manera independiente en el mismo lugar del genoma es altamente improbable (Batzer *et al.* 1994; Roy-Engel *et al.* 2002); *ii*) además son considerados selectivamente neutros, por lo que la distribución de sus frecuencias se encuentra afectada exclusivamente por el flujo génico y la deriva (Comas *et al.* 2000); *iii*) son marcadores altamente estables, debido a que no se conocen mecanismos de delección (Edwards y Gibbs, 1992), con una prácticamente nula probabilidad de una completa delección (Willard *et al.* 1987; Britten, 1997); y *iv*) se conoce el estado ancestral del polimorfismo, que es la ausencia de la inserción, de modo que el evento mutacional es la aparición de la misma.

Estas propiedades hacen de los elementos *Alu* unos marcadores eficientes y fiables para los estudios poblacionales (Batzer *et al.* 1994; Novick *et al.* 1998; Batzer y Deininger, 2002; Herrera *et al.* 2007; Bahri *et al.* 2013). En esta Tesis Doctoral se han analizado un total de 17 inserciones, repartidas por diferentes cromosomas, incluido un grupo situado en el Complejo Mayor de Histocompatibilidad.

Por otro lado, también se han utilizado varios microsatélites. Los microsatélites o STRs son marcadores caracterizados por la repetición en tándem, un número variable de veces, de un fragmento de entre 2 y 6 pares de bases (Hearne *et al.* 1992). Los STRs son marcadores muy informativos debido a su naturaleza hipervariable (Hearne *et al.* 1992; Gill *et al.* 1994; Meyer *et al.* 1995; Rowold y Herrera, 2005). Pueden resultar muy útiles para caracterizar eventos de flujo génico y deriva genética y por tanto, aportar información sobre algunos aspectos de los procesos microevolutivos más recientes (Bowcock *et al.* 1994; Brinkmann, 1996; Brinkmann *et al.* 1996; Zhivotovsky *et al.* 2003; García-Obregón *et al.* 2011). Esta capacidad se debe principalmente a la elevada tasa de mutación que presentan (entre 10^{-3} y 10^{-4} por locus por generación) (Brinkmann *et*

al. 1998; Zhivotovsky *et al.* 2000; Zhivotovsky *et al.* 2003; Mardini *et al.* 2013; Sun *et al.* 2014.). Estos marcadores, además, son muy abundantes y se encuentran distribuidos por todo el genoma (Jorde *et al.* 1997; Zhivotovsky *et al.* 2003).

Con el presente trabajo se ha pretendido profundizar en los factores que afectan al flujo génico y la deriva genética, promoviendo procesos de microevolución y modelando así el acervo genético de las poblaciones. Esto ha sido posible gracias a las peculiaridades de este subcontinente, anteriormente descritas, que han promovido grados de aislamiento variables entre las poblaciones de diferentes regiones (Mesa *et al.* 2000; Tarazona-Santos *et al.* 2001; Malhi *et al.* 2002; Rosenberg *et al.* 2002; Zhivotovsky *et al.* 2003; Schurr, 2004; Battilana *et al.* 2006; Wang *et al.* 2007; Roewer *et al.* 2013).

PROCESOS DE COLONIZACIÓN DEL CONTINENTE AMERICANO

América fue el último de los continentes en ser poblado por el hombre (Goebel *et al.* 2008). Los motivos para explicar este hecho son la lejanía respecto de África y las notables barreras geográficas, con condiciones climáticas extremas en muchos casos, que hay que superar.

Entre las poblaciones amerindias actuales se observa una notable paradoja, ya que muestran una gran variedad étnica y una elevada heterogeneidad genética interpoblacional, a pesar de que su diversidad genética es relativamente escasa (Rosenberg *et al.* 2002; Zhivotovsky *et al.* 2003; Battilana *et al.* 2006; Estrada-Mena *et al.* 2010). Así, la diversidad genética del ADNmt entre los nativos americanos modernos comprende básicamente unos pocos haplogrupos, conocidos como Haplogrupos Fundadores (A2, B2, C1, D1, X2a). Dentro de los cuatro primeros pueden distinguirse diez sub-haplogrupos monofiléticos (A2, B2, C1b, C1c, C1d, C4c, D1, D2A, D3 y D4h3a), a los que se debe añadir el haplogrupo X2a (Bandelt *et al.* 2003; Kemp *et al.* 2007; Mahli *et al.* 2007; Tamm *et al.* 2007; Achilli *et al.* 2008; Volodko *et al.* 2008; Malhi *et al.* 2010).

En el caso del cromosoma Y, tan sólo se han observado dos haplogrupos principales: El haplogrupo Q (principalmente Q1a3a y raramente el Q1a3*), que se encuentra distribuido por todo el continente; y el haplogrupo C, donde el C3b está aparentemente circunscrito al Norte de América (especialmente en los pueblos Na-Dene) y el C3* presente hasta la fecha en unos pocos individuos en Sudamérica (principalmente Ecuador) y tan solo uno en Alaska (Bortolini *et al.* 2003; Zegura *et al.* 2004; Malhi *et al.* 2008; Battaglia *et al.* 2013; Roewer *et al.* 2013).

Esta situación parece haber sido el resultado de una larga serie de procesos sucesivos de efecto fundador y cuello de botella, de intensidad moderada o alta, tanto durante el proceso de poblamiento de América, como después del mismo (Mesa *et al.* 2000; Rosenberg *et al.* 2002; Zhivotovsky *et al.* 2003; Battilana *et al.* 2006; Raff *et al.* 2011; Roewer *et al.* 2013). Pero también se explica por la existencia de importantes barreras geográficas y una amplia gama de biomas,

generados por una compleja orografía (Tarazona-Santos *et al.* 2001; Malhi *et al.* 2002; Schurr 2004), lo que ha propiciado un cierto grado de aislamiento entre las poblaciones de diferentes regiones.

El actual patrimonio genético del continente americano, considerando todas sus poblaciones, no sólo las nativas americanas, es en esencia el resultado de dos eventos de poblamiento fundamentales ocurridos a lo largo de su historia. Por un lado, el poblamiento inicial por el cual el homo sapiens anatómicamente moderno (AMHS) cuando colonizó por vez primera el continente; y por el otro lado, mucho más recientemente, la llegada de colonos europeos y esclavos africanos que originó numerosos procesos de mestizaje en grado variable entre los tres grupos poblacionales implicados así como otros fenómenos demográficos de gran escala: guerras, esclavitud, deportaciones masivas, epidemias, etc.

POBLAMIENTO INICIAL DEL CONTINENTE

A partir del registro fósil se ha inferido que el continente americano no fue poblado hasta la llegada del hombre moderno (AMHS), habiéndose mantenido alejado de las expansiones de estadios anteriores del género *Homo* (Fagan, 2004; Goebel *et al.* 2008; Fernández-Gómez y Velasco-Ortiz, 2011).

En cuanto al origen de los primeros pobladores, actualmente existe un amplio consenso acerca de que las poblaciones de Nativos americanos descienden de grupos de centro-asiáticos que colonizaron el Noreste de Siberia (*Figura 1; Figura 2*). Desde allí y a través del estrecho de Bering, alcanzaron el continente Americano (Tamm *et al.* 2007; Kitchen *et al.* 2008; Fagundes *et al.* 2008b). Lo hicieron durante la última glaciación, ya que el nivel del mar experimentó un notable descenso (Clark y Mix, 2002), lo que provocó que el estrecho de Bering fuera durante varios miles de años un continuo de tierra, denominado Beringia, que unía Asia con América (*Figura 3*). A pesar de la elevada latitud de esta región, existen estudios que sugieren que pudo ser habitable en ese periodo, ya que no estuvo cubierta completamente por el hielo y disfrutaba de un clima

relativamente benigno (Guthrie, 1990; Szathmary, 1993; Zazula *et al.* 2003; Brubaker *et al.* 2005; Hoffecker *et al.* 2014).

La hipótesis del origen centro-asiático de las poblaciones nativas americana está sustentada en evidencias obtenidas tanto a partir de análisis genéticos (marcadores nucleares, ADMmt y cromosoma Y) como de restos fósiles (Merriwether *et al.* 1995, 1996a, b, 2006; Kolman *et al.* 1996; Crawford MH, 1998; Karafet *et al.* 1999, 2006; Mulligan *et al.* 2004; Schurr, 2004; Fagan, 2004; Tamm *et al.* 2007; Achilli *et al.* 2008; Fagundes *et al.* 2008a; Goebel *et al.* 2008; Bourgeois *et al.* 2009; Perego *et al.* 2009, 2010; Schroeder *et al.* 2009). Así, prácticamente todos los linajes presentes en amerindios, tanto de ADNmt (A2, B2, C1, D1) como del Cromosoma Y (C y Q), se encuentran entre las poblaciones del sur de Siberia, desde el Macizo de Altai a las regiones del Amur (Derenko *et al.* 2001; Zegura *et al.* 2004; Starikovskaya *et al.* 2005; Battaglia *et al.* 2013). Únicamente el haplogrupo X2a se ha encontrado fuera de estas regiones (concretamente en poblaciones de Asia Central y Europa), si bien su presencia podría explicarse por su gran diversidad, perteneciendo probablemente a diferentes linajes (Derenko *et al.* 2001, 2007; Zegura *et al.* 2004; Starikovskaya *et al.* 2005; Merriwether, 2006; Karafet *et al.* 2006).

A pesar del amplio consenso mencionado, se han propuesto otras hipótesis alternativas que mencionan otros lugares de procedencia para la población nativa americana, en general escasamente fundamentadas. Cabe destacar entre ellas la hipótesis del Solutrense, que sitúa el origen de la cultura Clovis (considerada durante casi todo el Siglo XX como la cultura indígena más antigua del continente americano) en poblaciones del Paleolítico Superior presentes en la Península Ibérica (*Figura 1*) (Holden, 1999; Standford y Bradley, 2002; Bradley y Standford, 2004). Las evidencias aportadas se limitan a las similitudes encontradas entre las herramientas utilizadas en la industria lítica Solutrense y la cultura Clovis (Goebel *et al.* 2008). Durante mucho tiempo ha sido una hipótesis minoritaria, siendo finalmente desechada a partir de la secuenciación en fecha reciente del genoma de un individuo de hace 13.000 años asociado a la cultura Clovis. Por la

comparación de genomas de individuos de diferentes continentes, se ha observado que este resto (Anzick) es más similar a individuos actuales sudamericanos que a ningún otro individuo (Rasmussen *et al.* 2014). De forma similar, un esqueleto casi completo, de 12.000-13.000 años de antigüedad, encontrado en una cueva submarina del Yucatán (México) ha revelado unas características craneofaciales típicas de paleoamerindios y un ADNmt de tipo D1, derivado de Beringia. No mostrando, por tanto, ningún posible parentesco con poblaciones de origen europeo (Chatters *et al.* 2014).

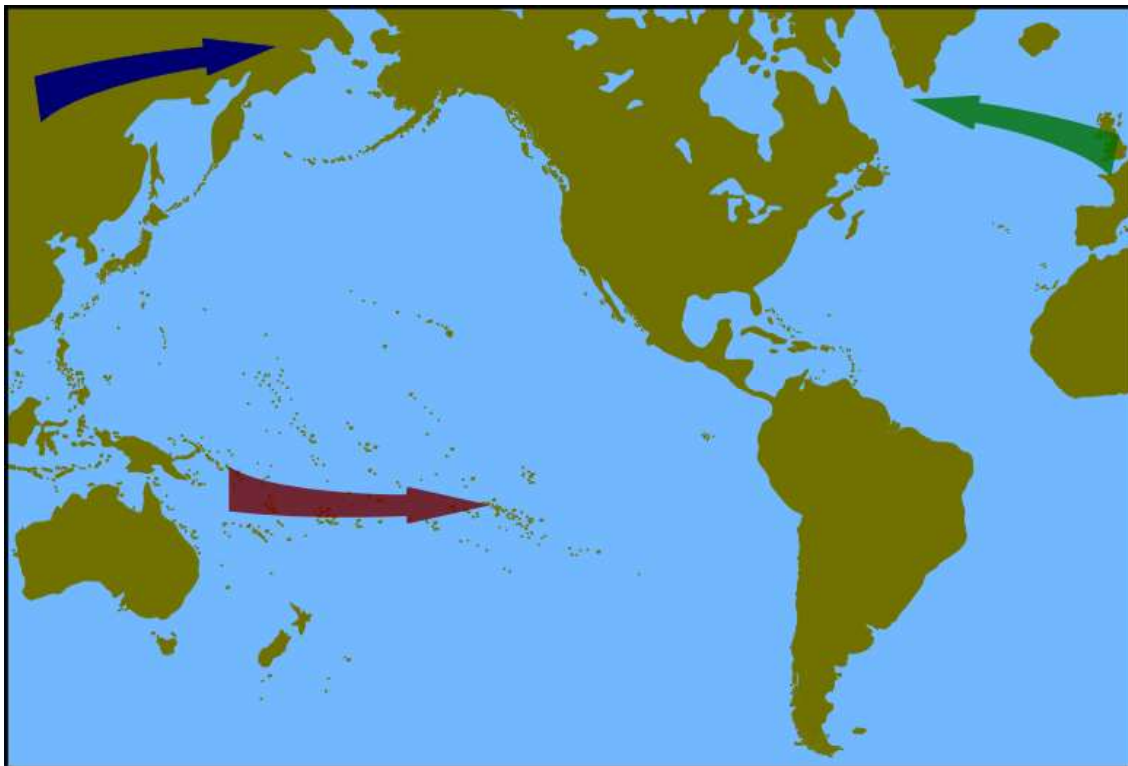


Figura 1. Posibles rutas de colonización del continente Americano. i) La flecha azul muestra la ruta a través de Beringia, actualmente la hipótesis generalmente aceptada. ii) La flecha roja muestra una posible ruta a través del Pacífico, por la que las poblaciones Nativas Americanas estarían emparentadas con poblaciones de la Polinesia. iii) La flecha verde muestra la ruta que habrían seguido las poblaciones europeas para alcanzar América, según la controvertida hipótesis del Solutrense.

Otras hipótesis proponen una ruta a través del Océano Pacífico (*Figura 1*), con poblaciones originarias provenientes de Australia o de las islas del Pacífico e incluso de la costa del sudeste

asiático (Cerna *et al.* 1993; León-S *et al.* 1996), que no presentan evidencias suficientes como para que sean consideradas como una alternativa plausible (Straus, 2000; Goebel *et al.* 2008).

Considerando las poblaciones centro-asiáticas como las antecesoras y Beringia como el punto de acceso más verosímil para el poblamiento de América, existe sin embargo todavía una considerable controversia acerca del número de migraciones, la fecha o fechas de entrada, la ruta seguida y el tamaño efectivo de los grupos colonizadores (Wallace y Torroni, 1992; Torroni *et al.* 1993; Forster *et al.* 1996; Kaufman y Golla, 2000; Goebel *et al.* 2003, 2008; Schurr y Sherry 2004; Wang *et al.* 2007; Waters y Stafford, 2007; Dillehay *et al.* 2008; Gilbert *et al.* 2008b; O'Rourke y Raff, 2010; Reich *et al.* 2012; Battaglia *et al.* 2013; Raghavan *et al.* 2014).

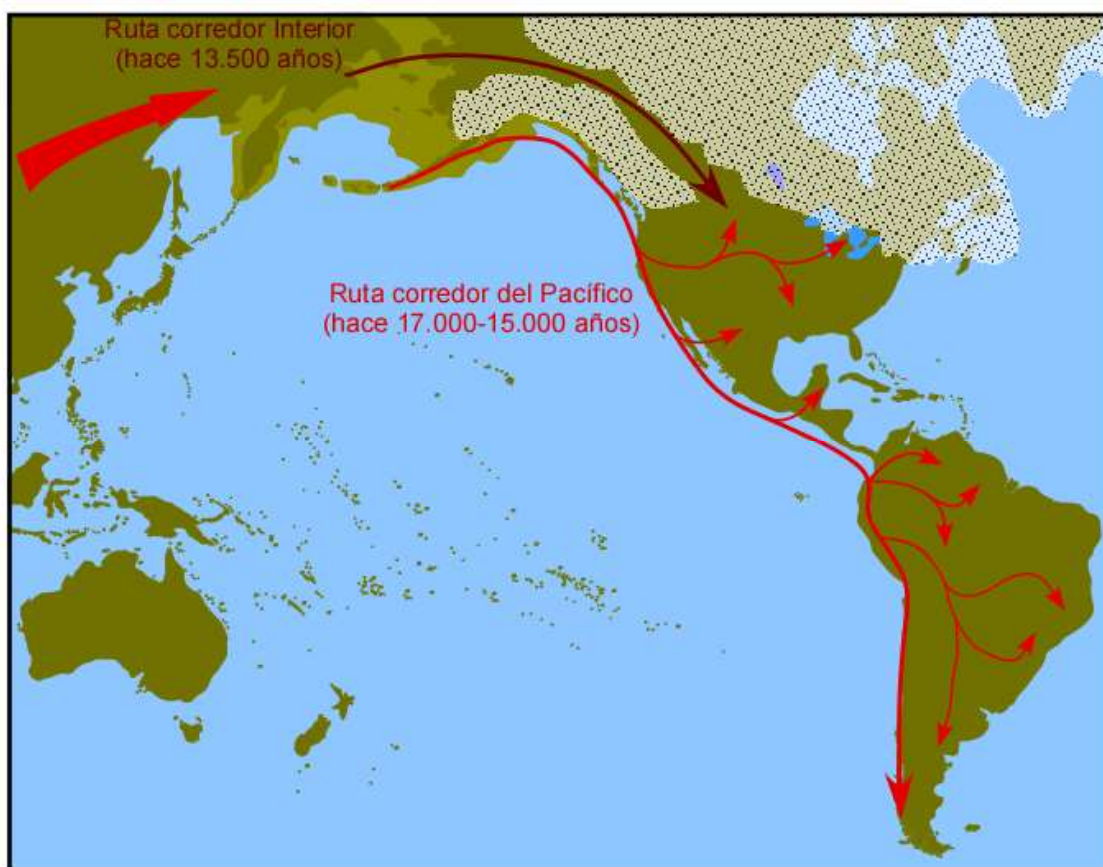


Figura 2. Posibles rutas de entrada en el continente americano tras la llegada a Beringia. La ruta interior (flecha marrón), pudo ser más tardía y precisaba de un paso entre los glaciares; la ruta costera pudo ser más temprana y utilizando una lengua de tierra actualmente sumergida o mediante embarcaciones. Modificado de Rothhammer y Dillehay (2009).

Entre las hipótesis que se han propuesto cabe mencionar las siguientes:

- I. Hipótesis de las tres oleadas migratorias (Greenberg *et al.* 1986). La entrada de los nativos americanos se habría producido en tres oleadas distintas. Se basa en la existencia de tres familias lingüísticas claramente diferenciadas en América, de modo que cada migración habría dado lugar a una de ellas. La primera migración correspondería a los Amerindios y se habría producido hace unos 12.000 años. Posteriormente se habría dado una entrada de los antepasados de los Na-dene, hace 8.000 años y por último, los Esquimo-Aleutianos, que habrían entrado hace 6.000 años. Tradicionalmente se la ha considerado sustentada por datos de algunos marcadores serológicos y el ADN mitocondrial (Torroni *et al.* 1993; Cavalli-Sforza *et al.* 1994).

Esta hipótesis estaba prácticamente desechada hasta que estudios recientes han aportado nuevos datos que parecen sustentarla, aunque con evidentes modificaciones. Entre ellos, cabe mencionar el trabajo de Reich *et al.* (2012) realizado sobre 52 poblaciones amerindias y 17 siberianas a partir de 364.470 SNPs. Sin embargo, sobre la hipótesis original de Greenberg varían las edades de entrada y se otorga una especial importancia a la primera migración. Además Reich *et al.* (2012) destaca que ha existido un importante flujo génico entre los tres grupos, dando lugar a un escenario mucho más complejo.

- II. Hipótesis del reemplazamiento (Powell y Neves, 1999). El poblamiento del continente americano se habría desarrollado en dos migraciones procedentes de diferentes regiones y en diferentes momentos. La primera de ellas, ocurrida hace entre 15.000 y 10.000 años, habría sido realizada por Paleoamerindios. Y hace alrededor de 4.000 años, los ancestros de los amerindios modernos habrían alcanzado el continente y habrían reemplazado a los primitivos pobladores. Esta hipótesis se propuso a partir del análisis de medidas de cráneos de los Paleoamerindios, más similares a Australoides y Melanesios, con una morfología diferente a la de los cráneos pertenecientes a los nativos americanos

actuales, más similares a poblaciones Asiáticas (Powell, 2005). Los análisis genéticos, tanto de ADN antiguo como actual han mostrado resultados diferentes (Smith *et al.* 2005; Kemp *et al.* 2007; Jenkins, 2007), sugiriendo que los paleoamerindios y los nativos actuales muestran una estrecha relación genética (Goebel *et al.* 2008). Además, en el trabajo mencionado anteriormente de Chatters *et al.* (2014), un esqueleto de 12.000-13.000 años de antigüedad mostraba características craneofaciales paleoamericanas además de un haplogrupo D1 (proveniente de Beringia), lo que llevó a concluir que las diferencias morfológicas entre paleoamericanos y amerindios actuales son debidas a evolución *in situ*.

- III. Modelo de 4 migraciones. Según Horai *et al.* (1993), cada haplogrupo mitocondrial (A, B, C, D) se habría introducido en América a través de una migración independiente, proveniente de cuatro poblaciones ancestrales diferentes del noreste de Asia. Estas poblaciones habrían emigrado de forma independiente hacia Beringia y habrían penetrado en el Nuevo Mundo en diferentes fases, estimando la primera migración hace 21.000-14.000 años. Esta hipótesis es poco probable, dado que los cuatro haplogrupos se encuentran prácticamente en todos los grupos étnicos a lo largo de América (Stone y Stoneking, 1998). Además, los tiempos de coalescencia encontrados para todos los haplogrupos presentan edades muy similares (Achilli *et al.* 2008; Fagundes *et al.* 2008b; Ho y Endicott, 2008), por lo que seguramente habrían formado parte de una sola población.
- IV. Modelo de migración única: Muchos autores se inclinan por un modelo con una migración mayoritaria procedente de Asia para explicar el origen de los Nativos Americanos (Merriwether *et al.* 1995; Forster *et al.* 1996; Bonatto y Salzano 1997a, b; Stone y Stoneking, 1998; Silva *et al.* 2002; Zegura *et al.* 2004; Schurr, 2004; Tamm *et al.* 2007; Achilli *et al.* 2008; Fagundes *et al.* 2008b; Goebel *et al.* 2008; Kitchen *et al.* 2008; Mulligan *et al.* 2008; Perego *et al.* 2010; Callaway, 2015).

Existen importantes evidencias que apoyan este modelo; pueden mencionarse la reducida variación genética encontrada al analizar el ADN mitocondrial y el cromosoma Y; la distribución ubicua de los principales haplogrupos (A2, B2, C1b, C1c, C1d, D1, and D4h3a) (Tamm *et al.* 2007; Achilli *et al.* 2008; Fagundes *et al.* 2008b; Perego *et al.* 2009, 2010); la existencia de variantes características en marcadores autosómicos (Mulligan *et al.* 2004; Wang *et al.* 2007), entre las que destaca el alelo 9RA del microsatélite D9S1120 (Wang *et al.* 2007; Phillips *et al.* 2008; Schroeder *et al.* 2009), que se encuentra en todas las poblaciones amerindias con frecuencias notables y que no aparece en poblaciones de otros continentes, excepto los Koryak y Chukchi del Oeste del Estrecho de Bering, o el haplogrupo del cromosoma Y Q1a3a-M3 (Bortolini *et al.* 2003; Karafet *et al.* 2008; Rasmussen *et al.* 2010). Hay otros indicios obtenidos de manera indirecta, que apoyan los anteriores argumentos, como las edades de divergencia entre los haplogrupos del cromosoma Y (Q y C), que son bastante similares y parecen haberse originado en algún lugar próximo a las Montañas Altai (Zegura *et al.* 2004) al igual que la edad de coalescencia (Achilli *et al.* 2008; Fagundes *et al.* 2008b; Ho y Endicott, 2008); asimismo, la escasa presencia de los haplogrupos mitocondriales característicos de América en poblaciones asiáticas minimiza la probabilidad de que varias migraciones independientes tuvieran una composición similar de haplogrupos mitocondriales (Kolman *et al.* 1996).

La presencia de variantes alélicas específicas o la homogeneidad de los haplogrupos fundadores apoyan un escenario en el que, junto a una migración principal mayoritaria, existió una estancia intermedia en la región de Beringia, lo que permitió la diferenciación de una serie de grupos, posiblemente por deriva genética. Una vez concluido un tiempo de estancia se habría dado una rápida migración hacia el sur a lo largo del nuevo continente (Tamm *et al.* 2007; Schroeder *et al.* 2009; Perego *et al.* 2010).

El modelo clásico *Clovis-First*, postulado por Hrdlicka en 1923 a raíz de los descubrimientos del yacimiento Clovis en Nuevo México y posteriormente modificado con el fin de incorporar nuevos datos cronológicos y culturales (Martin, 1973; Haynes, 2002), pertenece a la categoría de hipótesis de una sola migración. De hecho fue una hipótesis pionera en este sentido. Sin embargo, este modelo postula que los primeros pobladores del continente fueron individuos de la cultura Clovis y descubrimientos de finales del S.XX y principios del S.XXI parecen rebatir este supuesto, ya que han aparecido restos que superan el límite Clovis de antigüedad (Dillehay, 1997; Adovasio y Pedler 2004; Overstreet, 2005; Webb, 2005; Joyce, 2006; Jenkins, 2007; Dillehay *et al.* 2008; Fernández-Gómez y Velasco-Ortiz, 2011).

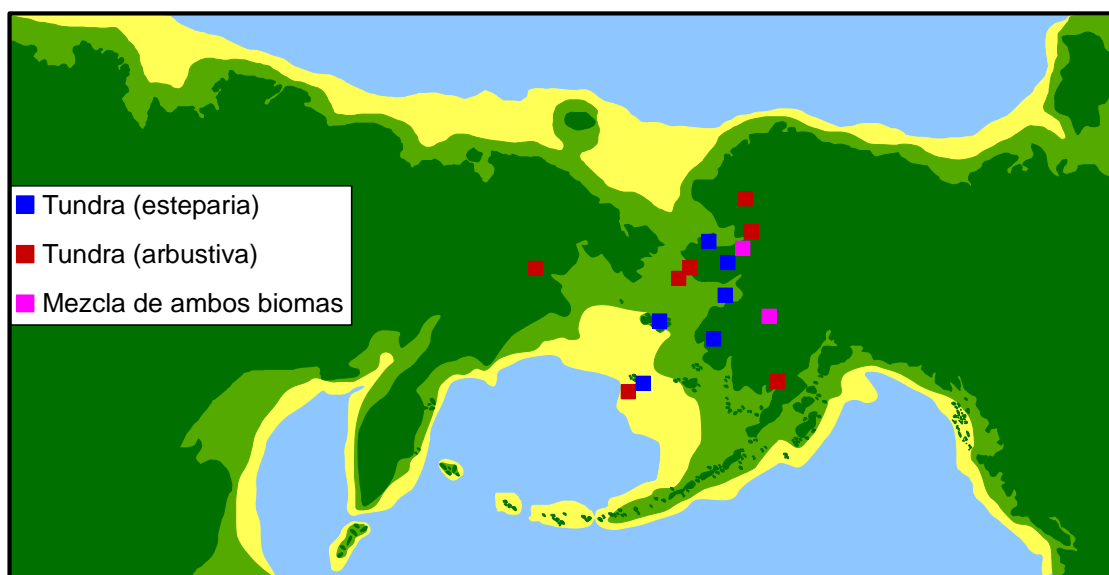


Figura 3. Representación de la tierra emergida en la región de Beringia (actualmente el estrecho de Bering) en diferentes épocas. Amarillo claro, hace 22.000 años durante el LGM; Verde claro, hace 15.000 años; Verde oscuro, en la actualidad. Los cuadrados representan los yacimientos donde se ha identificado el tipo de vegetación predominante durante el LGM según Hoffecker *et al.* (2014).

Durante la época del poblamiento del continente americano, el planeta se hallaba inmerso en la última glaciación. El último máximo glacial (LGM) de la glaciación Wisconsin se extendió entre hace 23.000 y 19.000 años (Mix *et al.* 2001) y muy posiblemente, durante ese periodo las placas de hielo de los glaciares de Norteamérica (Cordillera y Laurentino) cubrieran todo Canadá y el Norte

de EEUU, bloqueando la entrada en el continente. Una vez finalizado el LGM, habría comenzado el retroceso de las placas Cordillera y Laurentino, hasta dar lugar a la apertura de un paso libre de hielo, hace unos 14.000 años. Es el llamado corredor interior, situado en el valle del río Mackenzie, al Oeste de Canadá (Hoffecker *et al.* 1993; Mandryk *et al.* 2004; Dyke, 2004), que podría haber servido de paso entre Beringia y el norte de América. El modelo *Clovis-First*, proponía este camino como única vía de entrada de los primeros pobladores, los Clovis.

Sin embargo, estudios recientes postulan la posibilidad de que existiera un camino alternativo, el denominado corredor de la costa del Pacífico, que el *Homo sapiens* pudo haber utilizado mediante pequeñas embarcaciones, navegando muy cerca de la costa (Achilli *et al.* 2008; Fagundes *et al.* 2008b; Goebel, 2008; Bodner *et al.* 2012) (*Figura 2*). La apertura de este corredor ocurrió al menos 1.000 años antes que el corredor interior (Mandryk *et al.* 2004; Dyke, 2004), siendo incluso posible que fuera un acceso viable hace entre 17.000 y 15.000 años (Dixon, 2001).

Existen evidencias indirectas de presencia humana en América con una antigüedad superior a 14.000 años, fuera del límite Clovis, como las encontradas en Monte Verde (sur de Chile), en Schaefer/Hebior, Meadowcroft Rockshelter, Page-Ladson, Paisley Cave y Manis site (Dillehay, 1997; Aldovasio y Pedler 2004; Overstreet, 2005; Webb, 2005; Joyce, 2006; Jenkins, 2007; Waters *et al.* 2011) representadas en la *Figura 4*. La existencia del corredor costero daría una explicación plausible a la presencia de los yacimientos más antiguos, ya que era complicado explicar dichas dataciones con la entrada a través del corredor interior. Estas evidencias implicarían que los Clovis no habrían sido los primeros pobladores de América.

A tenor de los yacimientos y restos fósiles, la difusión desde el norte de América al resto del continente hasta llegar a Tierra del Fuego debió de ser rápida, probablemente desarrollada en unos pocos siglos. De hecho, uno de los yacimientos con dataciones más antiguas es Monte Verde, en el sur de Chile, con restos de hace unos 14.600 años (Dillehay, 1997; Bodner *et al.* 2012; Roewer *et al.* 2013; Callaway, 2015).



Figura 4. Localización de los yacimientos más relevantes donde se ha encontrado restos de humanos de los primeros pobladores o bien indicios de la presencia de estos en el continente Americano. Se incluyen las fechas (en miles de años) de las dataciones realizadas. En rojo se destaca la original de la cultura Clovis.

Una vez que el LGM se acercó a su final, el paso entre Asia y América a través de la región de Beringia quedó progresivamente sumergido, debido al deshielo y al consiguiente ascenso del nivel del mar, formando el actual estrecho de Bering (Figura 3). A pesar de ello, se ha especulado con posteriores contactos (migraciones menores y retromigraciones) entre las poblaciones americanas y asiáticas, una vez que América estuvo aislada. De esta forma, por ejemplo, se ha intentado explicar la presencia de algunos haplogrupos de ADNmt minoritarios (A2a, A2b, D2a, X2a) en poblaciones

de Norteamérica (Torrioni *et al.* 1992; Brown *et al.* 1998; Schurr y Sherry, 2004; Helgason *et al.* 2006; Achilli *et al.* 2008; Gilbert *et al.* 2008a; Perego *et al.* 2009; Raghavan *et al.* 2014). Algunos trabajos realizados a partir de datos genéticos y morfométricos han apoyado esta idea (González-José *et al.* 2008; Bourgeois *et al.* 2009; Rasmussen *et al.* 2010) ya que se ha observado una mayor similitud entre algunas poblaciones de Norteamérica con ciertas poblaciones asiáticas que con poblaciones sudamericanas. Como se ha comentado anteriormente, en un trabajo de Reich *et al.* (2012), y tras analizar 364.470 SNPs en 57 poblaciones americanas y 17 poblaciones siberianas, han propuesto que aproximadamente la mitad del patrimonio genético de las poblaciones eskimo-aleutianas provendría de una segunda oleada y que los Na-Dene habrían recibido en torno a un 10% de sus genes de una tercera oleada desde Asia. Estos autores postulan un escenario complejo con varias migraciones menores y flujo génico entre ellas. Incluso alguna otra migración menor (el caso de los Paleo-Eskimo) ha sido propuesta (Raghavan *et al.* 2014) para ser añadida a la hipótesis de Reich *et al.* (2012).

POBLAMIENTO DE ORIGEN EUROPEO Y AFRICANO

Una vez que finalizó la Glaciación, las poblaciones americanas nativas se mantuvieron aisladas del resto del mundo durante más de 10.000 años (Kitchen *et al.* 2008), durante los cuales se produjo una notable diversificación de las poblaciones en los diferentes entornos (estepario, selvático, insular, de altura,...) a la vez que aparecieron un gran número de culturas.

Este aislamiento se interrumpió con la llegada de los colonos europeos y el tráfico de esclavos. El segundo evento masivo de colonización del continente comenzó tras el descubrimiento de América por Colón en 1492, si bien la llegada de la mayor parte de los inmigrantes europeos y africanos ocurrió a partir del Siglo XVI, especialmente durante su segunda mitad.

Durante el siglo XV, las potencias europeas buscaban nuevos territorios por motivos tanto religiosos como económicos, en una incesante búsqueda de materias primas con las que

compensar los gastos militares generados por diferentes guerras y atender a un mercado con una demanda creciente de especias, telas, etc. Si en un primer momento fueron España y Portugal, rápidamente se unieron a ellos Inglaterra, Francia y los Países Bajos, sobre todo a partir del siglo XVII.

Se estima que entre los siglos XV y XIX, alrededor de 1 millón de españoles y portugueses llegaron a las regiones del centro y sur de América (Burkholder y Johnson, 2003). Además, junto con los colonos europeos llegaron individuos de origen africano, producto del tráfico de esclavos. Estos individuos eran capturados en las regiones del centro y sur de África y llevados a América con el fin de ser usados como mano de obra en diversos trabajos (agricultura, minería, etc.).

Se desconoce el número exacto de individuos africanos que fueron transportados al continente americano entre 1600 y 1850, para trabajar como esclavos. Las estimaciones ofrecen unas cifras de al menos 5 millones (Curtin, 1969; Thomas, 1997), hasta un máximo de 11,6 millones (Lovejoy, 1982).

Es evidente que la llegada de estas poblaciones produjo importantes cambios demográficos sobre las poblaciones nativas. Tanto los registros arqueológicos como los históricos indican que el contacto desencadenó una serie de procesos que tuvieron un gran impacto en la población nativa americana, fundamentalmente en la forma de aumento de la mortalidad por guerras, epidemias (viruela principalmente), trabajos forzados y desarraigo (Dobyns, 1993; Crosby, 2003, 2004; Livi-Bacci, 2006), lo que determinó una importante reducción de su tamaño (Cook y Borah, 1971; Ramenenovsky, 1987; Mulligan *et al.* 2004; Seymour *et al.* 2004; Halverson y Bolnick, 2008). Según algunos autores, la magnitud de ese impacto pudo ser muy importante (Bolnick y Smith, 2003; Wang *et al.* 2004; Shook y Smith, 2008; O'Fallon y Fehren-Schmitz, 2011) aunque otros autores restan importancia a dicha reducción poblacional (Atkinson *et al.* 2008; Mulligan *et al.* 2008; Ho y Endicott, 2008; Fagundes *et al.* 2008a). Lo cierto es que algunas estimaciones llegan a proponer que

en los primeros 130 años de colonización europea murió el 90-95% de la población total amerindia (Mann, 2006).

En todo caso, el probablemente elevado descenso demográfico experimentado por las poblaciones amerindias fue el principal motivo por el cual las potencias europeas recurrieron a un masivo reclutamiento de esclavos africanos para reemplazar la mano de obra indígena. Sin embargo, no pueden obviarse otros motivos, como el reconocimiento de los nativos por parte de la Corona Española como súbditos. Según las Leyes de Burgos de 1512, la esclavitud de los amerindios quedó abolida, aunque fue reemplazada por otros sistemas semiesclavistas, como la encomienda y el repartimiento (Sánchez-Domingo, 2012). En todo caso, la necesidad de una mano de obra barata y abundante impulsó la importación de esclavos procedentes del África subsahariana. En Norteamérica, en los territorios controlados por Francia e Inglaterra, la situación fue muy diferente, ya que la actitud con los nativos fue beligerante durante largo tiempo. Así, tras sucesivas guerras y desplazamientos, los nativos de Estados Unidos tuvieron que esperar para adquirir la nacionalidad y el derecho a voto hasta mediados del siglo XX, tras la promulgación de la *Indian Citizenship Act* de 1924 y la *Indian Relocation Act* de 1956.

El impacto de la llegada de los europeos y africanos fue también importante en el aspecto genético. La llegada de nuevos pobladores generó un gran número de procesos de flujo génico entre los grupos étnicos implicados. Esto originó una amplia serie de poblaciones con diferentes grados de mezcla entre amerindios, africanos y europeos (Carvajal-Carmona *et al.* 2000; Salzano y Bortolini, 2002; Bedoya *et al.* 2006; Salzano y Sans, 2014). En aquellas regiones controladas por españoles y portugueses, el contacto entre nativos y colonos fue importante. Las colonias inglesas, por el contrario, practicaron deportaciones masivas, aislando a los nativos en reservas y minimizando de este modo el flujo génico con ese grupo étnico. Los principales motivos de estos comportamientos fueron posiblemente i) religiosos, ya que la mayor parte de los inmigrantes británicos fueron protestantes calvinistas extremos (conocidos como puritanos), y ii) sociales,

debido a que la inmigración española y portuguesa estaba compuesta principalmente por hombres, en tanto que en la británica se integraban en una mayor proporción familias completas. Por ello, de manera general se puede afirmar que en la parte más septentrional de América apenas hubo mestizaje entre poblaciones nativas e inmigrantes y aunque escaso, ocurrió casi exclusivamente entre africanos y europeos. Sin embargo, en Centro y Sudamérica, así como en la parte meridional de Norteamérica, hubo un mestizaje importante entre los diversos grupos humanos, especialmente entre nativos y europeos (Sánchez-Albornoz, 1974; Bercovitch, 1975; Olson, 1980; Degler, 1984, 1986; Jones, 1996 Salzano y Sans, 2014).

El impacto genético de los procesos de flujo génico sobre las poblaciones nativas se vio influido también por las características de la región afectada. Las dificultades de los europeos para adaptarse a determinados ambientes y la riqueza en recursos de un territorio determinado condicionaron la densidad de los asentamientos. Las regiones montañosas, los desiertos, las selvas o las latitudes extremas fueron regiones con poco impacto migratorio, permaneciendo en algunos casos prácticamente intactas hasta una época reciente; y es precisamente en estos lugares donde se ha observado que las poblaciones nativas generalmente conservan un alto grado de componente amerindio en su genoma. En cambio, regiones más accesibles como las llanuras, zonas costeras o riberas de los ríos experimentaron una mayor influencia colonizadora, resultando en la actualidad en poblaciones muy mestizadas, incluso con presencia mayoritaria europea.

Los procesos de colonización habitualmente son procesos complejos, con muchos factores que afectan a las relaciones entre las poblaciones implicadas y habitualmente se extienden a lo largo de sucesivas generaciones, con lo que dichos factores pueden ir variando a lo largo del tiempo. En los análisis realizados en las diferentes poblaciones americanas se ha revelado un fenómeno singular que ocurrió durante la colonización de gran parte del continente americano, y que merece ser destacado por su efecto sobre el patrimonio genético de sus poblaciones, el mestizaje diferencial dependiente del sexo. De acuerdo con éste fenómeno, el mestizaje entre los

diferentes grupos humanos no ocurrió de forma equitativa entre ambos sexos, principalmente debido a la asimetría demográfica de los inmigrantes europeos principalmente españoles y portugueses, como se ha explicado anteriormente. Como ya se ha explicado, la mayoría de los individuos que llegaban al nuevo mundo eran varones, sobre todo durante los primeros años (Boyd-Bowman, 1976). Esto produjo que la mezcla ocurriera básicamente entre varones europeos y mujeres nativas americanas y un componente menor con individuos de origen africano, lo que se ha reflejado en numerosos análisis genéticos detectándose una marcada influencia europea en los marcadores del cromosoma Y, en tanto que se conserva una alta presencia de haplogrupos amerindios en los análisis de ADN mitocondrial (Merriwether *et al.* 1997; Sans *et al.* 1997; Dipierri *et al.* 1998; Bortolini *et al.* 1999; Mesa *et al.* 2000; Alves-Silva *et al.* 2000; Green *et al.* 2000; Carvajal-Carmona *et al.* 2003; Rojas *et al.* 2010; Marcheco-Teruel *et al.* 2014).

POBLACIONES ESTUDIADAS

En los diferentes trabajos que conforman esta memoria se han analizado una serie de muestras procedentes de tres regiones bien diferenciadas de Sudamérica. Dos de las muestras, compuestas por mestizos y afrocolombianos, se tomaron en el departamento de Antioquia, en Colombia. Otras muestras fueron recogidas en varias localidades situadas a diferentes altitudes en la provincia de Jujuy, en el Noroeste Argentino y se componen principalmente de individuos de origen amerindio. Por último, una muestra se tomó en la provincia de Pastaza, en Ecuador, e incluye individuos amerindios de la etnia Huaorani.

Las poblaciones representadas tenían *a priori* una serie de características específicas y claramente diferenciadas que las dotaba de un notable interés para el estudio de los procesos de microdiferenciación ocurridos en este continente.

- I. La muestra de mestizos de Antioquia, representa al mayor grupo étnico de Colombia y probablemente de toda Sudamérica y Centroamérica. Es resultado de los procesos de mestizaje entre amerindios y europeos, con un potencial aporte de genes de las poblaciones de origen africano. Por ello las expectativas implicaban una notable incidencia del flujo génico.
- II. Los Afrocolombianos son descendientes de individuos africanos que llegaron a América como esclavos transportados por los europeos desde diferentes regiones de África. Cabría, en todo caso, esperar un cierto grado de mestizaje con otras poblaciones vecinas.
- III. La población de Jujuy se encuentra dispersas entre la montaña andina y las quebradas que descienden desde ella. Es una región de difícil acceso, especialmente las regiones más elevadas de la provincia, estando sus comunidades notablemente aisladas. Jujuy se caracteriza por un asentamiento diferencial de europeos y afroamericanos, con preferencia por las zonas de menor altitud.

- IV. Por último, los Huaorani de Ecuador son una población nativa amerindia caracterizada por un comportamiento muy agresivo, con escasos contactos con las poblaciones circundantes, tanto nativas como foráneas. Esto es debido tanto a causas culturales como a un entorno, el de selva, que dificulta notablemente la movilidad.

A continuación, se hará una descripción pormenorizada de estas poblaciones.

PROVINCIA DE ANTIOQUIA (COLOMBIA)

Colombia se localiza en el Noroeste de Sudamérica, siendo el nexo de unión con Centroamérica. El departamento de Antioquia se encuentra en la zona Noroccidental del país. Tiene una superficie de 63.612 km² y cuenta con 5.962.885 habitantes. Limita al Norte con el mar Caribe y con el departamento de Córdoba, al Oeste con el departamento de Chocó, al Este con los departamentos de Bolívar, Santander y Boyacá y al Sur con los departamentos de Caldas y Risaralda (*Figura 5*).

Su capital es Medellín, situada en el valle de Aburrá. Es la segunda ciudad en importancia de Colombia, tanto en población como en economía. Se encuentran registrados 2.223.078 habitantes, de acuerdo al censo de 2005 (Departamento Administrativo Nacional de Colombia, DANE, <http://www.dane.gov.co/>).

Desde el punto de vista de su geografía, Antioquia es una región montañosa, de modo que aproximadamente el 80% de su superficie puede considerarse de este tipo mientras que tan sólo un 20% se puede describir como áreas llanas. Esto se debe a que las estribaciones de la cordillera de los Andes atraviesa Antioquia de Sur a Norte hasta desaparecer en la parte más septentrional. Fruto de ello la capital, Medellín, se encuentra situada a 1538 msnm. La fundación de Medellín (San Lorenzo de Aburrá, como era conocido en la época colonial), no ocurrió hasta 1675, posiblemente debido a los problemas derivados de su orografía.

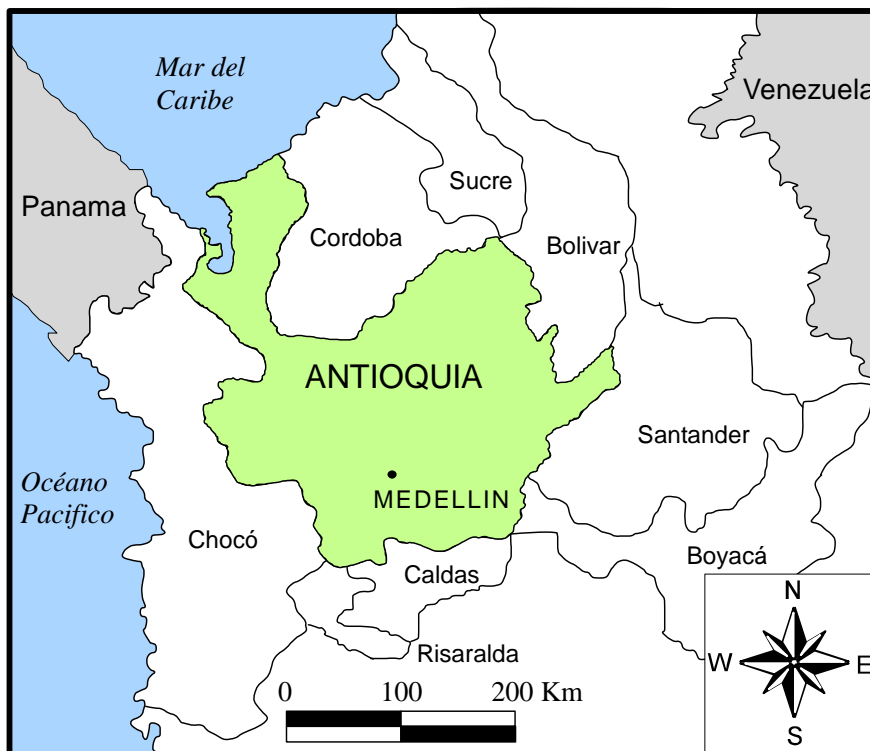


Figura 5. Mapa del Norte de Colombia. En él se muestra al departamento de Antioquia, junto con su capital, Medellín. También se muestran los departamentos limítrofes con Antioquia.

Según el último censo, realizado en 2005, la mayor parte de la población (88,66%) se caracterizó como integrada por mestizos y caucasoides, siendo minoritarios los afro-colombianos (10,83%) y aún menos abundantes los nativos americanos (0,51%) (DANE, 2005).

Antioquia, como el resto de América, experimentó cambios demográficos muy importantes tras el descubrimiento del continente por los europeos. Como se ha mencionado, la oleada migratoria de individuos europeos y de esclavos provenientes del Oeste y Centro de África dio lugar a nuevas poblaciones surgidas por diferentes procesos de mestizaje entre las 3 poblaciones parentales, amerindios, europeos y africanos (Carvajal-Carmona *et al.* 2000; Bedoya *et al.* 2006). Sin embargo, en esta región se estima que la población nativa era escasa cuando llegaron los conquistadores europeos, siendo mencionada por los historiadores la presencia de tan sólo algunos miles de individuos (Alvarez, 1996; Carvajal-Carmona *et al.* 2000; Bedoya *et al.* 2006).

El número exacto de individuos africanos que llegaron a Colombia se desconoce, aunque algunas estimas proponen que fueron entre 100.000 y 200.000, procedentes del África subsahariana (Colmenares, 1979; Del Castillo, 1981).

A partir de la primera época de su colonización y hasta el Siglo XIX, la región de Antioquia ha mantenido un cierto grado de aislamiento (Álvarez, 1996; Carvajal-Carmona *et al.* 2000; Carvajal-Carmona *et al.* 2003; Bedoya *et al.* 2006). A pesar de ello experimentó un importante aumento poblacional. La expansión demográfica, por tanto, se debió casi exclusivamente a su crecimiento vegetativo (Carvajal-Carmona *et al.* 2000; Bedoya *et al.* 2006).

Población Mestiza

Las poblaciones mestizas se componen de los descendientes de uniones entre individuos de origen europeo y amerindio. Pero lo cierto es que muestran generalmente un cierto flujo génico de origen afroamericano (Salzano y Sans, 2014) Como consecuencia, la composición genética de este grupo es muy variable.

En la bibliografía revisada se ha encontrado una gran variabilidad en relación a la composición del patrimonio genético de los mestizos de Antioquia; se han observado, asimismo, notables diferencias entre ambos sexos. A partir de marcadores del cromosoma Y se ha estimado que los mestizos de Antioquia muestran una composición mayoritaria europea (94%), con una proporción escasa de aporte africano (5%) y una fracción mínima de origen amerindio (1%) (Carvajal-Carmona *et al.* 2000). Por el contrario, en los trabajos que se han realizado sobre ADN mitocondrial, se ha encontrado un 90% de linajes con un origen amerindio, un 8% de origen africano y tan sólo un 2% de origen europeo (Carvajal-Carmona *et al.* 2000; Bedoya *et al.* 2006).

Estos resultados sugieren un modelo de mestizaje diferencial por sexos, de modo que se habrían dado preferencialmente uniones entre hombres inmigrantes y mujeres nativas (Mesa *et al.* 2000; Carvajal-Carmona *et al.* 2000; Bedoya *et al.* 2006; Rojas *et al.* 2010). Esta hipótesis tiene un

apoyo histórico bien documentado, como ya se ha mencionado, pues existen registros donde se describe que durante el periodo colonial (desde finales del S. XVI hasta principios del S. XIX) y particularmente durante los primeros años, el número de mujeres que llegaron al continente americano fue muy pequeño, aproximadamente un 10-20% de los inmigrantes (Boyd-Bowman, 1976; Sánchez-Albornoz, 1977).

Como ya se adelantó en la sección de *Colonización del Continente Americano*, el fenómeno de mestizaje diferencial dependiente del sexo ha sido relativamente frecuente y se ha descrito para otras poblaciones latinoamericanas en numerosos trabajos (Merriwether *et al.* 1997; Sans *et al.* 1997; Dipierri *et al.* 1998; Alves-Silva *et al.* 2000; Green *et al.* 2000; Carvajal-Carmona *et al.* 2003; Rojas *et al.* 2010; Marcheco-Teruel *et al.* 2014), siendo relativamente frecuente la unión entre mujeres nativas y varones de origen europeo. A pesar de la relativa frecuencia de esta situación, cabe destacar que los valores encontrados en Antioquia son muy llamativos, debido a la amplia diferencia entre la composición de origen amerindio en los linajes maternos (ADN mitocondrial) y paternos (cromosoma Y), así como lo minoritario del componente africano. De hecho, en las poblaciones mestizas de Colombia, los linajes masculinos amerindios revelados por marcadores del cromosoma Y prácticamente han desaparecido (Mesa *et al.* 2000).

Los resultados de marcadores monoparentales pueden contrastarse con los obtenidos mediante marcadores biparentales o autosómicos (heredados tanto por vía materna como paterna). En los trabajos realizados sobre estos polimorfismos se ha observado una proporción mayoritaria del componente europeo, en torno al 70% (Sandoval *et al.* 1993; Bedoya *et al.* 2006), notablemente más elevada de lo que cabría esperar teniendo en cuenta los resultados obtenidos mediante los marcadores uniparentales. Como una posible explicación a este hecho, puede pensarse en una población fundadora compuesta fundamentalmente por mujeres amerindias y varones europeos que recibiría aportes periódicos, más o menos continuados, de varones europeos inmigrantes, mientras que no habría nuevos aportes amerindios, por el reducido número inicial y la

posterior disminución de sus poblaciones. De este modo, la mayor parte de los linajes femeninos serían amerindios, los linajes masculinos serían fundamentalmente europeos, pero en promedio habría un mayor número de genes de origen europeo en el patrimonio genético resultante.

Población Afrocolombiana

Los afrocolombianos actuales son los descendientes de los esclavos que fueron capturados y transportados desde África durante la época colonial. Esta comunidad ha recibido una cantidad variable de genes de origen europeo y amerindio durante las últimas 15-20 generaciones (Bortolini *et al.* 1999).

En general, en Colombia se encuentran en una proporción minoritaria, siendo el 10,62% de la población, según el censo de 2005, excepto en algunas regiones de las costas pacífica y atlántica y particularmente en los departamentos de San Andrés y Chocó, donde son mayoritarios (*Figura 6*) (DANE, 2005). Esta situación corresponde a una parte del departamento de Antioquia, concretamente en las regiones mineras, incluyendo también el golfo de Urabá (la región caribeña de Antioquia) y toda la cuenca del río Cauca, que atraviesa el territorio antioqueño. En cualquier caso, tanto en este departamento como en toda Colombia, la presencia de individuos afrocolombianos se muestra en una proporción superior a la de los individuos amerindios, caracterizándose como la minoría más importante.

En cuanto a su composición genética, los procesos de mestizaje de los afroamericanos muestran unas características básicas similares, pero con peculiaridades propias de sus regiones de residencia. Es por ello una comunidad heterogénea y difícil de categorizar (Bortolini *et al.* 1992, 1995, 1997, 1998, 1999; Castro de Guerra *et al.* 1996; Da Silva *et al.* 1999; Martínez-Labarga *et al.* 1999). Cabe destacar que en general no muestran un desequilibrio tan notable como la población mestiza entre marcadores masculinos y femeninos, aunque existen algunas excepciones (Sans *et al.*

2002). Habitualmente las proporciones dependen más de las características de la comunidad analizada, así como del entorno en el que habitan (Rodas *et al.* 2003).

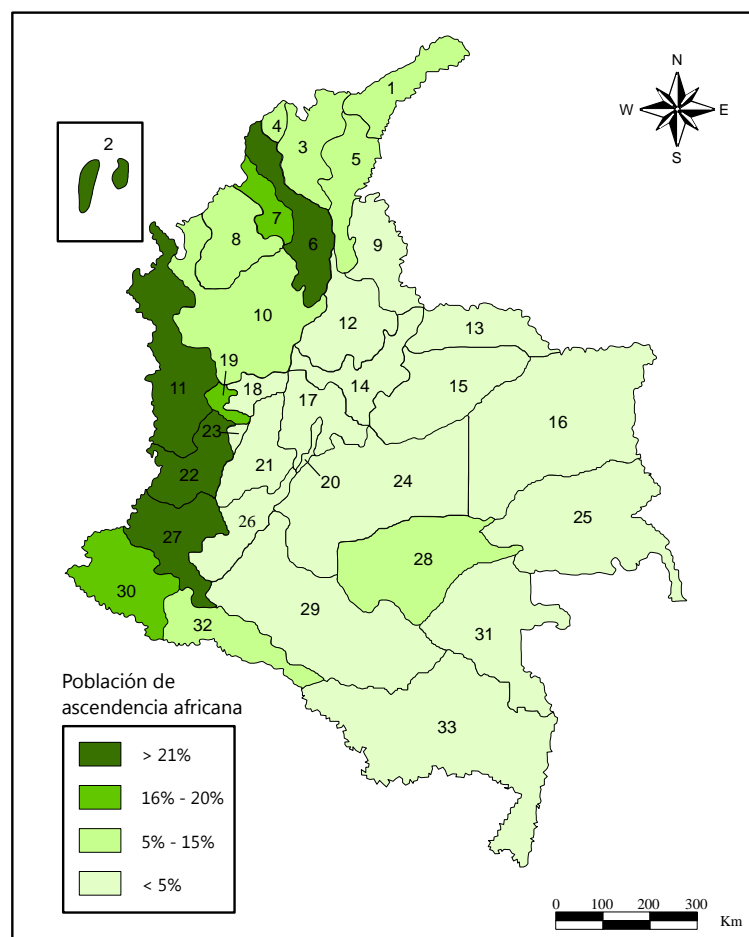


Figura 6. Mapa de Colombia donde se muestra la proporción de población de ascendencia africana en los distintos departamentos. Se observa que la mayoría de la población afrocolombiana se concentra en los departamentos del Noroeste del país, situados en la costa. Antioquia es el departamento nº 10. La leyenda para el resto de los departamentos se detalla en la *Tabla Anexo 1*.

Durante el tiempo en que se mantuvo la práctica de la esclavitud en Colombia (y por extensión en Sudamérica) fueron habituales los intentos de fuga; en algunas regiones los esclavos huidos (conocidos como cimarrones) fundaron comunidades que eran denominadas palenques o quilombos, en las que se refugiaban. A menudo fortificaban estos refugios, ya que ocasionalmente ocurrían enfrentamientos entre cimarrones y europeos esclavistas. Generalmente y por razones obvias, estas comunidades se construían en lugares de difícil acceso, de manera similar a fortalezas,

donde se defendían con violencia de los agresores europeos. De esta forma se mantenían a refugio de intentos de captura de los europeos y aislados consecuentemente para cualquier tipo de intercambio genético. A pesar de que estos asentamientos han existido por toda Colombia, la mayoría se han localizado en la parte septentrional y zona caribeña del país (Figura 7).

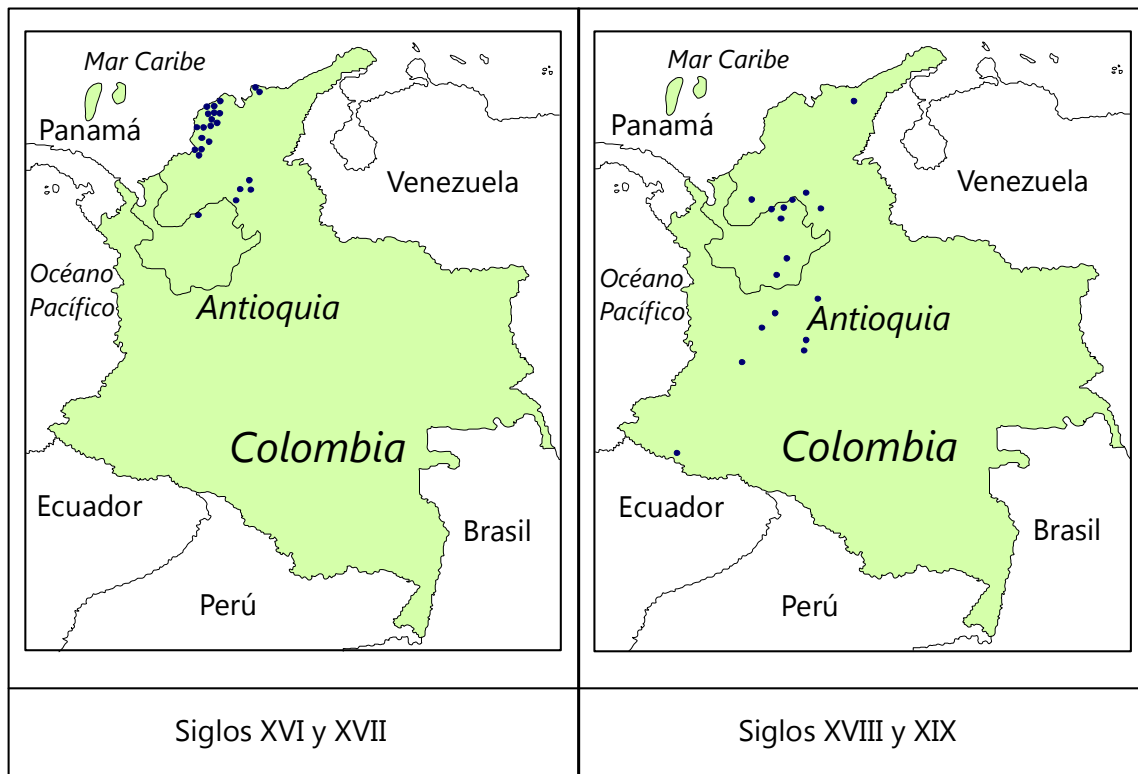


Figura 7. Palenques fundados en los siglos XVI al XIX en Colombia. Durante los dos primeros siglos, los palenques se concentran en la zona más al Norte, coincidiendo con la región más poblada por los colonos a su llegada. Posteriormente se fueron expandiendo por otros territorios (Modificado de Friedemann, 1993).

Una vez que la esclavitud fue abolida, una parte importante de los individuos afrocolombianos liberados se integró en la sociedad colombiana. Además, algunos grupos fundaron comunidades de libertos que, en todo caso, no mostraron el grado de aislamiento ni de hostilidad de los palenques. Estas comunidades se han mantenido hasta muy recientemente.

Han existido por lo tanto dos tipos de asentamientos típicos de afrocolombianos, ambos bien diferenciados por el tipo de contacto y relaciones con el resto de etnias de la zona, uno esencialmente aislado y el otro con un más alto grado de contacto.

PROVINCIA DE JUJUY (ARGENTINA)

Jujuy es una provincia del Noroeste Argentino. Limita al Oeste con la cordillera andina que la separa de Chile, al Norte con Bolivia y al Este y al Sur con la provincia de Salta, Argentina (*Figura 8*). La superficie de la provincia de Jujuy comprende un total de 53.219 km² y contiene una población de 673.307 habitantes, lo que le otorga una densidad poblacional de 12,65 hab/km² (INDEC, 2010).

San Salvador de Jujuy es la capital de la provincia; se localiza en el departamento Doctor Manuel Belgrano, que pertenece a la región conocida como Valle, en la parte sur de la provincia (*Figura 9*), en la confluencia de los ríos Grande y Xibi Xibi. Tiene una población de 265.249 individuos (INDEC, 2010) y se encuentra a una altitud media de 1.259 msnm.

La provincia de Jujuy presenta un relieve eminentemente montañoso, con importantes variaciones altimétricas, debido a la proximidad de la cordillera de los Andes. Se observa un patrón de aumento progresivo desde el sudeste, coincidiendo con el valle del río San Francisco y a una altitud mínima de 500 msnm, hacia el noroeste, con una altura máxima próxima a los 6.000 msnm en el nevado del Chañi (Bejarano *et al.* 2006). Estas características físicas, con sus correspondientes cambios climáticos, han influido en la distribución de las poblaciones humanas a lo largo del tiempo y también en las diferentes relaciones establecidas entre ellas (Carnese *et al.* 2010).

Es por esto que, a pesar de su pequeña extensión, la provincia de Jujuy está estructurada en tres áreas geográfico-ecológicas diferentes:

- I. El valle subtropical, con una altitud media de 1200 msnm. En esta área se encuentra la capital de la provincia.
- II. El área de la Pre-Puna, que incluye regiones situadas a unos 2000-2500 msnm de altitud.

III. La Puna de Jujuy que corresponde a la región situada a más de 3000 msnm.



Figura 8. Mapa de Sudamérica donde se destaca la localización de las provincias de Jujuy y Salta. Estas provincias se sitúan en el Noroeste Argentino.

Estas áreas se encuentran conectadas entre sí por estrechos pasillos o desfiladeros. El ejemplo más característico es la Quebrada de Humahuaca, de 170 km de largo, localizada en la Pre-Puna y que recorre en forma ascendente gran parte de la provincia de Jujuy en dirección Sur-Norte.

Administrativamente, la provincia de Jujuy se divide en 4 regiones (Bejarano *et al.* 2006) con la distribución por departamentos que se describe a continuación (Figura 10).

- I. La Puna: Se encuentra a una altitud de más de 3.000 msnm, siendo su altitud media superior a los 3.500 msnm. Incluye los departamentos de Susques, Rinconada, Santa Catalina, Yavi y Cochinoca. Es la región que tiene una mayor extensión, si bien se encuentra escasamente poblada, con una densidad poblacional de 1,34 hab/km².
- II. Quebrada: Esta región presenta una gran variación de altitud, que va desde los 600-700 msnm hasta los 3.700. Abarca los departamentos de Humahuaca, Tilcara y Tumbaya. Es la región con menor tamaño de población y una densidad poblacional muy baja, de 3,49 hab/km².
- III. Valle: Como se ha dicho anteriormente, es en esta región donde se encuentra la capital de la Provincia, San Salvador de Jujuy. A pesar de ser la región con menor extensión, es la más poblada y por lo tanto la que mayor densidad poblacional presenta (104,09 hab/km²). Se localiza a una altitud de unos 1.200 msnm. Esta región administrativa incluye los departamentos de Dr. Manuel Belgrano, El Carmen, Palpalá y San Antonio.
- IV. Selva: También conocida como Ramal, es la región que se encuentra a una altitud media más baja, rondando los 700 msnm. A esta región pertenecen los departamentos de Ledesma, San Pedro, Santa Bárbara y Valle Grande. Es la segunda región tanto en población (166.254 hab) como en densidad poblacional (15,38 hab/km²).

La provincia de Jujuy está situada en una posición estratégica en el contexto de Sudamérica, por lo que a pesar de la dificultades que presenta el terreno, pudo haber estado habitada desde fecha muy temprana. Tradicionalmente se ha estimado que el poblamiento de Jujuy comenzó hace entre 6.000 y 8.000 años, por grupos humanos procedentes de las regiones andinas de Perú y Bolivia (Rothhammer *et al.* 1984; Rothhammer y Silva, 1992). Otros autores han propuesto fechas de hasta 11.000 años cuando los primeros cazadores recolectores habrían colonizado la parte más

meridional del continente (Tarrago, 2000; Carnese *et al.* 2010). Las dataciones de los yacimientos más antiguos conocidos, particularmente el de Hornillos-2, en la Puna jujeña, reflejan una antigüedad de 9.710 ± 270 años (Yacobaccio *et al.* 2008).

La posición estratégica de Jujuy se debe a su localización en una de las dos posibles vías de acceso del cono sur de Sudamérica durante la primera colonización. En efecto, la Quebrada de Humahuaca (*Figura 10*) es una de las rutas posibles como alternativa a la ruta de la costa (Rothhammer y Dillehay, 2009). Posteriormente, fue una de las principales vías de paso del Imperio Inca. Del mismo modo, durante la época colonial española la Quebrada de Humahuaca formó parte de la transitada ruta que unía las minas de plata de Potosí en Bolivia con los puertos del Río de la Plata (Dipierri *et al.* 2000; Mendisco *et al.* 2014).

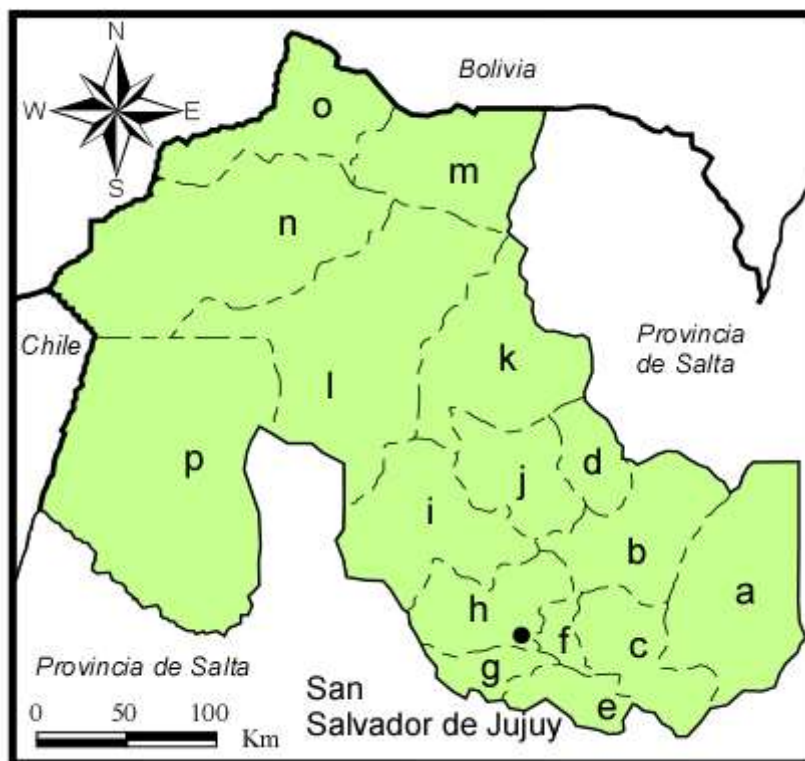


Figura 9. Mapa de la Provincia de Jujuy, con la localización de la Capital, San Salvador de Jujuy (circulo negro). Se muestra la división administrativa en departamentos de la provincia, detallados en la *Tabla Anexo 2*.

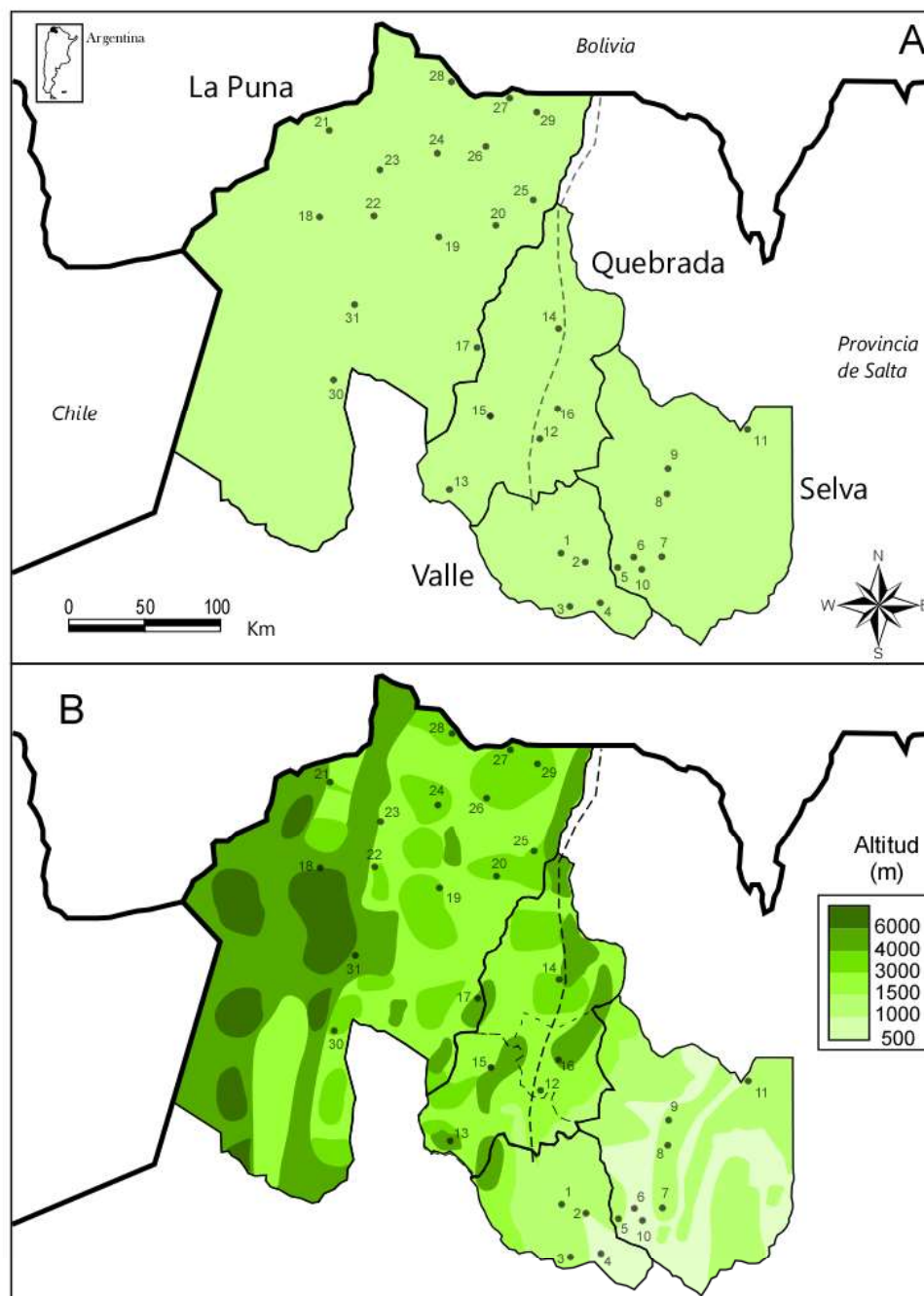


Figura 10. Características de la provincia de Jujuy. A) División administrativa de la provincia. B) Curvas de nivel para diferentes altitudes, con la Quebrada de Humahuaca en línea discontinua. En ambos mapas se muestran las localidades de muestreo (detalladas en la *Tabla Anexo 3*).

Inmediatamente antes de la llegada de los colonizadores españoles, Jujuy se encontraba poblada por varios grupos autóctonos, incluyendo atacamas, casabindos, chichas, chiriguano, cochinocas, diaguitas, jujuyes, maticos, mocovies, omaguacas, purmamarca, tilcaras y yalas, entre otros (Dipierri *et al.* 2000; Alfaro, 2010). Un reciente estudio de Mendisco *et al.* (2014) a través de

ADN antiguo destaca una composición heterogénea de esta población, pero con un posible origen Andino. Se ha registrado la presencia de varias lenguas amerindias, incluyendo aimara, chiripá, guaraní, quechua sudboliviano y wichí lhamtés vejoz, pertenecientes a familias lingüísticas tan diversas como aimara, matak, quechua y tupí (Lewis *et al.* 2014).

La provincia de Jujuy experimentó cambios importantes con la colonización europea y la correspondiente llegada masiva de inmigrantes al continente americano. De la misma forma que se ha descrito para Colombia, junto con los europeos llegaron un elevado número de africanos que fueron transportados para trabajar como esclavos en América (Curtin, 1969). Con su llegada se contribuyó a aumentar la diversidad genética de las poblaciones humanas que habitaban América en general y Jujuy en particular. Sin embargo algunas variantes genéticas de origen amerindias se perdieron debido al impacto demográfico y movimientos poblacionales resultantes de la llegada de los europeos (Mendisco *et al.* 2014)

La introducción de esclavos africanos en las colonias por el Río de la Plata comenzó en el Siglo XVI, concretamente en 1588. La mayor parte procedía de los territorios actuales de Angola, República Democrática del Congo, Guinea y el Congo, principalmente de sus regiones costeras. Como ya se ha mencionado, no se conoce el número exacto de individuos africanos capturados para trabajar como esclavos; las estimaciones abarcan un rango muy amplio, de entre 5 (Curtin, 1969; Thomas, 1997) y 11,6 millones (Lovejoy, 1982).

En todo caso, algunas fuentes muestran que Jujuy fue una de las provincias argentinas con menor presencia africana. Según el censo de 1778, menos del 8% de la población era de origen africano (*Tabla 1*). Además, su asentamiento en la provincia no ocurrió de manera uniforme, sino que se dio una mayor concentración en las regiones más bajas. Por lo tanto, las poblaciones situadas a mayor altitud tuvieron menores posibilidades de mestizaje que las poblaciones de las regiones más bajas de la provincia (Martínez-Sarasola, 1992; Dipierri *et al.* 1998, 2000; Gomes, 2006a, b).

Tabla 1. Distribución de los diferentes grupos étnicos por curatos en Jujuy, según el censo de 1778 (Dipierri et al. 2000).

Grupos Étnicos	Regiones (altitud)						Total	
	Jujuy, Perico (1500 msnm)		Tumbaya (2500 msnm)		Santa Catalina, Yavi, Rinconada y Cochinoca (3500msnm)			
	N	%	N	%	N	%	N	%
Ascendencia española	509	11,8	32	2,7	115	1,3	656	4,6
Mestizos	1751	40,7	520	44,2	460	5,2	2731	19,0
Nativos Americanos	1123	26,1	404	34,3	6692	75,2	8219	57,2
Mulatos	570	13,3	67	5,7	234	2,6	871	6,0
Africanos	317	7,4	9	0,8	24	0,3	350	2,4
Indeterminados	32	0,7	145	12,3	1369	15,4	1546	10,7
Total	4302		1177		8894		14373	

No solo la población africana se instaló preferentemente en las regiones más bajas. De un modo similar, los colonos europeos ocuparon preferencialmente esas zonas, por lo que las regiones de la Puna y la Quebrada quedaron habitadas fundamentalmente por nativos. Es muy posible que este comportamiento se debiera a las dificultades de adaptación a una elevada altitud. Pero también cabe pensar en la aridez de los terrenos de las regiones más altas, poco adecuadas en general para la agricultura y la ganadería, excepto para la cría de llamas. Así, en el censo de 1859 (*Tabla 2*), la mayor parte de la población de la Puna se dedicaba a este tipo de ganadería y los oficios derivados de la misma, hilanderos y tejedores (Gil, 2000-2001).

De acuerdo con los antecedentes históricos descritos, y al igual que en el caso de los afroamericanos, se ha observado que la distribución de los diferentes grupos humanos en esta región se encuentra estratificada, de modo que los nativos americanos son proporcionalmente más numerosos en las regiones más altas de Jujuy. Así, en el censo de 1778 ya se mencionan unas

diferencias muy notables de los diferentes componentes poblacionales en razón de su origen a diferentes altitudes (Tabla 1).

Tabla 2. Profesiones de los habitantes de la Puna de Jujuy por sexo y comarca, según el censo de 1859 (Gil, 2000-2001).

Profesión	Sta. Catalina		Cochinoca		Rinconada		Yavi	
	Mujer	Varón	Mujer	Varón	Mujer	Varón	Mujer	Varón
Sin datos	323	329	490	465	279	260	423	410
hilanderos	564	247	665	50	486	176	550	64
tejedores	6	165	7	65	22	66	112	94
ganaderos	4	35	191	515	11	40	82	283
sirvientes	16	9	19	15	21	12	40	23
en escuela	2	10	15	30	7	15	7	42
comerciantes	-	17	12	18	6	18	5	21
labradores	3	1	-	3	-	-	3	76
mineros	-	8	-	-	-	145	-	-
Total	918	821	1399	1161	832	732	1222	1013

ETNIA HUAORANI (ECUADOR)

La región Occidental de la selva Amazónica reúne una gran diversidad de culturas indígenas, entre las que se encuentran algunas de las más aisladas del mundo (Gamboa-Balbin, 2007; Finer *et al.* 2008). Una de ellas es la que identifica a la etnia Huaorani, o Waorani en su grafía anglosajona.

Tradicionalmente, el territorio de los Huaoranis se extendía a lo largo de un área de dos millones de hectáreas y abarcaba gran parte de las tres provincias más orientales de Ecuador (Napó, Orellana y Pastaza) según Wallis (1973) (Figura 11). Actualmente, habitan en la Reserva Étnica Huaorani, que comprende una pequeña parte de estas 3 provincias y se ubica junto al Parque Nacional Yasuní. Ambos territorios componen la Reserva Yasuní de la Biosfera (Figura 12), que protege un bosque templado y húmedo con una temperatura relativamente constante a lo largo del año de entre 18 y 24°C. Los ríos Napó al norte y Curaray y Vilano al sur delimitan este territorio (Cabodevilla, 1994).

Los Huaorani son también conocidos como Sabela, Auishiri y Waodani. La forma singular es Huao (Wao, en inglés). Dentro de la etnia Huaorani existen varios grupos diferenciados, (Toñampare, Quenahueno, Tihueno, Quihuaro, Damuintaro, Zapino, Tigüino, Huamuno, Dayuno, Quehueruno, Garzacocha, Quemperi, Mima, Caruhue, Tagaeri) que en conjunto pueden suponer no más de 1.800 a 2.000 individuos (Lewis *et al.* 2014).

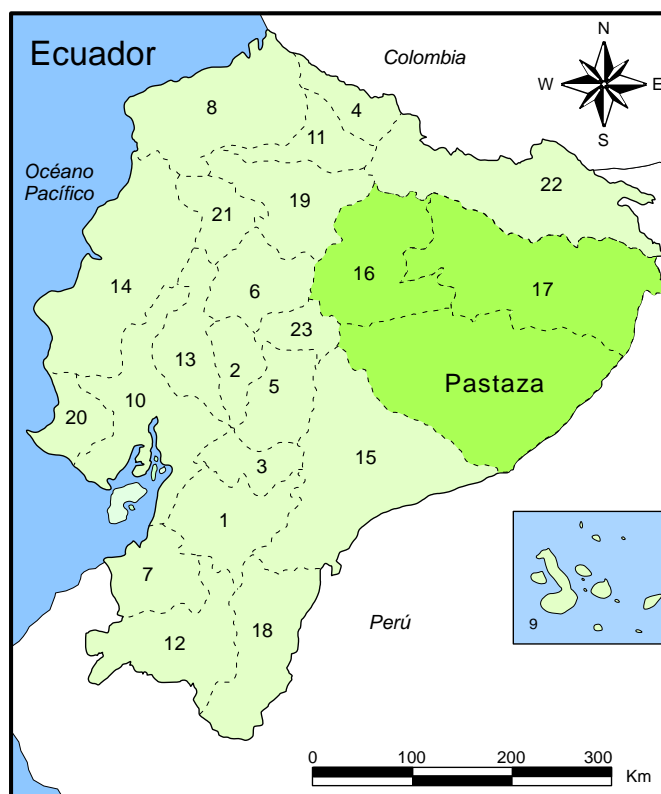


Figura 11. Mapa de Ecuador donde se detallan las provincias junto con los países limítrofes. La localización de las provincias que albergan el Parque Nacional Yasuní y la Reserva Huaorani (Pastaza, Napo y Orellana) se destaca con el color verde más intenso. La muestra analizada en este trabajo procede de la provincia de Pastaza. El nombre del resto de las provincias se detalla en la *Tabla Anexo 4*.

La historia del pueblo Huaorani se ha caracterizado por un alto grado de aislamiento hasta hace unas pocas décadas, lo que puede explicar que sus orígenes, tanto étnicos como lingüísticos sigan siendo fuente de especulación. Se ha propuesto que los Huaorani podrían derivar de grupos de origen tupi-guaraní (omaguas, yameos y oas) que se dispersaron en la zona entre los Siglos X al XVI. Estos grupos, que se desplazarían desde Brasil hacia el Pacífico, habrían terminado

estableciéndose finalmente en la selva amazónica, a lo largo de la ribera del río Napo, ante la imposibilidad de cruzar la Cordillera de los Andes (Cabodevilla, 1994).

En todo caso, hasta bien entrado el S. XX han practicado un modo de vida básicamente cazador-horticultor, alimentándose de la caza (sobre todo grandes monos y pecaríes), cultivos en claros del bosque y mediante la recolección de plantas silvestres; su carácter seminómada les lleva a cambiar sus asentamientos con una cierta frecuencia, dependiendo de las épocas de las cosechas. Actualmente están adaptándose de forma progresiva al sedentarismo y dejando atrás su modo de vida más tradicional (Rival, 2002; Papworth *et al.* 2013).

La unidad básica tradicional de organización social entre los Huaorani es el *Nanicabo*, que puede comprender entre seis y diez familias que viven bajo un mismo techo. Estos grupos domésticos son autosuficientes y autónomos. La unidad territorial o grupo local se conoce como *Huaomoni* y se corresponde con la unión de varios *Nanicabos*. Esos *Huaomoni* tienen un marcado carácter endogámico y dentro de él existen relaciones de alianza. Gran parte de las comunidades están formadas por dos o tres *Huaomonis* de diversa procedencia (Beckerman *et al.* 2009; González-Andrade *et al.* 2009).

Como consecuencia de su estilo de vida, la sociedad Huaorani es una población homogénea y tiene un alto grado de endogamia (Cardoso *et al.* 2008, 2012; González-Andrade *et al.* 2009; Baeta *et al.* 2009, 2013; Baeta, 2011). Existe un marcado predominio de los matrimonios bilaterales cruzados entre primos, que por lo general son organizados por los padres de la joven pareja (Robarchek y Robarchek, 1992; Beckerman *et al.* 2009). No obstante, en ocasiones, se autoriza la poligamia (poliginia y menos frecuentemente, poliandria). Esta situación se considera especial, y se debe habitualmente a los casos en los que existe algún desequilibrio demográfico debido a las guerras o cualquier otro evento catastrófico.

Los Huaorani habitan en una región caracterizada por los comportamientos violentos de sus pobladores, como es la cuenca occidental del Amazonas, existiendo todavía en la actualidad

constancia de episodios violentos de matanzas por venganza (Pappalardo *et al.* 2013). Los Huaorani participan de ello, ya que han sido descritos como una de las sociedades más belicosas del mundo, caracterizada por unas altas tasas de homicidios (Yost, 1981; Beckerman *et al.* 2009). Hasta donde se conoce, los Huaorani se han caracterizado tradicionalmente por sucesiones de conflictos, incluso entre grupos estrechamente relacionados. Los motivos han sido variados, incluyendo disputas sobre arreglos matrimoniales, represalias por homicidios anteriores, acusaciones de brujería, etc.

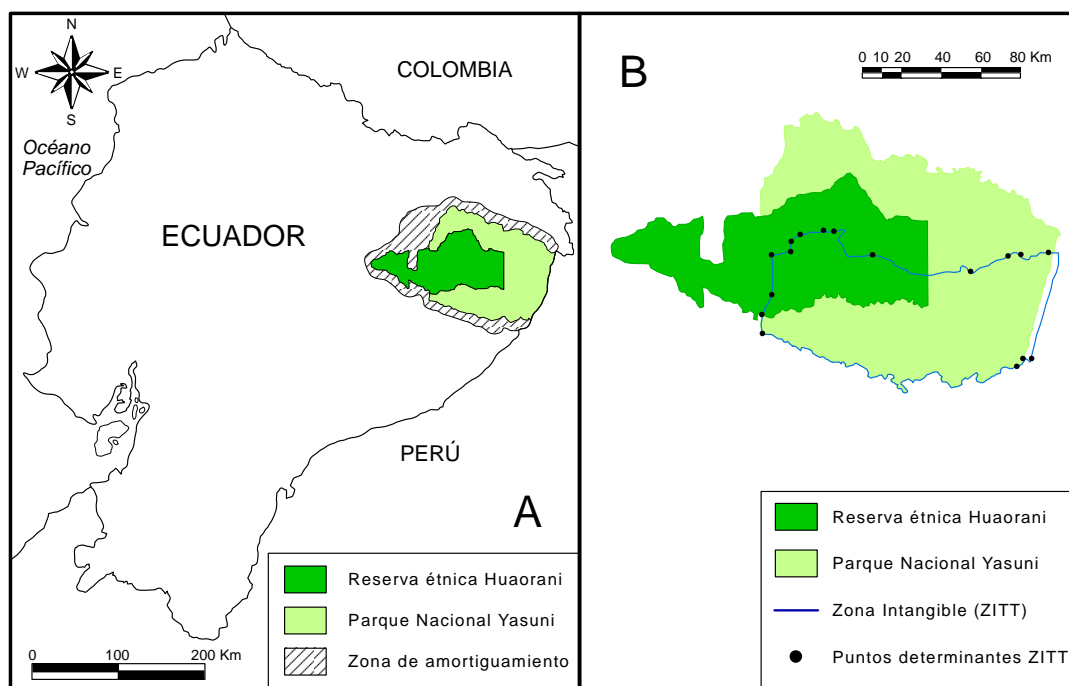


Figura 12. A) Localización de la Reserva Yasuní de la Biosfera, compuesta por la Reserva Étnica Huaorani y el Parque Nacional Yasuní. B) Zona Intangible Tagaeri Taromenane (ZITT), se fija mediante 17 puntos geográficamente unívocos declarados en el Decreto 2187 (Decreto Ejecutivo, 2007). Para asignar las fronteras de la ZITT es necesario unir los pares de coordenadas por segmentos rectilíneos y por tramos de cursos fluviales de acuerdo con las descripciones geomorfológicas incluidas en el Decreto. Modificado de Pappalardo *et al.* (2013).

En un trabajo realizado sobre genealogías de hasta cinco generaciones, Yost (1981) estimó que alrededor del 60% de las muertes de adultos fueron causadas por acciones de guerra; de éstas un 44% fue resultado de disputas internas. Por lo tanto, la violencia y la venganza han sido considerados como principales factores determinantes de la mortalidad entre el pueblo Huaorani

(Wallis, 1973; Robarchek y Robarchek, 1992; Boster *et al.* 2003; Beckerman *et al.* 2009). Es posible que debido a esta agresividad, los Huaorani no hayan sido absorbidos en el pasado por otros grupos.

Los Huaorani hablan un idioma (*Huao Terero*) aislado y sin clasificar, debido a que su construcción lingüística se considera única, sin similitudes con la fonología o estructura de otras familias lingüísticas de la Amazonia (Peeke, 1973).

El primer contacto pacífico de las culturas occidentales con la comunidad Huaorani fue realizado por misioneros norteamericanos en 1958 (Wallis, 1973). A partir de ese momento, se inició un proceso de reagrupación. A finales de la década de los años 70 del pasado siglo, el 90% de la población Huaorani había sido reubicada en asentamientos. Esta concentración trajo consecuencias negativas para sus habitantes (enfermedades, tensiones inter-grupales y agotamiento de recursos), provocando una importante reducción poblacional. Por este motivo se inició un proceso inverso de descentralización con la formación de varias comunidades nucleadas a lo largo de los ríos orientales de las provincias de Napo, Pastaza y Orellana. Posteriormente, ha existido un contacto limitado, pero creciente, con los extranjeros (Yost, 1981; Boster *et al.* 2003). No obstante, la mayor parte de los Huaorani conviven actualmente en contacto con otros pueblos indígenas vecinos (kichwas), así como con colonos y trabajadores de la industria petrolera y maderera (Fuentes, 1997). Aunque algunos grupos Huaorani aún viven completamente aislados en la selva amazónica (Yost, 1981; Boster *et al.* 2003) como los Tagaeri y Taromenane, dos tribus indígenas que han decidido vivir en un aislamiento voluntario (Cabodevilla, 2008; Colleoni y Proaño, 2010; Papworth *et al.* 2013). Para proteger el territorio donde habitan éstas tribus en la Reserva Yasuní, el Gobierno Nacional de Ecuador declaró en 1999 y asignó, en 2007, a través de dos decretos presidenciales la Zona Intangible Tagaeri Taromenane (ZITT) (Decreto Ejecutivo 1999, 2007) (Pappalardo *et al.* 2013).

MARCADORES GENÉTICOS

Etimológicamente, polimorfismo quiere decir "varias formas". En Biología, se define un polimorfismo como aquella ocurrencia, dentro de una población, de una variación discontinua de origen genético presente en un locus y cuya frecuencia no puede ser explicada por la mutación. Para que una variación genética no se pueda explicar mediante la mutación se considera que su frecuencia debe ser de, al menos, un 1%. Por lo tanto se puede decir que un polimorfismo debe tener al menos 2 variantes con una frecuencia igual o superior al 1% (Berovides-Alvarez y Alfonso-Sánchez, 1995).

Se define como marcador genético o marcador molecular a un segmento de ADN con una ubicación física identificable (locus) en un cromosoma y cuya herencia genética se puede rastrear. Un marcador puede ser un gen, una región del ADN sin función conocida, un nucleótido, etc.

En los siguientes párrafos se describirán con un cierto detalle los marcadores genéticos utilizados en los análisis y publicaciones de esta Tesis.

ELEMENTOS *ALU*

Los SINE (Short Interspersed Nuclear Elements) son elementos repetitivos nucleares dispersos y cortos (entre 100 y 500 pb) con un promotor interno de la enzima Polimerasa III, pero sin capacidad para codificar proteínas ni para transponerse (Batzer y Deininger, 2002; Witherspoon *et al.* 2009).

Los elementos *Alu* representan la familia más grande de SINEs presente en nuestra especie (Batzer y Deininger, 2002). Fueron descubiertos en el año 1979 como un componente de las curvas de renaturalización del ADN humano. Reciben este nombre porque en su secuencia existe un lugar de reconocimiento para la enzima de restricción *AluI* (AGCT) (Houck *et al.* 1979). El genoma humano contiene cerca de 1.100.000 repeticiones *Alu*, que representan aproximadamente un 11% del total del ADN nuclear (Lander *et al.* 2001; Ade *et al.* 2013; Lee *et al.* 2015). Aproximadamente un

tercio del total de los dinucleótidos CpG en el genoma humano está contenido en la familia *Alu* (Batzer y Deininger, 2002; Su *et al.* 2014).

Los elementos *Alu* son considerados elementos móviles, ya que el promotor interno les permite realizar copias de sí mismo que se insertan en diferentes lugares del genoma; pero el hecho de no codificar proteínas implica que necesiten de la acción de otros elementos, habitualmente a los elementos LINE (Konkel y Batzer, 2010; Ade *et al.* 2013; Lee *et al.* 2015).

Como otros SINE, las repeticiones *Alu* suelen estar ubicadas en regiones no codificantes (espacios intergénicos e intrones) (Batzer *et al.* 1990). Su distribución a lo largo del genoma no es al azar, ya que se encuentran preferentemente en regiones ricas en genes (Batzer y Deininger, 2002; Su *et al.* 2014).

Del mismo modo que con otros elementos repetitivos, las inserciones *Alu* fueron consideradas inicialmente como parásitos del genoma, por lo que se acuñó el término ADN egoísta o basura (*Selfish DNA* o *Junk DNA*, en inglés) para referirse a ellas.

Por el contrario, en la actualidad se considera que estos fragmentos de ADN han tenido una gran influencia en la evolución del genoma (Häsler y Strub, 2006; Konkel y Batzer, 2010). Así, debido a la abundancia de las repeticiones *Alu* y al alto grado de similitud de sus secuencias, podrían incrementar las tasas de recombinación (Witherspoon *et al.* 2009) o actuar como puntos de nucleación para la recombinación homóloga desigual (Deininger y Batzer, 1999), aumentando la variabilidad genética intraespecífica mediante reorganizaciones, duplicaciones y deleciones (Deininger *et al.* 2003). En un reciente estudio realizado *in vitro* destacan el efecto de *Alu* en la inestabilidad del genoma a través de la recombinación *Alu-Alu* y la reparación de las roturas en el ADN de doble hebra (Morales *et al.* 2015).

Otro efecto importante podría ser la aparición de genes duplicados y la consiguiente generación de nuevas proteínas que podrían llegar a realizar nuevas funciones (Batzer y Deininger, 2002; Grover *et al.* 2005). Además, los elementos *Alu* pueden introducir señales de poliadenilación,

splice sites alternativos, que pueden conducir a exonización con grandes efectos sobre el transcriptoma humano (Sela *et al.* 2007).

La especiación es otra de las funciones que se especula pueda tener el llamado ADN basura. Su rápida evolución podría crear incompatibilidades entre poblaciones reproductivamente aisladas, mediante la alteración del apareamiento de cromosomas homólogos, impidiendo que se reprodujeran (Ferree y Barbash, 2009; Ferree *et al.* 2014). En el caso de las inserciones *Alu*, podrían promover la recombinación entre diferentes cromosomas con secuencias repetitivas similares, dando lugar a nuevos fragmentos cromosómicos (Roy-Engel *et al.* 2008).

También podrían tener un efecto epigenético, silenciando genes o incluso regiones cromosómicas mediante mecanismos de metilación de la citosina (Slotkin y Martienssen, 2007; Rodríguez *et al.* 2008; Prendergast *et al.* 2014), habiéndose identificado ya algunos genes diana y el mecanismo de acción de varios elementos transponibles en *Arabidopsis* (McCue *et al.* 2013). Asimismo, se ha propuesto una relación entre la metilación del gen *POMC*, mediada por inserciones *Alu* y la obesidad infantil (Kuehnen y Krude, 2012).

Por otra parte, aparte del impacto positivo, también ha tenido consecuencias negativas sobre el genoma humano. Dado que las repeticiones *Alu* afectan a la composición, organización y expresión del genoma, juegan un papel importante en la aparición de enfermedades genéticas humanas. Los trastornos patológicos debidos a inserciones *Alu* se pueden dividir en tres clases (Deininger y Batzer, 1999; Grover *et al.* 2005): i) Los trastornos causados por la retrotransposición, ii) Los trastornos causados por la recombinación, iii) Los trastornos causados por la exonización. Entre los casos identificados pueden citarse el síndrome de Apert, la Neurofibromatosis tipo 1 y las Hemofilias A y B (Cordaux *et al.* 2006).

Origen y estructura de los elementos *Alu*

La aparición de las inserciones *Alu* se remonta a hace varios millones de años. Se estima que aparecieron al inicio de la radiación de los primates, lo que explicaría que sean específicos de este orden de mamíferos. Están consideradas como un acontecimiento evolutivo relativamente reciente (Batzer y Deininger, 2002; Roy-Engel *et al.* 2008) debido a su relativamente escasa antigüedad.

El origen de los elementos *Alu* se explica a través de duplicaciones del gen 7SL RNA (Ullu y Tschudi, 1984; Batzer *et al.* 1994, 1996a; Batzer y Deininger, 2002; Häsler y Strub, 2006). El gen 7SL RNA forma parte del complejo ribosómico y es el encargado de codificar para una partícula de reconocimiento de la señal del retículo endoplásmico (Ullu y Tschudi, 1984). Las estimas realizadas indican que la mayoría de los elementos *Alu*, comparte el 90% de su secuencia con el gen 7SL RNA (Novick *et al.* 1998).

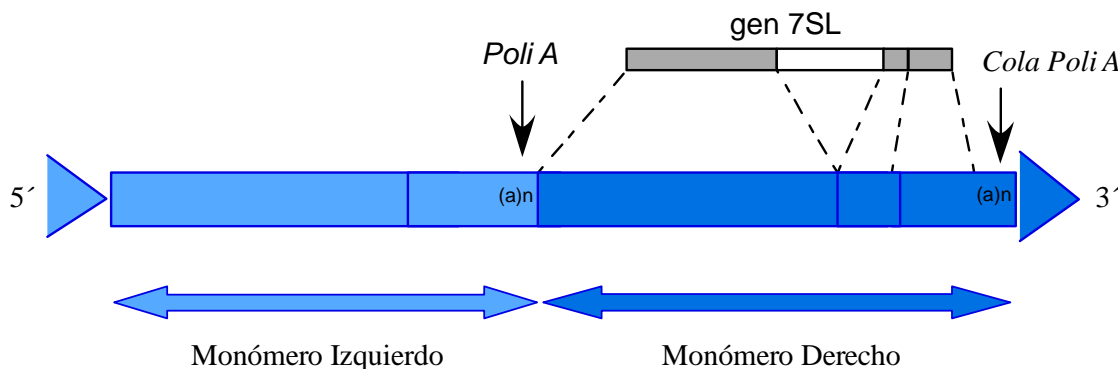


Figura 13. Representación de la estructura dímica de un elemento *Alu*. Se muestran las dos subunidades que lo componen y la correspondencia con el gen 7SL. Ambos monómeros están separados por una región rica en A y el extremo 3' contiene una cola poli A.

El elemento *Alu* moderno es un heterodímero (Figura 13) cuyas dos subunidades descienden de la misma secuencia ancestral (Roy-Engel *et al.* 2008; Konkel y Batzer, 2010) llamada FAM (*Fossil Alu Monomer*). Dicha secuencia *Alu* primitiva tenía una estructura monomérica. La secuencia *Alu* moderna surgió probablemente de la fusión cabeza-cola de dos FAM (Figura 14), lo que dio lugar a

la actual estructura dimérica, compuesta por dos subunidades similares, unidas por una región rica en Adenina y con una cola Poli-A (Figura 13). La secuencia *Alu*, además presenta unas repeticiones cortas flanqueando su secuencia (Economou *et al.* 1990; Quentin, 1992a, b; Rowold y Herrera, 2000; Häsler y Strub, 2006; Roy-Engel *et al.* 2008; Lee *et al.* 2015). El monómero izquierdo o 5' contiene dos elementos promotores de la RNA polimerasa III, las cajas A y B, de unos 10 pb cada uno (Figura 15). Mediante estos promotores se controla la transposición. En cambio, el monómero del extremo derecho contiene un segmento adicional de 31 pb (Jurka y Zuckerkandl, 1991; Quentin, 1992a). En promedio, las inserciones *Alu* son de aproximadamente 300 pb de longitud.

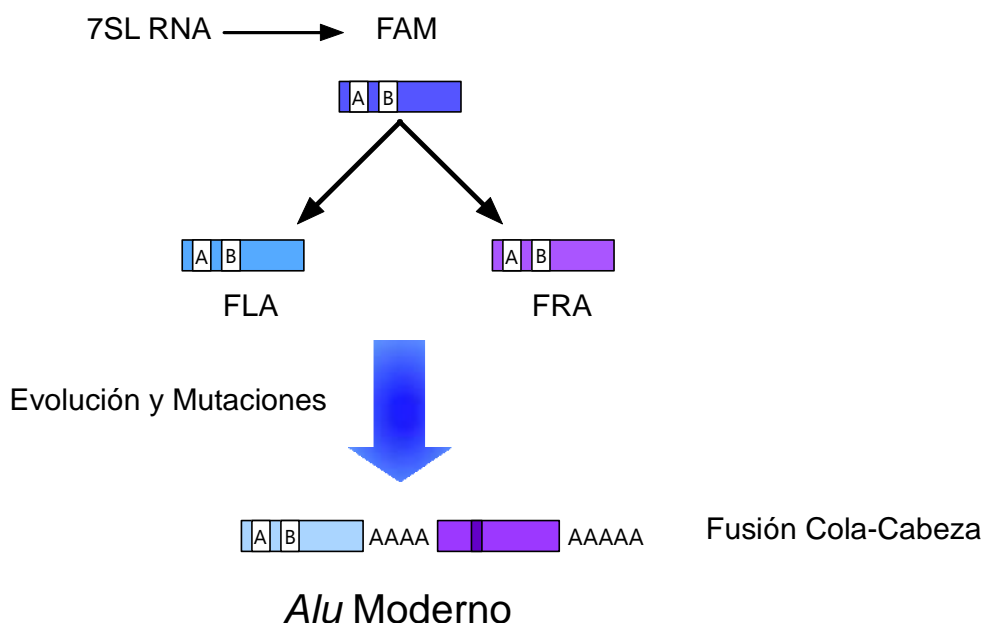


Figura 14. Representación del origen y formación del elemento *Alu*. Se cree que los elementos *Alu* surgieron a partir de un procesado del gen 7SL RNA que dio lugar al elemento ancestral, el monómero *Alu* fósil (FAM). Dicho FAM evolucionó al monómero izquierdo (FLA) y monómero derecho (FRA). Estos dos monómeros presentaban variaciones de secuencia entre sí y fueron acumulando mutaciones a lo largo del tiempo. El primer elemento *Alu* de estructura dimérica posiblemente surgió a través de la fusión de los monómeros FLA y FRA. (Modificado de Roy-Engel *et al.* 2008).

El origen de los elementos *Alu* se puede dividir en dos fases: i) La aparición del FAM, que se remonta al comienzo de la radiación de los mamíferos (~112 Ma.), llamada fase monomérica. ii) La fase dimérica, como consecuencia de la fusión de dos FAM, que se produjo en algún momento

entre la divergencia de roedores y primates (hace unos 100 Ma.) y la radiación de los primates, (hace aproximadamente 65 Ma.) (Kapitonov y Jurka, 1996; Roy-Engel *et al.* 2008).

La tasa de inserción de elementos *Alu* ha variado notablemente con el tiempo. Se ha estimado que los valores más altos se dieron hace unos 35 a 40 Ma. (Shen *et al.* 1991; Britten, 1994; Batzer y Deininger, 2002; Wu *et al.* 2014), con una tasa de aproximadamente una nueva inserción por cada nacimiento. Aparentemente, desde entonces ha ido disminuyendo progresivamente hasta alcanzar un valor 100 a 200 veces menor (Deininger y Batzer 1999). Pero hay constancia de que los elementos móviles han continuado propagándose en el genoma humano desde la divergencia con los chimpancés. De hecho se han encontrado evidencias de un reciente incremento en la tasa de expansión de los elementos móviles (elementos *Alu* entre ellas) en la línea humana comparada con la de chimpancés (Mills *et al.* 2006).

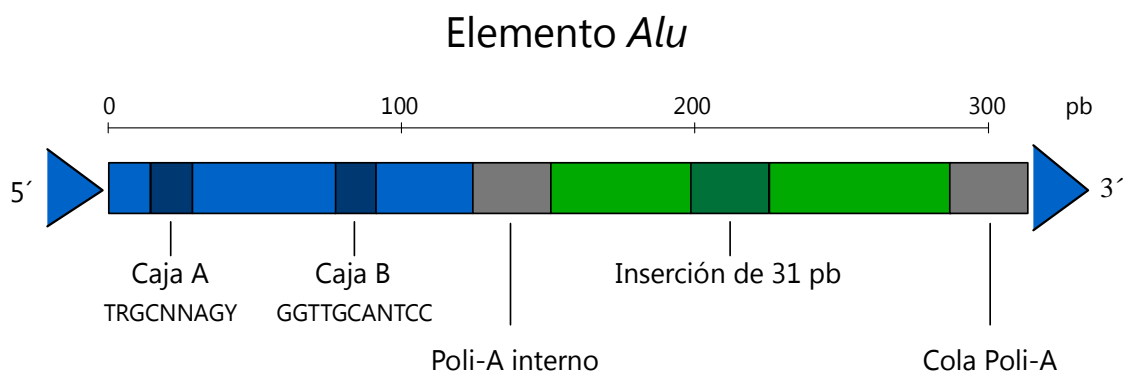


Figura 15. Estructura dimérica del elemento *Alu*. Los dos monómeros están unidos por una secuencia Poli-A interna. Existen diferencias entre ambos monómeros; el monómero derecho tiene un fragmento específico de 31 pb, mientras que el monómero izquierdo contiene el promotor para la RNA polimerasa III (Caja A y Caja B). Al final de elemento se encuentra otra secuencia Poli-A (modificado de Hasler y Strub, 2006)

Retrotransposición

Conforme a su cualidad de elementos móviles, las inserciones *Alu* se insertan en diferentes regiones del genoma utilizando un intermediario de ARN. El mecanismo por el cual se producen las copias que se insertan en otros lugares, se denomina retrotransposición o TPRT (*target primed*

reverse transcription). Este proceso implica la transcripción inversa de un transcrito *Alu* mediante una RNA polimerasa III (Luan *et al.* 1993; Novick *et al.* 1996; Batzer y Deininger, 2002; Ade *et al.* 2013; Wu *et al.* 2014).

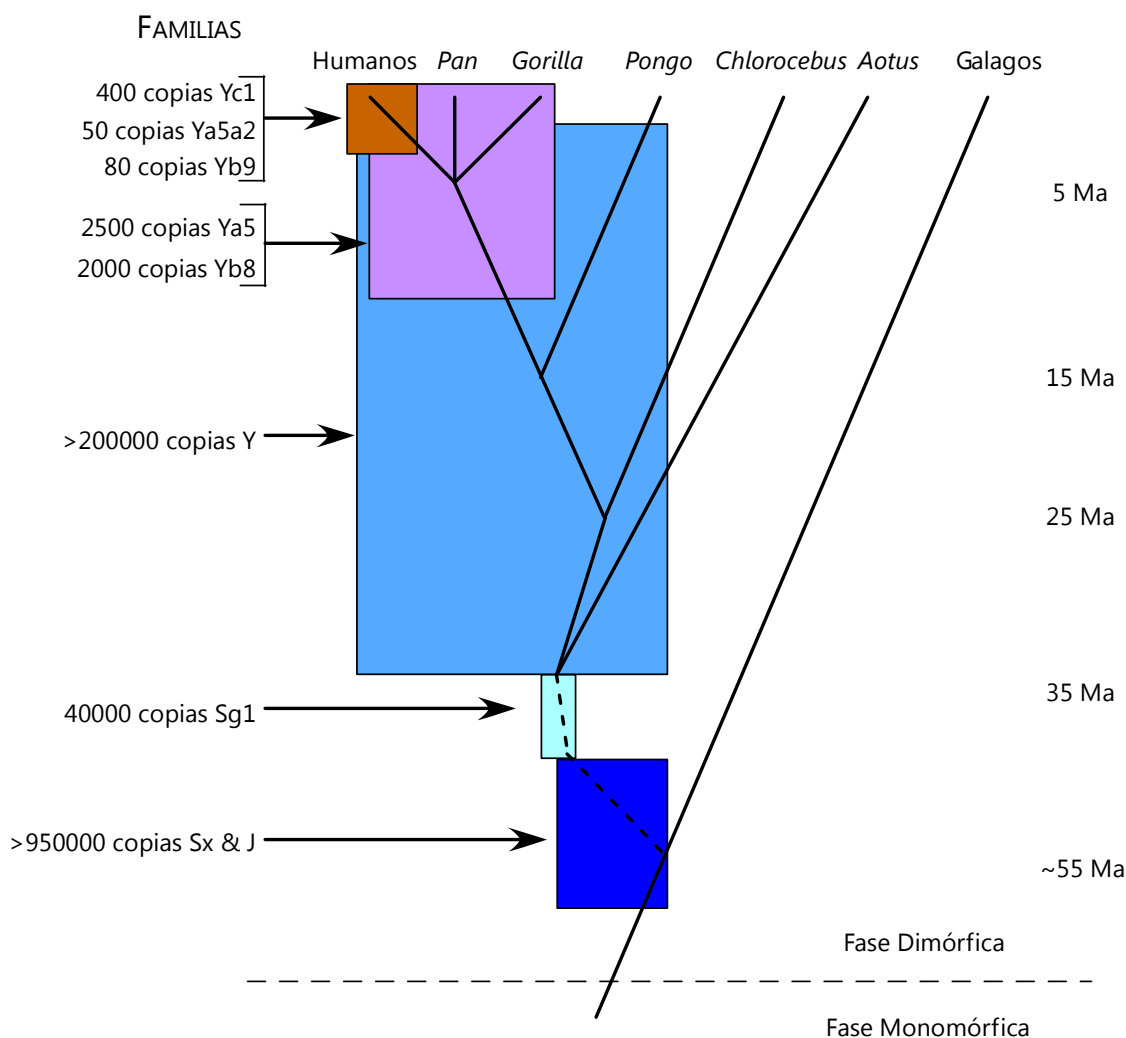


Figura 16. Árbol filogenético de las subfamilias *Alu* durante la radiación de los primates. Las subfamilias *Alu* más antiguas (J, SX y Sg1) fueron más activas hace alrededor de 35-55 millones de años, dando lugar a la mayoría de los elementos *Alu* presentes en la actualidad en el genoma humano. Las áreas enmarcadas representan el período potencial de la actividad máxima para cada subfamilia *Alu*. La tasa de amplificación de *Alu* disminuyó con el tiempo evolutivo como se observa por la reducción en el número de copias. Actualmente las subfamilias *Alu* jóvenes (Y, YA5, Yb8, YA5, Ya5a2 y YC1) constituyen todos los *Alu* polimórficos conocidos en el genoma humano (Tomado de Batzer y Deininger, 2002).

Los mecanismos necesarios para realizar la amplificación de elementos *Alu* requieren la presencia de dos enzimas específicas (Transcriptasa Inversa y Endonucleasa). Como se ha

mencionado anteriormente, los elementos *Alu* no codifican para estas enzimas. Por lo tanto, para que se realice el TPRT deben ser codificadas por otras secuencias, probablemente derivadas de elementos LINEs (Mathias *et al.* 1991; Batzer y Deininger, 2002; Häsler y Strub, 2006; Lee *et al.* 2015) y posiblemente de los L1 (Ade *et al.* 2013).

El proceso de TPRT puede dividirse en dos fases: i) La primera comprende la generación de un transcrito ARN complementario a la secuencia *Alu* mediante la acción de la RNA Polimerasa III. (Batzer *et al.* 1996a; Stoneking *et al.* 1997; Batzer y Deininger, 2002; Ade *et al.* 2013); ii) La segunda consiste en la utilización de este ARN como molde para generar un ADN complementario gracias a la acción de la Transcriptasa Inversa. Este ADN complementario es integrado en un nuevo lugar del genoma. Durante el proceso de inserción se producen las características repeticiones cortas que flanquean al elemento *Alu*, mencionadas anteriormente (Batzer y Deininger, 2002; Ade *et al.* 2013).

Con el tiempo, se han ido acumulando mutaciones en la secuencia de las sucesivas inserciones. Algunas de estas mutaciones han permitido caracterizar linajes y con ellos establecer una clasificación jerarquizada de las familias de inserciones *Alu* (Batzer *et al.* 1996b; Kapitonov y Jurka, 1996) (*Figura 16; Figura 17*).

Clasificación de las subfamilias *Alu*

Una familia es un grupo de elementos *Alu* que comparten una serie de mutaciones en lugares de su secuencia que se identifican como posiciones diagnósticas (Quentin, 1992a; Batzer *et al.* 1994; Batzer y Deininger, 2002; Häsler y Strub, 2006).

Se han definido 12 grandes familias de elementos *Alu*, en función del tiempo transcurrido desde su aparición (Kapitonov y Jurka, 1996), clasificadas como antiguas, intermedias y jóvenes (*Figura 17*). Los elementos pertenecientes a la misma familia tienen edades de inserción similares (Quentin, 1992a; Batzer *et al.* 1994; 1996b; Stoneking *et al.* 1997; Novick *et al.* 1998; Batzer y Deininger, 2002; Häsler y Strub, 2006). Las posiciones diagnósticas características de una nueva

familia se habrán originado mediante mutaciones sobre la secuencia de una inserción antecesora perteneciente a otra familia (Quentin, 1992a, b). Con el tiempo, las familias más recientes poseerán un número mayor de mutaciones en posiciones diagnósticas, siendo algunas de ellas específicas (Figura 18).

Las familias antiguas se denominan J_o y J_b y presentan edades de inserción de alrededor de 81 Ma. Se insertaron después de la separación entre roedores y primates (hace 100 millones de años) y antes de la radiación de estos últimos (65 Ma.) (Kapitonov y Jurka, 1996).

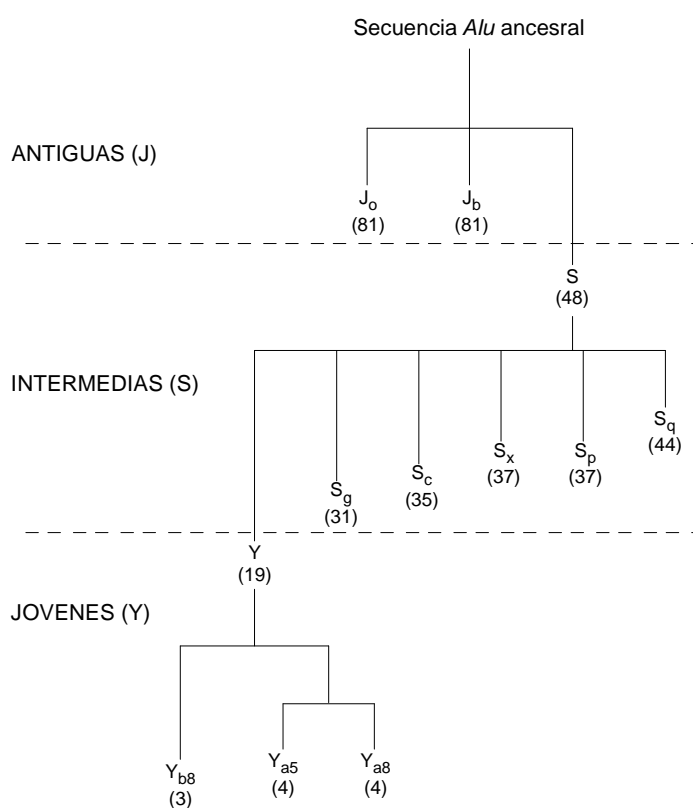


Figura 17. Representación de las 12 subfamilias *Alu*, con su clasificación en antiguas (J), intermedias (S) y jóvenes (Y). Los números entre paréntesis representan la antigüedad (en millones de años) de cada una de ellas (Mighell *et al.* 1997).

Las intermedias son las familias S (S_q , S_p , S_x , S_c , S_g) y su edad de inserción se encuentra entre 44 y 31 Ma. Comprenden más del 80% de los loci *Alu* en el genoma humano, producto de la alta tasa de retrotransposición que manifestaron durante ese periodo (Roy-Engel *et al.* 1998).

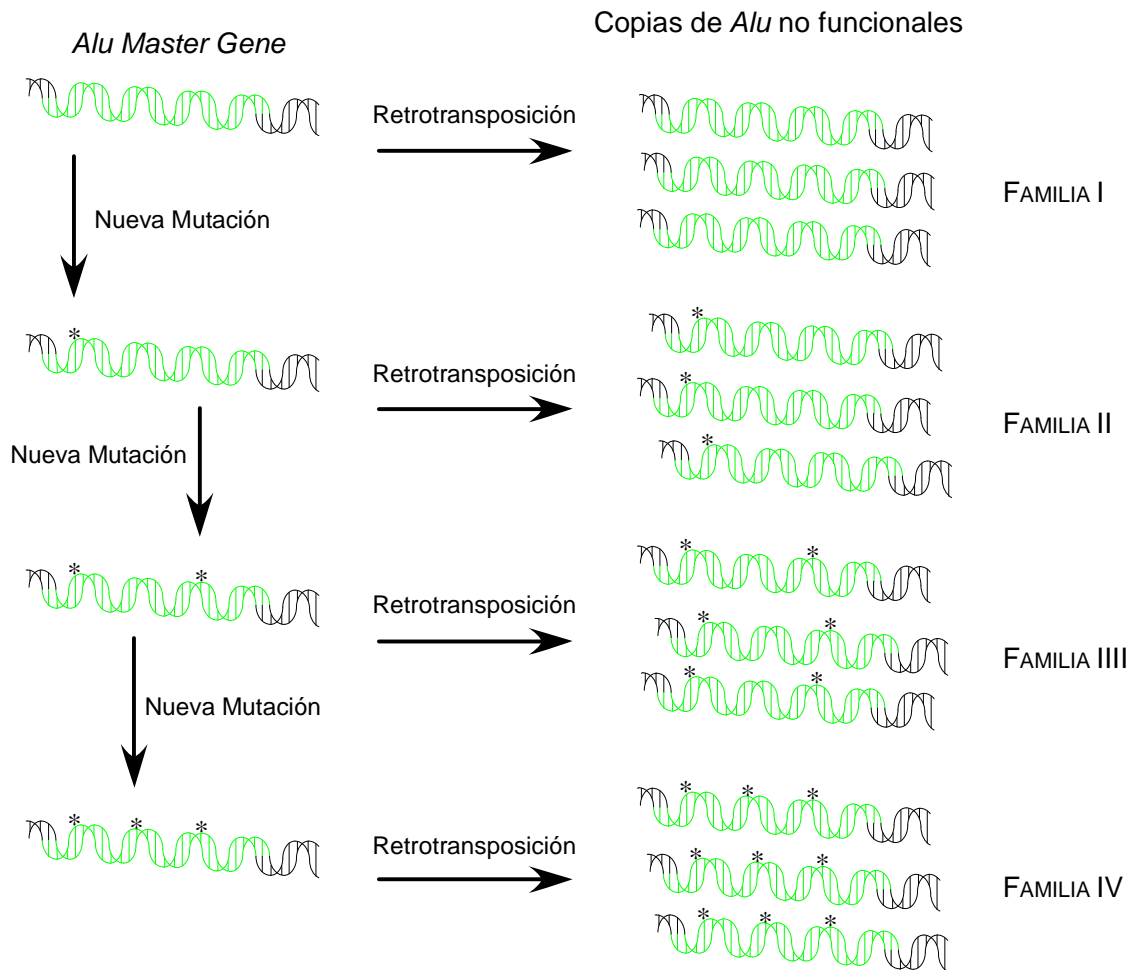


Figura 18. Esquema que representa la acumulación de mutaciones en posiciones diagnósticas de sucesivas subfamilias de elementos *Alu*. A medida que el *Alu Master Gene* va adquiriendo mutaciones, aparecen nuevas familias *Alu* que tienen en común esas nuevas mutaciones (posiciones diagnósticas) además de las que han acumulado anteriormente (modificado de Stoneking *et al.* 2001).

De entre todas las subfamilias, el tercer grupo es particularmente interesante. Se compone de las inserciones *Young* (Y), que son las más recientes. Se agrupan en varias subfamilias (Figura 16). Si bien la mayor parte de los elementos *Alu* se encuentran en estado monomórfico, excepcionalmente en la familia *Young* aproximadamente el 25% de sus elementos se ha retrotranspuesto tan recientemente que son polimórficos para la presencia/ausencia de la inserción (Batzer y Deininger, 2002). En un trabajo reciente, se han censado 5799 inserciones *Alu Young* polimórficas (Witherspoon *et al.* 2013).

Inserciones *Alu* en estudios Genético-Evolutivos

La utilidad de las inserciones *Alu* en estudios de Antropogenética se basa precisamente en el carácter polimórfico de los elementos de la familia Y.

Estos polimorfismos presentan una serie de características que los convierte en marcadores extraordinariamente útiles:

En primer lugar, su identidad por descendencia. Todas las copias de una inserción determinada son idénticas por descendencia, de modo que todos los individuos que la compartan serán descendientes de un mismo antecesor común. Esto es así porque la probabilidad de que se produzca la inserción de dos elementos *Alu* en un mismo lugar del genoma de manera independiente es insignificante. Por ello, se dice que los polimorfismos de inserción *Alu* están libres de homoplasia. En cambio, otros marcadores polimórficos (SNPs, STRs,...) son considerados idénticos por estado, ya que el mismo alelo puede aparecer de manera independiente un número indeterminado de veces; en este caso, dos individuos que compartan un mismo alelo no necesariamente tendrán una relación de parentesco. Por este motivo, los elementos *Alu* son muy fiables para representar relaciones filogenéticas (Batzer *et al.* 1994; Novick *et al.* 1998; Batzer y Deininger, 2002).

Otra característica clave es el estado ancestral conocido. La posibilidad de inferir el estado ancestral de los elementos *Alu* añade valor a este tipo de marcadores. Esto es posible porque la dirección del cambio mutacional es necesariamente hacia la adquisición del elemento. La ausencia de inserción es, por lo tanto, el estado ancestral (Batzer *et al.* 1994; Novick *et al.* 1998; Batzer y Deininger, 2002). Esto facilita enormemente su análisis estadístico, ya que entre otras cosas, es posible simular una población ancestral hipotética, con frecuencias génicas originales (100% de ausencia de inserción) a partir de la cual se puede enraizar un árbol filogenético, o conseguir una mayor exactitud en los análisis de relaciones poblacionales (Stoneking *et al.* 1997). Existen variaciones sobre este patrón, como el elemento *Alu* CD4, cuyo estado ancestral es la presencia de

la inserción. Esto se debe a que un elemento *Alu* fijado experimentó una delección parcial. Esta delección *a posteriori* es la que dio lugar al polimorfismo actual en las poblaciones humanas. En cualquier caso, la inferencia de su estado ancestral también es posible.

En tercer lugar, cabe destacar la estabilidad de la inserción. La baja probabilidad de que se realice una delección de un elemento *Alu* junto con la, también baja tasa de inserciones *de novo*, minimizan sobremanera la tasa de cambio. En las raras ocasiones en las que se produce una delección de un elemento, esta no ocurre de manera completa, de modo que quedan rastros en el ADN que permiten reconocer que en algún momento hubo una inserción *Alu* en ese lugar (Batzer *et al.* 1994; Novick *et al.* 1998; Batzer y Deininger, 2002).

Cabe destacar también que, en general, los elementos *Alu* son considerados neutrales en relación a la selección natural. Por lo tanto son polimorfismos cuyas frecuencias dependerán fundamentalmente de las interacciones entre los procesos de deriva genética y flujo génico (Comas *et al.* 2000).

Por último, hay que destacar que estos polimorfismos son fáciles de analizar, mediante técnicas sencillas, rápidas y poco costosas (generalmente PCR y electroforesis de agarosa) (Comas *et al.* 2000), aunque existen métodos más sofisticados, en el caso de que sea necesario analizar un gran número de inserciones (Asari *et al.* 2012).

Los polimorfismos de inserción *Alu* son marcadores genéticos robustos. Como una prueba indirecta de su robustez como marcadores filogenéticos, puede mencionarse que los resultados obtenidos con estos polimorfismos concuerdan con los obtenidos con otros, incluso en los casos en que se utilizó un número reducido de inserciones (Stoneking *et al.* 1997; Comas *et al.* 2000).

Por estos motivos, los polimorfismos de inserción *Alu* sirven eficazmente como marcadores de ADN nuclear para el estudio de la evolución humana (Stoneking *et al.* 1997). La gran mayoría de estos polimorfismos se encuentran en cromosomas autosómicos, por lo que son marcadores biparentales y reflejan tanto la historia paterna como la materna de una población (Nasidze *et al.*

2001). Existen algunos elementos *Alu* que se han insertado en el cromosoma Y (YAP), lo que adicionalmente permite ayudar a establecer linajes de haplogrupos de herencia patrilineal.

La gran mayoría de los análisis realizados en trabajos que utilizan los polimorfismos de inserción *Alu* se basan en las frecuencias de inserción. Sin embargo, existen otras formas de utilizar estos polimorfismos como marcadores. Así por ejemplo, la diversidad nucleotídica de los elementos *Alu* es un parámetro que se puede utilizar para examinar la estructura genética de las poblaciones humanas y también de datar el evento de inserción (Knight *et al.* 1996). Otro enfoque posible consiste en analizar haplotipos compuestos de polimorfismos de inserción *Alu* ligados, solos o en combinación con otros marcadores, como STRs o SNPs (González-Pérez *et al.* 2007, 2010; El Moncer *et al.* 2010) permitiendo así, por ejemplo, la comparación de la diversidad en cromosomas con y sin un elemento *Alu* determinado (Tishkoff *et al.* 2000).

Todos los elementos *Alu* analizados en los trabajos que conforman esta memoria tienen en común su ubicación en cromosomas autosómicos, pero pueden dividirse en dos grupos:

- I. Las que denominaremos *Alu* autosómicas, que comprenden un grupo de inserciones que se encuentra distribuido entre varios cromosomas. Se compone de 12 inserciones *Alu* independientes, sin desequilibrio de ligamiento entre ellas.
- II. Las MHC-*Alu*, que se localizan íntegramente en el Complejo Mayor de Histocompatibilidad, en el cromosoma 6 y que presentan un cierto desequilibrio de ligamiento tanto entre ellas como con algunos genes HLA.

Elementos *Alu* autosómicos analizados

Como se ha mencionado anteriormente, en esta Tesis se ha analizado una serie de inserciones *Alu* polimórficas distribuidas entre diferentes cromosomas autosómicos, de tal modo que no presentan desequilibrio de ligamiento entre ellas. Se compone de 12 inserciones: A25, ACE, APO, B65, D1, FXIII B, HS3.23, HS4.32, HS4.59, HS4.65, PV92 y TPA25.

Alu A25

Este elemento fue descrito por Arcot *et al.* (1995a). Se encuentra en el cromosoma 8 (8q21.3), en una región intergénica, a aproximadamente 500 kb en posición centromérica del gen MMP16 (Rajeevan *et al.* 2012). Dicho gen codifica para una proteína de la familia de las proteínas de matriz metaloproteinasas, relacionadas con la degradación de la matriz extracelular en procesos fisiológicos y patológicos (Sato *et al.* 1997).

La longitud de la inserción A25 es de 284 pb. Pertenece a la subfamilia *AluYa5* (Arcot *et al.* 1995a). Se ha identificado en un amplio número de poblaciones con el objeto de evaluar la variabilidad intra e interpoblacional de las poblaciones humanas (Arcot *et al.* 1995a; Stoneking *et al.* 1997; Deininger *et al.* 1999; Comas *et al.* 2000; Nasidze *et al.* 2001; Xiao *et al.* 2002).

Alu ACE

Esta inserción *Alu* se encuentra ubicado en el intrón 16 del gen ACE (del inglés, *Angiotensin-Converting Enzyme*), denominado también DCP1 (del inglés *Dipeptidyl Carboxypeptidase 1*) (Hubert *et al.* 1991). Está localizado en el cromosoma 17, en la región 17q23 (Mattei *et al.* 1989). La proteína ACE juega un papel importante en la regulación de la presión sanguínea a través del sistema renina-angiotensina-aldosterona, RAAS (del inglés *Renin-Angiotensin-Aldosterone System*). Además de este elemento *Alu*, en esta región se han identificado numerosos marcadores polimórficos (Zhu *et al.* 2001).

La inserción *Alu ACE* pertenece a la familia *AluYa5* (Batzer *et al.* 1996a, b), como la mayor parte de las inserciones de este grupo. Su tamaño es de 287 pb (Hubert *et al.* 1991; Tired *et al.* 1992).

Algunos trabajos han sugerido presiones selectivas para este elemento *Alu*, incluyendo su asociación con algunas enfermedades (Tired *et al.* 1994; Hingorani *et al.* 1995; Montgomery *et al.* 1998; Kehoe *et al.* 1999; Park *et al.* 1999; Danser y Schunkert, 2000; Cagliani *et al.* 2010; Uçer *et al.*

2011; Xie *et al.* 2014.). No obstante, son numerosos los estudios poblacionales realizados sobre esta inserción en los que se ha concluido que su comportamiento es, de manera general, selectivamente neutro, por lo que puede ser utilizado para establecer relaciones filogenéticas entre poblaciones (Batzer *et al.* 1996a; Stoneking *et al.* 1997; Novick *et al.* 1998; Rupert *et al.* 1999; Deininger *et al.* 1999; Majumder *et al.* 1999; Comas *et al.* 2000; de Pancorbo *et al.* 2001; Nasidze *et al.* 2001; Xiao *et al.* 2002; Antunez-de-Mayolo *et al.* 2002; Chakrabarti *et al.* 2002; Rupert *et al.* 2003).

Alu APO

Este polimorfismo fue descrito por Matera *et al.* en 1990. Esta secuencia *Alu* está próxima al clúster de genes APOA1, APOC3 y APOA4, localizados en el cromosoma 11 (11q23.3) (Karathanasis, 1985). Se localiza a 4.3kb en posición telomérica del gen APOA1 que codifica para la apolipoproteína A1. Esta proteína es el principal componente proteínico de las lipoproteínas de alta densidad (HDL), relacionadas con el metabolismo del colesterol (Breslow *et al.* 1982). Además, en esta región del genoma se han identificado otros muchos polimorfismos; sin embargo APO ha sido el único *Alu* polimórfico encontrado.

La inserción *Alu APO* es específica de la especie humana (Batzer *et al.* 1994) y pertenece a la subfamilia *Alu Ya5* (Batzer *et al.* 1996b). Tiene una longitud de 312 pares de bases y es una de las inserciones más utilizadas en estudios genético-evolutivos de poblaciones humanas (Batzer *et al.* 1996a; Stoneking *et al.* 1997; Novick *et al.* 1998; Deininger *et al.* 1999; Majumder *et al.* 1999; Comas *et al.* 2000; de Pancorbo *et al.* 2001; Nasidze *et al.* 2001; Watkins *et al.* 2001; Antunez de Mayolo *et al.* 2002; Chakrabarti *et al.* 2002; Xiao *et al.* 2002).

Alu B65

El polimorfismo de inserción *Alu B65* está localizado en una región intergénica del cromosoma 11 (11q14.2), cercana al gen CHORDC1 (del inglés *cysteine and histidine-rich domain-containing zinc binding protein 1*), a 1,76 kb de su exón 11. Este gen está implicado en la regulación de la duplicación del centrómero (Rajeevan *et al.* 2012). Fue descubierta por Arcot *et al.* (1995a).

Esta inserción *Alu* tiene una peculiaridad con respecto al resto de las mencionadas en este apartado. B65 se presenta en dos formas alternativas. En una de las dos variantes la inserción *Alu* incluye un microsatélite en su secuencia y cuenta con 14 Adeninas en la cola poli A. En la otra forma, esta inserción *Alu* carece tanto del microsatélite como de la cola poli A (Arcot *et al.* 1995b). De este modo, la inserción de este elemento *Alu* puede variar entre 342 pb y 313 pb.

El elemento *Alu* B65 pertenece a la subfamilia de elementos *Alu* conocida como Ya5 (Batzer *et al.* 1996b). También ha sido ampliamente analizado en diferentes poblaciones (Arcot *et al.* 1995a; Stoneking *et al.* 1997; Comas *et al.* 2000; Deininger *et al.* 1999; Nasidze *et al.* 2001; Watkins *et al.* 2001; Xiao *et al.* 2002).

Alu D1

El elemento *Alu* D1 está localizado en el cromosoma 3, en la región 3q26.32, aproximadamente a 200 kb del gen WIG1 (Geer *et al.* 2010; Rajeevan *et al.* 2012). Este gen codifica para una proteína cuya expresión es estimulada por el supresor tumoral p53 (Vilborg *et al.* 2011).

Esta inserción fue descrita por Batzer *et al.* (1995), aunque las condiciones de amplificación fueron publicadas por Arcot *et al.* (1995a). Pertenece a la subfamilia Yb8. La inserción consiste en una secuencia de 334 pb. Es interesante destacar que en el extremo final 3' se encuentra el comienzo de un microsatélite dinucleotídico con un motivo AT (Batzer *et al.* 1995). Se ha propuesto que pudo originarse por una transversión A->T en el oligo terminal dA de la inserción, seguida de una recombinación intra-alélica; este mecanismo puede haber sido relativamente frecuente en el origen de los STRs (Levinson y Gutman, 1987; Arcot *et al.* 1995a).

Como en los casos anteriores, esta inserción ha sido analizada en numerosas poblaciones para estudiar sus relaciones filogenéticas (Arcot *et al.* 1995a; Batzer *et al.* 1996a; Stoneking *et al.* 1997; Deininger *et al.* 1999; Majumder *et al.* 1999; Comas *et al.* 2000; Nasidze *et al.* 2001; de Pancorbo *et al.* 2001; Antunez de Mayolo *et al.* 2002; Chakrabarti *et al.* 2002; Xiao *et al.* 2002).

Alu FXIII B

El elemento *Alu FXIII B* está localizado en el cromosoma 1 (1q31-q32). Concretamente se encuentra en el intrón 10 del gen *FXIII B*. Este gen está compuesto por 12 exones y 11 intrones y codifica para una proteína del mismo nombre (Bottenus *et al.* 1990; Grundmann *et al.* 1990). Esta proteína forma parte del factor de coagulación XIII, un heterotetrámero compuesto por dos subunidades A y otras dos subunidades B. Como todos los factores de coagulación, el XIII se encuentra relacionado con la formación del coagulo sanguíneo, concretamente en la estabilización del coagulo de fibrina, siendo el último zimógeno activado en la cascada de coagulación de la sangre (Schwartz *et al.* 1973). Este gen posee más de 200 polimorfismos registrados en la base de datos del *National Center of Biotechnology Information*.

El elemento *Alu*, descubierto por Kass *et al.* en 1994, tiene una extensión de 290 pb y pertenece a la familia *AluYa5* (Batzer *et al.* 1996a).

FXIII B ha sido utilizada en numerosas ocasiones para caracterizar las poblaciones desde un punto de vista genético-evolutivo (Batzer *et al.* 1996; Stoneking *et al.* 1997; Novick *et al.* 1998; Deininger *et al.* 1999; Majumder *et al.* 1999; Comas *et al.* 2000; de Pancorbo *et al.* 2001; Nasidze *et al.* 2001; Antunez de Mayolo *et al.* 2002; Chakrabarti *et al.* 2002; Xiao *et al.* 2002).

Alu HS3.23

Esta inserción *Alu* pertenece a la subfamilia *Ya5*. Se localiza en el cromosoma 7, en una región intergénica próxima a *CNTNAP2* (*contactin-associated protein-like 2 precursor*) (7q35). Presenta en su extremo 3' una cola poli A de 16 nucleótidos de longitud. Está flanqueada a ambos lados por la misma secuencia de 14 nucleótidos de longitud (Arcot *et al.* 1996).

Alu HS4.32

Esta inserción *Alu* pertenece también a la subfamilia *Ya5*. Se localiza en el cromosoma 12, en las proximidades del gen *XRCC6BP1* (*XRCC6-binding protein 1* o *Mitochondrial inner membrane protease*

ATP23 homolog) en la región 12q14. Muestra en la parte terminal 3' una cola rica en A de 15 nucleótidos de longitud. Está flanqueada a ambos lados por la misma secuencia de 17 nucleótidos de longitud (Arcot *et al.* 1996).

Alu HS4.59

También pertenece a la subfamilia Ya5. Se encuentra en el cromosoma 9 (9q31.1). En su extremo terminal 3' porta una cola poli-A de 17 nucleótidos de longitud. Se encuentra flanqueada a ambos lados por la misma secuencia de 15 nucleótidos de longitud. Se localiza entre el gen OR13C9 (olfactory receptor 13C9) y el OR13D1 (olfactory receptor 13D1), ambos relacionados con receptores olfativos (Arcot *et al.* 1996).

Alu HS4.65

Pertenece a la subfamilia Ya5. Se localiza en el cromosoma 9, en la región 9q21.33-34. Muestra en el extremo terminal 3' una larga cola poli-A, de 25 nucleótidos de longitud aunque con presencia de otros nucleótidos. Está flanqueada en ambos extremos por la misma secuencia de 16 nucleótidos de longitud. Esta inserción se encuentra próxima al gen DAPK1 (death-associated protein kinase 1) (Arcot *et al.* 1996).

Alu PV92

Descrito por primera vez por Matera *et al.* (1990), este elemento *Alu* se encuentra en el cromosoma 16, en un intrón del gen que codifica para la Cadherina 13 (o Cadherina H, del inglés *Heart*, por su relación con el corazón) (Geer *et al.* 2010; Rajeevan *et al.* 2012). La Cadherina 13 es una proteína citoplasmática pero carece de un dominio citoplasmático, por lo que, a pesar de pertenecer a la superfamilia de las Cadherinas, no es una glicoproteína de adhesión célula-célula, calcio dependiente. Realiza una función de protección del endotelio vascular frente a la apoptosis causada por estrés oxidativo. También está asociada a la resistencia a la arterioesclerosis.

En esta región, existe una elevada presencia de marcadores polimórficos, además de esta inserción *Alu*.

La inserción *Alu* PV92 es específica de la especie humana (Batzer *et al.* 1994). Pertenece a la más joven de las subfamilias de elementos, Ya5, que se define a partir de 5 cambios en la secuencia consenso de la subfamilia *AluY* (Batzer *et al.* 1996b). Tiene una longitud de 314 pb y ha sido utilizada frecuentemente para estudios genético-evolutivos de poblaciones humanas (Batzer *et al.* 1996a; Stoneking *et al.* 1997; Novick *et al.* 1998; Deininger *et al.* 1999; Majumder *et al.* 1999; Comas *et al.* 2000; de Pancorbo *et al.* 2001; Nasidze *et al.* 2001; Watkins *et al.* 2001; Antunez de Mayolo *et al.* 2002; Chakrabarti *et al.* 2002; Xiao *et al.* 2002).

Alu TPA25

El elemento *Alu* TPA25 se encuentra en el intrón 8 del gen PLAT (del inglés *Plasminogen Activator Tissue*), que codifica para el activador del plasminógeno de tipo tisular (Tishkoff *et al.* 2000), una serina proteasa que transforma a la protoenzima plasminogeno inactiva en una proteinasa activa que está asociada a procesos como la fibrinólisis y la remodelación tisular (Benham *et al.* 1984). Este gen se extiende en el brazo corto del cromosoma 8 (8p11.2) (Degen *et al.* 1986; Yang-Feng *et al.* 1986). La región 5' que flanquea el gen y el propio gen contienen un total de 28 inserciones *Alu* (Degen *et al.* 1986).

La inserción *Alu* TPA25 consiste en una secuencia de 311 pb. Pertenece a la subfamilia Ya8 (Batzer *et al.* 1996a, b).

Al gen PLAT se le ha asociado con el infarto de miocardio, a través de la trombosis coronaria. Sin embargo, se ha demostrado que dicha relación no muestra correlación con la presencia/ausencia de la inserción *Alu* (Kluft, 1994; van den Eijnden-Schrauwen *et al.* 1995; Ridker *et al.* 1997; van der Bom *et al.* 1997; Ahmed *et al.* 2011). Por lo tanto, se considera a TPA25 un marcador neutro frente a la selección natural. Bajo esta premisa se ha empleado en numerosas ocasiones para el análisis genético-evolutivos de poblaciones humanas (Perna *et al.* 1992; Batzer *et*

al. 1996a; Tishkoff *et al.* 1996; Stoneking *et al.* 1997; Novick *et al.* 1998; Deininger *et al.* 1999; Majumder *et al.* 1999; Comas *et al.* 2000; de Pancorbo *et al.* 2001; Nasidze *et al.* 2001; Watkins *et al.* 2001; Antunez de Mayolo *et al.* 2002).

EL COMPLEJO MAYOR DE HISTOCOMPATIBILIDAD (MHC)

Además de las inserciones *Alu* mencionadas en el apartado anterior, se han analizado 5 polimorfismos de inserción *Alu* y 4 marcadores microsatélites situados en el MHC (*Figura 19*). Previamente se describirá de manera breve esta región, por sus peculiares características, especialmente en lo que se refiere a su elevada concentración de genes, muchos de los cuales se encargan de la respuesta inmune del organismo.

El MHC se encuentra en el brazo corto del cromosoma 6, en la región 6p21.31 y cubre una extensión de aproximadamente 3,6 Megabases (Trowsdale, 1995; The MHC sequencing consortium, 1999). Representa en torno al 0.13% del genoma humano completo y sin embargo contiene el 0.5% de genes codificantes y candidatos conocidos (Shiina *et al.* 2009). Es una de las regiones del genoma más estudiadas, lo que incluye trabajos sobre SNPs, variación haplotípica, expresión génica, diversidad intra e interespecífica, evolución de su organización estructural, etc., principalmente desde que se completó el mapa de su secuencia (The MHC Sequencing Consortium, 1999).

El gran interés que ha provocado tradicionalmente deriva de su función en la regulación de la inflamación, la cascada del complemento y en la respuesta inmune tanto innata como adquirida mediante los sistemas NK y de células T (Shiina *et al.* 2009).

Los genes del MHC contribuyen a limitar la interacción celular y la histocompatibilidad de los tejidos mediante la discriminación celular entre lo propio y lo ajeno, afectando en consecuencia a los trasplantes de tejidos y órganos (Claas y Duquesnoy, 2008) y las terapias de transfusión (Choo, 2007). Asimismo, el MHC juega un papel muy importante en los procesos de autoinmunidad

(Fernando *et al.* 2008) y en la respuesta del organismo a los agentes infecciosos (Martin y Carrington, 2005; Fellay *et al.* 2007).

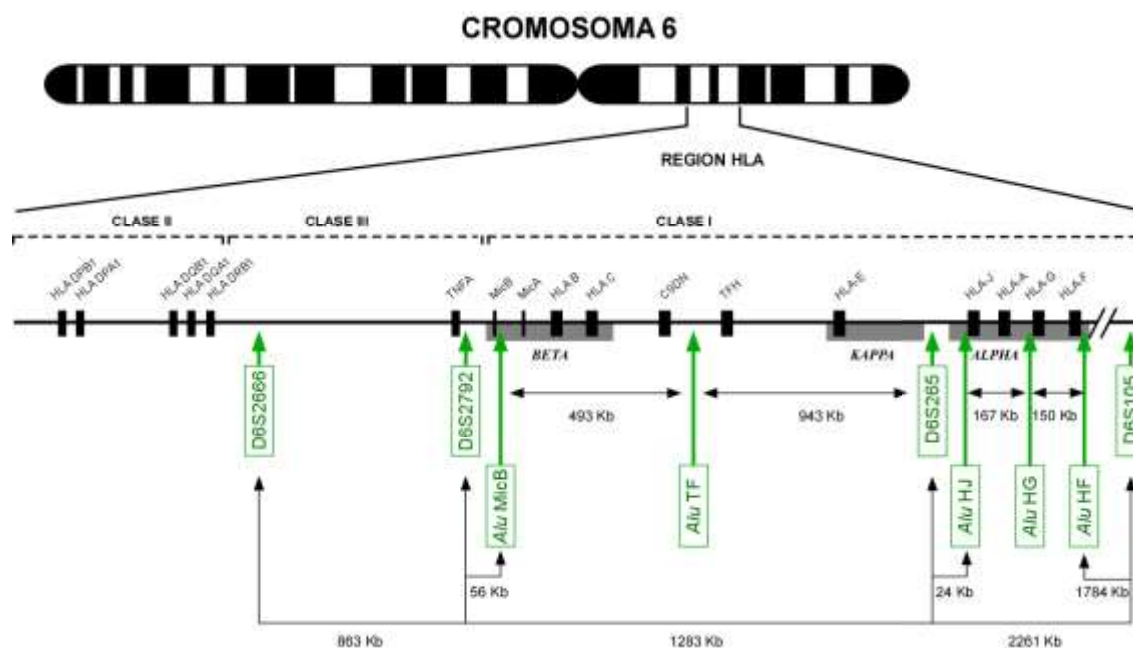


Figura 19. Esquema del cromosoma 6, donde se identifica la localización de los marcadores analizados pertenecientes a la región MHC. Además se muestran las distancias relativas entre los diferentes marcadores, en Kb. La mayoría de los marcadores se localiza en la Clase I de la región HLA, destacando los marcadores situados en el Bloque Alfa (D6S265, *Alu*HJ, *Alu*HG y *Alu*HF) por la proximidad que presentan entre ellos.

Además, la región genómica del MHC también ha mostrado una cierta influencia en el desarrollo y la plasticidad del sistema nervioso central (Xiao y Link, 1998; Huh *et al.* 2000; Boulanger y Shatz, 2004; Cullheim y Thams, 2007; Ohtsuka *et al.* 2008), la interacción entre células neurológicas (Matsuo *et al.* 2001; Patiño-López *et al.* 2006), la función sináptica y el comportamiento (Goddard *et al.* 2007; Tonelli *et al.* 2005), la especialización hemisférica cerebral (Legen *et al.* 2009) y las enfermedades neurológicas y psiquiátricas (O'Keefe *et al.* 2002; Raha-Chowdhury *et al.* 2005; Cohly y Panja, 2005; Bailey *et al.* 2006; McElroy y Oksenberg, 2008). Lógicamente, el hecho de que participen en funciones tan variadas se debe a que los genes del MHC no funcionan de forma aislada, sino que interactúan con otros genes tanto dentro del propio MHC como fuera de él (Shiina *et al.* 2009).

La gran variabilidad del MHC ha sido utilizada también a menudo en estudios de Antropogenética, de manera que se han empleado con asiduidad las frecuencias génicas y haplotípicas HLA para estudiar la estructura genética de las poblaciones humanas (Cavalli-Sforza *et al.* 1994; Sánchez-Mazas y Meyer, 2014).

En este sentido, la identificación de los alelos HLA es compleja y sus datos difíciles de tratar por la gran variabilidad existente, incluyendo diferentes niveles de resolución en los análisis. Por el contrario, el genotipado de otros polimorfismos localizados en el MHC (STRs, Inserciones *Alu*, SNPs,...) es frecuentemente mucho más sencillo y menos costoso y podrían ser útiles para el análisis de la diversidad poblacional de esta región del genoma.

Inserciones MHC-*Alu* analizadas

Las MHC-*Alu*, son un grupo de inserciones *Alu* localizadas en el Complejo Mayor de Histocompatibilidad, con una cierta proximidad entre ellas, por lo que son susceptibles de mostrar desequilibrio de ligamiento.

A pesar del interés de los elementos *Alu* como herramientas para describir los procesos evolutivos en las poblaciones humanas, el estudio de los elementos situados en el MHC es relativamente reciente (Kulski y Dunn, 2005; Dunn *et al.* 2005, 2007; Tian *et al.* 2008; Yao *et al.* 2009a, b, 2010).

Otro de los usos más interesantes que se les ha dado es el de contribuir a establecer un reloj molecular para dilucidar la organización genómica y la edad evolutiva de los genes duplicados dentro de las familias HLA clase I (Kulski *et al.* 1997; 1999; 2000), HLA clase II (Mnukova-Fajdelova *et al.* 1994) y MIC (Gaudieri *et al.* 1997), entre otras (Kulski *et al.* 2001).

Por otra parte, si se considera un grupo de elementos suficientemente próximos en la misma región, a las características descritas para las inserciones *Alu* en general, se le añade la posibilidad de utilizar los correspondientes haplotipos, tanto para la detección de procesos de flujo génico

como para detectar la existencia de posibles procesos de autostop genético y/o barrido selectivo (*Genetic Hitchhiking*).

En esta memoria se aborda el análisis de cinco elementos *Alu* situados a lo largo de una secuencia de 1,8 Mb de la región HLA clase I. Entre ellas, tres se encuentran en el denominado bloque Alfa y las dos restantes en el bloque Beta (*Figura 19*) (Kulski *et al.* 2001; Kulski *et al.* 2002; Dunn *et al.* 2002; Dunn *et al.* 2003). Hasta ahora, estos polimorfismos han sido estudiados en un número todavía pequeño de poblaciones humanas (Kulski y Dunn, 2005; Dunn *et al.* 2005, 2007; Tian *et al.* 2008; Yao *et al.* 2009a, b, 2010). Las características referentes al análisis de los polimorfismos seleccionados se detallaran en la sección de Material y Métodos.

Alu HF

Este elemento, junto con *Alu HJ* y *HG* conforma el bloque alfa, que ha sido utilizado en algunos trabajos para análisis antropogenéticos. Se encuentra próximo a una región rica en ATs, dentro de una secuencia HERV-16 (de *Human endogenous retrovirus*), localizada a 7,5 kb del gen HLA-F, del que toma su nombre (Kulski y Dunn, 2005; Dunn *et al.* 2005, 2007; Tian *et al.* 2008; Yao *et al.* 2009a, b, 2010). Su tamaño es de 145 pb.

Alu HG

Este elemento pertenece a la familia Yb8. Se encuentra entre los genes HLA-A y HLA-G, aproximadamente a 85 kb del gen HLA-A, en dirección 3' a 5', entre los restos de un fragmento del gen PERB11 (Mic) y del mismo modo que HF, próximo a una región rica en ATs (*Figura 19*) (Kulski *et al.* 2001).

Su nombre deriva del gen HLA-G, por su proximidad. Se cree que este gen puede jugar un papel importante en la tolerancia inmune durante el embarazo, de forma que se expresa cuando los genes HLA-A y HLA-B aún no se han expresado (Creasy *et al.* 2004). El tamaño de esta inserción es de 322 pb.

Alu HJ

Este elemento se encuentra situado aproximadamente a 18kb del pseudogen HJ, del cual toma su nombre. *Alu HJ* está insertado en una secuencia Mer65, rodeado por otros retroelementos (L1, LTR, Mer 65, entre otros) y del mismo modo que HF y HG, cercano a una secuencia rica en AT (Dunn *et al.* 2002). Su tamaño es de 332 pb.

Alu MicB

Este elemento se localiza en el primer intrón del gen MicB, de donde toma su nombre (figura 19). Posiblemente su inserción en esta región ocurrió hace unos 19 millones de años (Kapitonov y Jurka, 1996), bastante después de que ocurriera la duplicación del gen ancestral MicA/MicB, hace entre 37 y 44 millones de años (Gaudieri *et al.* 1997). Fue descrito por Kulski *et al.* en 2002. La secuencia ancestral de esta región, sin la inserción MicB incluye un elemento *Alu* perteneciente a la familia Y de 277 pb que se ha fijado en el transcurso de la evolución (Kapitonov y Jurka, 1996; Kulski *et al.* 2002). El alelo que presenta la inserción *Alu* moderna y polimórfica surgió, bien por duplicación parcial del elemento antiguo o por la inserción de un nuevo elemento en la misma región. El tamaño de MicB es de 162 pb. (Kulski *et al.* 2002). Perteneció a la familia Yb8 y por ello presenta una duplicación cuyo motivo es CAGTCCG (Kapitonov y Jurka, 1996). Actualmente se considera que es un marcador útil para el análisis de las relaciones filogenéticas humanas (Kulski *et al.* 2002).

Alu TF

El elemento *Alu TF* fue identificado y caracterizado por Dunn *et al.* en 2003. Se localiza entre los bloques beta y kappa de la región clase I del MHC, entre los genes CDSN y TFIIH. El nombre deriva del hecho de encontrarse próximo al gen TFIIH (factor de transcripción II H).

El elemento *Alu TF* se encuentra insertado dentro de una secuencia *AluSx* y cercano a un resto de un retroelemento L1. Perteneció a la familia Yb8 con cuya secuencia consenso tiene un 99,6% de similitud (Dunn *et al.* 2003). *Alu TF* tiene una longitud de 285pb.

MHC-STRs analizados

Como se ha mencionado anteriormente, en esta Tesis se han analizado también 4 microsatélites localizados en la región del Complejo Mayor de Histocompatibilidad.

Los microsatélites o STRs (*Short tandem repeat*) son fragmentos de ADN cuya estructura está basada en una secuencia nucleotídica o motivo que se repite en tándem un número determinado de veces (*Figura 20*). La unidad de repetición de los STRs se comprende entre 2 y 6 pb según Hearne *et al.* (1992), aunque algunos autores proponen un valor máximo de 7pb (Edwards *et al.*, 1991) y otros de 5 (Tamaki y Jeffreys, 2005). En todo caso, la mayor parte de los microsatélites son dinucleotídicos, trinucleotídicos o tetranucleotídicos (es decir 2, 3 o 4 pb en el motivo de repetición). El número de veces que dicho motivo se repite en tándem es variable y de ello depende la identificación de los diferentes alelos (Tamaki y Jeffreys, 2005). Basicamente su tamaño oscila entre 80 y 550 pb (Litt y Luty, 1989, Weber y May, 1989).

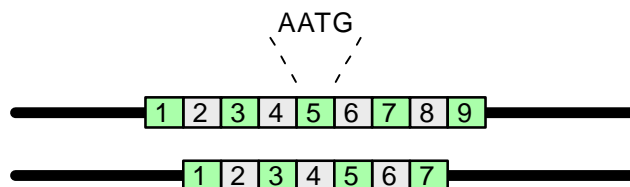


Figura 20. Esquema dos cromosomas donde se muestran la estructura de un microsatélite tetranucleotídico (motivo de repetición de 4 nucleótidos). En este caso se trata de un individuo heterocigoto, ya que los cromosomas presentan un número diferente de repeticiones.

Los microsatélites se encuentran distribuidos ampliamente por todo el genoma de forma aleatoria (Litt y Luty, 1989; Weber y May, 1989) y poseen un alto grado de variabilidad (Beckman y Weber, 1992; Hearne *et al.* 1992; Tautz y Schotterer, 1994; Foissac *et al.* 1997) lo cual los hace idóneos para el análisis en Genética Forense y Antropogenética. Una característica relacionada con su gran variabilidad es su elevada tasa de mutación, que oscila entre $2,3 \times 10^{-3}$ y 4.6×10^{-5} por locus y por generación (Zhivotovsky *et al.* 2000, 2003, 2004; Mardini *et al.* 2013; Sun *et al.* 2014.). El rango

de variación se atribuye a las diferencias existentes en las secuencias de los STRs, particularmente el tamaño de su unidad de repetición, pero también la complejidad de la estructura del STR y su localización en el genoma, entre otras causas (Brinkmann *et al.* 1998).

A pesar de que el motivo suele repetirse de manera casi perfecta, en ocasiones puede sufrir alteraciones, principalmente por inserciones o deleciones en la pauta de repetición. Por ello, las clasificaciones de los STRs a menudo se basan en la mayor o menor complejidad de su estructura. Así, Urquhart *et al.* (1994) realizó una clasificación en STRs simples, compuestos y complejos. Posteriormente, Brinkmann (1996) propuso una clasificación alternativa, aunque prácticamente equivalente: STRs con baja, intermedia y alta microvariabilidad.

- I. *STRs simples o de baja microvariación:* Están formados por sucesivas unidades de repetición contiguas e idénticas, tanto en longitud como en secuencia. Por lo tanto, los diferentes alelos se establecen en función de la variación en el número de repeticiones del motivo. La diferencia en tamaño entre los distintos alelos será siempre consecuencia de un diferente número de unidades de repetición. En este tipo de STR, la presencia de alelos intermedios es excepcional. Se consideran a estos STRs como unos marcadores de fácil tipaje. Tienen el inconveniente de una baja heterocigosidad, en relación a otros tipos de STRs. Se pueden clasificar en función de la longitud del motivo (dinucleotídicos, trinucleotídicos,...).
- II. *STRs compuestos o de microvariación intermedia:* Estos sistemas comprenden dos o más unidades de repetición contiguas con diferencias entre ellas. Esta variación puede deberse tanto a la secuencia como a la longitud de la propia unidad de repetición. Ello eleva el número de alelos y la heterocigosidad de estos marcadores.
- III. *STRs complejos o de alta microvariación:* Pueden contener varios bloques de unidades de repetición de longitud variable, con secuencias intermedias más o menos variables. Son frecuentes los alelos intermedios y las sustituciones simples de bases en algunas unidades, lo que ocasiona importantes variaciones tanto estructurales como en tamaño,

produciéndose problemas de designación de alelos. Son los STRs más difíciles de identificar y presentan una tasa de mutación más alta.

Con el tiempo se han acumulado evidencias acerca de que incluso los microsatélites más simples presentan algún grado de variabilidad interna en su estructura, por lo que la clasificación anterior es cuestionable. Existen STRs relativamente simples y extremadamente variables al mismo tiempo, como D12S391 y D1S1656 (Lareu *et al.* 1996a, b, 1998) en los que la asociación directa entre variabilidad y complejidad no es sostenible.

Los STRs constituyen elementos extraordinariamente útiles e interesantes para su uso en Antropogenética debido a una serie de características (Litt y Luty, 1989; Weber y May, 1989): *i*) presentan un elevado grado de polimorfismo y una gran variabilidad entre individuos y entre poblaciones; *ii*) tienen una herencia mendeliana simple; *iii*) son codominantes, con lo que se puede diferenciar los individuos homocigotos de los heterocigotos; *iv*) son relativamente fáciles de identificar de forma fiable y su análisis estadístico es sencillo; *v*) su pequeño tamaño facilita su amplificación y su resistencia a procesos degradativos; *vi*) comparados con otros marcadores de características similares (como los minisatélites) su tasa de mutación es relativamente baja, por lo que desde un punto de vista microevolutivo son más estables.

Los microsatélites generalmente se encuentran en regiones no codificantes del genoma y distribuidos de manera uniforme (Goldstein y Schlotterer, 1999; Aranguren-Méndez *et al.* 2005). En ocasiones, no obstante, pueden encontrarse en regiones codificantes. De hecho, se ha descrito su asociación con algunas enfermedades (Armour *et al.* 1994; Hancock, 1999; Tautz y Schlotterer, 1994), aunque su número es escaso.

En la región del Complejo Mayor de Histocompatibilidad se han descrito al menos 1.527 microsatélites, de los cuales 268 han sido utilizados como marcadores genéticos (Shiina *et al.* 2009). Incluso se ha establecido una nomenclatura específica para ellos (Gourraud *et al.* 2007). Considerando un grupo suficientemente próximo, con desequilibrio de ligamiento, será posible su

estudio en la forma de haplotipos. Incluso se pueden emplear en combinación con otros marcadores más estables, como las inserciones *Alu*. Si además se tiene en cuenta la presencia de numerosos genes, presumiblemente sometidos a presiones selectivas, el conjunto puede convertirse en una interesante herramienta de análisis genético.

En este trabajo se han analizado 4 MHC-STRs, que se describen brevemente a continuación (*Figura 19*). Las condiciones PCR y la secuencia de los cebadores se detalla en la sección Anexo.

D6S2666

Este microsatélite se encuentra situado entre los genes *BTNL2* y *HLA-DRA*, ubicados en la región clase III (*Figura 19*). Su unidad de repetición está formada por una secuencia dimérica CA_n (TG_n). Ha sido utilizado en estudios de asociación con patologías como el lupus eritematoso (Graham *et al.* 2007) y la artritis reumatoide (Lee *et al.* 2006).

D6S2792

Tradicionalmente denominado *TNF α* , su nombre se modificó siguiendo las recomendaciones internacionales de nomenclatura para los microsatélites del MHC de Gourraud *et al.* (2007). Su nombre original se debe a su localización próxima al gen *TNF α* , que se encuentra en la región clase III del MHC (*Figura 19*). La región donde se encuentra este gen se conoce como *TNF* (*Tumor necrosis factor*) y comprende tres genes, cada uno con 4 exones. El gen *TNF α* codifica para moléculas implicadas en reacciones inmunológicas e inflamatorias relacionadas con la proliferación celular, apoptosis, metabolismo lipídico, etc. (Hajeer y Huntchinson, 2000; Locksley *et al.* 2001).

Este microsatélite presenta una secuencia repetitiva dimérica $(AC/GT)_n$ (Carrol *et al.* 1987; Ragoussis *et al.* 1988).

Dada la proximidad física del microsatélite D6S2792 al gen *TNF α* , se han encontrado algunas asociaciones significativas entre ambos marcadores (McManus *et al.* 1996; Hajeer *et al.* 2000; Takeuchi *et al.* 2000; Castro *et al.* 2001; Biral *et al.* 2006; Törn *et al.* 2006). Es uno de los marcadores

genéticos más utilizados en la definición de haplotipos ancestrales del MHC (Dawkins *et al.* 1991; Alper *et al.* 1992; Abraham *et al.* 1993) y se ha caracterizado en algunos estudios poblacionales (Crouau-Roy *et al.* 1993; Charron, 1997; Peña *et al.* 2002; Dejean *et al.* 2004; Fernández-Mestre *et al.* 2006; Grubić *et al.* 2007; Sens-Abuázar *et al.* 2009).

D6S265

Este microsatélite se encuentra en la región Clase I del MHC, aproximadamente a 100kb del locus HLA-A (*Figura 19*). Su unidad de repetición está formada por secuencias diméricas AC_n (GT_n) y presenta algunas asociaciones con genes HLA (Worwood *et al.* 1994a, b; Pyper *et al.* 1995; Jazwinska *et al.* 1995). Algunas variantes alélicas de este STR se encuentran asociadas con alelos del locus HLA-A. Es el caso de la asociación entre el alelo *D6S265*1* con el alelo *HLA*A3* (Worwood *et al.* 1994a) estando este último asociado a su vez con la hemocromatosis (Simon *et al.* 1976). También se han descrito asociaciones con la mononucleosis infecciosa (McAulay *et al.* 2007), la artritis reumatoide (Barnetche *et al.* 2006) y la esclerosis múltiple (Palacio *et al.* 2002).

Ha sido empleado en algunos estudios de estructura genética de poblaciones humanas (Charron, 1997; Peña *et al.* 2002; Dejean *et al.* 2004; Fernández-Mestre *et al.* 2006).

D6S105

Este microsatélite es un marcador polimórfico telomérico situado dentro de la clase I de la región MHC. Se encuentra a aproximadamente 3cM del locus HLA-A (*Figura 19*). Está formado por secuencias repetitivas diméricas AC_n (GT_n) (Trent y Ziegler, 1993; Stone *et al.* 1994; Volz *et al.* 1994). Se han encontrado algunas asociaciones con genes HLA (Worwood *et al.* 1994a, b) y también con la hemocromatosis (Jazwinska *et al.* 1993). Se ha estudiado en un reducido número de poblaciones (Charron, 1997; Peña *et al.* 2002; Fernández-Mestre *et al.* 2006).

HIPÓTESIS DE TRABAJO

La hipótesis de trabajo de la que se ha partido para realizar esta Tesis puede describirse del siguiente modo:

Las diferencias en la intensidad de la acción de factores ambientales y socioculturales pueden afectar de forma muy variable al patrimonio genético de las poblaciones a una escala microevolutiva.

Son numerosos los factores ambientales (barreras orográficas, bosques tropicales, desiertos,...) y socioculturales (tradiciones culturales, barreras lingüísticas, conflictos históricos,...) que pueden afectar a los patrones de cruzamiento de una población, determinando la aparición de subpoblaciones aisladas en mayor o menor medida. Ello provocará variaciones en la intensidad de la acción de las principales fuerzas microevolutivas, la deriva y el flujo génico, sobre su patrimonio genético. Si estas presiones se mantienen en el tiempo, pueden desembocar en procesos de microdiferenciación entre las subpoblaciones y en definitiva, de microevolución.

Con el fin de analizar estos procesos, se han estudiado varias poblaciones sudamericanas en las que se han dado unas condiciones características y al mismo tiempo muy diferentes entre sí, incluyendo la confluencia de poblaciones procedentes de tres continentes que estuvieron aisladas reproductivamente durante largo tiempo, una gran variedad de biomas, una heterogeneidad en su orografía muy importante y notables diferencias culturales.

OBJETIVOS DE LA TESIS DOCTORAL

El objetivo general que se ha contemplado en esta Tesis ha sido *la valoración de los efectos del flujo génico y deriva genética sobre el patrimonio genético en poblaciones del continente americano con diferentes características, mediante el análisis de marcadores genéticos autosómicos.*

Como se ha mencionado previamente, el continente americano ha experimentado dos eventos fundamentales de poblamiento: un proceso inicial de colonización, con un efecto fundador a gran escala, y una posterior inmigración masiva de colonos europeos y esclavos africanos a partir del descubrimiento de América, que generó toda una serie de procesos de mestizaje de intensidad variable entre europeos, africanos y amerindios. A todo ello cabe añadir, entre otros procesos de microdiferenciación, los cuellos de botella experimentados por las poblaciones de amerindios originados por la disminución de sus poblaciones tras la llegada de los europeos.

Por ello, el continente americano reúne las condiciones idóneas para el estudio de los procesos de flujo génico y de deriva genética permitiendo observar en sus poblaciones procesos de microdiferenciación de diferente intensidad.

Cada una de las poblaciones seleccionadas para esta Tesis tiene una serie de características particulares que la diferencia del resto, permitiendo observar grados variables de aislamiento y de mestizaje, incluso en territorios de dimensiones reducidas.

En consecuencia, los objetivos específicos que se han propuesto son los que se describen a continuación.

- I. Debido a que la mayor parte de los análisis genéticos descritos en esta memoria se iban realizados con inserciones *Alu*, un objetivo previo fue *la valoración de la utilidad de las inserciones Alu como marcadores discriminantes entre poblaciones y grupos continentales, es decir, como AIMs, particularmente para el estudio de los procesos de mestizaje.*

Este objetivo se muestra particularmente en el Manuscrito I (*Utilidad de las inserciones Alu en los estudios de mestizaje*), pero también en el Manuscrito II (*Genetic admixture*

estimates by Alu elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia) de la sección Resultados. Asimismo, se ha valorado la utilidad de un grupo ligado de inserciones *Alu* del Complejo Mayor de Histocompatibilidad, en el Manuscrito V (*Young Alu insertions within the MHC class I region in Native American populations: insights into the origin of the MHC-Alu repeats*).

- II. Otro objetivo ha consistido en *examinar las consecuencias de la deriva genética sobre la diversidad de poblaciones sometidas a unas condiciones de prolongado aislamiento y disminución drástica del tamaño efectivo poblacional*. Con este fin se ha seleccionado una de las poblaciones más aislada de Sudamérica, la etnia Huaorani, para un análisis que se muestra en el Manuscrito IV de la sección Resultados (*The Waorani tribe from the Ecuadorian Amazon according to Alu polymorphisms, Young Alu insertions within the MHC class I region in Native American populations: insights into the origin of the MHC-Alu repeats*) Asimismo, la población residente en las comarcas más altas de la provincia de Jujuy ha resultado muy ilustrativa para el desarrollo de este objetivo, como se muestra en el Manuscrito VI (*Jujuy province (NW Argentina): STR loci unveil genetic footprints of microevolutionary responses to a steep orography*).
- III. Además, *se han evaluado los efectos de una barrera geográfica de gran tamaño sobre el flujo génico y la estructura genética de una población*. Para ello se ha seleccionado la población de Jujuy, ubicada en localidades con diferencias de hasta 3.000 m de altitud en las estribaciones de los Andes. Los resultados se muestran en los Manuscritos III (*Microevolutionary Processes due to Landscape Features in the Province of Jujuy, Argentina*) y VI (*Jujuy province (NW Argentina): STR loci unveil genetic footprints of microevolutionary responses to a steep orography*).

- IV. *Se han valorado las proporciones diferenciales de mestizaje entre grupos geográficamente próximos, sometidos en grado variable a una barrera a la panmixia, tanto cultural como orográfica. Este objetivo ha sido desarrollado en los Manuscritos II (Genetic admixture estimates by Alu elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia) y III (Microevolutionary Processes due to Landscape Features in the Province of Jujuy, Argentina).*

MATERIALES Y MÉTODOS

En esta sección se pretende ampliar la información acerca de los materiales y los métodos utilizados a lo largo del desarrollo de esta Tesis, los cuales por razones de espacio no pudieron ser explicados con suficiente detalle en las diferentes publicaciones que la componen.

MATERIALES

MUESTRAS POBLACIONALES

Mestizos de Colombia

La muestra de los mestizos colombianos se tomó en la ciudad de Medellín, en el departamento de Antioquia, Colombia (*Figura 21*). Estaba compuesta por muestras de ADN de 34 individuos seleccionados de acuerdo a su fenotipo morfológico y su auto-reconocimiento, siguiendo los criterios utilizados en el censo de 2005 (DANE, 2005). El criterio requerido para considerar a un individuo como originario de Antioquia fue que su lugar de nacimiento y el de todos sus ascendientes hasta la tercera generación se encontraran en dicho departamento. Los individuos eran voluntarios informados y sanos.

Afrodescendientes de Colombia

La muestra de afrocolombianos procedía de diferentes núcleos poblacionales del departamento de Antioquia, concretamente de la región limítrofe con la provincia de Chocó (*Figura 21*). Consistía en muestras de ADN de 64 individuos de ascendencia africana, voluntarios informados y sanos.

Del mismo modo que con la muestra de mestizos, la clasificación fue realizada en función de su fenotipo morfológico y su auto-reconocimiento, siguiendo los criterios utilizados en el censo de 2005 (DANE, 2005). Asimismo, el criterio requerido para considerar a los individuos como originarios del departamento tuvo en cuenta su lugar de nacimiento y el de sus ascendientes hasta la tercera generación.

Tanto las muestras de mestizos como las de afrocolombianos fueron tomadas por un equipo dirigido por la Dra. Bravo (GENES, Medellín, Colombia).

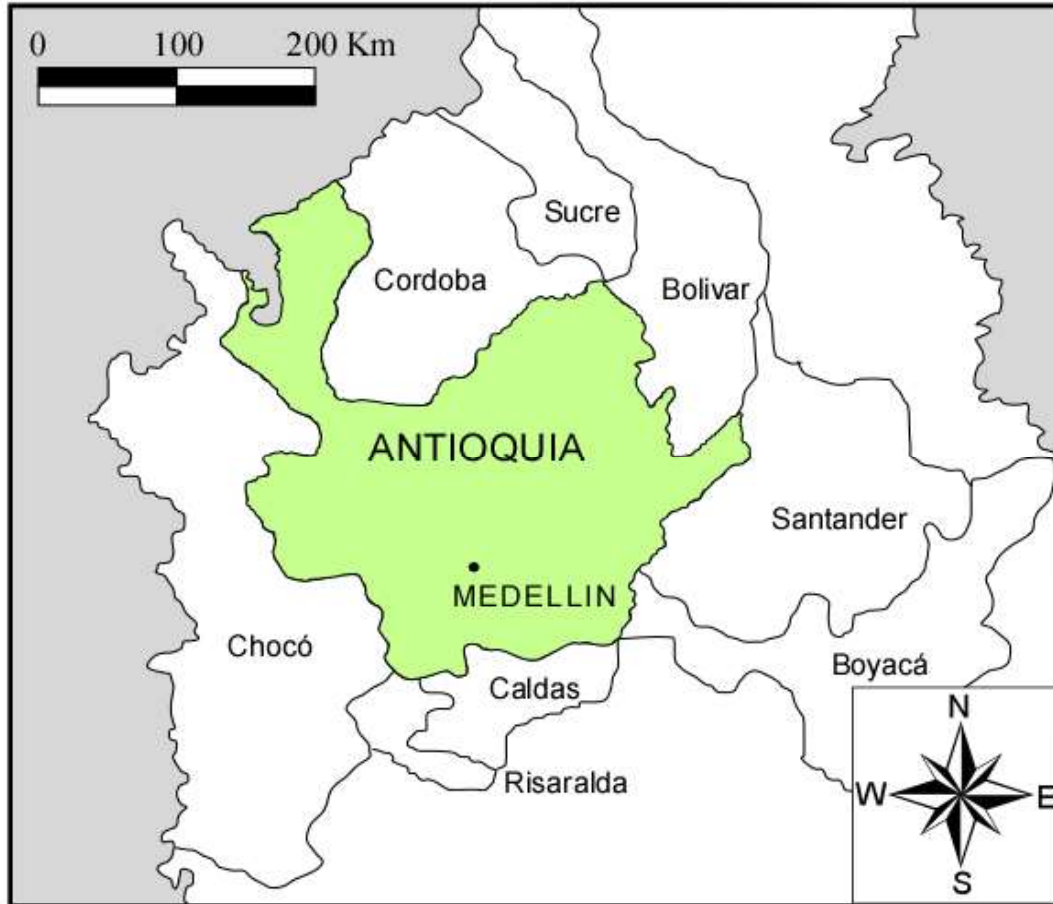


Figura 21. Ubicación de la provincia de Antioquia y su capital Medellín dentro de Colombia. La muestra de Mestizos procede de Medellín. La muestra de Afrocolombianos procedía de diferentes núcleos poblacionales del departamento de Antioquia.

Jujeños de Argentina

La muestra argentina consistió en manchas de sangre sobre papel (Whatman y de filtro) de 226 individuos procedentes de diferentes núcleos poblacionales de la provincia de Jujuy, en el Noroeste de Argentina (Figura 22). Todos los individuos eran voluntarios informados y sanos.

La selección de los individuos se realizó de acuerdo al lugar de nacimiento del donante y de sus ascendientes hasta la tercera generación.

La muestra se subdividió por comarcas de la siguiente forma:

- I 47 individuos procedentes de la Puna, residentes en las localidades de Mina El Aguilar, Mina Pirquitas, Cochinoca, Abra Pampa, Paicone, Pucará de Rinconada, Rinconada, Yavi, Chocoite y Pumahuasi (Yavi), Mina Pan de Azúcar, La Quiaca, Casira, Susques, Coranzulí, La altitud media de estas localidades es de 3700 msnm. Como curiosidad, cabe mencionar que El Aguilar es la 3ª localidad más alta del mundo, ubicada a 4.895 msnm.
- II 72 individuos nacidos en la Quebrada. Debido al amplio rango de altitudes de la región y de las localidades implicadas, esta muestra fue separada en dos grupos. Esta división no se corresponde administrativamente con ninguna región:
 - o 36 individuos de Quebrada Alta, cuyas localidades de muestreo fueron San Miguel de los Colorados, Tilcara, El Moreno y Humahuaca, situadas a una altitud promedio de 3300 msnm.
 - o 36 individuos de Quebrada Baja, cuya localidad de muestreo fue Maimara, localizada a una altitud de 2500 msnm.
- III 62 individuos procedentes de la región de Valle. Las muestras fueron recogidas de las localidades de El Carmen, Palpalá y San Salvador de Jujuy, con una altitud promedio de 1100 msnm.
- IV 45 individuos de la región de Selva, con una altitud media de 700 msnm. Las muestras proceden de las ciudades de San Lucas, San Pedro de Jujuy, Libertador General San Martín, La Esperanza, La Mendieta, El Talar y Fraile Pintado.

Las muestras de Jujuy fueron tomadas por un equipo dirigido por el Dr. Dipierri (Universidad Nacional de Jujuy, Argentina).



Figura 22. La provincia de Jujuy, fue dividida en 5 regiones para su análisis. Se especifica la posición de las localidades de muestreo. La poblaciones de muestreo se encuentran detallados en la Tabla Anexo 3.

Alto Jujuy

En algunos artículos se ha considerado, con un criterio operativo, una parte de las muestras de Jujuy. Debido a las particulares características de los departamentos situados a mayor altitud, se ha definido el Alto Jujuy como una agrupación de las comarcas de La Puna y Quebrada, incluyendo esta región por tanto las muestras correspondientes a Puna, Quebrada Alta y Quebrada Baja. El término utilizado en Inglés ha sido *Jujuy highlands*.

Etnia Huaorani de Ecuador

La muestra de Huaorani se obtuvo de 36 individuos de esta etnia procedentes de Ecuador y residentes en la segmento de la Reserva Étnica Huaorani que se localiza dentro de los límites administrativos de la provincia de Pastaza (Figura 23). Se ha estimado que estos individuos pertenecen al menos a 17 clanes familiares diferentes o nanicabos, como son denominados los grupos domésticos tradicionales. Esta distinción se ha podido realizar a partir de la composición de apellidos de la muestra. La selección de individuos de diferentes nanicabos pretendió minimizar los efectos de la consanguinidad más reciente, ya que *a priori* se esperaba un alto grado de endogamia.

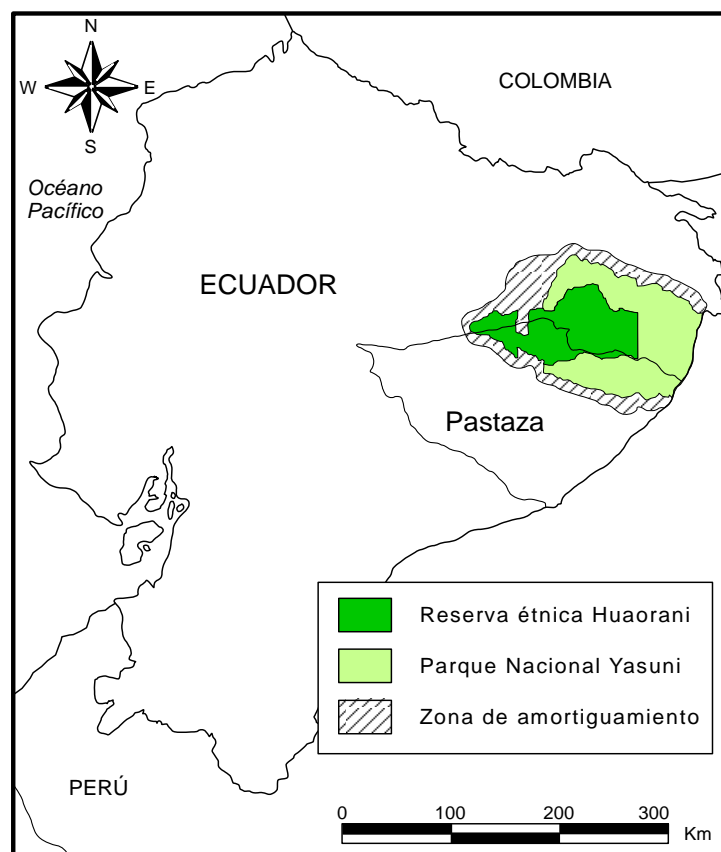


Figura 23. Mapa donde se muestra la provincia de Pastaza junto con la Parque Nacional Yasuní. La muestra Huaorani procede de la parte de la Reserva Huaorani perteneciente a la provincia de Pastaza.

Todos los individuos de los que se obtuvo una muestra fueron voluntarios informados acerca del propósito de la investigación y se encontraban sanos. Las muestras fueron tomadas por un equipo dirigido por la Dra. Martínez Jarreta, de la Universidad de Zaragoza.

En cada caso se siguieron las directrices éticas para la investigación con seres humanos de las instituciones involucradas (GENES, Medellín, Colombia; Universidad Nacional de Jujuy, Argentina; Universidad de Zaragoza; Universidad del País Vasco).

MÉTODOS

Extracción del ADN

El proceso de extracción del ADN en una parte de las muestras que se han analizado en esta Tesis ya se había realizado. Sin embargo, en las muestras de Jujuy se partía de manchas de sangre en papel, tanto Whatman como de filtro. En función de las características del soporte, se ha recurrido a dos métodos alternativos, el método *Fenol-Cloroformo-Isoamil Alcohol*, con alguna modificación sobre el protocolo original y el método de extracción en columna, mediante el kit comercial QiAmp® DNA Micro de Qiagen.

Método *Fenol-Cloroformo-Isoamil Alcohol*

El método *Fenol-Cloroformo-Isoamil Alcohol* se ha aplicado sobre las muestras que se habían recogido y almacenado mediante papel Whatman. Los reactivos utilizados han sido:

- SDS al 10%
- Tampón DLB
- Proteinasa K
- Fenol:Cloroformo:Isoamil Alcohol (25:24:1)
- Etanol 98%
- Agua Ultrapura (H₂O_{up})

El protocolo se describe a continuación.

- Se toma aproximadamente 1 cm² de mancha cortada en pequeños fragmentos y se introduce en un tubo de 1,5 ml. A continuación se resuspende en 500 µl de tampón DLB.
- Se añaden 50 µl de SDS al 10% y 10 µl de Proteinasa K (20mg/ml) y se agita para que se homogeneice.
- Se incuba durante 2 horas a 65°C. Una vez pasado este tiempo y para detener la reacción, se introduce el tubo en hielo durante 5 minutos.
- Se añaden 550 µl de Fenol:Cloroformo:Isoamil Alcohol (25:24:1), mezclando por inversión. Este paso se realiza en una campana de extracción de gases para evitar inhalar los vapores tóxicos del reactivo.
- Se centrifuga durante 7 minutos a 12.000 rpm, manteniendo la temperatura de la muestra a 7°C. Tras la centrifugación, en el tubo aparecen dos fases, separadas por una fina capa blanca. Se pasa el sobrenadante a otro tubo mediante una pipeta, procurando no resuspender la interfase blanca de leucocitos.
- Se añade al sobrenadante 1 ml de Etanol absoluto (almacenado a -20°C). A continuación se enfría la muestra durante 45 minutos a -20°C. Una vez transcurrido este tiempo, se centrifuga a 14.000 rpm durante 20 minutos, manteniendo la temperatura a 7°C.
- Se retira mediante una pipeta todo el sobrenadante para eliminar todos los restos de etanol asegurándose de no remover el precipitado. Después se deja el tubo a temperatura ambiente durante unos 15-20 minutos.
- Una vez que el tubo se encuentra perfectamente seco, se resuspende el precipitado en 50 µl de H₂O_{up}. Por último, se incuba durante 1 hora a 60°C.

La calidad y la concentración del ADN extraído se evaluaron mediante electroforesis en gel de agarosa y, posteriormente, a través de un fluorímetro. El almacenaje de las muestras se realizó a -20°C.

Extracción mediante Kit comercial en columna

La extracción mediante Kit comercial en columna se ha aplicado sobre las muestras que se habían obtenido y almacenado en papel de filtro, estando contraindicado para manchas de sangre en papel Whatman. La extracción de ADN mediante columna se ha realizado con el kit comercial QiAmp® DNA Micro de Qiagen. Para su realización se han utilizado los materiales y reactivos que proporciona dicho Kit:

- Columnas QIAamp MinElute™
- Tubos colectores
- Tampón ATL
- Tampón AL
- Tampón AW1: Para su utilización es necesario resuspenderlo con Etanol absoluto. 50 tests: 19 ml AW1 + 25 ml Etanol
- Tampón AW2: Para su utilización es necesario resuspenderlo con Etanol absoluto. 50 tests: 13 ml AW1 + 30 ml Etanol
- Proteinasa K
- Etanol Absoluto
- Tampón AE

Se ha utilizado el protocolo de extracción que proporciona el Kit QiAmp® DNA Micro, con modificaciones menores:

- Se toman 3-4 círculos de aproximadamente 3 mm de diámetro y se introducen en un tubo de 1,5 ml.
- Se añaden 180 µl de tampón ATL. A continuación se añaden 30 µl de Proteínasa K. Se mezclan con vórtex durante 15 s y se incuban durante 1 hora a 56°C con agitación.
- Una vez transcurridos este tiempo, se añaden 200 µl de tampón AL mezclando de nuevo con vórtex e incubar a 70°C con agitación durante 10 minutos.
- Una vez transcurridos los 10 minutos, se añaden 200 µl de Etanol a la mezcla y se mezclan con vórtex.
- Se transfiere cuidadosamente la solución a las columnas QIAamp MinElute™ sin mojar el cuello de la columna y se centrifuga durante 1,5 minutos a 8.000 rpm.
- Una vez realizado, se desecha la solución y se dispone la columna QIAamp MinElute™ en un nuevo tubo colector. Se añaden 500 µl de tampón AW1 y se centrifuga de nuevo a 8.000 rpm durante 1,5 minutos. Se realiza este paso dos veces, desechando la solución en cada uno.
- Se añaden 500 µl de tampón AW2 a la columna y se centrifuga a 14.000 rpm durante 4,5 minutos.
- Se desecha el filtrado y se transfiere la columna a un tubo eppendorf de 1,5 ml. Se añaden 50 µl de tampón de elución AE (o H₂O ultrapura) a la columna y se incuba a temperatura ambiente durante 5 minutos. Una vez transcurrido el tiempo, se centrifuga durante 1,5 minutos a 8.000 rpm.
- El líquido filtrado contiene el ADN eluído. Se puede hacer repetir el paso para recoger todo el ADN de la membrana de filtración añadiendo otros 50 µl en un nuevo tubo eppendorf de 1,5 ml.

Del mismo modo que con la extracción orgánica, la calidad y la concentración del ADN extraído se evaluaron mediante electroforesis en gel de agarosa y fluorímetro. Posteriormente, dichas muestras se almacenaron a -20°C.

Marcadores Genéticos

En esta sección se describen las principales características técnicas de los marcadores analizados en los trabajos de esta memoria. Sin embargo, por un criterio de fluidez del texto, una parte de esta información se desarrolla en la sección de Anexos.

Inserciones *Alu*

El análisis de los polimorfismos de inserción *Alu* es relativamente sencillo. En primer lugar, se realiza una amplificación PCR (Mullis *et al.* 1986) para después visualizar los resultados mediante una electroforesis en gel de agarosa.

Para llevar a cabo las diferentes reacciones de amplificación se ha utilizado un termociclador Perkin Elmer GeneAmp PCR system 9700 (Applied Biosystem). Las características específicas de los reactivos utilizados así como de las reacciones PCR se muestran en el anexo.

La asignación de los genotipos individuales en todos los casos se ha realizado mediante electroforesis en gel de agarosa. A pesar de no mostrar un poder de resolución tan alto como la poliacrilamida, la agarosa es eficaz para discriminar bandas con diferencias de tamaño superiores a 15-20 pb. Puesto que los amplificadores de los alelos de las inserciones *Alu* se diferencian entre 100 y 350 pb, esta técnica es perfectamente adecuada para dicho propósito, además de ser sencilla, rápida y poco costosa.

Tabla 3. Ubicación, secuencia de los cebadores, tamaño del amplímero y temperatura de anillamiento de las inserciones *Alu* analizadas.

Locus	Localización Cromosómica	Secuencia de los Primers (5'-3')	Tamaño Amplímero	Tª Anillamiento
A25	8q21.3	F: CCACAAATAGGCTCATGTAGAAC R: TATAATATGGCCTGGATTATACC	<i>Alu</i> (+): 552pb <i>Alu</i> (-): 268pb	58°C
ACE	17q23	F: CTGGAGACCACTCCCATCCTTTCT R: GATGTGGCCATCACATTCGTGAGAT	<i>Alu</i> (+): 490pb <i>Alu</i> (-): 190pb	55,5°C
APO	11q23.3	F: AAGTGCTGTAGGCCATTTAGATTAG R: AGTCTTCGATGACAGCGTATACAGA	<i>Alu</i> (+): 409pb <i>Alu</i> (-): 97pb	47°C
B65	11q14.2	F: ATATCCTAAAAGGGACACCA R: AAAATTTATGGCATGCGTAT	<i>Alu</i> (+): 423pb <i>Alu</i> (-): 81pb	57,5°C
D1	3q26.32	F: TGCTGATGCCAGGGTTAGTAAA R: TTTCTGCTATGCTCTCCCTCTC	<i>Alu</i> (+): 670pb <i>Alu</i> (-): 333pb	56,5°C
FXIIIB	1q31-q32	F: TCAACTCCATGAGATTTTCAGAAGT R: CTGGAAAAATGTATTCAGGTGAGT	<i>Alu</i> (+): 700pb <i>Alu</i> (-): 410pb	58°C
HF	6p21.31	F: GCCTCATGGCCTGAATCTGCCAGTGCCTT R: GTAACCTGACCTGCCCTCTATAGTATAGTCT	<i>Alu</i> (+): 602pb <i>Alu</i> (-): 457pb	62°C
HG	6p21.31	F: CAGGACAACCAGTAAAGATGCTGG R: GCTTCAGTTAACATGCAAGTTTATGCC	<i>Alu</i> (+): 540pb <i>Alu</i> (-): 218pb	64°C
HJ	6p21.31	F: AAGAACCATAAECTCACTTG R: TGTGTCCAGGTTAACTTCAG	<i>Alu</i> (+): 494pb <i>Alu</i> (-): 162pb	62°C
HS3.23	7q35	F: GGTGAAGTTTCCAACGCTGT R: CCCTCCTCTCCCTTTAGCAG	<i>Alu</i> (+): 498pb <i>Alu</i> (-): 200pb	60°C
HS4.32	12q14.1	F: CTTTATTGGGCTAACCTGGG R: TGACCTGCTAACTGTACTTTAACCC	<i>Alu</i> (+): 601pb <i>Alu</i> (-): 289pb	60°C
HS4.59	9q31.1	F: TTTCTGAACCTGAAGACAGG R: TCTCAATTTCTTCTCATAATGCA	<i>Alu</i> (+): 461pb <i>Alu</i> (-): 143pb	60°C
HS4.65	9q21.33-34	F: TGAAGCCAATGGAAAGAGAG R: ACAGGAGCATCTAAACCTTGG	<i>Alu</i> (+): 650pb <i>Alu</i> (-): 329pb	61°C
MicB	6p21.31	F: GCCTTCCAATGCCATTACAG R: CTCAGCCCTGCTTTCCCATCT	<i>Alu</i> (+): 665pb <i>Alu</i> (-): 502pb	68°C
PV92	16q23.3	F: AACTGGGAAAAATTTGAAGAGAAAAGT R: TGAGTTCTCAACTCCTGTGTGTTAG	<i>Alu</i> (+): 443pb <i>Alu</i> (-): 129pb	57,8°C
TF	6p21.31	F: GTGCCTGGTAAAAATTTAAGAGCTGTA R: TGCACCCGGCCTAAAACCACTGGTT	<i>Alu</i> (+): 711pb <i>Alu</i> (-): 426pb	65°C
TPA25	8p11.2	F: GTAAGAGTTCCGTAAACAGGACAGCT R: CCCACCCCTAGGAGAACTTCTCTTT	<i>Alu</i> (+): 424pb <i>Alu</i> (-): 113pb	63°C

La concentración de agarosa utilizada en los geles ha sido en todos los casos del 1,5 % en peso. Para la visualización de las bandas de ADN se ha utilizado en la mayor parte de los análisis una tinción de Bromuro de Etidio, un intercalante con mucha afinidad por la molécula de ADN, pero altamente tóxico. En los últimos análisis ha sido sustituido por GelRed, un reactivo igualmente eficaz y en absoluto tóxico, ya que a pesar de ser también un agente intercalante, no penetra las membranas celulares.

Para llevar a cabo la migración de las moléculas de ADN, los geles han sido sometidos a una diferencia de potencial y tiempo de migración acorde a las características del marcador analizado. La visualización de los resultados se ha llevado a cabo mediante un transiluminador de luz UV tanto para el Bromuro de Etidio como para GelRed (Figura 24). Los resultados han sido documentados mediante un sistema de fotografía digital (Kodak Digital Science EDAS 120).

Para la identificación del tamaño de los amplificados se han utilizado dos marcadores: DNA molecular weight marker V (Roche, 0,25 µg/µL), con un rango de 8 pb a 587 pb; y Amerhsam™ 100 Base pair ladder (GE Healthcare, 1 µg/µL), con un rango de 100 pb a 2000 pb.

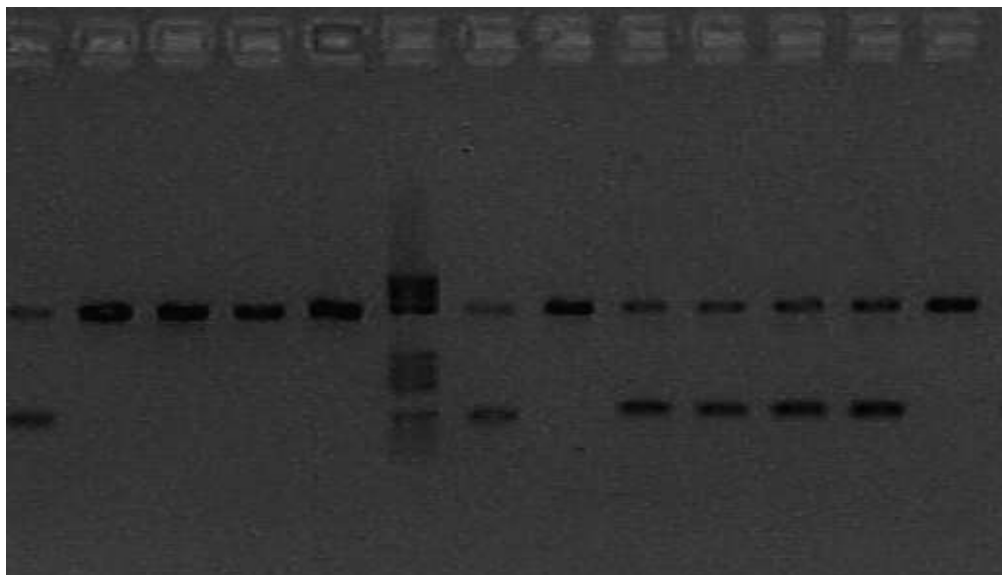


Figura 24. Electroforesis en gel de agarosa con amplificadores del marcador *Alu* TPA25. En el centro, el marcador de tamaño del ADN (DNA Molecular Weight Marker V, 8-587 pb, Roche).

Microsatélites

Para los microsatélites también se han realizado amplificaciones PCR, si bien en este grupo de marcadores un cebador de cada par llevaba incorporado un fluorocromo, con el fin de poder identificar posteriormente los genotipos en un analizador de ADN. Las características específicas de los reactivos utilizados así como de las reacciones PCR se muestran en el anexo.

Tabla 4. STRs analizados, secuencia de los cebadores, tamaño del amplímero y temperatura de anillamiento.

Locus	Secuencia de los Primers (5'-3')	Rango Amplímero	T ^a anillamiento
D6S105	F: GCCCTATAAAATCCTAATTAAC R: GAAGGAGAATTGTAATCCG	100pb – 140pb	55°C
D6S265	F: ACGTTCGTACCCATTAACCT R: ATCGAGGTAAACAGCAGAAA	110pb – 150pb	57°C
D6S2666	F: GGTGAGTCTAAATTGTCCCAGGAG R: TCATGGATGACTGGAGCCTATCC	140pb – 180pb	60°C
D6S2792	F: GCCTCTAGATTTTCATCGAGCCACA R: CCTCTCCCTCCTGCAACACACA	90pb – 130pb	60°C

Una vez realizada la técnica PCR amplificadas, se debe preparar las muestras para su análisis en el analizador genético, mediante la preparación del siguiente *mix*:

- Amplificado PCR, 1 µl por normal general.
- Formamida desionizada, 20 µl por muestra.
- Marcador de tamaño estándar con fluorocromo (Genescan 500 Tamra, Applied Biosystems), que se utiliza como referencia para extrapolar el tamaño de los fragmentos amplificados. Se utilizaron 0,5 µl por muestra.

Esta solución se somete a 95°C durante 5 minutos, con el fin de desnaturalizar el ADN. A continuación se estabiliza a -20°C, durante 2-3 minutos.

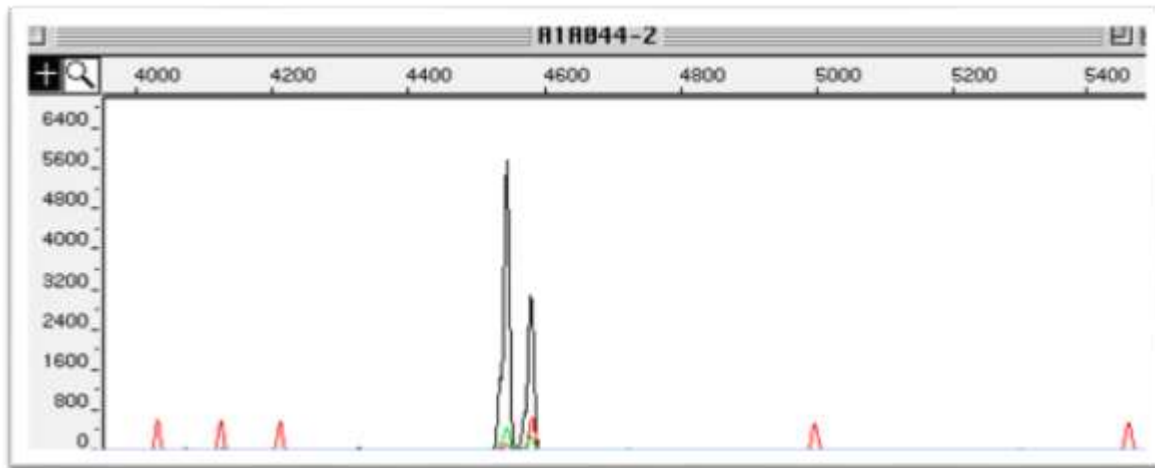


Figura 25. Electroferograma del microsatélite D6S2666. Los picos negros corresponden con el STR y los rojos con el marcador de tamaño (Genescan 500 TAMRA Size standard).

Con el fin de identificar los alelos de los genotipos para los microsatélites, se ha llevado a cabo un análisis de fragmentos mediante un Analizador Genético por electroforesis capilar ABI PRISM 310 (Applied Biosystems). En este equipo las muestras han migrado durante 15-30 minutos, a una diferencia de potencial de 15,0 kV, después de un tiempo de inyección variable (que habitualmente oscila entre 5 y 10 segundos).

El análisis de fragmentos se ha realizado mediante el programa GeneMapper Analysis v3.2 (Figura 25), en el que la identificación del tamaño de los alelos se obtiene por extrapolación a partir de un marcador con tamaños de ADN conocidos (Genescan 500 TAMRA Size standard).

Análisis estadísticos realizados

En los siguientes párrafos se describen con un cierto detalle los métodos estadísticos utilizados en los artículos que componen la sección Resultados.

- **Frecuencias Alélicas.** Las frecuencias alélicas es la proporción relativa de cada alelo frente al total de alelos existentes. El cálculo de las frecuencias y sus errores estándar se ha realizado mediante un programa desarrollado en el laboratorio (STRs) para tal propósito.

- **Frecuencias haplotípicas.** Las frecuencias haplotípicas, al igual que las alélicas, se calculan como la proporción relativa de cada haplotipo frente al total de haplotipos existentes. No obstante, para los haplotipos autosómicos es imposible conocer la fase gamética de los alelos, lo que determina que su cálculo se realice mediante inferencias por el método de máxima verosimilitud. Se ha realizado mediante el programa ARLEQUIN.

- **Error Estándar de la Media.** Para el cálculo de las frecuencias haplotípicas y alélicas se ha estimado el error estándar. Este estadístico cuantifica las oscilaciones potenciales de la media muestral alrededor de la media poblacional. El EEM (o *SEM*) se estima generalmente dividiendo la desviación estándar de la población entre la raíz cuadrada del tamaño de la muestra

$$SE = \frac{S}{\sqrt{n}}$$

donde *S* es la desviación estándar de la población y *n* es el tamaño muestral.

$$S = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

- **Diversidad alélica.** Se define como la probabilidad de que, en un locus dado, dos alelos de dos cromosomas tomados al azar en la población sean diferentes entre sí. Es sinónimo de Heterocigosidad esperada. También puede considerarse como la probabilidad de que un individuo tomado al azar posea dos alelos diferentes en un locus determinado. La Heterocigosidad promedio calculada para un grupo de loci es una estimación adecuada del grado de variabilidad genética de una población.

- **Diversidad haplotípica.** Es equivalente al concepto de diversidad alélica, aplicado a las frecuencias haplotípicas. La diversidad haplotípica (Nei, 1987) se define como la probabilidad de que dos haplotipos (del mismo grupo de loci) de dos cromosomas tomados al azar en una

población, sean diferentes. Tanto la diversidad alélica como la haplotípica y sus correspondientes varianzas se calculan como:

$$H = \frac{n}{(n-1)} \left(1 - \sum_i p_i^2 \right)$$

$$V = \frac{2}{n(n-1)} \left\{ 2(n-2) \left[\sum_i p_i^3 - \left(\sum_i p_i^2 \right)^2 \right] + \sum_i p_i^2 - \left(\sum_i p_i^2 \right)^2 \right\}$$

donde n es el número total de individuos analizados y p_i es la frecuencia del i-ésimo alelo o haplotipo.

Para estimar las diversidades se ha utilizado el programa Arlequin v3.5 (Excoffier y Lischer, 2010).

- **Prueba de Kolmogórov-Smirnov (prueba K-S).** Es una prueba no paramétrica que se utiliza para determinar la bondad del ajuste de la distribución observada de los datos a una distribución esperada de probabilidades (prueba K-S de una muestra) o para comparar la distribución de dos muestras entre sí (prueba K-S de dos muestras). La distribución nula se calcula bajo la hipótesis nula de que la muestra se ajuste a la distribución de referencia (Prueba K-S de una muestra) o que ambas muestras tomen de la misma distribución (prueba K-S de dos muestras). La distribución de los datos F_n con n observaciones con valores y_i se define como:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1 & \text{si } y_i \leq x \\ 0 & \text{alternativa} \end{cases}$$

Para las dos colas, el estadístico viene dado por:

$$D_n^+ = \max(F_n(x) - F(x))$$

$$D_n^- = \max(F(x) - F_n(x))$$

donde $F(x)$ es la distribución presentada como hipótesis.

En nuestro caso, se realizó la prueba K-S de una muestra.

Se ha utilizado para su cálculo el programa SPSS v18.

- **Desequilibrio de ligamiento.** Existen varias pruebas estadísticas diseñadas con el fin de detectar alteraciones en las frecuencias genotípicas debido al ligamiento que puede existir entre dos o más marcadores. Se ha utilizado una extensión del test de probabilidad exacta de Fisher (Slatkin, 1994a). Se basa en una tabla de contingencia con las frecuencias de los haplotipos observados, con $k_1 \times k_2$ entradas que se corresponden con el número de alelos del locus 1 y locus 2, respectivamente. Se comprueba la probabilidad de obtener una tabla con las mismas sumas de frecuencias en filas y columnas que aparezca con una frecuencia menor o igual que la tabla observada. Bajo la hipótesis nula de no asociación entre los dos loci analizados la probabilidad de la tabla observada será:

$$L_0 = \frac{n!}{\prod_{i,j} n_{ij}} \prod_i (n_{i*}/n)^{n_{i*}} \prod_i (n_{*i}/n)^{n_{*i}},$$

donde n_{ij} denota el número de los haplotipos que tienen el alelo i en el primer locus y el alelo j en el segundo locus, n_{i*} es la frecuencia general del alelo i en el primer locus ($i=1, \dots, k_1$) y n_{*i} es el recuento del alelo i en el segundo locus.

En lugar de analizar la probabilidad de todas las posibles tablas de contingencia, puede utilizarse la cadena de Markov para explorar de manera eficiente una muestra representativa del espacio de dichas tablas. Sin embargo, cuando la fase haplotípica es desconocida, como es el caso de los loci autosómicos analizados en esta Tesis Doctoral, el ensayo basado en la cadena de Markov no es válido porque la composición haplotípica de la muestra es desconocida y se trata sólo de una estimación. Por lo tanto, el desequilibrio de ligamiento entre un par de loci para datos genotípicos se comprueba mediante una prueba de razón de verosimilitud, cuya distribución empírica se obtiene por un procedimiento de permutación (Slatkin y Excoffier, 1996).

Debe tenerse en cuenta que esta prueba de desequilibrio de ligamiento asume que los genotipos siguen las proporciones Hardy-Weinberg, por lo que el rechazo de la prueba puede ser también debido a desviaciones en las proporciones del equilibrio de Hardy-Weinberg (Excoffier y Slatkin, 1998).

El cálculo del desequilibrio de ligamiento y la estima de las frecuencias haplotípicas se ha realizado mediante el programa Arlequin v3.5 (Excoffier y Lischer, 2010).

- **Índice Garza-Williamson.** Este estadístico ha demostrado ser sensible a cuellos de botella poblacionales, ya que en reducciones recientes del tamaño poblacional, el número de alelos se reduce de manera más importante que el rango, de tal manera que la distribución de los diferentes alelos de un locus mostrará posiciones ausentes en todo el rango de alelos encontrados. Por lo tanto, mostrará valores bajos en aquellas poblaciones que hayan sufrido cuellos de botella, mientras que en las poblaciones estacionarias tendrá un valor mayor, cercano a uno.

Según Garza y Williamson (2001), se calcula como

$$G - W = \frac{k}{R + 1}$$

donde k es el número de alelos de un locus dado en una muestra poblacional y R es el rango de alelos.

Se ha calculado mediante el programa Arlequin v3.5 (Excoffier y Lischer, 2010).

- **Test de neutralidad de Ewens-Watterson.** Es una prueba de neutralidad selectiva basada en la teoría de muestreo de Ewens en una población en equilibrio (Ewens, 1972), de acuerdo a un modelo de infinitos alelos. Se puede utilizar tanto para casos de selección balanceada como de alelos ventajosos.

La distribución nula del estadístico (F) se genera mediante la simulación de muestras aleatorias neutras, que contienen el mismo número de genes y el mismo número de haplotipos,

utilizando el algoritmo de Stewart (1977). Se registra la probabilidad de observar muestras aleatorias con valores de F idénticos o menores que la muestra original.

Se ha calculado mediante el programa Arlequin v3.5 (Excoffier y Lischer, 2010).

- **Networks (Redes de Haplotipos).** Son construcciones gráficas, con un objetivo filogenético, que muestran las relaciones de ancestro-descendiente entre un grupo de haplotipos. Se utilizan para reconstruir las redes filogenéticas, inferir los tipos ancestrales, los tipos posibles, las ramificaciones evolutivas y proporcionar estimaciones de la edad del ancestro común de los haplotipos seleccionados.

Se han realizado mediante el programa Network v4.6 (fluxus-engineering.com; Bandelt *et al.* 1999).

- **AMOVA.** El Análisis Molecular de la Varianza (*Analyses of MOlecular VAriance*) es un análisis jerarquizado análogo al ANOVA, aplicado a datos moleculares (secuencias de ADN, frecuencias alélicas o frecuencias haplotípicas) donde los coeficientes de correlación intraclases se hacen depender del proceso mutacional. El método fue desarrollado por Excoffier *et al.* (1992). Se utiliza como medida de la subdivisión poblacional intraespecífica sin asumir ningún modelo espacial *a priori*, de modo que la estructura de los grupos la definen las poblaciones que los componen. Del mismo modo que en el ANOVA, se estiman los estadísticos F mediante correlaciones intraclass definidas como proporciones de los componentes de la varianza.

El uso principal para el que fue diseñado es la valoración del grado de estructuración que presentan uno o varios grupos de poblaciones. Para ello se utilizan tres componentes o niveles jerarquizados de diversidad, dentro de la población (WP), entre poblaciones dentro de grupos (AP/WG) y entre grupos (AG), como se describen en la *tabla 5* (Excoffier, 2007).

Los cálculos se han realizado mediante el programa Arlequin v3.35 (Excoffier y Lischer, 2010).

Tabla 5. Diseño general para el análisis jerárquico de la varianza molecular (AMOVA).

Fuente de Variación	Grados de libertad	MSD	MSD esperada
Entre Grupos	G-1	MSD (AG)	$\sigma_c^2 + n'\sigma_b^2 + n''\sigma_a^2$
Entre poblaciones dentro de grupos	$\sum_{g=1}^G I_g - G$	MSD (AP/WG)	$\sigma_c^2 + n'\sigma_b^2$
Entre individuos dentro de cada población	$N - \sum_{g=1}^G I_g - G$	MSD (WP)	σ_c^2
Total	N-1		

donde N es el número de individuos totales

I_g es el número de poblaciones en el grupo g para $g= 1, \dots, G$ grupos, σ son los componentes de la varianza descritas a partir de los estadísticos F de Wrigth (Wrigth, 1950, 1951, 1965) o Φ según la terminología de Excofier *et al.* (1992), n se refiere al tamaño medio poblacional para los 3 diferentes niveles jerárquicos del modelo y MSD se corresponde con la media de las desviaciones estándar al cuadrado (*Mean Squared Deviation*), que son obtenidas a partir del cociente de SSD (suma de las desviaciones estándar al cuadrado) entre los grados de libertad

- **Método del Centroide.** Se trata de un modelo diseñado para valorar la intensidad del flujo génico que ha recibido una población, en relación al resto de poblaciones que se incluyen en el análisis (Harpending y Ward, 1982). Se basa en el establecimiento de una relación lineal simple entre la Heterocigosidad observada de cada población y la distancia de cada una de las poblaciones al centroide.

La distancia al centroide (r_i), que equivale a la aportación de cada población a la varianza de Wahlund del conjunto, se calcula como la diferencia al cuadrado entre la frecuencia alélica de cada población y la media aritmética de las frecuencias alélicas del conjunto de poblaciones analizadas, dividida por el valor máximo de la varianza,

$$r_i = \frac{(p_i - P)^2}{P(1 - P)}$$

donde p_i es la frecuencia del alelo en la población i y P es la frecuencia promedio del alelo en la población total.

El cálculo se realiza para cada alelo y después es promediado para todos los loci analizados. Para realizar una valoración visual, se representan en un gráfico los valores de r_i en el eje de abscisas y la heterocigosidad observada en el de ordenadas.

Este método se ha realizado mediante el programa GeDis v1.8 (Peña *et al.* 2009).

- **Estimas del mestizaje.** Los análisis de mestizaje se utilizan para estimar las proporciones de poblaciones parentales que se encuentran representadas en el acervo genético de una población mestiza. Para llevar a cabo este cálculo hay varias aproximaciones metodológicas, como el método de la Máxima verosimilitud (Wang, 2003), el de los Mínimos cuadrados (Long *et al.* 1991) y el de la Identidad génica (Chakraborty, 1985).

En general, se busca la solución más adecuada a la función:

$$p_{ih} = \sum_{j=1}^J p_{ij} \mu_j$$

donde p_{ih} es la frecuencia del alelo i en la población híbrida, p_{ij} denota la frecuencia del alelo i en la población de referencia j ($j = 1, J$) y μ_j es la contribución proporcional de la población de referencia j a la población híbrida. El sumatorio de las contribuciones proporcionales debe ser igual a 1 y todas deben tomar valores positivos,

$$\sum_{j=1}^J \mu_j = 1$$

$$\mu_j > 0, \forall j$$

Para las diferentes estimas de tasas de mestizaje realizadas en los trabajos de esta tesis se han utilizado los programas Admix2 (Chakraborty, 1985), Admix (Long *et al.* 1991) y Leadmix (Wang, 2003).

- **Clinas.** Se denomina clina a la variación asociada a una componente geográfica de una característica determinada genéticamente. Su origen puede deberse a variaciones en la eficacia biológica de diferentes genotipos a lo largo de un territorio en respuesta a factores ambientales, siempre que estos sean duraderos y estables en el tiempo. Algunos ejemplos de factores con variabilidad geográfica y por tanto susceptibles de generar clinas son la altitud, la latitud, la insolación y la humedad.

Además de aparecer como resultado de presiones selectivas, las clinas pueden ser generadas por otros procesos, como la migración, cuando dos poblaciones con orígenes diferentes se asientan en puntos distantes de una región, de modo que la migración de individuos procedentes de ambas hacia la zona intermedia y el correspondiente mestizaje crearán una variación gradual (Hartl y Clark, 1997). Además de la migración, las clinas también puede aparecer debido a la acción de la Deriva Genética.

Para la detección de potenciales clinas, en esta tesis se han calculado los coeficientes de correlación entre las frecuencias génicas y la posición geográfica de las poblaciones analizadas, respecto a un sistema de coordenadas móviles que rota 360 grados. Las posiciones geográficas de las poblaciones se proyectan perpendicularmente sobre el eje, calculando la correlación entre las coordenadas y las frecuencias. En sucesivas iteraciones, se rota el eje un grado y se calcula nuevamente la correlación.

Cuando se encuentra un eje que presenta una correlación significativa, su orientación es considerada como una clina. A menudo se observan varias correlaciones significativas en grados consecutivos. En ese caso, el algoritmo selecciona aquella orientación para la cual la significación es máxima (Pérez-Miranda *et al.* 2003, 2004; Peña, *et al.* 2009).

La detección de clinas se ha realizado mediante el programa GeDis v1.8 (Peña *et al.* 2009).

- **Coefficientes de Distancia Genética.** La distancia genética se puede definir como una medida de la afinidad existente entre diferentes taxones, poblaciones o individuos, calculada a partir de datos con un origen genético. En esta memoria se han utilizado dos coeficientes de distancia:

- Distancia genética de Reynolds (Reynolds *et al.* 1983). Asume que no existen mutaciones y que la deriva genética es el único factor responsable del cambio en las frecuencias de los haplotipos. A diferencia de otras distancias como la distancia genética de Nei y Li (1979), esta medida no asume un tamaño efectivo de población constante e igual para todas las poblaciones. La distancia genética de Reynolds se calcula mediante la fórmula:

$$D^2 = \frac{\sum_m \sum_i (p_{1mi} - p_{2mi})^2}{2 \sum_m (1 - \sum_i p_{1mi} - p_{2mi})}$$

donde m se refiere al conjunto de los loci, i al conjunto de alelos, p_{1mi} las frecuencias del alelo i del locus m en la población 1 y p_{2mi} las frecuencias del alelo i del locus m en la población 2.

- Distancia genética de Harpending y Jenkins (Harpending y Jenkins, 1973). Se basa en la matriz R de parentesco, que se define como:

$$r_{ij} = \sum_k \frac{(P_{ik} - \bar{P}_k)(P_{jk} - \bar{P}_k)}{\bar{P}_k(1 - \bar{P}_k)}$$

Donde p_{ik} , p_{jk} son las frecuencias del alelo k en las poblaciones i y j y \bar{P}_k es la frecuencia promedio del alelo k .

A partir de la matriz R de parentesco se construye una distancia genética mediante la fórmula:

$$d_{ij} = r_{ii} + r_{jj} - 2r_{ij}$$

Las matrices de distancia genética entre poblaciones para los dos coeficientes se han obtenido mediante los programas Phylip v3.69 (Felsenstein, 1989) y GeDis v1.8 (Peña *et al.* 2009).

- **Diferenciación poblacional.** Se estima mediante estadísticos que valoran la significación de la heterogeneidad genética existente entre poblaciones, tomadas dos a dos.

- Test exacto de diferenciación poblacional (Raymond y Rousset, 1995; Goudet *et al.* 1996). Es el estadístico más utilizado para comprobar las diferencias entre poblaciones. Es análogo al test exacto de Fisher en una tabla de contingencia 2x2, pero extendido a una tabla de contingencia rxk.

La hipótesis nula postula que la distribución de las frecuencias alélicas es independiente a través de las poblaciones, lo que corresponde a la ausencia de diferenciación poblacional, de forma que mediante un método no paramétrico se calcula una estimación no sesgada de valor p de aceptación o rechazo de la hipótesis nula.

Se analizan todos los estados posibles de la matriz, estimando si existe una probabilidad igual o menor de que aparezca y se determina para cada caso si su probabilidad de ocurrencia es mayor o menor que p (en relación a la matriz observada en el supuesto de panmixia). Para que esta estimación sea no sesgada se realiza un proceso de aleatorización, de modo que la cadena de Markov comienza a partir de un lugar al azar de todo el espacio de matrices posibles (Guo y Thompson, 1992). La propiedad de Markov asegura que el proceso no tiene memoria, lo que implica que la distribución de probabilidad condicional de estados futuros sólo depende del estado

actual, y no de la secuencia anterior de acontecimientos. Las cadenas markovianas se caracterizan por usar matrices de tipos infinitesimales que cuantifican las probabilidades de intercambio de caracteres discretos en un intervalo de tiempo infinitamente pequeño (Bielejec *et al.* 2014).

El método de la cadena de Markov genera, en este caso, una distribución de probabilidad exacta bajo la hipótesis nula, que no está sesgada por la presencia de alelos raros o por pequeños tamaños muestrales. Mediante este método se obtiene la probabilidad de diferenciación genética entre dos o más poblaciones para cada polimorfismo.

- Test F_{ST} Pairwise. En este método, la distribución nula de los valores F_{ST} Pairwise bajo la hipótesis de no diferenciación entre poblaciones se obtiene a través de la permutación de los alelos frente a las poblaciones. Frente a esa distribución nula se comparan los valores de distancia de Reynolds, dando como resultado unos valores de probabilidad del test, que muestran la proporción de permutaciones que llevan a un valor de F_{ST} igual o mayor que el valor observado en la distancia (Reynolds *et al.* 1983; Slatkin, 1995).

Ambos métodos se han realizado mediante el programa Arlequin v-3.5 (Excoffier y Lischer, 2010).

- **Análisis de escalamiento multidimensional (MDS)**. Es un análisis multivariante mediante el cual se realiza una interpretación gráfica de una matriz de disimilaridad. Mediante un algoritmo de tanteo, basado en el concepto de preordenación de las distancias, se asigna una ubicación para cada uno de los elementos en un espacio de N dimensiones, donde N es un valor especificado previamente y habitualmente bajo.

Se ha utilizado el método no paramétrico del Escalamiento Multidimensional, debido a que los coeficientes de distancia genética habitualmente no siguen una distribución paramétrica.

El objetivo del MDS es transformar los datos de distancias genéticas en un conjunto de distancias euclídeas. La solución, es decir, la configuración final debe optimizar el ajuste entre el orden de las distancias euclídeas en el gráfico y el orden de las distancias genéticas en la matriz original, dispuestas de menor a mayor.

El grado de fiabilidad de la representación resultante se evalúa mediante el diagrama de Sephard y el coeficiente de estrés. En el primer caso, se representan las distancias euclídeas frente a las distancias genéticas; en el segundo caso, se valora el resultado de la fórmula:

$$Stress = \left[\frac{\sum_i \sum_j (d_{ij} - \hat{d}_{ij})^2}{\sum_i \sum_j d_{ij}^2} \right]^{1/2}$$

donde d_{ij} es la distancia euclídea entre las poblaciones i y j y \hat{d}_{ij} su distancia genética.

Los análisis MDS se han realizado mediante los programas SPSS v18, Past v2.16 (Hammer *et al.* 2001), GeDis v1.8 (Peña *et al.* 2009) y mediante el software Phylip v3.69 (Felsenstein, 1989) para el cálculo de la distancia y SPSS v18 para la representación así como para la obtención del valor RSQ y del Stress.

- **Dendrogramas.** Este grupo de métodos muestra la información de las matrices de distancias mediante diagramas en forma de árbol. Su objetivo puede ser evolutivo, donde se trata de reconstruir la historia filogenética de un grupo de taxones o meramente descriptivo, como es el caso de los realizados en esta tesis, buscando explicar las relaciones entre un grupo de poblaciones. Son herramientas muy útiles en estudios que incluyan gran cantidad de datos por su capacidad para resumir la información de una matriz de distancias, si bien a menudo una consecuencia indeseada es la excesiva simplificación del resultado final.

Existen numerosos algoritmos para obtener dendrogramas (UPGMA, Neighbor-joining,...), algunos con aproximaciones conceptuales muy diferentes y un gran número de variantes para

realizar las representaciones. Únicamente se describirá el método Neighbor-joining, que ha sido el empleado en alguno de los trabajos de esta Tesis.

El Neighbor-joining (NJ) es un método iterativo desarrollado por Saitou y Nei (1987) para la obtención de fenogramas (dendrogramas sin raíz). Una de sus principales ventajas es que no se asume la misma tasa de evolución para todos los linajes. Se basa en el criterio de mínima evolución, de modo que se seleccionan sistemáticamente las distancias más pequeñas. Además, puede utilizarse para la interpretación de matrices de distancias no euclideas. Este método se basa en el mantenimiento de un registro de nodos. Se parte de un árbol cuya topología corresponde a una red en forma de estrella (*Figura 26*). Después se construye una matriz de distancia modificada (matriz Q), en la que la separación entre los pares de nodos es ajustada en función de su media de divergencia con respecto al resto de nodos. En cada etapa se reducen dos nodos a uno. El proceso se completa cuando únicamente existan dos nodos con una rama que los une.

La secuencia de construcción del árbol NJ sigue los siguientes pasos:

1. Se calcula la matriz Q, a partir de la matriz de distancias original.

$$Q_{(i,j)} = (r - 2)d(i, j) - \sum_{k=1}^r d(i, k) - \sum_{k=1}^r d(j, k)$$

donde $d(i,j)$ es la distancia entre el taxón i y el taxón j, siendo r el número de taxones totales.

2. Se encuentra y se une al par de taxones cuya distancia en la matriz Q es menor. Se calcula la distancia entre ambos taxones y el nodo como:

$$d(f, u) = \frac{d(f, g)}{2} + \frac{1}{2(r - 2)} \left[\sum_{k=1}^r d(f, k) - \sum_{k=1}^r d(g, k) \right]$$

y

$$d(g, u) = d(f, g) - d(f, u)$$

donde f y g son los taxones cuya distancia es mínima.

3. Se calcula la distancia del resto de taxones respecto al nuevo nodo.

$$d(u, k) = \frac{1}{2} [d(f, k) + d(g, k) - d(f, g)]$$

donde u es el nuevo nodo.

4. Se repiten los pasos 1 y 2, considerando al par de taxones f y g como uno solo, con sucesivas iteraciones hasta que se define la distancia entre el último par de taxones.

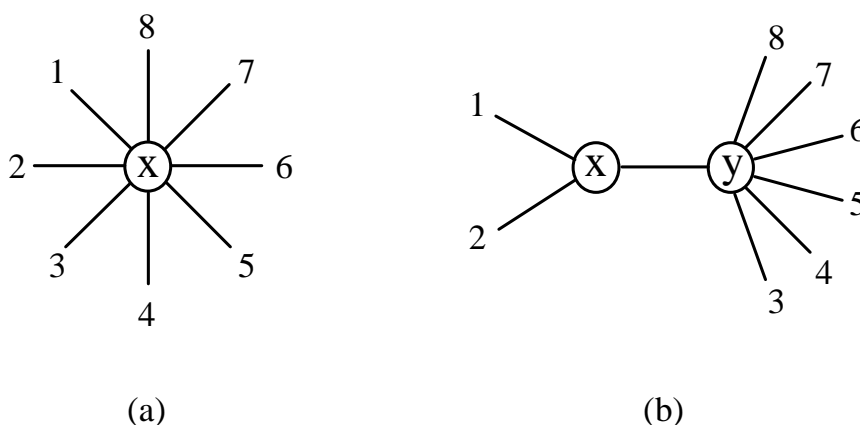


Figura 26. Ejemplo de aplicación del Neighbour-Joining en el que el fenograma inicial (a) es modificado (b) después de considerar las distancias entre las poblaciones 1 y 2 (Tomado de Nei y Kumar, 2000).

Para la realización de los diferentes dendrogramas se ha utilizado el programa Phylip v3.69 (Felsenstein, 1989). El programa TREEVIEW se ha empleado para la representación del dendrograma. Además se ha completado con el método bootstrap (Efron, 1979) para comprobar la robustez de los resultados obtenidos.

- **Test de Mantel.** Este método valora la significación de la correlación entre dos matrices. Dado que los elementos de una matriz son interdependientes, no es posible conocer la distribución esperada de probabilidades de un coeficiente de correlación calculado entre matrices. Por ello, mediante un procedimiento de permutación iterativo se obtiene una distribución empírica nula del coeficiente de correlación.

Dadas dos matrices cuadradas $X = \{x_{ij}\}$ e $Y = \{y_{ij}\}$ de dimensión N , su correlación puede calcularse como:

$$r_{XY} = \frac{SP(X, Y)}{\sqrt{SS(X) \cdot SS(Y)}}$$

donde $SP(X, Y)$ es el producto cruzado de X e Y y $SS(X) \cdot SS(Y)$ es el producto de la suma de los cuadrados.

El denominador de esta ecuación no se ve afectado por eventuales permutaciones, de tal manera que sólo el numerador cambiará. Se puede demostrar que la única cantidad que se verá modificada por intercambios de valores entre filas y columnas en una matriz es el producto:

$$Z_{XY} = X * Y = \sum_{i=1}^N \sum_{j=1}^i X_{ij} Y_{ij}$$

El procedimiento del test de Mantel aplicado a dos matrices consistirá en calcular la cantidad Z_{XY} de las matrices originales, permutar aleatoria y repetitivamente algunos valores de una matriz, siguiendo un patrón de remuestreo bootstrap, manteniendo la otra constante, obteniendo cada vez un nuevo valor de Z_{XY} , que será comparado con el original (Smouse *et al.* 1986).

En definitiva, se obtiene una serie de valores mediante las permutaciones que permite definir una distribución nula respecto de la que comparar el valor observado del producto de las matrices originales y obtener de esa forma una probabilidad asociada al coeficiente de correlación.

Para su realización se ha utilizado el programa Past v2.16 (Hammer *et al.* 2001).

- **Bootstrap.** Es un método no paramétrico de remuestreo propuesto por Efron (1979), con el fin de obtener medidas de la verosimilitud de determinados estimadores muestrales. Mediante este método se obtiene, de forma sencilla, una distribución muestral empírica de un estadístico, cuando no es posible conocerla de forma teórica.

Para obtener dicha distribución, se obtienen un número suficientemente amplio de conjuntos de datos, del mismo tamaño que el original, pero con reemplazamiento al azar de una parte de sus valores. Calculado el estadístico correspondiente, se ordenan los valores y se sitúa entre ellos el valor obtenido para el conjunto de datos original. La posición en que este quede es indicativa de su verosimilitud y puede expresarse en términos de probabilidad.

Se utiliza habitualmente para i) valorar el sesgo y el error muestral de un estadístico calculado a partir de una muestra, ii) establecer un intervalo de confianza para un parámetro estimado, o iii) realizar una prueba de hipótesis respecto a uno o más parámetros poblacionales.

En esta tesis se ha utilizado para valorar la verosimilitud de las diferentes ramas de los dendrogramas Neighbor-joining mediante el programa Phylip 3.69 (Felsenstein, 1989) y para estimar la probabilidad asociada al coeficiente de correlación del test de Mantel, mediante Past v2.16 (Hammer *et al.* 2001).

- **Análisis Discriminante.** Es un análisis estadístico multivariante cuya finalidad es evaluar si existen diferencias significativas entre dos o más grupos de objetos (poblaciones) respecto a un conjunto de variables presuntamente discriminantes medidas sobre dichos grupos. En el caso de que existan diferencias significativas, se trata de explicar en qué sentido se dan y proporcionar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados. Con este fin se diseñan una serie de funciones lineales (funciones discriminantes), construidas a partir de las variables originales que presentan un mayor poder de discriminación (Fisher, 1936; Rao, 1948).

Para su realización se ha utilizado los programas SPSS v18 y PAST (Hammer *et al.* 2001).



RESULTADOS

En este apartado se describen los trabajos realizados a lo largo del proyecto de Tesis Doctoral, siguiendo un orden cronológico de publicación. En cada caso la exposición consta de varias secciones: i) una breve descripción del artículo, enumerando los objetivos concretos del trabajo, ii) una copia de la publicación, iii) una pequeña discusión complementaria cuando se ha considerado conveniente y iv) en algún caso material complementario de interés para una mejor comprensión de los resultados.

Se incluyen los siguientes trabajos:

- I. Gómez-Pérez L, Alfonso-Sánchez MA, Pérez-Miranda AM, de Pancorbo MM, Peña JA. 2007. Utilidad de las inserciones *Alu* en los estudios de mestizaje. *Antropo* 14, 29-36.
- II. Gómez-Pérez L, Alfonso-Sánchez MA, Pérez-Miranda AM, García-Obregón S, Builes JJ, Bravo ML, De Pancorbo MM, Peña JA. 2010. Genetic admixture estimates by *Alu* elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia. *Ann Hum Biol* 37, 488-500.
- III. Gómez-Pérez L, Alfonso-Sánchez MA, Dipierri JE, Alfaro E, García-Obregón S, De Pancorbo MM, Bailliet G, Peña JA. 2011. Microevolutionary Processes due to Landscape Features in the Province of Jujuy (Argentina). *Am J Hum Biol* 23, 177-184.
- IV. Gómez-Pérez L, Alfonso-Sánchez MA, Sánchez D, García-Obregón S, Espinosa I, Martínez-Jarreta B, De Pancorbo MM, Peña JA. 2011. *Alu* Polymorphisms in the Waorani Tribe from the Ecuadorian Amazon Reflect the Effects of Isolation and Genetic Drift. *Am J Hum Biol* 23, 790-795.
- V. Gómez-Pérez L, Alfonso-Sánchez MA, Dipierri JE, Sánchez D, Espinosa I, De Pancorbo MM, Peña JA. 2013. Young *Alu* insertions within the MHC class I region in Native American populations: insights into the origin of the MHC-*Alu* repeats *Am J Hum Biol* 25, 359-365.

- VI. Gómez-Pérez L, Alfonso-Sánchez MA, Dipierri JE, Peña JA. 2015. Jujuy province (NW Argentina): STR loci unveil genetic footprints of microevolutionary responses to a steep orography. *Enviado*



MANUSCRITO I

Utilidad de las inserciones *Alu* en los estudios de mestizaje

Gómez-Pérez L, Alfonso-Sánchez MA, Pérez-Miranda AM, de Pancorbo MM, Peña JA. 2007.

Antropo, 14, 29-36.

INTRODUCCIÓN

El objetivo de este artículo ha consistido en valorar la aplicabilidad de las inserciones *Alu* en estudios de heterogeneidad genética intra e interpoblacional, principalmente en relación con su capacidad de discriminación entre poblaciones de diferentes continentes.

Puesto que el objetivo principal del proyecto de tesis ha sido valorar los efectos del flujo génico sobre el patrimonio genético en poblaciones sudamericanas con diferentes características, se ha tratado de caracterizar las inserciones *Alu* como marcadores adecuados para ello.

Con este fin, se han seleccionado las inserciones más utilizadas en estudios poblacionales y se han recopilado sus frecuencias en poblaciones de diferentes continentes. Casualmente, el menos representado fue el continente americano, con tan sólo 3 poblaciones, ya que si bien existían otras poblaciones incluidas en diferentes trabajos, tan sólo en los Muskokes, Mayas y Yanomamos se habían analizado las 8 inserciones seleccionadas, consideradas *a priori* como el mínimo de marcadores a emplear.



Utilidad de las inserciones *Alu* en los estudios de mestizaje

The usefulness of Alu insertions in admixture studies

Luis Gómez-Pérez¹, Miguel A. Alfonso-Sánchez¹, Ana M. Pérez-Miranda¹, M. M. de Pancorbo², Jose A. Peña¹

¹ Departamento de Genética, Antropología Física y Fisiología Animal, Facultad de Ciencia y Tecnología, Universidad del País Vasco, 48080 Bilbao, Spain.

² Departamento de Zoología y Dinámica Celular Animal, Facultad de Farmacia, Universidad del País Vasco, 48080 Bilbao, Spain

Dirección para correspondencia: Luis Gómez-Pérez. Departamento de Genética, Antropología Física y Fisiología Animal, Facultad de Ciencia y Tecnología, Universidad del País Vasco, Aptdo. 644, 48080 Bilbao, Spain. E-mail: luis.gomez@ehu.es

Palabras clave: Inserciones *Alu*, análisis discriminante, mestizaje.

Keywords: *Alu* insertions, discriminant analysis, admixture.

Resumen

Los polimorfismos de inserción *Alu* son unos marcadores muy interesantes para los estudios de la evolución humana, porque se producen por un único evento mutacional y se conoce su estado ancestral.

Sobre una base de datos de frecuencias génicas de 8 inserciones *Alu* en 47 poblaciones de diferentes continentes, se ha realizado un análisis discriminante, con el objetivo de evaluar la capacidad de estos marcadores para clasificar poblaciones y en consecuencia para estimar el grado de mestizaje de poblaciones híbridas. Para las 6 regiones establecidas, se ha obtenido un porcentaje de correcta clasificación del 91,5%, permitiendo distinguir razonablemente bien entre poblaciones del Africa subsahariana, Africa del Norte y Próximo Oriente, Europa, Medio Oriente, Asia y América. Se confirma por tanto la utilidad de las inserciones *Alu* para los estudios de mestizaje.

Abstract

Alu insertion polymorphisms are robust markers for human evolutionary studies because they have a unique mutational mechanism, with a known ancestral state.

On a data base of gene frequencies for 8 *Alu* insertions in 47 populations of different continents, a discriminant analysis has been performed, with the aim of evaluating the capacity of these markers to classify populations and consequently to estimate the degree of admixture of hybrid populations. For the 6 established regions, a percentage of correct classification of 91.5% has been obtained, allowing to distinguish reasonably well between populations of Sub-Saharan Africa, North Africa and Middle East, Europe, India and Central Asia, East Asia and America. The utility of the *Alu* insertions is confirmed therefore for the admixture studies.

Introducción

Los elementos móviles pueden constituir aproximadamente el 45% del genoma del hombre. Entre ellos, el grupo más abundante lo compone la familia de elementos *Alu*, que supone más del 10% de nuestro ADN con más de un millón de copias. Las inserciones *Alu* son marcadores bialélicos que se definen por la presencia o ausencia de unas secuencias de aproximadamente 300 pares de bases, que con variaciones más o menos importantes derivan del gen 7SL RNA (el cual forma parte del complejo ribosomal) por retrotransposición. Son especialmente interesantes en Antropogenética por aparecer hace unos 65 millones de años y encontrarse asociadas con el origen y expansión de los primates (Batzer y Deininger, 2002).

La mayor parte de las inserciones *Alu* son relativamente recientes, por lo que en general son polimórficas, pudiendo tomar diferentes valores de frecuencias en diferentes poblaciones humanas (Stoneking *et al.*, 1997).

Es altamente improbable que la misma inserción *Alu* ocurra dos veces independientemente en el mismo lugar, por lo que se considera que son eventos mutacionales únicos (Batzer *et al.*, 1994). Sin embargo, se han observado excepciones, como es el caso de la inserción PV92, en la que al menos algunos individuos del País Vasco y del Norte de África muestran una inserción dentro de una inserción (Comas *et al.*, 2001; García-Obregón *et al.*, 2007).

Por otra parte, salvo algunas excepciones, la selección no debiera tener una influencia considerable sobre las secuencias *Alu*, ya que no son codificantes. No obstante, en algunos casos pueden encontrarse fuertemente ligadas a un gen y verse afectadas por las presiones selectivas que afecten al mismo o pueden incluso determinar alteraciones en un gen próximo facilitando la aparición de enfermedades, como la neurofibromatosis tipo 1, el síndrome de Apert y varios tipos de cáncer (Wallace *et al.* 1991; Miki *et al.* 1996; Deininger y Batzer 1999; Oldridge *et al.* 1999).

En general, puede afirmarse que una de las principales características de las inserciones *Alu* es su estabilidad, debido a la escasa influencia de la mutación y de la selección sobre estos polimorfismos (García-Obregón *et al.*, 2007). Por ello, cabe esperar que a partir de un número limitado de inserciones *Alu* pueda obtenerse una buena discriminación entre poblaciones de diferentes continentes, lo que sustentaría el argumento de que estos marcadores son interesantes para los estudios de mestizaje.

En este trabajo se han recopilado las frecuencias de 8 inserciones *Alu* (A25, ACE, APO, B65, D1, FXIII B, PV92 y TPA25) en muestras representativas de diferentes continentes, con el objetivo de contrastar la capacidad de discriminación de estos marcadores y en consecuencia su interés para evaluar el grado de mestizaje de una población metamórfica.

Materiales y métodos

A partir de trabajos publicados previamente se han recopilado las frecuencias alélicas de 8 inserciones *Alu* (A25, ACE, APO, B65, D1, FXIII B, PV92 y TPA25) en 47 poblaciones de 4 continentes, que se distribuyeron en 6 regiones definidas con una cierta coherencia desde un punto de vista antropogenético: 1) África subsahariana, 2) Norte de África y Próximo Oriente, 3) Europa, 4) Medio Oriente, 5) Asia y 6) América (Tabla 1). La ausencia de poblaciones de Oceanía se explica por la inexistencia de muestras analizadas para las 8 inserciones en un número representativo. Por otra parte, América se encuentra representada por tan sólo 3 poblaciones, pero se ha considerado conveniente incluirlas, por ser el continente donde mayor número de poblaciones mestizas existen. La localización geográfica, las referencias de origen de los datos y el código en las figuras de las 47 muestras poblacionales se muestran en la tabla 1 y la figura 1.

Considerando las 6 regiones geográficas mencionadas, se analizó la heterogeneidad genética intra e interpoblacional mediante un análisis molecular de la varianza (AMOVA) con el programa Arlequin versión 3.0 (Excoffier *et al.*, 2005). Las distancias genéticas entre poblaciones se establecieron a partir del coeficiente R de Harpending y Jenkins (1973). Sobre la matriz de distancias se realizó un análisis de escalamiento multidimensional (MDS) (Kruskal 1964), con el paquete estadístico SPSS versión 13.0 (SPSS, Chicago, Ill.). Se obtuvieron 6 dimensiones, para cuyos vectores propios se contrastó la hipótesis de normalidad de su distribución, mediante el test de Kolmogorov-Smirnov. Finalmente, se realizó sobre dichos vectores un análisis discriminante, también con el programa SPSS.

Gómez-Pérez *et al.*, 2007. Antropo, 14, 29-36 www.didac.ehu.es/antropo

Muestra	Región	Referencia	Código
!Kung	Africa subsahariana	Stoneking <i>et al.</i> , 1997	1
Kenia (Bantús)	Africa subsahariana	Terreros <i>et al.</i> , 2005	2
Nguni	Africa subsahariana	Stoneking <i>et al.</i> , 1997	3
Nigeria	Africa subsahariana	Stoneking <i>et al.</i> , 1997	4
Pigmeos Biaka	Africa subsahariana	Bowcock <i>et al.</i> , 1987	5
Pigmeos Mbuti	Africa subsahariana	Bowcock <i>et al.</i> , 1987	6
Sotho-Tswana	Africa subsahariana	Stoneking <i>et al.</i> , 1997	7
Argelia	N. Africa y Próximo Oriente	Comas <i>et al.</i> , 2000	8
Marruecos (Norte)	N. Africa y Próximo Oriente	Comas <i>et al.</i> , 2000	9
Marruecos (Oeste)	N. Africa y Próximo Oriente	Comas <i>et al.</i> , 2000	10
Marruecos (Sudeste)	N. Africa y Próximo Oriente	Comas <i>et al.</i> , 2000	11
Sáhara Occidental	N. Africa y Próximo Oriente	Comas <i>et al.</i> , 2000	12
Túnez	N. Africa y Próximo Oriente	Comas <i>et al.</i> , 2000	13
Armenia	N. Africa y Próximo Oriente	Nasidze <i>et al.</i> , 2001	14
Azerbaiján	N. Africa y Próximo Oriente	Nasidze <i>et al.</i> , 2001	15
Emiratos Arabes	N. Africa y Próximo Oriente	Stoneking <i>et al.</i> , 1997	16
Siría	N. Africa y Próximo Oriente	Nasidze <i>et al.</i> , 2002	17
Turquía	N. Africa y Próximo Oriente	Romualdi <i>et al.</i> , 2004	18
Albania	Europa	Comas <i>et al.</i> , 2004	19
España (Andalucía)	Europa	Comas <i>et al.</i> , 2000	20
España (Cataluña)	Europa	Comas <i>et al.</i> , 2000	21
España (País Vasco)	Europa	García-Obregón <i>et al.</i> , 2007	22
España (Valencia)	Europa	García-Obregón <i>et al.</i> , 2006	23
Francia	Europa	Stoneking <i>et al.</i> , 1997	24
Grecia	Europa	Comas <i>et al.</i> , 2004	25
Macedonia	Europa	Comas <i>et al.</i> , 2004	26
Moldavia	Europa	Varzari <i>et al.</i> , 2007	27
Rumanía	Europa	Comas <i>et al.</i> , 2004	28
Suiza	Europa	Stoneking <i>et al.</i> , 1997	29
La India (Cristianos)	Medio Oriente	Stoneking <i>et al.</i> , 1997	30
La India (Hindús)	Medio Oriente	Stoneking <i>et al.</i> , 1997	31
La India (Musulmanes)	Medio Oriente	Stoneking <i>et al.</i> , 1997	32
Paquistán	Medio Oriente	Stoneking <i>et al.</i> , 1997	33
Pastunes	Medio Oriente	Stoneking <i>et al.</i> , 1997	34
Tamiles	Medio Oriente	Stoneking <i>et al.</i> , 1997	35
China (Han)	Asia	Xiao <i>et al.</i> , 2002	36
China (Sibo)	Asia	Xiao <i>et al.</i> , 2002	37
China (Uygures)	Asia	Xiao <i>et al.</i> , 2002	38
Filipinas	Asia	Stoneking <i>et al.</i> , 1997	39
Indonesia (Nusa Tenggara)	Asia	Stoneking <i>et al.</i> , 1997	40
Java	Asia	Stoneking <i>et al.</i> , 1997	41
Malasia	Asia	Stoneking <i>et al.</i> , 1997	42
Molucas	Asia	Stoneking <i>et al.</i> , 1997	43
Taiwan	Asia	Stoneking <i>et al.</i> , 1997	44
Mayas	América	Stoneking <i>et al.</i> , 1997	45
Muskokes	América	Stoneking <i>et al.</i> , 1997	46
Yanomamos	América	Romualdi <i>et al.</i> , 2004	47

Tabla 1. Procedencia de las muestras consideradas. El código numérico se corresponde con la Figura 1.**Table 1.** Location of the samples. The numerical code is the same in the Figure 1.

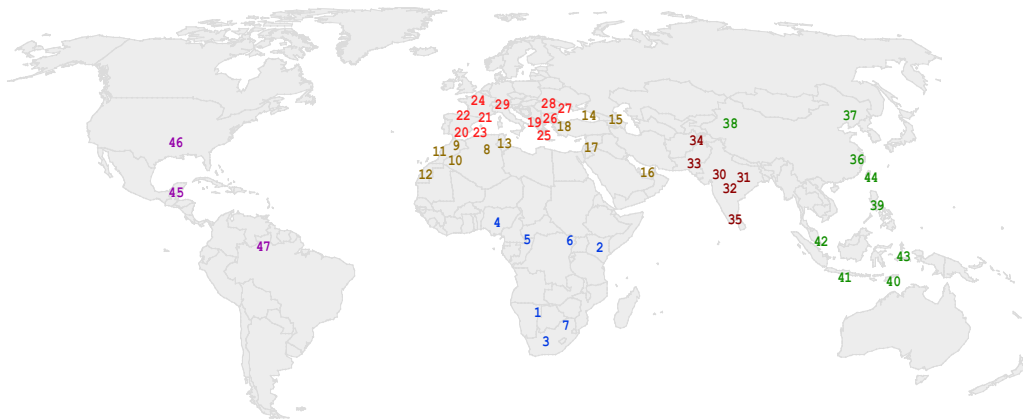


Figura 1. Origen de las muestras. El código numérico se corresponde con la Tabla 1.
Figure 1. Location of the samples. The numerical code is the same in the Table 1.

Resultados

En la tabla 2 se muestran los valores medios y las varianzas estandarizadas de las frecuencias de las 8 inserciones, calculados para el conjunto de poblaciones consideradas.

El valor más alto de frecuencia alélica media corresponde a APO y el más bajo a A25; el resto de inserciones muestran unos valores no muy dispares, dentro de un rango medio de frecuencias. Los valores más altos de F_{ST} , por el contrario, corresponden a FXIIIB y PV92 y los más bajos a TPA25, D1 y B65. Esta disparidad entre ambos estadísticos se pone de manifiesto al calcular el cociente entre la media y la varianza estandarizada, en el que se observan dos grupos de inserciones claramente diferenciados. Por un lado, A25, PV92 y FXIIIB, con unos valores del índice próximos a 1 (1,1, 1,6 y 1,7 respectivamente). Por otro, el resto de inserciones con valores claramente más altos (ACE: 4,4; D1: 4,6; B65: 6,7; APO: 7,8 y TPA25: 8,1).

Inserción	Posición	Media	F_{ST}
A25	8q21.3	0,146	0,132
ACE	17q23	0,453	0,104
APO	11q23-q24	0,870	0,112
B65	11q14.2	0,520	0,078
D1	3q26.32	0,297	0,064
FXIIIB	1q31-q32.1	0,473	0,284
PV92	16q24.2	0,396	0,244
TPA25	8p11.2	0,488	0,060

Tabla 2. Posición en el genoma, valor medio y varianza estandarizada de las frecuencias de las 8 inserciones *Alu* para las 47 poblaciones consideradas.

Table 2. Chromosome location, average frequencies and standardized variance of the 8 *Alu* insertions for the 47 populations.

Los resultados del análisis molecular de la varianza se muestran en la Tabla 3. Las 3 componentes del AMOVA resultaron significativas para las 8 inserciones. En relación a las diferencias entre las 6 regiones en que se agruparon las 47 poblaciones, el valor medio del porcentaje de la varianza entre grupos fue del 10,08%, con un valor mínimo del 1,33% para D1 y un valor máximo del 22,53% para PV92. Obviamente, los valores del porcentaje de la varianza entre grupos y de F_{ST} se encuentran claramente relacionados. Por otra parte, si se considera la varianza dentro de poblaciones, aquellas que presentan unos porcentajes más bajos son precisamente las que mostraron una relación más baja entre los valores medios y la varianza estandarizada de las frecuencias alélicas, aunque dicho resultado entra asimismo dentro de lo esperado.

Puesto que las frecuencias génicas no son variables adecuadas para realizar un análisis discriminante, se obtuvieron previamente unos vectores propios sobre los que aplicarlo. Con este fin, se calculó un coeficiente de distancia (distancia R de Harpending y Jenkins, 1973) entre cada

Gómez-Pérez *et al.*, 2007. Antropo, 14, 29-36 www.didac.ehu.es/antropo

par de poblaciones. Se añadió una población hipotética ancestral, en la que las frecuencias de todas las inserciones eran 0. Sobre la matriz de distancias se hizo un análisis de escalamiento multidimensional (MDS), obteniéndose 6 dimensiones, puesto que eran 6 los grupos poblacionales que se pretendía discriminar. Consecuentemente el estrés fue muy bajo, de tan sólo un 2,1% para una varianza explicada del 99,8%. Se realizó un test de Kolmogorov-Smirnov para comprobar la normalidad de los 6 vectores propios obtenidos, encontrándose una probabilidad no significativa en todos los casos (con unos valores del coeficiente Z de 0,923, 0,767, 0,595, 0,602, 0,668 y 0,886 respectivamente).

Inserción	Dentro de poblaciones		Entre poblaciones dentro de grupos		Entre grupos	
	Varianza (%)	ϕ_{ST}	Varianza (%)	ϕ_{SC}	Varianza (%)	ϕ_{CT}
A25	85,74	0,14**	6,37	0,07**	7,89	0,08**
ACE	88,36	0,12**	2,89	0,03**	8,75	0,09**
APO	87,71	0,12**	4,71	0,05**	7,58	0,08**
B65	91,78	0,08**	5,04	0,05**	3,18	0,03**
D1	93,53	0,06**	5,14	0,05**	1,33	0,01*
FXIIIB	68,72	0,31**	6,21	0,08**	25,07	0,25**
PV92	72,93	0,27**	4,54	0,06**	22,53	0,23**
TPA25	93,41	0,07**	2,27	0,02**	4,32	0,04**

Tabla 3. Análisis molecular de la varianza (AMOVA). ** p<0,001; * p<0,05

Table 3. Analysis of molecular variance (AMOVA). ** p<0,001; * p<0,05

Sobre los vectores propios se realizó un análisis discriminante. Siendo 6 los vectores, se obtuvieron 5 funciones, de las cuales las 2 primeras explicaron el 94,7% de la varianza.

En la Figura 2 se representan los coeficientes de las dos primeras funciones discriminantes en las 47 poblaciones analizadas y la población ancestral hipotética, que se utilizó como testigo. De una forma perfectamente coherente, las más cercanas a la población ancestral son las africanas. Entre éstas y las europeas se encuentran las muestras representativas del Norte de Africa y Próximo Oriente. Asimismo, en una zona intermedia entre las poblaciones africanas y asiáticas aparecen dispuestas las muestras del Medio Oriente. Por último, las poblaciones americanas se localizan en la proximidad de las asiáticas.

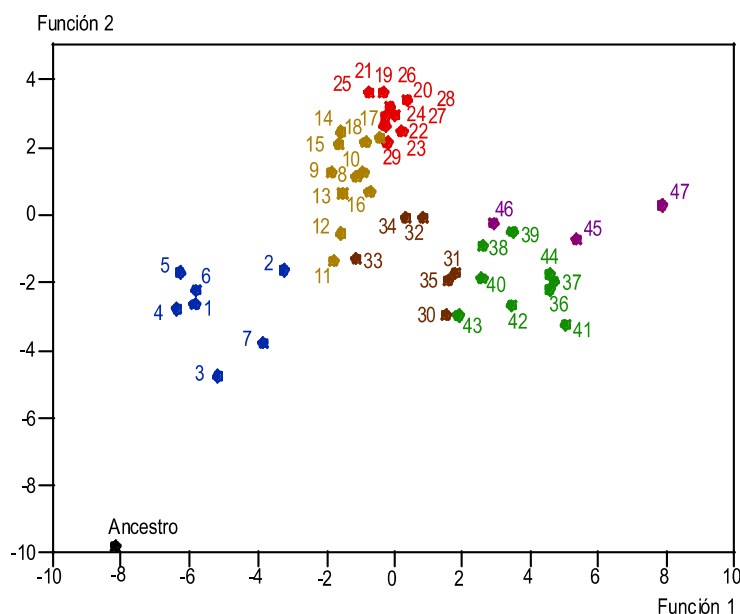


Figura 2. Representación de las 47 poblaciones comparadas, más la población ancestral hipotética, respecto a las 2 primeras funciones discriminantes.

Figure 2. Chart showing the 47 populations plus the hypothetical ancestral population, with respect to the 2 first discriminant functions.

El porcentaje de clasificación correcta considerando las 5 funciones discriminantes fue del 91,5%, quedando incorrectamente asignada una población asiática (Nusa Tenggara, de Indonesia), que se incluyó dentro del grupo de Medio Oriente y tres poblaciones del Próximo Oriente (Armenia, Siria y Turquía), que se incluyeron dentro del grupo europeo.

Discusión

Sobre una base de datos de 47 poblaciones de todo el mundo analizadas, para 8 inserciones *Alu*, se ha comparado la variabilidad genética observada entre inserciones y entre regiones continentales.

Teniendo en cuenta que, como se ha dicho, las inserciones *Alu* pueden ser consideradas como eventos mutacionales únicos y que no han podido ocurrir procesos de flujo génico y deriva diferentes entre las 8 inserciones, pues las poblaciones utilizadas han sido las mismas, es verosímil la hipótesis de que las diferencias observadas en la relación entre el valor medio y la varianza estandarizada de las frecuencias génicas puedan deberse a los efectos de presiones selectivas. Así, un valor medio más alto debiera implicar en general una mayor antigüedad de la inserción, por lo que debiera ir acompañado de un elevado valor de F_{ST} . El mayor valor de dicho índice se observa en TPA25, un marcador que se encuentra en las proximidades del gen que codifica para el activador tisular del plasminógeno (t-PA, del inglés *tissue-type Plasminogen Activator*), una proteasa que se produce especialmente en el cerebro y tiene que ver con la degeneración neuronal. También presenta un valor elevado APO, que se encuentra a 4,3 kb del gen que codifica para la Apolipoproteína A-I, implicada en el metabolismo del colesterol. En general, pueden encontrarse genes más o menos próximos a todas las inserciones *Alu* estudiadas y sin embargo, salvo casos muy concretos, no se han publicado evidencias de efectos de la selección sobre las frecuencias de inserciones *Alu*. Lamentablemente, desde un punto de vista metodológico el test de Ewens y Watterson no es informativo, por tratarse de marcadores dialélicos.

Por otra parte, la clasificación realizada por el análisis discriminante ha sido realmente efectiva, no sólo por obtenerse un porcentaje de asignación errónea muy bajo, sino porque la representación respecto de las 2 primeras funciones ha resultado ser una excelente descripción del proceso de dispersión del hombre moderno por los diferentes continentes. Todo ello obtenido tan sólo con 8 marcadores dialélicos.

En cuanto a las poblaciones incorrectamente clasificadas, la población de Nusa Tenggara, situada en el Extremo Sudeste de Indonesia aparece incluida dentro del grupo de Medio Oriente, junto a poblaciones de La India, Paquistán y Afganistán. Es una posición difícil de explicar, dado que es precisamente la población asiática más alejada geográficamente del Medio Oriente. Sin embargo, en el trabajo original realizado sobre esta población (Stoneking *et al.*, 1997), los resultados obtenidos en forma de dendrograma, son similares a los descritos aquí. Cabe pensar, por tanto, en un efecto de la deriva genética por sus características históricas y geográficas. Un posible efecto fundador durante el proceso de colonización de la zona, junto con la dispersión y el consiguiente aislamiento de las subpoblaciones, todo ello condicionado por un entorno salpicado de islas, podría haber determinado cambios al azar en el patrimonio genético de esta población.

La posición de Armenia, Siria y Turquía, clasificadas como poblaciones europeas, tiene su razón de ser si se tiene en cuenta que, como se observa en la Figura 2 (puntos 14, 17 y 18 respectivamente), su posición no se encuentra en el interior del grupo europeo, sino en sus proximidades, reflejando más bien una variación gradual en las frecuencias alélicas sin solución de continuidad entre Próximo Oriente y Europa.

Sea como fuere, se ha constatado la gran eficacia de las 8 inserciones *Alu* consideradas para la clasificación de poblaciones de diferentes continentes y por tanto para la estima de tasas de mestizaje en poblaciones mestizas.

Agradecimientos. Este trabajo ha sido financiado por el programa de subvenciones a Grupos de Investigación de la Universidad del País Vasco (GIU 05/51). El trabajo de Luis Gómez-Pérez ha sido financiado por una beca predoctoral de la Universidad del País Vasco.

Bibliografía

- Batzer, M.A., y Deininger, P.L., 2002, *Alu* repeats and human genomic diversity. *Nature Reviews Genetics*, 3,370-379
- Batzer, M.A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D., Shaikh, T., Novick, G., Ioannou, P., Scheer, D., Herrera, R., Deininger, P., 1994, African origin of human-specific polymorphic *Alu* insertions. *Proc. Natl. Acad. Sci. USA*, 91,12288–12292.
- Bowcock, A.M., Bucci, C., Hebert, J.M., Kidd, J.R., Kidd, K.K., Friedlaender, J.S., Cavalli-Sforza, L.L., 1987, Study of 47 DNA markers in five populations from four continents. *Gene Geogr.*, 1, 47-64.
- Comas, D., Calafell, F., Benchemsi, N., Helal, A., Lefranc, G., Stoneking, M., Batzer, M.A., Bertranpetit, J., Sajantila, A., 2000, *Alu* insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar straits. *Hum. Genet.*, 107, 312-319
- Comas, D., Plaza, S., Calafell, F., Sajantila, A., Bertranpetit, J., 2001, Recent insertion of an *Alu* element within a polymorphic human-specific *Alu* insertion. *Mol. Biol. Evol.*, 18,85–88
- Comas, D., Schmid, H., Braeuer, S., Flaiz, C., Busquets, A., Calafell, F., Bertranpetit, J., Scheil, H.G., Huckenbeck, W., Efremovska, L., Schmidt, H., 2004, *Alu* insertion polymorphisms in the Balkans and the origins of the Aromuns. *Ann. Hum. Genet.*, 68,120–127.
- Deininger, P.L. and Batzer, M.A., 1999, *Alu* repeats and human disease. *Mol. Genet. Metab.*, 67,183–193
- Excoffier, L., Laval, G., Schneider, S., 2005, ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis, *Evol Bioinf Online*, 1,47–50
- García-Obregón, S., Alfonso-Sánchez, M.A., Pérez-Miranda, A.M., de Pancorbo, M.M., Peña, J.A., 2007, Polymorphic *Alu* insertions and the genetic structure of Iberian Basques. *J. Hum. Genet.*, 52,317-327
- García-Obregón, S., Alfonso-Sánchez, M.A., Pérez-Miranda, A.M., Vidales, C., Arroyo, D., Peña, J.A., 2006, Genetic position of Valencia (Spain) in the Mediterranean basin according to *Alu* insertions. *Am. J. Hum. Biol.*, 18,187–195
- Harpending, H., Jenkins, T., 1973, Genetic distance among southern African populations. En *Methods and Theories of Anthropological Genetics*, editado por M. Crawford y P. Workman (Albuquerque: University of México Press) pp. 177–99.
- Kruskal, J.B., 1964, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika*, 29,1–27
- Maca-Meyer, N., Villar, J., Pérez-Méndez, L., de León, A. Cabrera, Flores, C., 2004, A Tale of Aborigines, Conquerors and Slaves: *Alu* Insertion Polymorphisms and the Peopling of Canary Islands. *Annals of Human Genetics*, 68, 600-605
- Miki, Y., Katagiri, T., Kasumi, F., Yoshimoto, T., and Nakamura, Y., 1996, Mutation analysis in the BRCA2 gene in primary breast cancers. *Nat. Genet.*, 13,245–247
- Nasidze, I., Risch, G.M., Robichaux, M., Sherry, S.T., Batzer, M.A., Stoneking, M., 2001, *Alu* insertion polymorphisms and the genetic structure of human populations from the Caucasus, *Eur. J. Hum. Genet.*, 9,267-72
- Oldridge, M., Zackai, E.H., McDonald-McGinn, D.M., Iseki, S., Morriss-Kay, G.M., Twigg, S.R., Johnson, D., Wall, S.A., Jiang, W., Theda, C., *et al.*, 1999, De novo *Alu*-element insertions in FGFR2 identify a distinct pathological basis for Apert syndrome. *Am. J. Hum. Genet.*, 64,446–461
- Romualdi, C., Balding, D., Nasidze, I.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M.A., Barbujani, G., 2002, Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms, *Genome Research*, 12,602-612
- Stoneking, M., Fontius, J., Clifford, S., Soodyall, H., Arcot, S., Saha, N., Jenkins, T., Tahir, M., Deininger, P., Batzer, M.A., 1997, *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.*, 7,1061–1071
- Terreros, M.C., Martinez, L., Herrera, R.J., 2005, Polymorphic *Alu* insertions and genetic diversity among African populations, *77*, 675-704.
- Varzari, A., Stephan, W., Stephanov, V., Raicu, F., Cojocaru, R., Roschin, Y., Glavce, C., Dergachev, V., Spiridonova, M., Schmidt, H.D., Weiss, E., 2007, Population history of the

Gómez-Pérez *et al.*, 2007. *Antropo*, 14, 29-36 www.didac.ehu.es/antropo

Dniester–Carpathians: evidence from *Alu* markers. *J. Hum. Genet.*, 52, 308-316

Xiao, F.X., Yang, J.F., Cassiman, J.J., Decorte, R., 2002, Diversity at eight polymorphic *Alu* insertion loci in Chinese populations shows evidence for European admixture in an ethnic minority population from northwest China, *Hum. Biol.*, 74,555-68

Wallace, M.R., Andersen, L.B., Saulino, A.M., Gregory, P.E., Glover, T.W., y Collins, F.S., 1991, A de novo *Alu* insertion results in neurofibromatosis type 1. *Nature*, 353, 864–866

PRINCIPALES RESULTADOS Y MATERIAL COMPLEMENTARIO

En este trabajo se ha podido concluir que los marcadores evaluados son adecuados para el análisis del mestizaje, ya que con 8 inserciones *Alu* se ha obtenido una clasificación correcta del 91,5% de las poblaciones mediante un análisis discriminante.

Tan sólo 4 poblaciones se han clasificado de forma errónea; entre ellas se encuentran 3 poblaciones de Próximo Oriente, una región en la que se observa un alto grado de solapamiento con el patrimonio genético de las poblaciones europeas por tres motivos principales: i) proximidad geográfica, de modo que desde un punto de vista antropológico se podría considerar la misma región (como el Norte de África); ii) La influencia de las poblaciones del Próximo Oriente en el poblamiento de Europa durante el neolítico (Ammerman y Cavalli-Sforza, 1971, 1973, 1984; Richards *et al.* 2000; Torrioni *et al.* 2001; Ammerman y Biagi, 2003; Richards, 2003; Sjödin y François, 2011; Cardoso *et al.* 2013b); iii) El posterior intercambio genético, a través de movimientos migratorios de diferentes civilizaciones en los que el Mediterráneo ha representado un papel importante (Martin, 1992; García-Obregón *et al.* 2006; Soares *et al.* 2010; Sjödin y François, 2011). Entre estas poblaciones se encuentra Turquía, una población situada en la frontera entre Europa y Asia, con una parte de su territorio en cada continente, comportándose como un puente para los movimientos de poblaciones humanas, incluso presumiblemente como un refugio durante el LGM (Cinnioğlu *et al.* 2004; Berkman *et al.* 2008).

Como se ha mencionado anteriormente, la presencia de poblaciones americanas en este trabajo es escasa, incluyendo tan sólo Muskokes, Mayas y Yanomamos. Puesto que desde la elaboración del manuscrito han aparecido frecuencias de otras poblaciones de este continente, algunas de ellas fruto de este proyecto, se ha realizado un nuevo análisis con el fin de observar la disposición del conjunto de poblaciones americanas con una más adecuada representación. Se han incluido asimismo dos poblaciones de inuits que inicialmente se desecharon por una cuestión de

representatividad. Las poblaciones añadidas, con sus correspondientes códigos se muestran en la *Figura 27*.

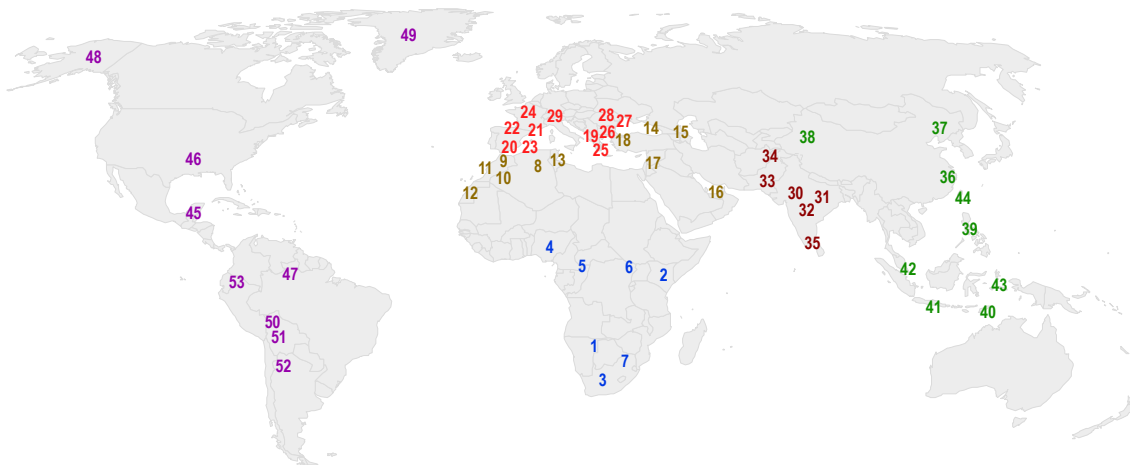


Figura 27. En este mapa se muestra la localización de las muestras empleadas. Actualización (*Figura 1, Manuscrito I*) donde además de las utilizadas en el manuscrito original, se han añadido Inuits de Alaska (48) y de Groenlandia (49) (Stoneking *et al.* 1997) Quechuas (50) y Aymaras (51) de Bolivia (Gaya-Vidal *et al.* 2010), Jujeños de Argentina (52) (Gomez-Pérez *et al.* 2011a) y Huaoranis de Ecuador (53) (Gomez-Pérez *et al.* 2011b).

El resultado obtenido explica el 93,8% de la varianza. Mediante las nuevas funciones discriminantes se clasifican correctamente 49 poblaciones de las 53 incluidas (un 92,5%). Aparecen mal clasificadas las mismas poblaciones que en el trabajo original (Nusa Tenggara, Armenia, Siria y Turquía).

Por lo tanto se corroboran los resultados obtenido para el artículo con una mayor presencia de poblaciones americanas.

Todas las poblaciones americanas se han clasificado correctamente, formando un grupo con un alto grado de dispersión, como se observa en la *Figura 28*. A pesar de esta dispersión, aparecen relaciones con el grupo asiático, de acuerdo a la teoría ampliamente reconocida del origen asiático de los nativos americanos (Fagundes *et al.* 2008a; Goebel *et al.* 2008; Bourgeois *et al.* 2009; Perego *et al.* 2009; 2010; Schroeder *et al.* 2009). Las poblaciones procedentes de Sudamérica aparecen separadas de las norteamericanas, que quedan más próximas al grupo asiático, lo que se podría

relacionar con intercambios posteriores a la migración principal que habría dado lugar a los amerindios (Reich *et al.* 2012; Raghavan *et al.* 2014).

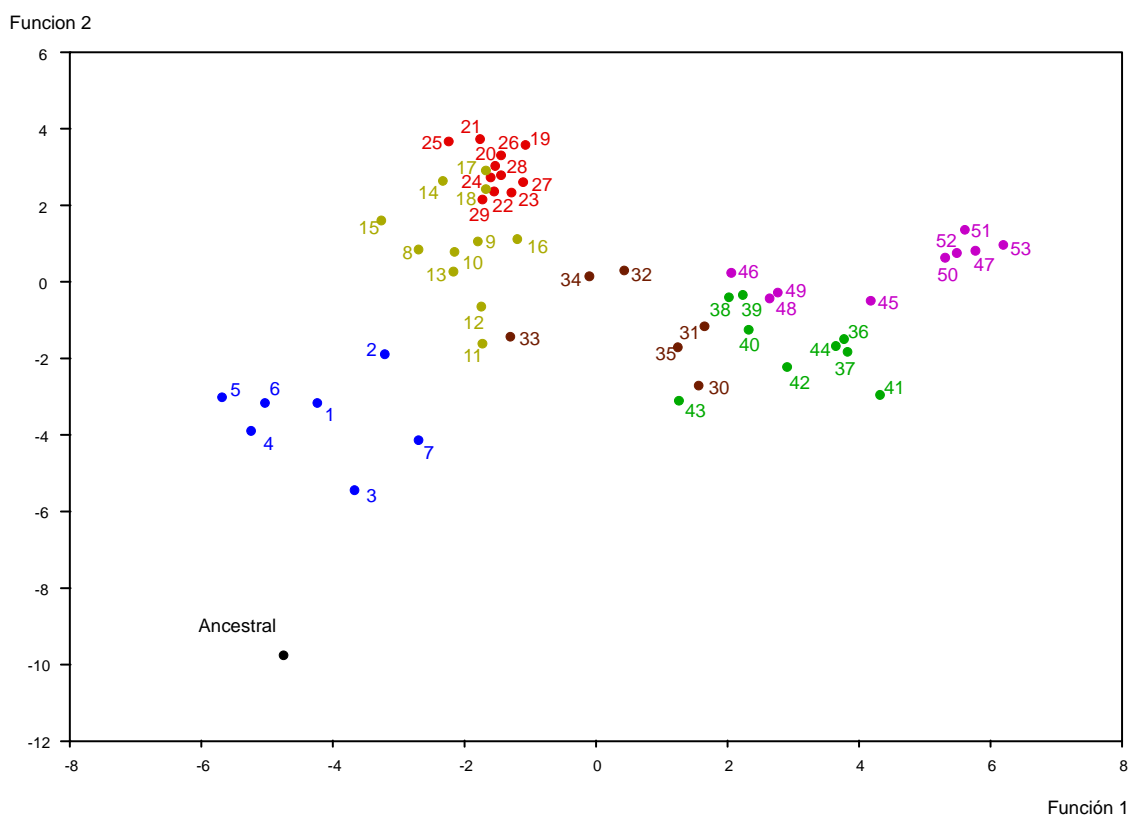
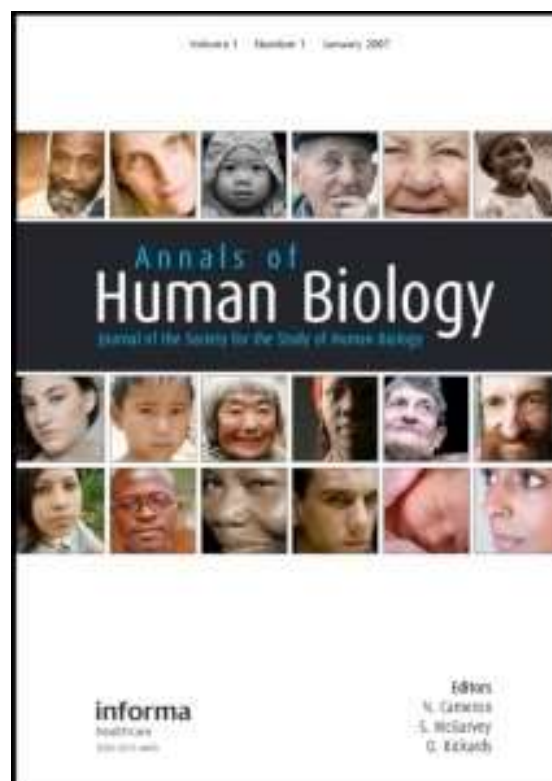


Figura 28. Gráfico de dispersión de las dos primeras funciones discriminantes recalculadas para los dos vectores principales de las 53 poblaciones empleadas obtenidos de un análisis MDS. Los códigos se corresponden con la descripción de la Figura 27 (Leyenda en Tabla 1, Manuscrito I; Figura 27).



MANUSCRITO II

Genetic admixture estimates by *Alu* elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia

Gómez-Pérez L, Alfonso-Sánchez MA, Pérez-Miranda AM, García-Obregón S, Builes JJ, Bravo ML, De Pancorbo MM, Peña JA. 2010.

Annals of Human Biology 37, 488-500

INTRODUCCIÓN

Una vez comprobada la utilidad de las inserciones *Alu* para los estudios de mestizaje, se ha procedido a analizar el grupo de inserciones *Alu* seleccionadas en el trabajo anterior (A25, ACE, APO, B65, D1, FXIII B, PV92 y TPA25) en diferentes poblaciones. En este caso, se incluyeron dos muestras procedentes de Antioquia, en Colombia (afroamericanos y mestizos), con el fin de estimar su grado de mestizaje.

Asimismo, se ha evaluado la capacidad de este grupo de inserciones *Alu* como marcadores informativos de ancestralidad (*AIMs*). Los *AIMs* pueden ser utilizados para identificar el origen de una población o un individuo. Se ha pretendido detectar la existencia de diferencias significativas entre las frecuencias alélicas en poblaciones de diferentes continentes con el objetivo de completar los paneles existentes.

Se han caracterizado específicamente estas dos poblaciones por ser representativas de la población actual de Colombia, cuyo patrimonio genético se ha definido a partir de las poblaciones originales de amerindios y los posteriores colonizadores europeos, junto a los esclavos africanos que trajeron consigo. Los mestizos son el componente mayoritario de la población, en tanto que los afrocolombianos son la principal minoría. La muestra de mestizos analizada procedía de Medellín (departamento de Antioquia) en tanto que la de afrocolombianos fue recogida en diferentes regiones del departamento de Antioquia colindantes con el departamento adyacente de Chocó.

Por definición, los mestizos son el resultado de emparejamientos entre nativos amerindios y europeos, así como con otros mestizos. A pesar de ello, en su patrimonio genético también puede encontrarse una proporción variable de genes de origen africano.

Los afrocolombianos, por su parte, son los descendientes de individuos africanos llevados para trabajar como esclavos en las haciendas de los europeos. Probablemente muestren un cierto

grado de mestizaje entre las diferentes etnias africanas de las que procedían, así como con europeos y en menor medida, con amerindios. Por su condición, sus duras circunstancias de vida y su marginación, incluso tras la abolición de la esclavitud, una parte de la población afrocolombiana evitó el contacto, especialmente con los europeos. De hecho, una parte de los afrocolombianos, compuesta por los esclavos fugados (cimarrones) crearon sus propias comunidades, denominadas palenques o quilombos, donde durante generaciones se mantuvieron aislados de los colonizadores. Con el fin de recuperar a los esclavos fugados, en algunas ocasiones se dieron enfrentamientos entre los colonos y los cimarrones, por lo que muchos de estos poblados adquirieron características de fortalezas, construidas en lugares poco accesibles, aisladas y amuralladas.

ORIGINAL ARTICLE

Genetic admixture estimates by *Alu* elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia

LUIS GÓMEZ-PÉREZ¹, MIGUEL A. ALFONSO-SÁNCHEZ^{1,2},
ANA M. PÉREZ-MIRANDA¹, SUSANA GARCÍA-OBREGÓN¹,
JUAN J. BUILES³, MARIA L. BRAVO³, MARIAN M. DE PANCORBO^{2,4} &
JOSÉ A. PEÑA¹

¹Departamento de Genética, Antropología Física y Fisiología Animal, Universidad del País Vasco (UPV/EHU), Bilbao, Spain, ²Servicio de Investigación Genómica: Banco de ADN, Universidad del País Vasco (UPV/EHU), Vitoria, Spain, ³GENES Ltda, Laboratorio de Genética Forense y Huellas Digitales del DNA, Medellín, Colombia, and ⁴Departamento de Zoología y Biología Celular Animal, Universidad del País Vasco (UPV/EHU), Vitoria, Spain

(Received 31 October 2008; accepted 12 October 2009)

Abstract

Aim: This work was intended to gain insights into the admixture processes occurring in Latin American populations by examining the genetic profiles of two ethnic groups from Antioquia (Colombia).

Subjects and methods: To analyse the genetic variability, eight *Alu* insertions were typed in 64 Afro-Colombians and a reference group of 34 Hispanics (Mestizos). Admixture proportions were estimated using the Weighted Least Squares and the Gene Identity methods. The usefulness of the *Alu* elements as Ancestry Informative Markers (AIMs) was evaluated through differences in weighted allelic frequencies (δ values) and by hierarchical analysis of the molecular variance (AMOVA).

Results: The Afro-Colombian gene pool was largely determined by the African component (88.5–88.8%), but the most prominent feature was the null contribution of European genes. Mestizos were characterized by a major European component (60.0–63.8%) and a comparatively low proportion of Amerindian (19.2–20.7%) and African (17.0–19.3%) genes. Five of the *Alu* loci examined (ACE, APO, FXIIIIB, PV92 and TPA25) showed an adequate resolving power to differentiate between continental groups, as indicated by δ values and AMOVA results.

Conclusions: The peculiarity of the Afro-Colombian gene pool seems to be associated with intense genetic drift episodes that occurred in isolated communities founded by small groups of runaway slaves. ACE, APO, FXIIIIB, PV92 and TPA25 could be efficiently utilized in studies dealing with demographic history and biogeographical ancestry in human populations.

Keywords: Genetic admixture, *Alu* insertions, Ancestry Informative Markers, Colombia, Afro-Colombian

Correspondence: Dr José A. Peña, Departamento de Genética, Antropología Física y Fisiología Animal, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain. E-mail: joseangel.pena@ehu.es

ISSN 0301-4460 print/ISSN 1464-5033 online © 2010 Informa UK Ltd.
DOI: 10.3109/03014460903433810

Introduction

In Colombia, the arrival of Spanish colonists started at the beginning of the 16th century. The genetic influence of South-Saharan Africa was present from the very earliest stage of the colonization process. Between approximately 1620 and 1850, several millions of indigenous West and Central Africans from coastal regions were kidnapped and transported to work as slaves in the Americas (Curtin 1969). Afro-Colombians represented 5.0% of the total Colombian population in 1778, 4.4% in 1827, 6.4% in 1912 and 10.6% in the most recent demographic census (year 2005). Traditionally, the largest proportion of the population in Colombia has been of European-derived and Mestizo people, constituting 80.0% of the total census in 1778, 87.0% in 1827, 86.8% in 1912 and 86.0% in 2005. The corresponding figures for Native Americans are 15.0% (1778), 8.6% (1827), 6.8% (1912) and 3.4% in 2005 (Departamento Administrativo Nacional de Colombia, <http://www.dane.gov.co/censo/>).

The current population of the department of Antioquia (north-western Colombia) is the result of the mixture of some hundreds of Native Americans with Europeans, mostly Spanish immigrants, and an indeterminate number of African slaves (Carvajal-Carmona et al. 2000; Bedoya et al. 2006). According to data from the most recent demographic census (2005), 88.7% of the present-day Antioquia population corresponds to Mestizos and 10.8% to Afro-Colombians, whereas Amerindians represents only a 0.5% of the total population.

A number of investigations have focused on the admixture events that occurred in Antioquia. All of them found a major European genetic contribution (60–75%), whereas the Native American and African components are much less important (Sandoval et al. 1993; Bedoya et al. 2006; Wang et al. 2008). All the cited investigations centred on the Mestizo population, while admixture proportions in Afro-Colombians remain relatively unknown, despite of being the principal minority of Antioquia.

Alu insertions are estimated to represent around 11% of the human genome, being present in an excess of 1 000 000 copies per haploid genome with an average distribution of one copy every four kilobases (Lander et al. 2001). *Alu* insertions are characterized by a dimeric pattern of insertion–lack of insertion, lack of the insertion being the ancestral state. Because the probability of two independent *Alu* elements being inserted by chance in the same genomic position is practically zero, the *Alu* insertions shared by different individuals should be identical by descent (Batzer et al. 1994; Roy-Engel et al. 2002). *Alu* elements are considered to be selectively neutral, so that their frequency distribution in human populations will be determined only by random processes (genetic drift), as well as gene flow between populations (Comas et al. 2000). Furthermore, the extremely low probability that a complete deletion of an *Alu* element would occur is another advantage of these molecular markers over other polymorphisms. All the above-mentioned properties of the *Alu* insertions (known ancestral state, identity by descent, selective neutrality and polymorphism stability) make them efficient and reliable markers for population genetic studies associated with the reconstruction of human evolutionary history (Batzer et al. 1994; Novick et al. 1998; Batzer and Deininger 2002; Herrera et al. 2007).

The present work was mainly focused on the genetic characterization of a sample of Afro-Colombians from Antioquia based on the analysis of eight polymorphic *Alu* insertions, to gain insights into the admixture processes that occurred in South America from the 17th century onward. To that end, we constructed an *Alu* database including representative populations of three different continents from previously published *Alu* frequencies, which was also utilized for phylogenetic analysis and for evaluating the potential usefulness of the *Alu* loci examined as Ancestry Informative Markers (AIMs). In order to complement the

490 L. Gómez-Pérez et al.

information provided by the Afro-Colombian gene pool, a Mestizo sample from the targeted region was also included in the study as a reference group.

Materials and methods

The Department of Antioquia is located in North West Colombia (Figure 1). From the geographical viewpoint, Antioquia is 80% mountainous, due to the presence of the Andes mountain range.

With a view to analysing the genetic variability of two of the main ethnic groups of Antioquia, eight *Alu* insertions (A25, ACE, APO, B65, D1, FXIIB, PV92 and TPA25) were typed in a sample of 98 unrelated and healthy individuals: 64 Afro-Colombian and 34 Mestizos. Voluntary donors were classed depending on their morphological phenotype, following the same criteria applied in the most recent demographic census, in the year 2005 (Departamento Administrativo Nacional de Colombia, <http://www.dane.gov.co/censo/>). The birthplace of an individual, surveyed up to the third generation back (the grandparents), was the criterion to decide whether a person was native or not from Antioquia. Ethical guidelines for research with human beings were adhered to as stipulated by each of the institutions involved in the study (Universidad del País Vasco, Spain, and Laboratorio de Genética Forense y Huellas Digitales del DNA, Medellín, Colombia). The study protocol was approved by the Institutional Review Board from Universidad del País Vasco.



Figure 1. Map showing the geographic location of the study area (Antioquia, Colombia).

Genomic DNA was extracted from blood stains using a Qiagen kit (QIAmp DNA Micro Kit) and stored at -20°C . PCRs were performed in a final volume of $10\ \mu\text{L}$, starting from $20\ \text{ng}$ of DNA. The PCR mix was composed of 1X PCR buffer ($50\ \text{mM}$ of KCl , $1.5\ \text{mM}$ of MgCl_2 , $10\ \text{mM}$ of Tris-HCl , and 0.01% gelatin), $1.25\ \text{mM}$ of MgCl_2 , $1\ \text{mM}$ of each dNTP, $0.50\ \mu\text{M}$ of each primer, Glycerol 5% , $5\ \text{U}$ Taq Polymerase and ultrapure H_2O . The PCR amplification process was carried out in a GeneAmp PCR System 9700 Thermal Cycler (PE Applied Biosystems, Norwalk, CT, USA). The specific amplification conditions, annealing temperatures and primers for each *Alu* insertion can be found in previous papers (García-Obregón et al. 2006, 2007). All PCR products were analysed by electrophoresis in 1.5% agarose gels. DNA bands were visualized by staining with ethidium bromide ($0.5\ \mu\text{L}/\text{mL}$) and photographed under UV light. To control the quality of the PCR and electrophoresis we included a positive control (homozygote for the insertion) as well as a negative control containing all the components of the PCR mix except the template DNA. Allelic identification was performed using a Roche DNA Molecular Weight Marker V ladder ($8\text{--}587\ \text{bp}$).

Statistical analysis

Allelic frequencies and gene diversity for the eight *Alu* elements analysed were calculated using version 3.0 of the Power Marker program (Liu and Muse 2004). To test for Hardy–Weinberg equilibrium (HWE), a Fisher’s exact probability test was conducted to estimate p -values (Guo and Thompson 1992) using Arlequin version 3.11 (Excoffier et al. 2005).

The genetic affinities of the study samples with populations worldwide were explored by computing Reynolds F_{ST} genetic distances (Reynolds et al. 1983) between pairs of populations with the Phylip v3.2 program (Felsenstein 1989). To that end, *Alu* frequency data were compiled from previously published works (Table I). To determine the contribution of each individual *Alu* loci to the genetic similarity or divergence between population groups and evaluate the effectiveness of the *Alu* elements screened as AIMs, we firstly estimated weighted *Alu* insertion frequencies (i.e. mean frequencies weighted by sample size) for three geographical groups taken as reference in the admixture analysis (Native American, South-Saharan Africa and Europe). These weighted frequencies were then used to estimate the parameter δ between population clusters. For dimorphic markers, $\delta = |p_1 - p_2|$, where p_1 and p_2 are the weighted frequencies of the insertion in populations groups 1 and 2, respectively. Previous studies have established the threshold δ as a frequency differential of 0.30 (Bonilla et al. 2004; Luizon et al. 2008). In addition, to ascertain the fraction of the genetic variability due to differences within and between human groups, genetic variance was hierarchically apportioned through the analysis of molecular variance, AMOVA using the Arlequin program (Excoffier et al. 2005). The AMOVA tests were performed by considering the three major geographic population clusters mentioned above.

Admixture proportions in the gene pool of the Afro-Colombian and Mestizo samples from Antioquia were estimated through the Weighted Least Squares method (Long et al. 1991), and the obtained results were further corroborated using the Gene Identity method (Chakraborty 1985). μ is the proportionate contribution of the j -the reference gene pool to the hybrid population, so that:

$$\sum_{j=1}^J \mu_j = 1.$$

For comparative purposes, admixture proportions estimated for both the samples examined herein and other hybrid populations of the Americas were represented by means of a Ternary Plot, using the PAST program package (Hammer et al. 2001).

Table I. Populations represented in the multidimensional scaling (MDS) analysis.

Geographic region	Population	Reference
South-Saharan Africa	Bantu	Romualdi et al. (2002)
	Congolese (Biaka)	Stoneking et al. (1997)
	Nguni	Stoneking et al. (1997)
	Northeast Zaire (Mbuti)	Stoneking et al. (1997)
	Sotho-Tswana	Stoneking et al. (1997)
	!Kung-San	Stoneking et al. (1997)
North Africa	Algeria	Comas et al. (2000)
	Northern Morocco	Comas et al. (2000)
	Southeast Morocco	Comas et al. (2000)
	West Morocco	Comas et al. (2000)
	Sahara	Comas et al. (2000)
	Tunisia	Comas et al. (2000)
Europe	Andalusia	Comas et al. (2000)
	Catalonia	Comas et al. (2000)
	Albania	Comas et al. (2004)
	Greece	Comas et al. (2004)
	Macedonia	Comas et al. (2004)
	Valencia	García-Obregón et al. (2006)
	Basques	García-Obregón et al. (2007)
	Canary Isles	Maca-Meyer et al. (2004)
	France	Romualdi et al. (2002)
	Swiss	Stoneking et al. (1997)
North America	Mvskoke	Stoneking et al. 1997
Central America	Mayan	Stoneking et al. 1997
South America	Ayoreo (average frequencies)	Dornelles et al. (2004)
	Ache	Battilana et al. (2006)
	Cinta Larga	Battilana et al. (2006)
	Gaviaõ	Battilana et al. (2006)
	Guarani	Battilana et al. (2006)
	Kaigang	Battilana et al. (2006)
	Quechua (Arequipa region)	Battilana et al. (2006)
	Quechua (Tayacaja province)	Battilana et al. (2006)
	Wai-wai	Battilana et al. (2006)
	Xavante	Battilana et al. (2006)
	Zoro	Battilana et al. (2006)
Yanomamo	Romualdi et al. (2002)	

Results

Allelic frequencies (\pm standard errors, SE) and gene diversity (GD) values for the eight *Alu* insertions examined in the Afro-Colombian and Mestizo samples from Antioquia are listed in Table II. All *Alu* markers proved to be polymorphic. In the Afro-Colombian collection, allelic frequencies oscillated between 0.154 (D1) and 0.637 (APO). The highest *Alu* frequency in Mestizos was also detected for the APO locus (0.924), whereas the lowest was observed in A25 (0.088). Gene diversity values varied between 0.260 (D1) and 0.499 (B65) in the Afro-Colombian sample, and between 0.140 and 0.499 (APO and ACE, respectively) in Mestizos. The great majority of the *Alu* loci showed no significant departure

Admixture rates from *Alu* insertions in Colombia 493Table II. *Alu* insertion frequencies with standard errors (\pm SE) and gene diversity (GD) in Afro-Colombian and Mestizo samples from Antioquia.

<i>Alu</i> marker	Population					
	Afro-Colombian Frequency			Mestizo Frequency		
	2N*	(\pm SE)	GD	2N*	(\pm SE)	GD
A25	122	0.213 (\pm 0.037)	0.335	68	0.088 (\pm 0.030)	0.161
ACE	104	0.442 (\pm 0.044)	0.493	68	0.485 (\pm 0.053)	0.499
APO	124	0.637 (\pm 0.044)	0.462	66	0.924 (\pm 0.029)	0.140
B65	124	0.524 (\pm 0.046)	0.499	66	0.591 (\pm 0.063)	0.483
D1	104	0.154 (\pm 0.042)	0.260	66	0.227 (\pm 0.053)	0.351
FXIIIB	112	0.232 (\pm 0.041)	0.356	68	0.441 (\pm 0.057)	0.493
PV92	124	0.315 (\pm 0.037)	0.431	68	0.368 (\pm 0.064)	0.465
TPA25	126	0.333 (\pm 0.038)	0.444	66	0.561 (\pm 0.052)	0.493

*2N, sample size in number of chromosomes analysed.

from HWE expected proportions, excepting D1 in Afro-Colombian ($p < 0.001$) and PV92 in Mestizos ($p < 0.05$). After applying the conservative Bonferroni correction for multiple testing, the Bonferroni-corrected significance threshold (5%) rendered a p value of 0.00313; therefore, only D1 did not meet HWE expectations.

The multidimensional scaling (MDS) (coefficient of stress: 5.4%) was largely determined by geography (Figure 2), in accord with previous human diversity studies based on *Alu* insertions (Stoneking et al. 1997; Antunez-de-Mayolo et al. 2002; Cotrim et al. 2004). Roughly, three main population clusters seem to be generated by the two-dimensional representation of the genetic distance matrix: (i) Native American groups all plotted in the negative segment of dimension I, without any definite pattern, (ii) a second major cluster showed the European populations tightly grouped in the upper right quadrant of the MDS plot, defined by the positive segments of both dimensions. In the same quadrant, North African populations (Sahara, Tunisia, Algeria and the different Moroccan samples) appeared at an intermediate position between Europe and South-Saharan Africa, and (iii) a population clade in the lower right quadrant formed by samples from South-Saharan Africa.

Regarding the two populations under study, Afro-Colombian (ACOL) segregated close to other collections from South-Saharan Africa, which suggests a high degree of genetic kinship between them and, at the same time, a low level of admixture with other human groups. Mestizos (MEST) plotted together with European populations, but very close to the centroid of the two-dimensional representation, perhaps reflecting genetic admixture processes with both Native American and African groups.

To determine which *Alu* elements contribute the most to the genetic heterogeneity revealed by the MDS plot, weighted allele frequencies and their standard errors were estimated for the eight *Alu* loci in the three geographical clusters taken as reference groups from previously published data. Likewise, differences in weighted frequencies (δ values) between the European (EUR), South-Saharan Africa (AFR) and Native American (NAM) groups were also computed (Table III). A total of nine comparisons showed $\delta \geq 0.30$ (the threshold value established to differentiate population groups), in seven (77.7%) of which the Native American cluster was involved. Average δ values (not shown) for the comparisons EUR–NAM, EUR–AFR and NAM–AFR were 0.23, 0.17 and

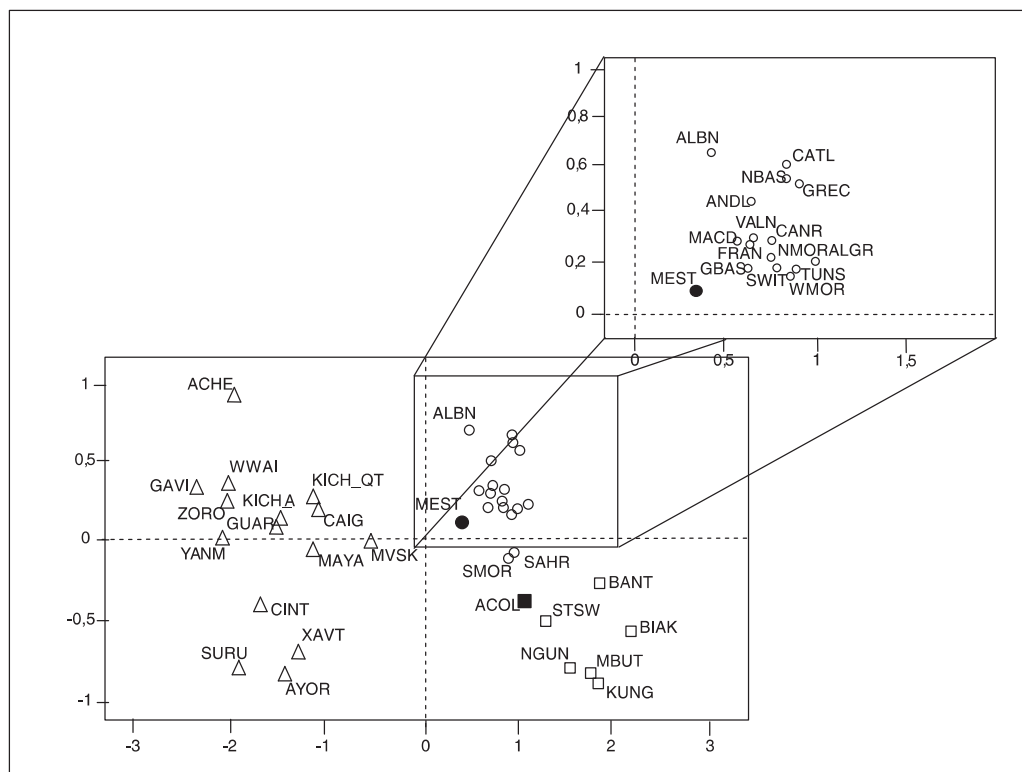


Figure 2. Nonmetric multidimensional scaling (MDS) plot representing a Reynold's F_{ST} genetic distance matrix for 40 European, African and Native American populations. Genetic distances were computed based on the insertion frequencies of six *Alu* elements. The study populations, Afro-Colombian (ACOL) and Mestizo (MEST) from Antioquia are highlighted (solid square and full circles, respectively). European and North African populations are represented by circles, Native American by triangles and collections from South-Saharan Africa by squares. Population key: Biaka (BIAK) and Mbuti (MBUT) Pygmies, Kung San (KUNG), Nguni (NGUN), Sotho-Tswana (STSW), Bantu (BANT), southeast Morocco (SMOR), west Morocco (WMOR), north Morocco (NMOR), Algeria (ALGR), Sahara (SAHR), Tunisia (TUNS), France (FRAN), Canary Isles (CANR), Basques from Guipúzcoa (GBAS) and Navarre (NBAS), Valencia (VALN), Catalonia (CATL), Andalusia (ANDL), Greece (GREC), Macedonia (MACD), Albania (ALBN), Switzerland (SWIT), Ayoreo (AYOR), Caigang (CAIG), Cinta Larga (CINT), Gavião (GAVI), Guaraní (GUAR), Quechua from Arequipa (KICH_A) and from Tayacaja (KICH_QT), Mvskoke (MVSK), Maya (MAYA), Suruí (SURU), Wai-Wai (WWAI), Xavante (XAVT), Yanomamo (YANM) and Zoró (ZORO).

0.36, respectively, corroborating that the population cluster with the greatest heterogeneity is NAM. As can be noted, the lowest intergroup differentiation appeared between EUR–AFR, and the highest between NAM–AFR. With respect to individual *Alu* markers, δ values of 0.30 or higher were obtained for TPA25, ACE, PV92, APO and FXIIIIB, the latter with δ values above 0.30 in all intergroup comparisons.

Based on MDS results, which indicate a certain overlapping of *Alu* diversity with geography, we analyse how the observed genetic heterogeneity is spatially structured by hierarchical analysis of molecular variance (AMOVA). Upon assignment of the populations within the three major geographic clusters, AMOVA analyses were performed for the six *Alu* insertions included in the MDS analysis (Table IV). According to values of the fixation indices, we found highly significant differences ($p < 0.001$) between population clusters (F_{CT}) for all *Alu* insertions examined, excepting A25 ($p < 0.05$), thereby confirming the existence of geographic structuring of the genetic heterogeneity among continental groups.

Admixture rates from *Alu* insertions in Colombia 495

Table III. Weighted insertion frequencies (\pm standard error) for eight *Alu* markers in three population groups classed according to geography, and differences in weighted frequencies (δ) between them.

<i>Alu</i> marker	Weighted allelic frequencies			Difference in frequencies (δ)		
	EUR	NAM	AFR	EUR/NAM	EUR/AFR	NAM/AFR
A25	0.132 (± 0.009)	0.149 (± 0.007)	0.289 (± 0.008)	0.016	0.157	0.140
ACE	0.385 (± 0.032)	0.700 (± 0.022)	0.369 (± 0.016)	0.316	0.016	0.332
APO	0.976 (± 0.009)	0.958 (± 0.019)	0.673 (± 0.029)	0.018	0.303	0.285
B65	0.527 (± 0.016)	0.288 (± 0.019)	0.510 (± 0.024)	0.240	0.017	0.223
D1	0.371 (± 0.029)	0.415 (± 0.038)	0.263 (± 0.023)	0.044	0.108	0.152
FXIII B	0.468 (± 0.012)	0.929 (± 0.032)	0.099 (± 0.018)	0.462	0.368	0.830
PV92	0.195 (± 0.016)	0.840 (± 0.054)	0.271 (± 0.012)	0.644	0.076	0.569
TPA25	0.582 (± 0.011)	0.662 (± 0.013)	0.283 (± 0.018)	0.079	0.299	0.379

Population groups are: EUR, Europe (Basques, Catalonia, Andalusia and Valencia); NAM, Native American (Maya and Yanomamo); AFR, South-Saharan Africa (Nguni, Sotho-Tswana and Bantu).

Bold type indicates $\delta > 0.30$.

Data sources are given in the Materials and Methods section.

On the other hand, genetic variation among populations within groups (F_{SC}) proved to be nonsignificant for APO, FXIII B, PV92 and TPA25, indicating that the allelic frequencies of these *Alu* markers are relatively homogeneous within each geographic cluster.

Admixture level estimates are shown in Table V. As expected, the Afro-Colombian genetic background was largely determined by the African component, whereas only 11.2–11.5% was contributed by Amerindian genes. Interestingly, no European genetic footprint was detected in the Afro-Colombian gene pool by the admixture estimation methods

Table IV. Fixation indices (F_{ST} , F_{SC} and F_{CT}) generated from a hierarchical analysis of molecular variance (AMOVA) for six polymorphic *Alu* insertions, considering three populations groups† classed according to geography.

<i>Alu</i> insertion	Fixation indices (%)		
	F_{ST}	F_{SC}	F_{CT}
A25	8.57***	3.95***	4.82*
ACE	19.13***	1.05***	18.28***
APO	23.81***	-1.52 NS	23.92***
FXIII B	41.44***	-1.13 NS	41.74***
PV92	43.09***	1.05 NS	43.02***
TPA25	8.91***	-0.08 NS	8.98***

†Groups are: Europe (Basques, Catalonia, Andalusia, Valencia and France), South-Saharan Africa (Nguni, Sotho-Tswana and Bantu) and Native American (Maya, Yanomamo, Cinta Larga, Quechua A, Quechua QT, Wai-Wai and Surui).

F_{ST} , genetic variation within populations; F_{SC} , genetic variation among populations within groups; F_{CT} , genetic variation among groups. NS, non-significant.

*Statistical significance for $p < 0.05$; ***statistical significance for $p < 0.001$.

Table V. Estimated admixture proportions (\pm standard error) from European, African and Native American in the gene pool of two ethnic groups (Afro-Colombian and Mestizo) from Antioquia.

Population	Admixture proportions		
	μ_1 (European) \ddagger	μ_2 (African) \S	μ_3 (Native American) \P
Afro-Colombian*	–	0.885 (\pm 0.138)	0.115 (\pm 0.060)
Afro-Colombian \dagger	–	0.888 (\pm 0.022)	0.112 (\pm 0.009)
Mestizo*	0.600 (\pm 0.139)	0.193 (\pm 0.159)	0.207 (\pm 0.069)
Mestizo \dagger	0.638 (\pm 0.052)	0.170 (\pm 0.050)	0.192 (\pm 0.220)

*Admixture estimates by the Weighted Least Squares method (Long et al. 1991).

\dagger Admixture estimates by the Gene Identity method (Chakraborty 1985).

\ddagger European populations: Basques, Catalonia, Andalusia and Valencia.

\S African populations: Nguni, Sotho-Tswana and Bantu.

\P Native American populations: Maya, Yanomamo, Cinta Larga, Quechua from Arequipa and Tayacaja, Wai-Wai and Suruí.

utilized. Conversely, the Mestizo gene pool exhibited a primarily European component ($\geq 60.0\%$) and the estimated contribution of African and Amerindian genes oscillated between 17.0 and 19.3% and 19.2–20.7%, respectively. These admixture proportions were represented in a ternary plot (Figure 3). In addition to the studied collections, we also represented other hybrid central- and south-American samples to evaluate the magnitude of gene flow processes in Antioquia. Data on the populations included in this comparative analysis can be consulted in supplementary Table S1 (online version only). The Afro-Colombian sample plotted very close to the ‘African pole’, reflecting its low admixture proportion. Most African–American populations grouped in an area that represents between 35% and 55% of non-African gene pool, indicating a substantially higher admixture level than the Afro-Colombians from Antioquia. In general terms, Mestizo populations showed a variable genetic admixture structure, oscillating from an elevated proportion of European genes to a major contribution of the Amerindian gene pool. The Mestizo population from Antioquia proved to be much more hybridized than most of other Mestizo populations of the continent, with a comparatively high component of the African gene pool.

Discussion

This work provides new data on the genetic background of American admixed populations from the perspective of polymorphic *Alu* insertions, through the analysis of one of the most important ethnic groups from Antioquia (Afro-Colombians) and its comparison with a reference sample (Mestizos) from the same region. The global picture derived from the distance analysis (MDS plot, Figure 2) is indicative that the human genetic diversity tends to be spatially structured, in accordance with historical patterns of human migrations (gene flow) and random phenomena (genetic drift). Presumably, genetic drift has played a major role in the modelling of the human genome, due to the partial isolation of the human populations during much of their evolutionary history (Jorde and Wooding 2004).

Bearing in mind the wide geographical distribution of the populations included in the analysis, the level of genetic differentiation among the different human groups was expected to be maximized. Five (TPA25, ACE, PV92, APO and FXIIIIB) out of the eight *Alu* markers exhibited significant δ values (≥ 0.30) between continental groups. In addition, F_{CT} values generated by the AMOVA analysis also indicated that these *Alu* loci are robust and reliable enough to discriminate between continental groups, and therefore, for evaluating the genetic

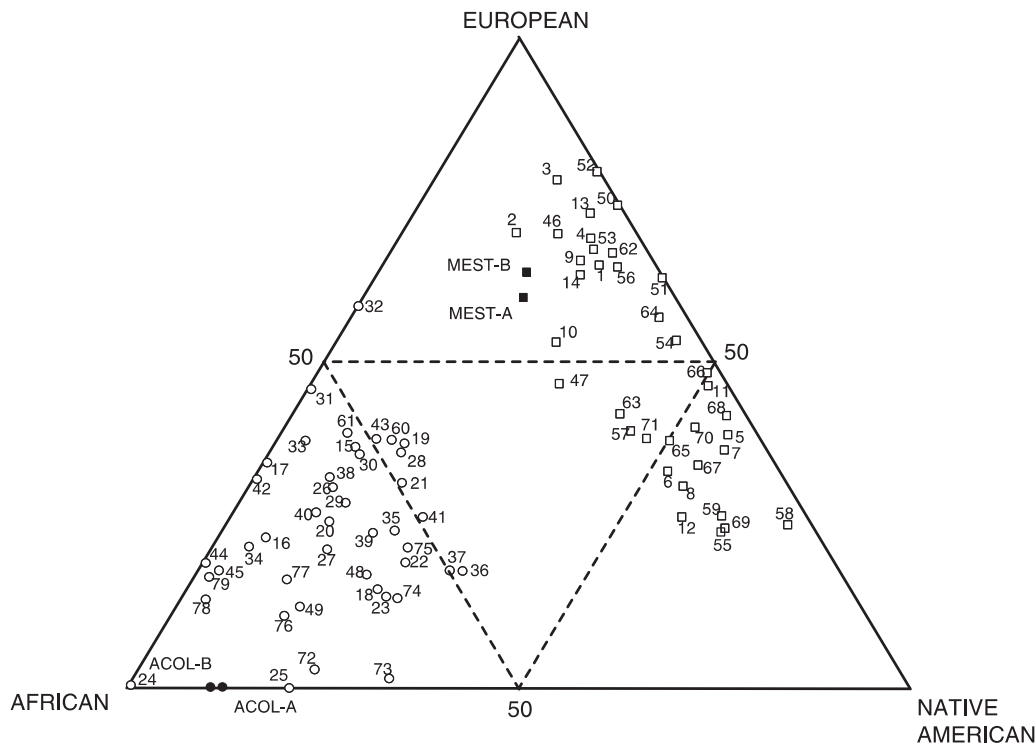


Figure 3. Ternary plot showing the admixture level of different Mestizo (squares) and African-derived (circles) populations from central and south America. Triangle vertices correspond to parental populations (European, African and Native American). Positions in the plot represent the relative proportion of each parental population's gene pool in the samples included in the analysis. The two highlighted points for the samples examined (MEST, Mestizos; ACOL, Afro-Colombians) represent admixture rates obtained by different methods: the Weighted Least Squares method (MEST-A and ACOL-A) and the Gene Identity method (MEST-B and ACOL-B). Populations are numbered according to data presented in Table S1 (online version only).

admixture level in a given sample. Of them, FXIIIB and PV92 have been already included in some AIM panels (Bonilla et al. 2004; Luizon et al. 2008).

Data obtained from the genetic admixture analysis (Table V) indicated that the African component is highly predominant (88.5–88.8%) in the Afro-Colombian collection. We also found a contribution of the Native American gene pool, although modest (11.2–11.5%). The most notable finding in this population was, however, the lack of traces of European genetic contribution. As can be verified by consulting the genetic admixture data compiled for other African-derived populations in America (Table S1; online version only), there are very few populations featuring a European component below 3%.

A very different situation was found for the Mestizo reference sample, which showed a predominantly European genetic background ($\geq 60\%$). The proportion of African genes in the genome of the Mestizos from Antioquia was relatively high (Figure 3), in accord with the findings of a previous study on admixture levels of Mestizo populations from Colombia and Brazil (Wang et al. 2008). As for the Amerindian component, the estimated proportion tended to be lower than the reported ones for other adjacent Mestizo populations (Sandoval et al. 1993; Zabala-Fernandez et al. 2005; Bedoya et al. 2006; Wang et al. 2008). Specifically, the Antioquian collection analysed by Wang et al. (2008) showed the lowest admixture rates with Native Americans ($\approx 25.0\%$) among all Mestizo populations included in their

study. This is consistent with the relatively reduced presence of Native American groups in Antioquia since the beginnings of the Spanish conquest until present day (Carvajal-Carmona et al. 2000; Bedoya et al. 2006).

A plausible explanation to the null European contribution to the Afro-Colombian gene pool could be associated with the harsh slavery conditions and the way in which many Afro-Colombian communities emerged. Very often, such communities were founded by African individuals who escaped from the work plantations and mines where they worked as slaves of the European colonists. The runaway slaves were called 'cimarrones', and the communities they inhabited, 'palenques' (hiding places).

Some authors have suggested that the European component in the gene pool of cimarron-derived populations must be rather scarce, due to the traditional hostility of the inhabitants of the palenques to the European colonists, which effectively kept them isolated from cultural and genetic contacts with this segment of the society (Castro 1993; Bortolini et al. 1995; Martínez-Labarga et al. 1999; Cotrim et al. 2004). Nevertheless, several studies on African-American populations founded by fugitive slaves have reported variable proportions of genes of European origin, which has led to the hypothesis that isolation was only partial (Arpini-Sampaio et al. 1999; Bortolini et al. 1999; Da Silva et al. 1999; Sans et al. 2002; Cotrim et al. 2004). Such findings could be indicative of variability in the behaviour of these communities. In fact, admixture estimates based on *Alu* markers have revealed a certain degree of gene flow between runaway slaves and Native American populations, as witnessed by the presence of Amerindian genes in the Afro-Colombian gene pool (see Table V). Anyway, the presumably long genetic isolation of the cimarrones in combination with the scarce number of founders of the palenques most likely promoted intense genetic drift episodes by founder effects and population bottlenecks, which would explain the incongruities between the results of the different studies.

In summary, the African American gene pool is largely determined by the African component, with a lower contribution of Amerindian and European genes. Within this group, the Afro-Colombians from Antioquia stand out for the very limited contribution of Amerindian genes and a null proportion of European genes. These peculiarities might have been connected with the way their communities (palenques) appeared, founded by small groups of runaway slaves characterized by a great reluctance to interact with strangers to the group, all of which would have promoted and/or reinforced the effect of the genetic drift. On the other hand, Mestizos are characterized by variable proportions of European and Amerindian genes, with a generally minor fraction of genes of African origin. Specifically, the Mestizos from Antioquia featured a rather scarce proportion of Amerindian genes, which is in agreement with the demographic history of this Colombian region. Furthermore, the set of *Alu* markers analysed in this study has shown an adequate resolving power to differentiate between continental groups. In particular, ACE, APO, FXIIB, PV92 and TPA25 could be efficiently utilized in investigations focusing the relationships between genetic profile, demographic history and biogeographical ancestry in human populations.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, de Pancorbo MM, Caeiro JL, Puzyrev VP, Herrera RJ. 2002. Phylogenetics of worldwide human populations as determined by polymorphic *Alu* insertions. *Electrophoresis* 23:3346–3356.

- Arpini-Sampaio Z, Costa MCB, Melo AA, Carvalho MFVA, Deus MSM Simões AL. 1999. Genetic polymorphisms and ethnic admixture in African-derived black communities of northeastern Brazil. *Hum Biol* 71:69–85.
- Battilana J, Fagundes NJ, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E, Munkhbat B, Munkhtuvshin N, Krylov M, Benevolenskaia L, Arnett FC, Batzer MA, Deininger PL, Salzano FM, Bonatto SL. 2006. Alu insertion polymorphisms in Native Americans and related Asian populations. *Ann Hum Biol* 33:142–160.
- Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL. 1994. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci USA* 91:12288–12292.
- Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. *Nat Rev Genet* 3:370–379.
- Bedoya G, Montoya P, Garcia J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, Ruiz-Linares A. 2006. Admixture dynamics in Hispanics: A shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci USA* 103:7234–7239.
- Bonilla C, Parra EJ, Pfaff CL, Dios S, Marshall JA, Hamman RF, Ferrell RE, Hoggart CL, McKeigue PM, Shriver MD. 2004. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann Hum Genet* 68:139–153.
- Bortolini MC, Weimer TA, Salzano FM, Callegari-Jacques SM, Schneider H, Layrisse Z, Bonatto SL. 1995. Evolutionary relationships between black South American and African populations. *Hum Biol* 67:547–559.
- Bortolini MC, Da Silva WA Junior W, De Guerra DC, Remonato G, Mirandola R, Hutz MH, Weimer TA, Silva MC, Zago MA, Salzano FM. 1999. African-derived South American populations: A history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. *Am J Hum Biol* 11:551–563.
- Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, Ruiz-Linares A. 2000. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet* 67:1287–1295.
- Castro D. 1993. Relación entre polimorfismos genéticos e historia en dos poblaciones negras venezolanas. *Boletín Sociedad Española Antropología Biológica* 14:21–29.
- Chakraborty R. 1985. Gene identity in racial hybrids and estimation of admixture rates. In: Neel JV, Ahuja Y, editors. *Genetic microdifferentiation in man and other animals*. Delhi: Indian Anthropological Association, pp. 171–180.
- Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A. 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: Evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* 107:312–319.
- Comas D, Schmid H, Brauer S, Flaiz C, Busquets A, Calafell F, Bertranpetit J, Scheil HG, Huckenbeck W, Efremovska L, Schmidt H. 2004. Alu insertion polymorphisms in the Balkans and the origins of the Aromuns. *Ann Hum Genet* 68:120–127.
- Cotrim NH, Auricchio MT, Vicente JP, Otto PA, Mingroni-Netto RC. 2004. Polymorphic Alu insertions in six Brazilian African-derived populations. *Am J Hum Biol* 16:264–277.
- Curtin PD. 1969. *The Atlantic slave trade: A census*. Milwaukee: University of Wisconsin Press.
- Da Silva WA Jr, Bortolini MC, Meyer D, Salzano FM, Elion J, Krishnamoorthy R, Schneider MP, De Guerra DC, Layrisse Z, Castellano HM, Weimer TD, Zago MA. 1999. Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. *Am J Phys Anthropol* 109:425–437.
- Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, Salzano FM. 2004. Mitochondrial DNA and Alu insertions in a genetically peculiar population: The Ayoreo Indians of Bolivia and Paraguay. *Am J Hum Biol* 16:479–488.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.
- Felsenstein J. 1989. PHYLIP: Phylogeny inference package (version 3.2). *Cladistics* 5:164–166.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, Vidales C, Arroyo D, Pena JA. 2006. Genetic position of Valencia (Spain) in the Mediterranean basin according to Alu insertions. *Am J Hum Biol* 18:187–195.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, de Pancorbo MM, Pena JA. 2007. Polymorphic Alu insertions and the genetic structure of Iberian Basques. *J Hum Genet* 52:317–327.
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.
- Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4:9 pp.
- Herrera RJ, Rojas DP, Terreros MC. 2007. Polymorphic Alu insertions among Mayan. *J Hum Genet* 52:129–142.

500 L. Gómez-Pérez et al.

- Jorde LB, Wooding SP. 2004. Genetic variation, classification and 'race'. *Nat Genet* 36:S28–S33.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, the Human Genome Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
- Long JC, Williams RC, McAuley JE, Medis R, Partel R, Tregellas WM, South SF, Rea AE, McCormick SB, Iwaniec U. 1991. Genetic variation in Arizona Mexican Americans: Estimation and interpretation of admixture proportions. *Am J Phys Anthropol* 84:141–157.
- Liu K, Muse SV. 2004. PowerMarker: New genetic data analysis software. Version 3.0. Free program distributed by the author over the Internet from <http://www.powermarker.net>.
- Luizon MR, Mendes-Junior CT, De Oliveira SF, Simões AL. 2008. Ancestry informative markers in Amerindians from Brazilian Amazon. *Am J Hum Biol* 20:86–90.
- Maca-Meyer N, Villar J, Perez-Mendez L, Cabrera de Leon A, Flores C. 2004. A tale of aborigines, conquerors and slaves: Alu insertion polymorphisms and the peopling of Canary Islands. *Ann Hum Genet* 68:600–605.
- Martínez-Labarga C, Rickards O, Scacchi R, Corbo RM, Biondi G, Peña JA, Varas de Vieira C, Guevara AE, Santurino MS, de Stefano GF. 1999. Genetic population structure of two African–Ecuadorian communities of Esmeraldas. *Am J Phys Anthropol* 109:159–174.
- Novick GE, Novick CC, Yunis J, Yunis E, Antunez-de-Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ. 1998. Polymorphic Alu insertions and the Asian origin of Native American populations. *Hum Biol* 70:23–39.
- Reynolds J, Weir BS, Cockerman CC. 1983. Estimation of the coancestry coefficient: Basis for a short term genetic distance. *Genetics* 105:767–779.
- Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA, Barbujani G. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res* 12:602–612.
- Roy-Engel AM, Carroll ML, El-Sawy M, Salem AH, Garber RK, Nguyen SV, Deininger PL, Batzer MA. 2002. Non-traditional Alu evolution and primate genomic diversity. *J Mol Biol* 316:1033–1040.
- Sandoval C, De la Hoz A, Yunis EJ. 1993. Estructura genética de la población colombiana. *Análisis de mestizaje. Rev Fac Med UN Colombia* 43:3–14.
- Sans M, Weimer TA, Franco MHL, Salzano FM, Bentancor N, Alvarez I, Bianchi NO, Chakraborty R. 2002. Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am J Phys Anthropol* 118:33–44.
- Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. 1997. Alu insertion polymorphisms and human evolution: Evidence for a larger population size in Africa. *Genome Res* 7:1061–1071.
- Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camarena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L, Ruiz-Linares A. 2008. Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet* 4:e1000037.
- Zabala-Fernandez WM, Borjas-Fajardo L, Fernandez Salgado E, Castillo C, Socca L, Portillo MG, Sanchez MA, Delgado W, Morales-Machin A, Layrisse Z, Pineda Bernal L. 2005. Use of short tandem repeats loci to study the genetic structure of several populations from Zulia State, Venezuela. *Am J Hum Biol* 17:451–459.

Supplementary Table S1. Estimated admixture proportions in central- and south-American populations. Genetic markers analyzed and data sources are also given.

Nº	Population	Type ^a	Admixture Proportions ^b			Markers	Source
			μ_1^b (European)	μ_2^b (African)	μ_3^b (Native American)		
Colombia							
1	Cundinamarca	M	0.650	0.070	0.280	classical markers	Sandoval et al. (1993)
2	Antioquia	M	0.700	0.150	0.150	classical markers	Sandoval et al. (1993)
3	Antioquia	M	0.790	0.060	0.160	autosomal markers	Bedoya et al. (2006)
4	Antioquia	M	0.690	0.060	0.250	chromosome X markers	Bedoya et al. (2006)
5	Pasto	M	0.392	0.036	0.572	autosomal microsatellites	Wang et al. (2008)
6	Pasto	M	0.331	0.143	0.526	X-chromosomal microsatellites	Wang et al. (2008)
7	Peque	M	0.365	0.053	0.582	autosomal microsatellites	Wang et al. (2008)
8	Peque	M	0.311	0.133	0.557	X-chromosomal microsatellites	Wang et al. (2008)
9	Medellin	M	0.656	0.091	0.253	autosomal microsatellites	Wang et al. (2008)
10	Medellin	M	0.529	0.185	0.285	X-chromosomal microsatellites	Wang et al. (2008)
11	Cundinamarca	M	0.464	0.025	0.511	autosomal microsatellites	Wang et al. (2008)
12	Cundinamarca	M	0.264	0.159	0.577	X-chromosomal microsatellites	Wang et al. (2008)
Venezuela							
13	Maracaibo	M	0.730	0.040	0.230	autosomal microsatellites	Zabala Fernández et al. (2005)
14	Isla de Toas	M	0.630	0.100	0.260	autosomal microsatellites	Zabala Fernández et al. (2005)
15	Venezuelan Mixed Communities	AD	0.370	0.520	0.110	autosomal markers	Da Silva et al. (1999)
16	Curiepe	AD	0.232	0.704	0.064	autosomal markers	Bortolini et al. (1999)
17	Curiepe	AD	0.352	0.648	0.000	classical markers	Bortolini et al. (1998)
18	Birongo	AD	0.153	0.601	0.246	autosomal markers	Bortolini et al. (1999)
19	Birongo	AD	0.375	0.456	0.169	classical markers	Bortolini et al. (1998)
20	Sotillo	AD	0.256	0.610	0.134	autosomal markers	Bortolini et al. (1999)
21	Sotillo	AD	0.347	0.537	0.216	classical markers	Bortolini et al. (1998)
22	Panaquire	AD	0.193	0.546	0.261	autosomal markers	Bortolini et al. (1999)
23	Panaquire	AD	0.140	0.590	0.260	classical markers	Castro de Guerra et al. (1996)
24	San José de Heras	AD	0.000	1.000	0.000	autosomal microsatellites	Zabala Fernández et al. (2005)
25	Ganga	AD	0.000	0.790	0.210	classical markers	Castro de Guerra et al. (1996)
26	Patanemo	AD	0.310	0.580	0.110	classical markers	Castro de Guerra et al. (1996)
27	Tapija	AD	0.210	0.630	0.150	classical markers	Castro de Guerra et al. (1996)
Brazil							
28	Curiepe	AD	0.362	0.467	0.171	classical markers	Bortolini et al. (1995)
29	Curiepe	AD	0.285	0.575	0.140	classical markers	Bortolini et al. (1992)
30	Brazilian Mixed Communities	AD	0.360	0.520	0.120	autosomal markers	Da Silva et al. (1999)
31	Porto Alegre	AD	0.464	0.536	0.000	autosomal markers	Bortolini et al. (1999)
32	Porto Alegre	AD	0.590	0.410	0.000	classical markers	Bortolini et al. (1997)
33	Salvador	AD	0.380	0.580	0.040	classical markers	Bortolini et al. (1997)
34	Ribeirao Preto	AD	0.218	0.731	0.051	autosomal markers	Bortolini et al. (1999)
35	Cametá	AD	0.242	0.534	0.224	autosomal markers	Bortolini et al. (1999)
36	Cametá	AD	0.179	0.480	0.341	classical markers	Bortolini et al. (1995)
37	Cametá	AD	0.180	0.496	0.323	classical markers	Bortolini et al. (1992)
38	Trombetas	AD	0.325	0.576	0.099	autosomal markers	Bortolini et al. (1999)
39	Trombetas	AD	0.238	0.564	0.198	classical markers	Bortolini et al. (1995)
40	Trombetas	AD	0.270	0.620	0.110	classical markers	Schneider et al. (1987)
41	Cajueiro	AD	0.262	0.488	0.250	autosomal markers	Bortolini et al. (1999)
42	Cajueiro	AD	0.326	0.674	0.000	classical markers	Bortolini et al. (1998)
43	Paredão	AD	0.380	0.492	0.128	autosomal markers	Bortolini et al. (1999)
44	Paredão	AD	0.198	0.802	0.000	classical markers	Bortolini et al. (1995)
45	Paredão	AD	0.181	0.791	0.028	classical markers	Bortolini et al. (1992)
46	Rio Grande do Sul	M	0.696	0.099	0.204	autosomal microsatellites	Wang et al. (2008)
47	Rio Grande do Sul	M	0.466	0.213	0.321	X-chromosomal microsatellites	Wang et al. (2008)
48	Mimbó	AD	0.165	0.611	0.224	classical markers/autosomal microsatellites	Arpini-Sampaio et al. (1999)
49	Sítio Velho	AD	0.124	0.715	0.160	classical markers/autosomal microsatellites	Arpini-Sampaio et al. (1999)
Argentina							
50	Bahia Blanca Original	M	0.743	0.000	0.257	Alu insertions	Resano et al. (2007)
51	Bahia Blanca Mixed	M	0.632	0.000	0.368	Alu insertions	Resano et al. (2007)
52	Bahia Blanca General	M	0.791	0.000	0.209	Alu insertions	Resano et al. (2007)
53	La Plata City	M	0.676	0.065	0.259	biallelic markers	Martinez-Marignac et al. (2004)
54	Catamarca	M	0.532	0.030	0.437	autosomal microsatellites	Wang et al. (2008)
55	Catamarca	M	0.238	0.119	0.643	X-chromosomal microsatellites	Wang et al. (2008)
56	Tucuman	M	0.646	0.049	0.305	autosomal microsatellites	Wang et al. (2008)
57	Tucuman	M	0.395	0.158	0.447	X-chromosomal microsatellites	Wang et al. (2008)
58	Salta	M	0.252	0.030	0.719	autosomal microsatellites	Wang et al. (2008)
59	Salta	M	0.264	0.109	0.627	X-chromosomal microsatellites	Wang et al. (2008)
Uruguay							
60	Melo	AD	0.376	0.469	0.155	classical markers	Sans et al. (2002)
61	Melo	AD	0.392	0.521	0.087	classical markers	Sans et al. (2002)
Costa Rica							
62	Central Valley of Costa Rica	M	0.667	0.046	0.287	autosomal microsatellites	Wang et al. (2008)
63	Central Valley of Costa Rica	M	0.420	0.159	0.420	X-chromosomal microsatellites	Wang et al. (2008)
Mexico							
64	Mexico City	M	0.568	0.034	0.398	autosomal microsatellites	Wang et al. (2008)
65	Mexico City	M	0.383	0.116	0.501	X-chromosomal microsatellites	Wang et al. (2008)
Chile							
66	Quetalmahue	M	0.485	0.013	0.502	autosomal microsatellites	Wang et al. (2008)
67	Quetalmahue	M	0.342	0.098	0.561	X-chromosomal microsatellites	Wang et al. (2008)
68	Paposo	M	0.420	0.023	0.557	autosomal microsatellites	Wang et al. (2008)
69	Paposo	M	0.244	0.114	0.642	X-chromosomal microsatellites	Wang et al. (2008)
Guatemala							
70	Oriente	M	0.399	0.073	0.528	autosomal microsatellites	Wang et al. (2008)
71	Oriente	M	0.383	0.142	0.474	X-chromosomal microsatellites	Wang et al. (2008)
72	Livingstone	AD	0.024	0.752	0.224	classical markers	Crawford et al. (1981)
Nicaragua							
73	Bluefields	AD	0.010	0.660	0.330	classical markers	Biondi et al. (1988)
Ecuador							
74	Viche	AD	0.14	0.584	0.276	classical markers	Martinez-Labarga et al. (1999)
75	Viche	AD	0.218	0.531	0.251	classical markers	Martinez-Labarga et al. (1999)
76	Rio Cayapas	AD	0.111	0.743	0.145	classical markers	Martinez-Labarga et al. (1999)
77	Rio Cayapas	AD	0.168	0.712	0.121	classical markers	Martinez-Labarga et al. (1999)
Haiti							
78	Haiti	AD	0.139	0.830	0.031	classical markers	Basu et al. (1976)
79	Haiti	AD	0.172	0.808	0.019	classical markers	Basu et al. (1976)

^a Type of population is Mestizo (M) or African-derived (AD)

^b Proportionate contribution of the reference gene pool to the hybrid population (μ_1 , μ_2 and μ_3)

Sources

Arpini-Sampaio et al. (1999) Hum Biol 71: 69-85
 Basu et al. (1976) Hum Biol 48: 245-269
 Bedoya et al. (2006) PNAS 103: 7234-9
 Biondi et al. (1988) Ann Hum Biol 15: 399-412
 Bortolini et al. (1992) Gene Geogr 6: 1-16
 Bortolini et al. (1995) Hum Biol 67: 547-59
 Bortolini et al. (1997) Int J Anthropol 12: 5-16
 Bortolini et al. (1998) Hum Biol 70: 443-61
 Bortolini et al. (1999) Am J Hum Biol 11: 551-63
 Castro de Guerra et al. (1996) Hum Hered 46: 323-8
 Crawford et al. (1981) Hum Biol 53: 87-103
 Da Silva et al. (1999) Am J Phys Anthropol 109: 425-37
 Martinez-Labarga et al. (1999) Am J Phys Anthropol 109:159-174
 Martinez-Marignac et al. (2004) Hum Biol 76: 543-57
 Resano et al. (2007) Am J Hum Biol 19: 827-35
 Sandoval et al. (1993) Rev Fac Med UN Colombia 43: 3-14
 Sans et al. (2002) Am J Phys Anthropol 118: 33-44
 Schneider et al. (1987) Rev Bras Genet 10: 565-74
 Wang et al. (2008) PLoS Genet 21: e1000037
 Zabala Fernández et al. (2005) Am J Hum Biol 17: 451-9

PRINCIPALES RESULTADOS Y MATERIAL COMPLEMENTARIO

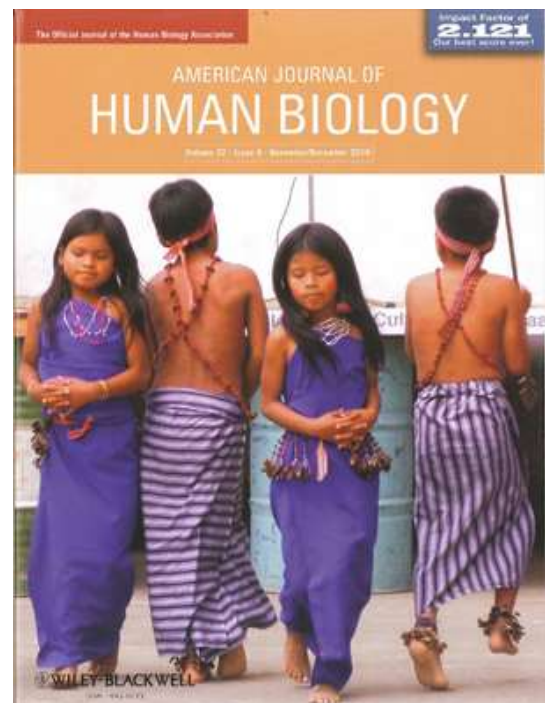
Como cabría esperar, en la muestra de Afrocolombianos el componente africano es predominante. Sin embargo, resulta sorprendente que no se haya encontrado rastro alguno de contribución genética europea en su patrimonio genético. La explicación más plausible para esta situación se basa en el efecto que pudo tener el componente social de la esclavitud sobre algunas poblaciones de origen africano. Como se menciona en la introducción de esta memoria y en el propio manuscrito, algunas de las comunidades de afrocolombianos actuales tienen su origen en individuos que escaparon de los lugares donde se encontraban esclavizados, los denominados cimarrones. Estos cimarrones constituyeron comunidades, conocidas como palenques o quilombos, que eran reticentes al contacto con los europeos y con los que mantenían habitualmente conflictos armados. De hecho, existe algún otro caso registrado de poblaciones afroamericanas en las que no se ha encontrado prácticamente ningún aporte europeo en su patrimonio genético, como San José de Heras y Ganga, dos localidades de Venezuela (Castro de Guerra *et al.* 1996; Zabala Fernández *et al.* 2005), que aparecen citadas en la Tabla suplementaria 1 del manuscrito.

Por su parte, la muestra de Mestizos ha mostrado lógicamente una contribución europea predominante, en torno al 60%, con aportes significativos de las otras dos poblaciones parentales. La mayor parte de las poblaciones mestizas americanas estudiadas ha mostrado un patrón similar, con variaciones más o menos importantes en las proporciones de los diferentes grupos étnicos que las conforman.

En un trabajo reciente (Ibarra *et al.* 2014) realizado con SNPs, se muestran unos valores no muy diferentes para la provincia de Antioquia, con una proporción europea menor (47%) y una amerindia más alta (34%). Por otra parte, si bien no analizan una población claramente africana, incluyen la población general de Choco, de origen mayoritariamente africano (12% nativo, 82.7% africano y 5.42 europeo según el DANE, 2005). Lógicamente, encuentran un claro componente

mayoritario de origen africano, igual que en Mulaló, población del departamento del Valle de Cauca. En un análisis MDS realizado en este trabajo, ambas poblaciones tienen comportamiento similar que la población afrocolombiana de esta Tesis Doctoral, apareciendo en las proximidades de las poblaciones africanas.

Además, se ha observado que varios de los marcadores *Alu* analizados (ACE, APO, FXIII B, PV92 y TPA25) pueden ser considerados como AIMs, ya que exhiben valores de δ significativos ($\delta \geq 0,30$) entre las poblaciones consideradas.



MANUSCRITO III

Microevolutionary Processes due to Landscape Features in the Province of Jujuy (Argentina)

Gómez-Pérez L, Alfonso-Sánchez MA, Dipierri JE, Alfaro E, García-Obregón S, De Pancorbo MM, Bailliet G, Peña JA. 2011.

American Journal of Human Biology 23, 177-184

INTRODUCCIÓN

Si en el trabajo anterior se han considerado dos poblaciones con un grado de mestizaje potencialmente alto, en éste se ha caracterizado la población de Jujuy, en el Noroeste de Argentina, con un potencial de mestizaje bajo debido a las condiciones orográficas de su territorio. En efecto, la elevada altitud de algunas regiones como la Puna jujeña las convierte en hábitats inhóspitos, con dificultades de aclimatación para los no autóctonos, debido a un ambiente con baja presión de oxígeno. Asimismo, las grandes diferencias de altura entre las regiones de Jujuy ofrecen una gran variedad de entornos, que han condicionado en grado variable el asentamiento de europeos y africanos en el territorio, como se desprende del censo de Carlos III de 1778 (Introducción, *Tabla 1*).

La provincia de Jujuy une las regiones del Gran Chaco, las pampas y en general el cono sur americano con el altiplano de los Andes. Es una vía de paso natural, debido a la existencia de corredores como la Quebrada de Humahuaca, que asciende desde 2.000 m hasta 4.000 m de altitud. Ha sido la ruta por la que han transitado diferentes culturas en época precolombina, como se ha mencionado anteriormente. También fue parte de la vía de transporte de la plata boliviana hacia el Mar de Plata durante el imperio español y es posible que fuera una de las vías de entrada del primer poblamiento humano de esta región del continente.

A partir de las mismas 8 inserciones *Alu* utilizadas en los trabajos anteriores, se ha analizado la influencia de la orografía sobre el poblamiento diferencial del territorio por parte de nativos, descendientes de europeos y afroamericanos y en definitiva sobre el patrimonio genético de las poblaciones implicadas.

Este trabajo ha sido utilizado por Relethford (2012) en su manual sobre Genética de Poblaciones Humanas como ejemplo de análisis de la estructura de las poblaciones humanas, junto a dos trabajos sobre las islas Aland y los Gainj y Kalam de Nueva Guinea.

Original Research Article

Microevolutionary Processes due to Landscape Features in the Province of Jujuy (Argentina)

LUIS GÓMEZ-PÉREZ,¹ MIGUEL A. ALFONSO-SÁNCHEZ,¹ JOSÉ E. DIPIERRI,² EMMA ALFARO,² SUSANA GARCÍA-OBREGÓN,¹ MARIAN M. DE PANCORBO,³ GRACIELLA BAILLIET,⁴ AND JOSÉ A. PEÑA^{1*}

¹Departamento de Genética y Antropología Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco, 48080 Bilbao, Spain

²Instituto de Biología de la Altura, Universidad Nacional de Jujuy, 4600 San Salvador de Jujuy, Argentina

³Departamento de Zoología y Dinámica Celular Animal, Facultad de Farmacia, Universidad del País Vasco, 01006 Vitoria, Spain

⁴Laboratory of Human Molecular Population Genetics, IMBICE, La Plata, Argentina

Objectives: We seek to evaluate the influence of a diverse and rugged physical environment on the genetic background of human populations.

Methods: We analyzed eight polymorphic *Alu* insertions in 226 individuals from Jujuy province (Argentina), which is composed of several regions with well-defined geographical features and marked contrasts between them associated with differences in altitude (range: 700–3300 m). This regional division was used to assess the spatial variation of the *Alu* diversity.

Results: Deviations from Hardy-Weinberg Equilibrium expectations resulting from heterozygous deficit were found for FXIIIB and PV92 in the highest subpopulations. Several *Alu* elements showed genetic heterogeneity between the highest region (La Puna) and the lowest regions (Valle and Selva). Similarly, a decreasing trend of the average heterozygosity according to altitude was found. Both the centroid method and the admixture analysis unveiled a gene flow above the average in lowland populations, indicating a higher proportion of foreign genes introduced by immigrants of European and African ancestry. Furthermore, several *Alu* frequency clines fitting the orientation of the altitude gradient were detected.

Conclusions: Our study reveals a spatial patterning of the *Alu* diversity in Jujuy, most likely determined by disparities in landscape and environmental features between the different subregions. Differences in the physical environment would have drastically reduced the homogenizing effects of the gene flow and would have promoted genetic drift episodes in the highest subpopulations. Microevolutionary processes detected in Jujuy have played an important role in the shaping of the gene pool of the populations from this sub-Andean zone from Argentina. *Am. J. Hum. Biol.* 23:177–184, 2011. © 2010 Wiley-Liss, Inc.

Geographical barriers are among the paramount factors affecting gene flow, genetic drift, and local adaptation processes, thus contributing to the shaping of the gene pool of the populations especially at a microevolutionary scale. A thorough analysis of the spatial distribution of the genetic heterogeneity in a subdivided population then requires the exhaustive knowledge of how landscape characteristics structure populations. Consequently, simultaneous collection and analysis of geographical and genetic data can be a suitable strategy to interpret accurately the microevolutionary processes that have generated the spatially structured genetic heterogeneity of the populations (Manel et al., 2003). During the 70s and 80s of the past century, numerous studies on the genetic structure of human populations based on demographic and geographical data were focused on those aspects of population's microevolution that could lead to inferences on human macroevolution (Jorde, 1982; Relethford, 1991). However, studies of this nature based on DNA molecular markers are extremely scarce.

The province of Jujuy, located in northwest Argentina, could be an interesting area to evaluate the influence of a diverse and rugged physical environment on the genetic background of the human populations living there. Jujuy borders Bolivia, Chile, and the Argentinian province of Salta (see Fig. 1). It spreads over a surface of 53,219 km² and according to the demographic census of 2001, its population was >608,000 inhabitants (INDEC, 2001). The population nuclei from Jujuy feature notable altitude differences due to the proximity of the Andes mountain

range. For that reason, Jujuy is divided in three different geographical/ecological areas, depending on altitude above sea level. These three areas are well differentiated by their environmental conditions: (1) the subtropical valley (~1200 m), which includes the province's capital city (San Salvador de Jujuy), (2) the Pre Puna (~2500 m), and (3) La Puna of Jujuy (>3300 m). The different areas are connected by gorges or narrow corridors, such as the Quebrada de Humahuaca (170 km long) located in the Pre Puna region, which ascends across the Jujuy province with a south-north orientation. The human presence in Jujuy is partly accounted for its strategic geographical position, since it communicates the Andean highlands with the extensive South American plains. Thus, Jujuy region would be inserted in one of the main routes for the first settlers on their colonization of the southern end of the continent. Specifically, the Quebrada de Humahuaca was part of one of the main roads of the Inca Empire. Fur-

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Programa de Subvención a Grupos Consolidados, Departamento de Educación, Universidades e Investigación' of the Basque Government (IT-424-07), and University of the Basque Country (to L.G.-P.)

*Correspondence to: Dr. José A. Peña, Departamento de Genética, Antropología Física y Fisiología Animal, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain. E-mail: joseangel.pena@ehu.es

Received 28 May 2010; Revision received 3 August 2010; Accepted 9 August 2010

DOI 10.1002/ajhb.21098

Published online 10 November 2010 in Wiley Online Library (wileyonlinelibrary.com).



Fig. 1. The province of Jujuy (Argentina). Highlighted are the five regions considered in this study. Dashed line represents the Quebrada de Humahuaca, which is the main communication route in the region. Numbered dots represent the sampling localities: San Salvador de Jujuy (1), Palpalá (2), El Carmen (3), Perico (4), La Mendieta (5), San Pedro (6), La Esperanza (7), Fraile Pintado (8), Libertador General San Martín (9), San Lucas (10), Tilcara (11), Maimara (12), El Moreno (13), Humahuaca (14), Susques (15), Coranzuli (16), Mina El Aguilar (17), Mina Piriquitas (18), Cochinoca (19), Abra Pampa (20), Paicone (21), Guadalupe (22), Rinconada (23), Chocoite (24), Mina Pan de Azúcar (25), Pumahuasi (26), La Quiaca (27), Casira (28), and Yavi (29).

thermore, during Spanish colonial times Jujuy was part of the famous route that linked the silver mines of Potosí with the River Plate harbors (Dipierrri et al., 2000).

Some estimates indicate that the peopling of Jujuy started between 6,000 and 8,000 years ago by human groups stemming from Peruvian and Bolivian Andean regions (Rothhammer et al., 1984; Rothhammer and Silva, 1992). During colonial times, between the 16th and 19th centuries, the arrival of European and African individuals contributed to the genetic diversity of human populations inhabiting Jujuy. Several millions of indigenous West and Central Africans from coastal regions were kidnapped and transported to work as slaves in the Americas (Curtin, 1969). The introduction of African slaves in the River Plate colonies started in the 16th century, ~1,588. The great majority of the African individuals carried to Argentina were captured in the current territories of Angola, Democratic Republic of the Congo, Guinea, and Congo. The volume of exports from Africa across the Atlantic has been calculated at ~11.6 million slaves, whereas imports into the Americas and most other parts of the Atlantic basin has been estimated in ~9.8 million slaves (Lovejoy, 1982). Specifically in Argentina, the population census of 1778 revealed that Jujuy was one of the Argentinian provinces with lowest African presence (<13%), with a decreasing proportion in the population nuclei from the highest altitudes (Gomes, 2006). Immigrants concentrated mainly in those zones located at the lowest altitude levels (Martínez-Sarasola, 1992). This may be the reason why some authors have found differential gene flow and admixture rates between the different human settlements of Jujuy province (Dipierrri et al., 1998, 2000). Because of this historical background, Native Americans are more

numerous in the highest regions of Jujuy, where even at present the Amerindian genetic component is highly predominant (Dipierrri et al., 2000). On the other hand, Jujuy region has been characterized by substantially low population densities over time, notably in those areas of more difficult access, which obviously can have promoted genetic drift processes.

Polymorphic *Alu* insertions constitute a family of genetic markers featuring a series of properties that make them unique loci to investigate human evolutionary history and conduct population genetic studies. Recently inserted human-specific *Alu* repeats are characterized by a biallelic, codominant pattern of insertion-lack of insertion inheritance reflecting common ancestry, the ancestral state being the absence of the insertion (Batzer et al., 1994; Stoneking et al., 1997; Batzer and Deininger, 2002). Accordingly, *Alu* markers shared by different individuals should be identical by descent not just by state: if two individuals share an *Alu* insertion, it is most likely that they share a common ancestor in whom the insertion took place. In addition, the extremely low probability that a complete deletion of an *Alu* loci occurs is another advantage of these molecular polymorphisms. No known mechanism exists for the complete and specific removal of an *Alu* element (Edwards and Gibbs, 1992); therefore, the lack of insertion represents in all likelihood the ancestral state. Furthermore, bearing in mind that the rate of insertion and fixation of new *Alu* repeats has been estimated in ~100-200 per Myr (Willard et al., 1987; Britten, 1997), the probability of two independent *Alu* elements being inserted by chance in the same genomic position is practically zero (Batzer et al., 1994; Roy-Engel et al., 2002). Finally, *Alu* markers are considered to be selectively neu-

tral, so that their frequency distribution in human populations will be conditioned solely by genetic drift and gene flow (Battilana et al., 2006, Terreros et al., 2009).

This work was intended to explore the potential influence of physical landscape on the genetic profile of the human populations. To that end, we analyzed a set of eight polymorphic *Alu* insertions in a subdivided human settlement from the province of Jujuy, which is composed of several population centers with well-defined geographical characteristics and marked contrasts between them mainly associated with differences in altitude above sea level.

MATERIALS AND METHODS

Subjects

With the purpose of characterizing the genetic variability of Jujuy, 226 blood samples from healthy and unrelated individuals were collected in different localities of the targeted province. The ancestry of all voluntary donors was verified by biographical information traced back at least three generations (mainly birthplaces of the donor's parents and grandparents). Ethical guidelines for research with human beings were adhered to as stipulated by each of the institutions involved in the study (Universidad del País Vasco, Spain, and Universidad Nacional de Jujuy, Argentina). The study protocol was approved by the Institutional Review Board from Universidad del País Vasco.

To analyze the genetic structure of Jujuy province with a high resolving power, samples were grouped according to geographical origin in five regions. This regional division was the reference in our work to assess the spatial variation of the *Alu* diversity within the province of Jujuy (Fig. 1). Thus, the subtropical valley was divided in two regions namely Selva ($N = 45$), with a mean altitude of 700 m, and Valle ($N = 62$, mean altitude: 1,200 m). Likewise, the Pre Puna region was divided in two zones, Quebrada Baja ($N = 36$, mean altitude: 2,500 m) and Quebrada Alta ($N = 36$, mean altitude: 3,300 m). Finally, those individuals living at a mean altitude >3300 m were all grouped in La Puna region ($N = 47$).

DNA amplification

A set of eight polymorphic *Alu* insertions (A25, ACE, APO, B65, D1, FXIIIIB, PV92, and TPA25) was analyzed. Genomic DNA was extracted from blood stains using the standard phenol:chloroform procedure (Sambrook et al., 1989) and stored at -20°C when not in use. Polymerase chain reactions (PCRs) were performed in a final volume of 10 μL , starting from 20 ng of DNA. The amplification conditions, annealing temperatures and primers for each *Alu* insertion can be consulted in a previous publication (García-Obregón et al., 2006). Analysis of the PCR products was performed by electrophoresis in 1.5% agarose gels. DNA bands were visualized by staining with ethidium bromide (0.5 $\mu\text{L}/\text{mL}$) and photographed under UV light. The quality of both the PCR and the electrophoresis was checked by including a positive control (homozygote for the insertion) and a negative control containing all the components of the PCR mix excepting the template DNA. Allelic identification was carried out using a Roche DNA Molecular Weight Marker V ladder. Because of the variable quantity and quality of the DNA samples, we could

not type the whole set of *Alu* insertions in all cases, thus obtaining discrepant sample sizes in different *Alu* markers.

Statistical analysis

Allelic frequencies and their standard errors for the eight *Alu* loci in the five populations (regions from Jujuy) examined, as well as observed and expected heterozygosity were calculated using version 3.0 of the Power Marker program (Liu and Muse, 2004). To test for Hardy-Weinberg Equilibrium (HWE), a Fisher's exact probability test was conducted to estimate *P*-values (Guo and Thompson, 1992) using Arlequin software, version 3.0 (Excoffier et al., 2005).

Genetic relationships among the collections from Jujuy were evaluated by computing Reynolds F_{ST} unbiased genetic distances (Reynolds et al., 1983) between all pairs of populations. Subsequently, the resultant F_{ST} genetic distance matrix was used to construct a dendrogram based on the Neighbor-Joining (NJ) method (Saitou and Nei, 1987), using the Phylip v3.2 program (Felsenstein, 1989). The statistical robustness of the consensus NJ tree was ascertained by bootstrapping with 1,000 replicates (Felsenstein, 1985). The genetic distance matrix was also used to perform a population pairwise F_{ST} analysis with the Arlequin software.

We further investigated the spatial trend of the allele frequency distributions in a system of moving coordinates to detect potential frequency clines. This analysis is based on the calculation of linear regressions between the allelic frequencies in the populations and their geographical coordinates with regard to an axis that is rotated in successive iterations. In those cases where, for the same allele, more than one statistically significant association is obtained between allele frequencies and spatial position, we select that direction of the axis for which the coefficient of determination (R^2) produces its maximum value (Peña et al., 2009). In addition, Mantel tests of matrix correspondence (Mantel, 1967) were applied to assess the concordance between genetic vs. geographical distances and between genetic distances vs. differences in altitude for the Jujuy collections examined.

To assess the relative influence of isolation, gene flow and genetic drift on the gene pool of each of the targeted populations from Jujuy, the expected Hardy-Weinberg proportions of heterozygotes of each sample were plotted against the distance from the centroid (Harpending and Ward, 1982). Bearing in mind the potential effects of genetic drift, a strong negative relationship is expected between population divergence and heterozygosity. Therefore, it is assumed that populations that plot above the theoretical prediction given by the regression line have experienced above average levels of gene flow, whereas those that fall below the regression line have remained relatively isolated.

To assess the impact of the gene flow in the gene pool of the different Jujuy subpopulations, admixture proportions were estimated through the Weighted Least Squares method (Long et al., 1991). To that end, allele frequency data on American, European and African populations were compiled from previously published works (Gómez-Pérez et al., 2010; Gayà-Vidal et al., 2010).

Finally, we analyzed the genetic relationships of Jujuy in a broader geographical context by compiling *Alu* data

TABLE 1. Allele frequencies with standard errors ($\pm SE$), observed (H_o) and expected (H_e) heterozygosities and average heterozygosity for eight polymorphic *Alu* insertions in five regions from the province of Jujuy (Argentina)

Locus		ACE	TPA25	PV92	APO	FXIIIIB	D1	A25	B65	Average
Population	2N ^a	90	94	90	84	94	90	94	88	
La Puna	Freq	0.844	0.713	0.944	1.000	0.926	0.322	0.085	0.182	
	SE	0.039	0.052	0.022	0.000	0.035	0.049	0.029	0.045	
	H _o	0.222	0.234	0.111	0.000	0.021	0.378	0.128	0.182	0.160
	H _e	0.263	0.410	0.105	0.000	0.138	0.437	0.156	0.298	0.226
	2N	72	72	72	72	72	72	72	72	
Quebrada Alta	Freq	0.792	0.625	0.819	0.986	0.903	0.444	0.042	0.319	
	SE	0.053	0.058	0.053	0.013	0.030	0.044	0.022	0.058	
	H _o	0.194	0.361	0.028	0.028	0.194	0.667	0.083	0.306	0.233
	H _e	0.330	0.469	0.296	0.027	0.176	0.494	0.080	0.435	0.288
	2N	72	70	68	70	70	72	72	70	
Quebrada Baja	Freq	0.792	0.543	0.809	1.000	0.829	0.500	0.069	0.371	
	SE	0.046	0.051	0.043	0.000	0.036	0.057	0.027	0.043	
	H _o	0.306	0.571	0.324	0.000	0.343	0.472	0.139	0.629	0.348
	H _e	0.330	0.496	0.309	0.000	0.284	0.500	0.129	0.467	0.315
	2N	122	122	124	124	124	124	124	124	
Valle	Freq	0.738	0.672	0.734	0.976	0.782	0.476	0.137	0.387	
	SE	0.034	0.041	0.037	0.013	0.037	0.043	0.031	0.041	
	H _o	0.459	0.426	0.403	0.048	0.307	0.468	0.210	0.484	0.351
	H _e	0.387	0.441	0.391	0.047	0.341	0.499	0.237	0.475	0.352
	2N	90	90	90	90	90	90	90	90	
Selva	Freq	0.756	0.667	0.700	1.000	0.800	0.600	0.122	0.367	
	SE	0.048	0.048	0.049	0.000	0.043	0.041	0.034	0.045	
	H _o	0.267	0.400	0.333	0.000	0.267	0.622	0.200	0.511	0.325
	H _e	0.369	0.444	0.420	0.000	0.320	0.480	0.215	0.464	0.339

^a2N, sample size in number of chromosomes analyzed.

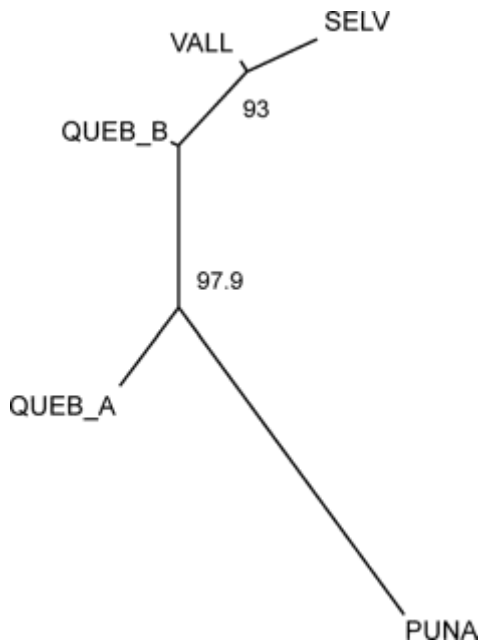


Fig. 2. Neighbor-Joining tree based on the F_{ST} genetic distance matrix. Genetic distances were computed from allelic frequencies of eight *Alu* insertions in five regions from the province of Jujuy (Argentina). Figures are percentage bootstrap values, estimated from 1,000 replicates. Population abbreviations are SELV (Selva), VALL (Valle), QUEB_B (Quebrada Baja), QUEB_A (Quebrada Alta), and PUNA (La Puna).

for other Native American groups from previously published works. Populations included in the phylogenetic analysis were Greenland Inuit, Alaskan Inuit, Mvskoike, Maya (Stoneking et al., 1997), Yanomamo (Romualdi

et al., 2002), Quechua and Aymara (Gayà-Vidal et al., 2010). To that end, nonmetric multidimensional scaling (MDS) analysis was conducted to represent the F_{ST} genetic distance matrix in two-dimensional space using the SPSS v.16 statistical package (SPSS Inc., Chicago, IL).

RESULTS

Allele frequencies, standard errors ($\pm SE$), and the corresponding heterozygosity values are summarized in Table 1. The highest frequencies were found for APO, which was the only *Alu* element fixed for the insertion, specifically in La Puna, Quebrada Baja, and Selva. Likewise, the lowest frequency values were observed in A25. Observed heterozygosity values ranged between zero (APO locus) in the three regions where it was monomorphic for the insertion, and 0.67 for D1 in Quebrada Alta. Significant departures from the HWE expectations were found through the Fisher's exact probability test in five of the 40 tests performed (12.5%). Yet, after applying the conservative Bonferroni correction for multiple testing only two tests (5.0%) remained in significant disagreement with the expected HWE frequency values (Bonferroni corrected 5% threshold value: 0.00125): FXIIIIB in La Puna and PV92 in Quebrada Alta. As can be inferred from the heterozygosity values, heterozygous deficit seems to be the causative agent of the disequilibrium in both cases. Interestingly, it was also detected a decreasing trend of the mean heterozygosity by region as the altitude increased. Both these findings could be indicative of the low genetic variability of the populations living at the highest altitudes (La Puna and Quebrada Alta) relative to those settled at intermediate and low altitude levels (Quebrada Baja, Valle, and Selva), probably reflecting a certain degree of genetic isolation of the highest population nuclei.

Genetic affinities among populations were assessed using NJ analysis based on F_{ST} genetic distances (see Fig. 2). The resultant topology of the phylogenetic tree



Fig. 3. Frequency clines found in Jujuy province (Argentina) for five *Alu* insertions. Arrows indicate the direction of the increasing gradients of allele frequencies. Also represented are the coefficient of determination of the linear regression (R^2) and the statistical significance of the regression. The length of the lines is proportional to the Wahlund variance of each *Alu* loci. Circumference radius represents the highest value estimated for the Wahlund variance, found in PV92.

notably overlapped the altitudinal distribution of the different collections from Jujuy. Thus, one distinctive cluster grouped those population samples located at the lowest altitude levels (Valle and Selva), whereas a second major clade grouped the subpopulations from Quebrada Alta and La Puna, the highest human settlements. The position of both clusters on the dendrogram was statistically supported by the high bootstrap values obtained for their corresponding tree nodes (93.0 and 97.9%, respectively), based on 1,000 iterations. Quebrada Baja segregated in an approximately intermediate position between the two main clades.

A population pairwise F_{ST} analysis gives further support to the genetic relationships depicted in the NJ tree. Thus, statistically significant differences were found in the following comparisons: (i) La Puna vs. Quebrada Baja for B65 ($F_{ST} = 0.064$, $P < 0.05$), (ii) La Puna vs. Valle for B65 ($F_{ST} = 0.077$, $P < 0.05$), D1 ($F_{ST} = 0.042$, $P < 0.05$), and PV92 ($F_{ST} = 0.115$, $P < 0.0001$), and (iii) La Puna vs. Selva for B65 ($F_{ST} = 0.069$, $P < 0.05$), PV92 ($F_{ST} = 0.144$, $P < 0.001$), and D1 ($F_{ST} = 0.136$, $P < 0.05$). In all other comparisons no statistically significant differences were found (see Supplementary Table S1). These results point

to the existence of genetic heterogeneity among the subpopulations from the province of Jujuy, most likely associated with the altitude gradient that their human settlements delineate.

For a thorough examination of the geographical distribution of the *Alu* frequencies, we analyze how the observed genetic heterogeneity is spatially oriented by computing linear regressions between *Alu* allele frequencies (dependent variable) and the geographical coordinates of the populations relative to an axis that rotates in successive iterations. Allele frequency clines were detected for 5 of the *Alu* markers, namely ACE, FXIII B, B65, PV92 and D1 (Fig. 3). On the other hand, TPA25, APO, and A25 showed frequency distributions that seem to be largely determined by random processes. The frequency gradients obtained showed, for the most part, a southeast-northwest orientation, consistent with the spatial trend of the increase in altitude above sea level in Jujuy province. The exception was the frequency cline for FXIII B, with a predominant East-West orientation. The highest frequencies for FXIII B, ACE and PV92 were observed in the highest regions, characterized by a more marked Native American genetic component. This is in ac-

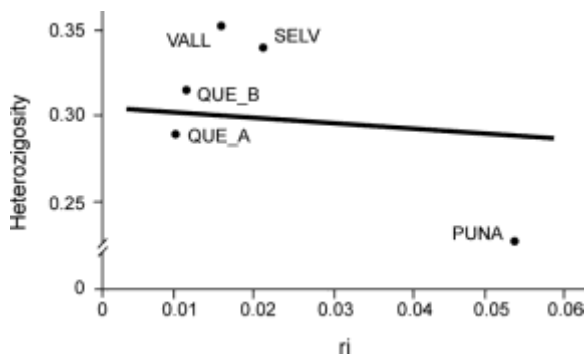


Fig. 4. Plot of heterozygosity versus distance from genetic centroid for five regions from the province of Jujuy (Argentina). The line represents the expected relationship predicted by the model of Harpending and Ward (1982). Population abbreviations are SELV (Selva), VALL (Valle), QUE_B (Quebrada Baja), QUE_A (Quebrada Alta), and PUNA (La Puna).

cordance with the fact of these three *Alu* markers being more frequent in Amerindians than in European and African groups (Stoneking et al., 1997; Battilana et al., 2006; Gaya-Vidal et al., 2010). The same reasoning applies to explain the orientation of the allele frequency gradient detected for B65, with lower frequencies in Native Americans than in Europeans and Africans. The slight difference in the orientation of FXIIIIB might be explained by small local fluctuations on account of genetic drift. Finally, frequency range observed in Jujuy for D1 proved to be greater than the reported one between continents, so that the frequency gradient could be consequence of high isolation and inbreeding levels in the highest Jujuy subpopulations. Worthy of mention is the fact that the *Alu* elements showing the highest values of Wahlund variance (PV92, B65, and D1) were also those that rendered statistically significant F_{ST} differences between some of the Jujuy subpopulations.

We further explored whether the spatial distribution of the *Alu* diversity was more influenced by the geographical position of the different regions from Jujuy or by their differences in altitude. To that end, two Mantel tests of matrix correspondence were performed: “ F_{ST} genetic distances” vs. “geographical distances” and “ F_{ST} genetic distances” vs. “altitude differences.” Geographical position and altitude above sea level of the more representative demographic centers of each region were considered in all cases. The F_{ST} genetic distances proved to be correlated with both the geographical position ($R = 0.951, P < 0.01$) and the differences in altitude ($R = 0.497, P < 0.05$).

The amount of gene flow experienced by the different regions from Jujuy was assessed by the centroid method (Fig. 4). The regression line represents the expected heterozygosity, so that subpopulations plotting above the line are postulated to have received above average levels of gene flow, whereas those positioned under the regression line would have been less impacted by gene flow processes and, therefore, would have evolved in a relatively isolated milieu. Considering the positions of the different subpopulations relative to the line, it can be inferred that the weakest impact of the gene flow on the populations’ gene pools has taken place in La Puna and Quebrada Alta samples, in accordance with previously described findings.

TABLE 2. Admixture proportions (\pm standard error) from Native American, European, and African groups in the gene pool of five populations from the province of Jujuy (Argentina)

Populations	Admixture proportions ^a		
	μ 1 (Native American) ^b	μ 2 (European) ^c	μ 3 (African) ^d
La Puna	1.000 (\pm 0.117)	—	—
Quebrada Alta	0.920 (\pm 0.053)	0.047 (\pm 0.115)	0.033 (\pm 0.122)
Quebrada Baja	0.875 (\pm 0.072)	—	0.125 (\pm 0.122)
Valle	0.771 (\pm 0.036)	0.163 (\pm 0.078)	0.065 (\pm 0.083)
Selva	0.774 (\pm 0.071)	0.226 (\pm 0.153)	—

^aEstimated by the Weighted Least Squares method (Long et al. 1991).

^bNative American populations: Yanomamo (Romualdi et al. 2002), Maya (Stoneking et al. 1997), Aymara and Quechua from Tinguipaya (Gaya-Vidal et al. 2010).

^cEuropean populations: Basques, Valencia (García-Obregón et al., 2006, 2007), Catalonia, and Andalusia (Comas et al., 2000).

^dAfrican populations: Nguni, Sotho-Tswana (Stoneking et al. 1997) and Bantu (Romualdi et al. 2002).

Specifically, we found a conspicuous negative relationship between heterozygosity level and distance from the genetic centroid for La Puna collection, as expected for populations that have diverged mainly by genetic drift effects (Harpending and Ward, 1982). The centroid plot indicates that the rest of the subpopulations (Valle, Selva, and Quebrada Baja) have received above average levels of gene flow.

The lowest admixture rates were obtained for the highest regions (Quebrada Alta and La Puna), as revealed by the genetic admixture analysis (Table 2). In these populations, proportions of Native American genes were 92 and 100%, respectively. In general terms, a growing importance of both the African, and particularly, the European contribution to the gene pool of Jujuy was perceptible as the altitude decreased.

Genetic relationships of Jujuy with other Amerindian groups were assayed by using nonmetric MDS based on the F_{ST} genetic distance matrix. To that end, we selected those subpopulations from Jujuy characterized by a predominance of the Native American genetic component (La Puna and Quebrada Alta) and seven additional Native American collections with availability of data for the same set of *Alu* markers: Greenland Inuit, Alaskan Inuit, Myskoke, Maya, Yanomamo, Quechua, and Aymara. MDS results are depicted in Fig. 5. The two-dimensional representation of the F_{ST} matrix accounted for 98.5% of the total variance, with a coefficient of stress of 0.055. Populations were discriminated along dimension I according to geographical origin. Thus, the Amerindian samples from South America (Puna, Quebrada Alta, Yanomamo, Aymara, and Quechua) plotted in the negative values of dimension I, whereas Central and North American groups segregated at the positive segment of this axis. On the other hand, dimension II was mainly determined by the genetic heterogeneity between populations from the Andean plateau and tropical human groups, with the subpopulations from Jujuy placed in an intermediate position.

DISCUSSION

Landscape genetics is a discipline addressed to the understanding of how geographical and environmental features can structure genetic variation mainly at the population level (Manel et al., 2003). The two major goals

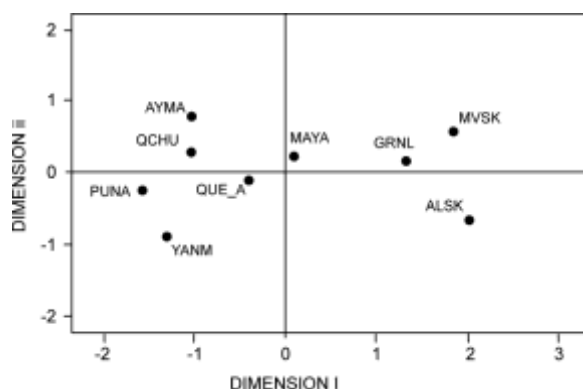


Fig. 5. Two-dimensional representation of a nonmetric MDS based on the Reynolds (F_{ST}) genetic distance matrix. Genetic distances were computed from allelic frequencies for eight *Alu* insertions in 10 Native American collections. The total variance accounted for is 98.5% with a coefficient of stress of 0.055. Populations abbreviations are: Alaskan Inuits (ALSK), Greenland Inuits (GRNL), Maya (MAYA), Mvskoke (MVSK), Yanomamo (YANM), Aymara (AYMA), Quechua (QCHU) and Quebrada Alta (QUE_A), and Puna (PUNA) from the province of Jujuy (Argentina).

of landscape genetics are the detection of genetic heterogeneity and its correlation with landscape and environmental features, such as physical barriers (e.g., mountains). The province of Jujuy (Argentina) is extraordinarily interesting for the study of the interaction between landscape features and microdifferentiation processes. The great variation in altitude above sea level between the population nuclei of the region have notably influenced on the genetic structure, probably as a result of the existence of geographical barriers to random mating (panmixia). This departure from panmixia has led to a certain degree of genetic isolation of some subpopulations on account of a restricted gene flow, thereby promoting genetic microdifferentiation because of the differential effects of the genetic drift according to altitude. Similarly, the differential impact of the gene flow according to altitude, which would be associated with differences in habitability conditions, would have resulted in a variable degree of genetic admixture across the province of Jujuy.

The microevolutionary processes that have occurred in Jujuy seem to be reflected in the significant genetic heterogeneity detected for several *Alu* insertions (B65, PV92, and D1) between the highest region (La Puna) and the lowest regions, especially Valle and Selva. Interestingly, PV92 has been reported to be an efficient ancestry informative marker to discriminate continental groups, notably between Europeans and Native Americans and between Africans and Native Americans (Gómez-Perez et al., 2010). The sensitivity and resolving power of various *Alu* markers, mainly PV92, proved to be effective enough to detect the differences in admixture rates among the subpopulations from Jujuy.

The average population density of Jujuy province is rather low (12.8 inh/km²). In addition, this population density is unequally distributed over the surveyed area, so that it is substantially lower in the most inaccessible regions, such as La Puna and Quebrada Alta (≈ 1.3 hab/km²), than in the rest of the subpopulations (≈ 27.8 hab/km²) (INDEC, 2001). The combined effects of genetic drift and a relatively high endogamy level have resulted in a loss of genetic variability in the highest human settle-

ments, which has led to a decrease of their heterozygous proportions. Thus, the average heterozygosity for the eight *Alu* insertions proved to be lower in La Puna and Quebrada Alta than in those regions located at the lowest altitudes. In all probability, the greater proportion of endogamous matings in the highest villages could have altered the HWE genotypic proportions due to heterozygous deficit, as found for FXIIB in La Puna and PV92 in Quebrada Alta. These results are consistent with the ones reported by Alfaro and Dipierri (1996), on having analyzed the characteristics of the civil marriages recorded in the province of Jujuy during 1990. These authors observed a direct relationship between geographical altitude, coefficient of consanguinity and local endogamy and an inverse relationship of the altitude with the exogamy and the marital distance.

In accordance with the findings discussed earlier, data generated by the centroid method unveiled a gene flow below the average in the subpopulations living at the highest altitudes, whereas those subpopulations located at the lowest altitudes were characterized by gene flow levels above the average. This is indicative of a higher proportion of foreign genes in the subpopulations from the lowest regions, which were predominantly introduced by immigrants of European origin and, to a lesser extent, by individuals with African ancestry (Table 2). Admixture levels obtained from *Alu* insertions proved to be similar to the estimated ones in previous studies using different types of genetic markers, such as immunoglobulin allotypes, serum polymorphisms (Dipierri et al., 2000), mitochondrial DNA and Y-chromosome polymorphisms (Dipierri et al., 1998). Specifically, Dipierri et al. (2000) observed a virtually identical pattern of admixture rates decrease according to altitude, albeit with slightly lower values. The immigrants primarily settled in those regions situated at the lowest altitude levels, which propitiated a greater degree of genetic admixture between the different ethnic groups. On the contrary, autochthonous, Native American individuals clearly predominate in the highest and more uncomfortable regions from Jujuy. This important bias in the spatial distribution of the immigrant population underlies the altitudinal admixture gradient and the gradual genetic microdifferentiation pattern, both of them reflected in the several frequency clines identified for the *Alu* insertions. Indeed, four of the five *Alu* frequency clines found in our study showed a southeast-northwest orientation, coinciding with the direction of altitude increase in Jujuy and not with the tracing of the region's major road, the Quebrada de Humahuaca, which is oriented from south to north. Nevertheless, the geographical distance between the subpopulations cannot be ruled out as an important influential factor on the genetic microdifferentiation processes that have occurred in the province of Jujuy. Along these lines, results of the Mantel test of matrix correspondence revealed that genetic distances were more strongly correlated with geographical distances than with altitude differences. This finding is supportive of the *Alu* frequencies being distributed in Jujuy province according to the isolation by distance model, although partially masked by the peculiarities of the physical environment of the region.

Another major consequence of the differential impact of the gene flow is that the human groups settled in La Puna and Quebrada Alta have preserved the Native American substrate of their gene pools. Both of these subpopulations showed strong genetic affinities with the other South American collections involved in the phylogenetic analysis (Fig. 5).

As we have previously seen, subpopulations from Jujuy plotted midway between the Andean samples (Quechua and Aymara) and the Yanomamo, most likely because of the geographical position of Jujuy, in a natural corridor between the Andean plateau and the South American forests.

The findings of our study reveal a spatial patterning of the *Alu* diversity in Jujuy, most likely determined by disparities in landscape and environmental features between the regions. Such differences in the physical environment would have drastically reduced the homogenizing effects of the gene flow and, at the same time, would have promoted severe genetic drift episodes in those populations living at the highest altitudes (mainly in La Puna region). In all likelihood, divergence in habitability conditions between the lowest and highest population nuclei influenced the spatial distribution of the immigrants (European colonists and African slaves), which mostly concentrated in the villages settled at the lowest altitude levels. This bias in the spatial distribution of the non-native genes underlies the gradient of genetic admixture rates, characterized by a decreasing trend according to altitude. From all the earlier, it can be concluded that the microevolutionary process detected in Jujuy results from the combined effect of differential gene flow and genetic drift effects, which have played an important role in the shaping of the gene pool of the populations currently living in this sub-Andean zone from Argentina. On the other hand, *Alu* insertions have proved to be robust markers for evolutionary studies, allowing the detection of spatial genetic patterns (clines) and sound assumptions of migration patterns and the differential impact of the gene flow.

ACKNOWLEDGMENTS

We thank all voluntary donors, who cooperated generously to the development of this study.

LITERATURE CITED

- Alfaro E, Dipierri JE. 1996. Isonimia, endogamia, exogamia y distancia marital en la provincia de Jujuy. *Rev Arg Antropol Biol* 1:41–56.
- Batzler MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL. 1994. African origin of human-specific polymorphic *Alu* insertions. *Proc Natl Acad Sci USA* 91:12288–12292.
- Batzler MA, Deininger PL. 2002. *Alu* repeats and human genomic diversity. *Nat Rev Genet* 3:370–379.
- Battilana J, Fagundes NJ, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E, Munkhbat B, Munkhtuvshin N, Krylov M, Benevolenskaia L, Arnett FC, Batzler MA, Deininger PL, Salzano FM, Bonatto SL. 2006. *Alu* insertion polymorphisms in Native Americans and related Asian populations. *Ann Hum Biol* 33:142–160.
- Britten RJ. 1997. Mobile elements inserted in the distant past have taken on important functions. *Gene* 205:177–182.
- Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzler MA, Bertranpetit J, Sajantila A. 2000. *Alu* insertion polymorphisms in NW Africa and the Iberian Peninsula: Evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* 107:312–319.
- Curtin PD. 1969. *The Atlantic slave trade: A census*. Milwaukee: University of Wisconsin Press. p338.
- Dipierri JE, Alfaro E, Martínez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO. 1998. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol* 70:1001–1010.
- Dipierri JE, Alfaro E, Peña JA, Constans J, Dugoujon JM. 2000. GM, KM immunoglobulin allotypes and other serum genetic markers (HP, GC, PI, and TF) among South American populations living at different altitudes (Jujuy Province, Argentina): Admixture estimates. *Hum Biol* 72:305–319.
- Edwards MC, Gibbs RA. 1992. A human dimorphism resulting from loss of an *Alu*. *Genomics* 14:590–597.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinf Online* 1:47–50.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.
- Felsenstein J. 1989. PHYLIP: Phylogeny inference package (version 3.2). *Cladistics* 5:164–166.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, Vidales C, Arroyo D, Pena JA. 2006. Genetic position of Valencia (Spain) in the Mediterranean basin according to *Alu* insertions. *Am J Hum Biol* 18:187–195.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, de Pancorbo MM, Pena JA. 2007. Polymorphic *Alu* insertions and the genetic structure of Iberian Basques. *J Hum Genet* 52:317–327.
- Gayà-Vidal M, Dugoujon JM, Esteban E, Athanasiadis G, Rodríguez A, Villena M, Vasquez R, Moral P. 2010. Autosomal and X chromosome *Alu* insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. *Am J Hum Biol* 22:154–162.
- Gomes MV. 2006. La presencia negrafricana en la Argentina: pasado y permanencia, por. Boletín digital de la Biblioteca del Congreso, N° 9.
- Gómez-Pérez L, Alfonso-Sánchez MA, Pérez-Miranda AM, García-Obregón S, Builes JJ, Bravo ML, De Pancorbo MM, Peña JA. 2010. Genetic admixture estimates by *Alu* elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia. *Ann Hum Biol* 37:488–500.
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.
- Harpending HC, Ward RH. 1982. Chemical systematics and human populations. In: Nitecki M, editors. *Biochemical aspects of evolutionary biology*. Chicago: University of Chicago. p 213–256.
- INDEC. 2001. Censo Nacional de Población y Vivienda 1991 y Censo Nacional de Población, Hogares y Viviendas. Available at: <http://www.indec.gov.ar/webcenso/index.asp>.
- Jorde LB. 1982. The genetic structure of the Utah Mormons: migration analysis. *Hum Biol* 54:583–597.
- Liu K, Muse SV. 2004. PowerMarker: New genetic data analysis software. Version 3.0. Available at: <http://www.powermarker.net>.
- Long JC, Williams RC, McAuley JE, Medis R, Partel R, Tregellas WM, South SF, Rea AE, McCormick SB, Iwaniec U. 1991. Genetic variation in Arizona Mexican Americans: Estimation and interpretation of admixture proportions. *Am J Phys Anthropol* 84:141–157.
- Lovejoy PE. 1982. The volume of the Atlantic slave trade: A synthesis. *J Afr Hist* 23:473–501.
- Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol* 18:189–197.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220.
- Martínez-Sarasola C. 1992. Nuestros paisanos los indios. Buenos Aires: EMECE.
- Peña JA, Alfonso-Sánchez MA, Pérez-Miranda AM, García-Obregón S, Gómez-Pérez L. 2009. GeDis: Un programa para análisis de datos en Antropogenética. *Antropo* 20:49–59. Available at: <http://www.didac.edu.es/antropo>.
- Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: bases for a short term genetic distance. *Genetics* 105:767–779.
- Relethford J. 1991. Effect of changes in population size on genetic microdifferentiation. *Hum Biol* 63:629–641.
- Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzler MA, Barbujani G. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res* 12:602–612.
- Rothhammer F, Cocilovo JA, Quevedo S. 1984. El poblamiento temprano de Sudamérica. *Chungara*. 13:99–108.
- Rothhammer F, Silva C. 1992. Gene geography of South America: testing models of population displacement based on archeological evidence. *Am J Phys Anthropol* 89:441–446.
- Roy-Engel AM, Carroll ML, El-Sawy M, Salem AH, Garber RK, Nguyen SV, Deininger PL, Batzler MA. 2002. Non-traditional *Alu* evolution and primate genomic diversity. *J Mol Biol* 316:1033–1040.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzler MA. 1997. *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* 7:1061–1071.
- Terreros MC, Alfonso-Sánchez MA, Novick GE, Luis JR, Lacau H, Lowery RK, Regueiro M, Herrera RJ. 2009. Insights on human evolution: an analysis of *Alu* insertion polymorphisms. *J Hum Genet* 54:603–611.
- Willard C, Nguyen HT, Schmid CW. 1987. Existence of at least three distinct *Alu* subfamilies. *J Mol Evol* 26:180–186.

Supplementary Table S1. Statistical significance values (*P* values) for the *F_{ST}* pairwise comparisons of *Alu* allele frequencies between five regions of Jujuy province (Argentina)

A25 <i>Alu</i> Insertion				
	PUNA	QUE ALTA	QUE BAJA	VALLE
PUNA	*			
QUE ALTA	0.658 ± 0.047	*		
QUE BAJA	0.676 ± 0.037	0.991 ± 0.003	*	
VALLE	0.450 ± 0.033	0.171 ± 0.025	0.234 ± 0.041	*
SELVA	0.721 ± 0.053	0.414 ± 0.031	0.469 ± 0.057	0.784 ± 0.039

APO <i>Alu</i> Insertion				
	PUNA	QUE ALTA	QUE BAJA	VALLE
PUNA	*			
QUE ALTA	0.757 ± 0.039	*		
QUE BAJA	0.784 ± 0.043	0.991 ± 0.003	*	
VALLE	0.198 ± 0.040	0.514 ± 0.027	0.469 ± 0.035	*
SELVA	0.451 ± 0.041	0.694 ± 0.049	0.757 ± 0.031	0.991 ± 0.003

ACE <i>Alu</i> Insertion				
	PUNA	QUE ALTA	QUE BAJA	VALLE
PUNA	*			
QUE ALTA	0.432 ± 0.045	*		
QUE BAJA	0.991 ± 0.003	0.991 ± 0.003	*	
VALLE	0.991 ± 0.003	0.991 ± 0.003	0.991 ± 0.003	*
SELVA	0.991 ± 0.003	0.487 ± 0.049	0.991 ± 0.003	0.991 ± 0.003

B65 <i>Alu</i> Insertion				
	PUNA	QUE ALTA	QUE BAJA	VALLE
PUNA	*			
QUE ALTA	0.234 ± 0.034	*		
QUE BAJA	0.036 ± 0.020	0.658 ± 0.039	*	
VALLE	0.009 ± 0.009	0.496 ± 0.069	0.991 ± 0.003	*
SELVA	0.045 ± 0.020	0.721 ± 0.043	0.991 ± 0.003	0.991 ± 0.003

D1 <i>Alu</i> Insertion				
	PUNA	QUE ALTA	QUE BAJA	VALLE
PUNA	*			
QUE ALTA	0.333 ± 0.039	*		
QUE BAJA	0.072 ± 0.023	0.865 ± 0.034	*	
VALLE	0.045 ± 0.015	0.784 ± 0.036	0.991 ± 0.003	*
SELVA	0.018 ± 0.012	0.162 ± 0.049	0.550 ± 0.039	0.189 ± 0.029

FXIIB <i>Alu</i> Insertion				
	PUNA	QUE ALTA	QUE BAJA	VALLE
PUNA	*			
QUE ALTA	0.991 ± 0.003	*		
QUE BAJA	0.279 ± 0.034	0.288 ± 0.042	*	
VALLE	0.090 ± 0.024	0.144 ± 0.041	0.784 ± 0.033	*
SELVA	0.117 ± 0.024	0.234 ± 0.036	0.802 ± 0.044	0.991 ± 0.003

PV92 <i>Alu</i> Insertion				
	PUNA	QUE ALTA	QUE BAJA	VALLE
PUNA	*			
QUE ALTA	0.090 ± 0.024	*		
QUE BAJA	0.072 ± 0.023	0.991 ± 0.003	*	
VALLE	0.000 ± 0.000	0.451 ± 0.053	0.559 ± 0.056	*
SELVA	0.009 ± 0.009	0.460 ± 0.050	0.487 ± 0.043	0.802 ± 0.035

TPA25 <i>Alu</i> Insertion				
	PUNA	QUE ALTA	QUE BAJA	VALLE
PUNA	*			
QUE ALTA	0.460 ± 0.029	*		
QUE BAJA	0.108 ± 0.030	0.667 ± 0.051	*	
VALLE	0.667 ± 0.043	0.649 ± 0.038	0.234 ± 0.034	*
SELVA	0.694 ± 0.058	0.856 ± 0.039	0.378 ± 0.045	0.991 ± 0.003

Significant values (*p* < 0.05) are highlighted in bold.

PRINCIPALES RESULTADOS

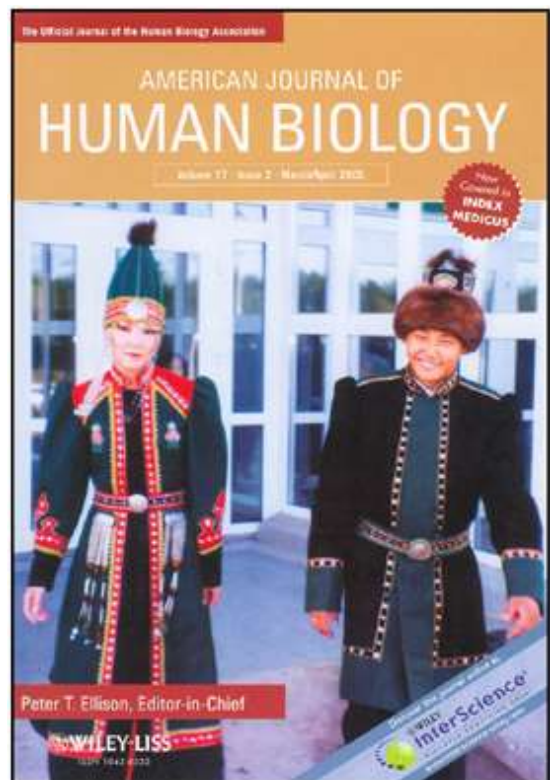
En general, se observa que las poblaciones de las diferentes comarcas analizadas son en su mayor parte de origen amerindio, siempre en una proporción superior al 77%, con aportes bajos y en algunos casos nulos de genes europeos y africanos, algo detectado mediante otros marcadores analizados (Dipierri *et al.* 2000; Cardoso *et al.* 2013a). El grado de mestizaje es variable, de modo que en las poblaciones situadas a mayor altitud, el mestizaje es menor que en las que se encuentran a más baja altitud. Esto se debe principalmente al asentamiento diferencial de los colonos europeos y los esclavos africanos, como ya se ha detallado anteriormente.

En las poblaciones procedentes de las regiones más altas se observa una menor diversidad genética, pudiendo establecerse una relación inversa entre altitud y diversidad genética. Las causas más probables serán el aislamiento secular de estas poblaciones debido a lo inaccesible del territorio y una muy baja densidad de población.

Precisamente por la asociación de estos factores se observa también una ausencia de equilibrio H-W en algunos marcadores (FXIII B y PV92 en Puna y Quebrada Alta respectivamente), fruto muy posiblemente de un elevado grado de endogamia, lo que puede alterar las frecuencias genotípicas de las poblaciones.

Todo ello revela una tendencia a la microdiferenciación entre las subpoblaciones de las diferentes regiones de Jujuy, con diferencias estadísticamente significativas entre las comarcas altas y las bajas (test de diferenciación F_{ST} *pairwise*) y una serie de clinas encontradas en varios alelos, con una clara orientación, en función de la altitud, en dirección Sudeste/Noroeste.

En definitiva, en este trabajo se muestran los efectos que una barrera geográfica importante puede causar sobre el patrimonio genético de una población subdividida, revelándose los procesos microevolutivos generados por una intensa acción de la deriva genética y un escaso flujo génico.



MANUSCRITO IV

***Alu* Polymorphisms in the Waorani Tribe from the Ecuadorian Amazon Reflect the Effects of Isolation and Genetic Drift**

Gómez-Pérez L, Alfonso-Sánchez MA, Sánchez D, García-Obregón S, Espinosa I, Martínez-Jarreta B, De Pancorbo MM, Peña JA. 2011.

American Journal of Human Biology 23, 790-795

INTRODUCCIÓN

La etnia Huaorani es una población de origen amerindio con unas características específicas, que la hacen particularmente interesante para su estudio en Antropogenética.

Y es que ésta población, que habita en la región amazónica de Ecuador, es probablemente una de las más aisladas de Sudamérica. Los Huaorani constituyen una de las sociedades humanas más belicosas, lo que se refleja en unas tasas de homicidios muy altas incluso entre grupos estrechamente relacionados. Los motivos varían entre disputas sobre acuerdos matrimoniales, venganzas por homicidios anteriores o acusaciones de brujería.

Debido a su agresividad y también a lo inaccesible de su entorno, los Huaorani no han tenido contactos importantes tradicionalmente con otros grupos de amerindios o europeos. De hecho, el primer contacto pacífico conocido con la cultura occidental no tuvo lugar hasta 1958 y algunos grupos todavía viven completamente aislados en la selva amazónica.

Por ello, se ha considerado que esta población completa el rango de posibilidades de interacción entre flujo génico y deriva representado por mestizos, afrocolombianos y jujeños, poblaciones que como los Huaorani, han sido analizados para A25, ACE, APO, B65, D1, FXIII B, PV92 y TPA25. En esta población se analizaron además los polimorfismos de inserción *Alu* HS3.23, HS4.32, HS4.59 y HS4.65, con el fin de aumentar la resolución de los análisis, aunque lamentablemente el número de poblaciones con las que establecer comparaciones se redujo considerablemente.

Original Research Article

Alu Polymorphisms in the Waorani Tribe from the Ecuadorian Amazon Reflect the Effects of Isolation and Genetic Drift

LUIS GÓMEZ-PÉREZ,¹ MIGUEL A. ALFONSO-SÁNCHEZ,¹ DORA SÁNCHEZ,² SUSANA GARCÍA-OBREGÓN,¹ IBONE ESPINOSA,¹ BEGOÑA MARTÍNEZ-JARRETA,³ MARIAN M. DE PANCORBO,⁴ AND JOSÉ A. PEÑA^{1*}

¹Departamento de Genética y Antropología Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), Spain

²Hospital Metropolitano, Laboratorio de Genética Molecular, Quito, Ecuador

³Departamento de Medicina Legal, Universidad de Zaragoza, Spain

⁴BIOMICS Research Group, Centro de Investigación y Estudios Avanzados "Lucio Lascaray", Universidad del País Vasco (UPV/EHU), Vitoria-Gasteiz, Spain

Objectives: The Amazon basin is inhabited by some of the most isolated human groups worldwide. Among them, the Waorani tribe is one of the most interesting Native American populations from the anthropological perspective. This study reports a genetic characterization of the Waorani based on autosomal genetic loci.

Methods: We analyzed 12 polymorphic *Alu* insertions in 36 Waorani individuals from different communal long-houses settled in the Yasuní National Park.

Results: The most notable finding was the strikingly reduced genetic diversity detected in the Waorani, corroborated by the existence of four monomorphic loci (ACE, APO, FXIII B, and HS4.65), and of other four *Alu* markers that were very close to the fixation for the presence (PV92 and D1) or the absence (A25 and HS4.32) of the insertion. Furthermore, results of the centroid analysis supported the notion of the Waorani being one of the Amerindian groups less impacted by gene flow processes.

Conclusions: The prolonged isolation of the Waorani community, in conjunction with a historically low effective population size and high inbreeding levels, have resulted in the drastic reduction of their genetic diversity, because of the effects of severe genetic drift. Recurrent population bottlenecks most likely determined by certain deep-rooted sociocultural practices of the Waorani (characterized by violence, internal quarrels, and revenge killings until recent times) are likely responsible for this pattern of diversity. The findings of this study illustrate how sociocultural factors can shape the gene pool of human populations. *Am. J. Hum. Biol.* 23:790–795, 2011. © 2011 Wiley Periodicals, Inc.

The western Amazon region in South America is home to a great diversity of indigenous ethnic groups, including some of the world's last uncontacted peoples that have lived in voluntary isolation (Finer et al., 2008). Among these isolated human groups, one of the most interesting from the anthropological viewpoint is the Waorani people.

The Waorani are an indigenous tribe that inhabits much of the lowland Amazon tropical forests (Wallis, 1973), with a demographic size between 1,500 and 2,000 individuals at present (Beckerman et al., 2009). The traditional Waorani territory was spread over an area of approximately 2 millions hectares, between the right bank of the Napo River and the left bank of Curaray River, in the eastern Ecuadorian provinces of Orellana, Napo, and Pastaza (see Fig. 1). This region includes the Yasuní National Park, a Pleistocene forest refugium of the Amazon basin, which together with the Waorani Ethnic Reserve delineates what the Ecuadorian government has called the Yasuní Biosphere Reserve.

The Waorani are descendants from hunter-gatherer groups, and they continue to subsist primarily on hunting and horticulture with a semi-nomadic lifestyle. While the traditional Waorani groups composed of 6–10 families practiced a nomadic lifestyle, moving regularly from one camp to another, current way of life is becoming increasingly sedentary (Rival, 2002). Regarding social organization, the Waorani society is a highly inbred and homogeneous population, according to findings from studies on the polymorphism of red cell enzymes, immunoglobulin

allotypes, and dermatoglyphics (Larrick et al., 1985). This is probably the consequence of a marked predominance of bilateral cross-cousin marriages, which have been usually arranged by the parents of the young couple (Robarchek and Robarchek, 1992; Beckerman et al., 2009). They nevertheless allowed, on occasion, polygamy (polygyny and, less frequently, polyandry), if there was any demographic imbalance because of warfare or any other catastrophic event (Holt et al., 2004).

The Waorani speak the *Wao Tededo*, an unclassified and isolated language without any congener even in the Amazon region (Peeke, 1973, 1979). This language is also known as Waorani, Wao Tiriro, Huao, Auishiri, and Sabela, whereas the autonym is Huao Tededo. Wao Tededo is considered unique in its linguistic construction, with no known similarities with Zaparoan phonology or structure (Orr et al., 1991; Lu, 2001; Rival, 2002).

Funding: This work was funded by the 'Programa de Subvención a Grupos Consolidados, Departamento de Educación, Universidades e Investigación' of the Basque Government (IT-424-07). L. Gómez-Pérez was supported by a grant of the University of the Basque Country.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Dr. José A. Peña. E-mail: joseangel.pena@ehu.es
Received 2 December 2010; Revision received 1 August 2011; Accepted 4 August 2011

DOI 10.1002/ajhb.21216

Published online 28 September 2011 in Wiley Online Library (wileyonlinelibrary.com).



Fig. 1. Map showing the geographic location of the Waorani Ethnic Reserve, the Yasuni National Park, and the Pastaza province in Ecuador.

Settled in a violent region of the world, the western Amazon basin, the Waorani have been considered one of the most warlike societies because of their strikingly high homicide rates (Yost, 1981; Beckerman et al., 2009). Scattered over their vast territory, the Waorani fought each other almost incessantly, on account of disputes over marriage arrangements, retaliations from past killings or indictments of sorcery even among closely related groups. In a study based on extensive genealogies (up to five generations back), Yost (1981) estimated that more than a half of adult deaths (> 60%) were caused by warfare, and specifically, 44% resulted from internal quarrels. Therefore, violence and revenge have been regarded as major determinants of population mortality among the Waorani (Wallis, 1973; Robarchek and Robarchek, 1992; Boster et al., 2003; Beckerman et al., 2009).

Because of their aggressiveness, the Waorani were not colonized by other neighboring indigenous groups or by foreigners. North American missionaries made the pioneering peaceful contact of western cultures with the Waorani community in 1958 (Wallis, 1973). Subsequently, they have had limited but increasing contact with outsiders, though some Waorani groups still live completely isolated in the Amazonian forest (Yost, 1981; Boster et al., 2003).

Obviously, both the behavioral and the linguistic features of the Waorani tribe have contributed to significant population isolation, which should be mirrored in the pattern of gene diversity in this Native American population. Here, we report the genetic characterization of the Waorani tribe through the analysis of a set of polymorphic *Alu* insertions. These data are used to explore the effects of the isolation and a very low demographic size in the genetic diversity of the population. A sizeable number of previous publications have thoroughly discussed the ad-

vantageous properties of the *Alu* family of repetitive elements in molecular studies with evolutionary perspective. Essentially, these properties are selective neutrality of polymorphic *Alu* insertions, identity by descent not just by state, lack of insertion as the ancestral state, and no known mechanism for the complete and specific removal of an *Alu* element (Batzer and Deininger, 2002; Terreros et al., 2009; Gómez-Pérez et al., 2011, among others). All these characteristics indicate that *Alu* loci are stable and conservative markers that reflect unique evolutionary events, and hence their usefulness in the reconstruction of the human demographic and evolutionary history.

MATERIALS AND METHODS

To assess the genetic diversity of the Waorani population, we screened 12 polymorphic *Alu* insertions (A25, ACE, APO, B65, D1, FXIIIIB, PV92, TPA25, HS3.23, HS4.32, HS4.59, and HS4.65). The sample consisted of 36 individuals (11 males and 25 females) living in different communal longhouses or *nanicabos* (traditional domestic groups) settled in the Yasuni National Park (Ecuador). The Yasuni National Park is a Neotropical rain forest, characterized by a warm and humid climate with an approximately constant temperature throughout the year, ranging between 18° and 24°C. We have estimated that the Waorani individuals included in the study belong to at least 12 different longhouses, bearing in mind the surname composition of the sample examined. Ethical guidelines for research with human beings were adhered to as stipulated by each of the institutions involved in the study. The Institutional Review Board from the University of the Basque Country approved the study protocol.

DNA amplification

Genomic DNA was extracted from bloodstains using the standard phenol–chloroform procedure (Sambrook et al., 1989) and stored at –20°C when not in use. PCRs were performed in a final volume of 10 µL, starting from 10 ng of DNA. Amplification conditions, annealing temperatures and primers for each *Alu* insertion can be consulted in previously published papers (Arcot et al., 1996; García-Obregón et al., 2006). PCR products were further electrophoresed in 1.5% agarose gels. The DNA bands obtained by electrophoresis were ethidium bromide-stained (0.5 µl/ml) for a better visualization and photographed under UV light. Allele identification was performed using a Roche DNA Molecular Weight Marker V ladder. Likewise, to control the quality of both the PCR and the electrophoresis, we included a positive control (homozygous for the insertion) as well as a negative control containing all the components of the PCR mix except the template DNA.

Statistical analysis

Allele frequencies with standard errors, as well as observed and expected heterozygosity for each *Alu* marker were calculated using Arlequin program v3.5 (Excoffier et al., 2010). Hardy–Weinberg Equilibrium (HWE) was tested by conducting a Fisher's exact probability test to estimate *P*-values (Guo and Thompson, 1992) with Arlequin.

To examine the genetic affinities of the Waorani with other Amerindian groups, *Alu* frequency data were compiled from the relevant literature. Populations included in

subsequent comparative analyses were Greenland Inuit, Alaskan natives, Mvskoke, Maya, Aché, Cinta Larga, Gaviao, Guarani, Kaingang, Quechua from Arequipa and from Tayacaja province in Peru, Surui, Xavante, Wai-Wai, Yanomami, and Zoró (Battilana et al. 2002, 2006). The allele frequency database was utilized to compute F_{ST} genetic distances (Reynolds et al., 1983) between all pairs of populations with the Phylip program (Felsenstein, 1989). The genetic distance matrix obtained was then represented in a two-dimensional space by applying nonmetric multidimensional scaling (MDS) analysis, using the SPSS v.16 (SPSS Inc., Chicago, IL) statistical package. Additionally, genetic heterogeneity among populations was surveyed by the exact test of population differentiation (Raymond and Rousset, 1995; Goudet et al., 1996) in the Arlequin program, to check for differences in allele frequencies between the Native American groups included in the analysis.

The relative influence of isolation, gene flow, and genetic drift on the gene pool of each of the Amerindian populations considered was explored by plotting the expected Hardy–Weinberg proportions of heterozygotes of each sample against the distance from the centroid, according to the methodology devised by Harpending and Ward (1982). Taking the potential effects of genetic drift into consideration, a strong negative relationship is expected between population divergence and heterozygosity. Consequently, it is assumed that populations plotting above the theoretical prediction given by the regression line have experienced a higher than average gene flow level, whereas those that fall below the regression line have experienced limited gene flow.

RESULTS

Allele frequencies for the *Alu* insertions typed in the Waorani sample are listed in Table 1. Three *Alu* elements (ACE, APO and FXIIIB) were monomorphic for the presence of the insertion, while HS4.65 was monomorphic for the lack of the insertion. The frequencies of the remaining *Alu* loci were distributed in a wide range (0.074–0.968). Among them, four *Alu* markers were close to the fixation for the presence (PV92 and D1) or the absence (A25 and HS4.32) of the insertion. Accordingly, average values for the observed (H_o : 0.207) and expected (H_e : 0.209) heterozygosity proved to be relatively low. Hardy–Weinberg equilibrium (HWE) was assessed by an exact test to calculate the P -value using the Markov-chain Monte Carlo method (Guo and Thompson, 1992). No significant departure from HWE expectations was detected in most of the eight unfixated *Alu* insertions, excepting PV92 ($P = 0.017$). After applying Bonferroni correction for multiple testing,

the significance threshold (P -value) proved to be 0.00625, so that PV92 also conformed to HWE expectations.

Results of the exact test of population differentiation show a notable genetic heterogeneity between the Waorani and the rest of Native American samples included in the comparative analysis (see Supporting Information Table S1). The greatest number of significant differences was observed between the Waorani and the North American collections notably Mayan and Greenland Inuit, with dissimilar *Alu* frequencies for 7 and 8 loci (out of 12 markers). On the other hand, South American populations showed the lowest number of significantly different *Alu* frequencies relative to the Waorani sample, specifically the Zoró and Surui groups.

Results of nonmetric MDS applied to the F_{ST} genetic distance matrix to evaluate genetic relationships among populations are depicted in Figure 2. Given the paucity of allele frequency data for B65 and D1 loci in Native American populations, these *Alu* markers were not considered in subsequent analyses. The two-dimensional representation of the F_{ST} distance matrix accounted for 89.4% of the total variance, with a coefficient of stress of 15.2%. MDS plot roughly reflected the geographic distribution of the Amerindian groups in subcontinents. In this way, North American populations (Greenland Inuit, Alaskan natives, and Mvskoke) appeared in a clearly differentiated cluster at the negative end of dimension I. Another substantial part of the genetic heterogeneity detected in the MDS plot was introduced by some South American populations, such as Aché, Wai-Wai, Gaviao, and Yanomami, as well as Waorani. Such Amerindian groups appeared in a peripheral position, somewhat separated from the rest, probably on account of their genetic distinctiveness. Finally, the main cluster was formed by the Mayan and most of the South American collections, which plotted close to the centroid of the distribution or in the positive semiaxis of dimension 1.

The amount of gene flow experienced by the Waorani was examined by the centroid method (Fig. 3). In the resultant figure, the regression line represents the expected heterozygosity, so that populations plotting above the line would have received a higher-than-average gene flow, whereas those positioned under the regression line would have been less impacted by gene flow processes and, therefore, would have evolved in a relatively isolated milieu. As expected, the Waorani population plotted far below the regression line, indicating that the gene flow experienced by this indigenous community was substantially lower than the average computed for the whole set of Native American populations included in this work. Coinciding with MDS results, other populations that showed gene flow levels below the average proved to be the Aché, Wai-Wai, Gaviao, and the Yanomami. In con-

TABLE 1. Allele frequencies with standard errors ($\pm SE$), observed heterozygosity (H_o) and expected heterozygosity (H_e) for 12 polymorphic *Alu* insertions in Waorani population (Ecuadorian Amazon)

Locus	ACE	TPA25	APO	PV92	FXIIIB	D1	A25	B65	HS3.23	HS4.32	HS4.59	HS4.65
Freq ^a	1.000	0.500	1.000	0.968	1.000	0.900	0.074	0.414	0.552	0.107	0.652	0.000
S.E.	0.000	0.045	0.000	0.031	0.000	0.041	0.028	0.060	0.052	0.035	0.057	0.000
H_o	0.000	0.667	0.000	0.000	0.000	0.133	0.147	0.371	0.552	0.214	0.394	0.000
H_e	0.000	0.500	0.000	0.062	0.000	0.180	0.136	0.485	0.495	0.191	0.454	0.000
P^b	/	0.093	/	0.017	/	0.242	1.000	0.173	0.268	1.000	0.457	/

^aEstimated for the presence of the *Alu* insertion.

^b P -values for the Fisher's exact test (Hardy–Weinberg equilibrium).

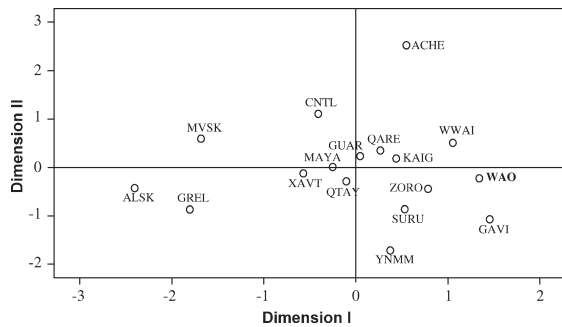


Fig. 2. Nonmetric multidimensional scaling (MDS) plot applied on a Reynolds F_{ST} genetic distance matrix for 17 Amerindian populations. Genetic distances were computed based on insertion frequencies for 10 *Alu* markers (A25, ACE, APO, FXIIB, HS3.23, HS4.32, HS4.59, HS4.65, PV92, and TPA25). Populations key: Waorani (WAO), Mvskoke (MVSK), Alaskan Inuits (ALSK), Greenland Inuits (GREL), Maya from Yucatan (MAYA), Yanomami from Venezuela (YNMM), Quechua from Arequipa (QARE) and from Tayacaja (QTAY) provinces in Peru, Aché from Paraguay (ACHE), and Cinta Larga (CNTL), Gavião (GAVI), Surui (SURU), Wai-Wai (WWAI), Zoró (ZORO), Guarani (GUAR), Kaingang (KAIG), and Xavante (XAVT) from Brazil.

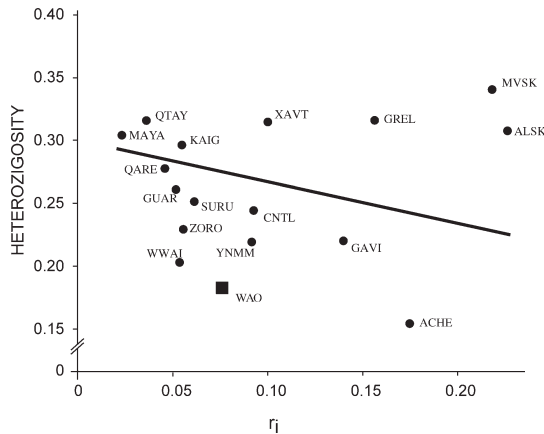


Fig. 3. Plot of heterozygosity versus distance from genetic centroid for 17 Native American populations. The line represents the expected relationship predicted by the model of Harpending and Ward (1982). Population key: Waorani (WAO), Mvskoke (MVSK), Alaskan Inuits (ALSK), Greenland Inuits (GREL), Maya from Yucatan (MAYA), Yanomami from Venezuela (YNMM), Quechua from Arequipa (QARE) and from Tayacaja (QTAY) provinces in Peru, Aché from Paraguay (ACHE), and Cinta Larga (CNTL), Gavião (GAVI), Surui (SURU), Wai-Wai (WWAI), Zoró (ZORO), Guarani (GUAR), Kaingang (KAIG), and Xavante (XAVT) from Brazil.

trast, gene flow levels above the expected ones were observed in the populations from the “northern cluster”, i.e., Mvskoke, Alaskan natives, and Greenland Inuit.

DISCUSSION

This work provides new data on the genetic background of the Waorani, one of the most isolated and anthropologically interesting Native American populations, from the perspective of classical polymorphic *Alu* insertions.

Paternal (Y-chromosome) and maternal (mtDNA) lineages have been widely analyzed in human evolutionary studies to characterize human groups in terms of origin, demographic landmarks (founder effects, population bottlenecks), and genetic variability. Yet, the reconstruction of the history of human populations from genetic data is a multifaceted task that also needs information from the recombining parts of the nuclear DNA, that is, the autosomes (Kidd et al., 2000). Polymorphic *Alu* markers constitute an essential source of nuclear genetic diversity. *Alu* repetitive elements are highly stable markers that are not affected by substantial mutation rates. Therefore, *Alu* repeats form a rich molecular fossil record that is faithfully recorded in the human genome from generation to generation, which constitutes an invaluable feature in studies with evolutionary perspectives. Furthermore, *Alu* elements are widely dispersed throughout the human genome, subject to very limited amounts of gene conversion, and selectively neutral (Batzer et al. 1996). In terms of selection, most young *Alu* elements can be considered as neutral residents of the human genome (Cordaux et al. 2006). With mutation and selection events ruled out, analysis of *Alu* insertions should exclusively reflect past processes of interaction between gene flow and genetic drift.

The most remarkable finding of this study is the strikingly low genetic diversity detected in the Waorani population. The substantial loss of genetic diversity might have been caused by intense genetic drift episodes associated with both founder effects during the early settlement of the New World (Novick et al., 1998), and with subsequent recurrent population bottlenecks probably conditioned by some behavioral peculiarities of this Amerindian community. Until very recently, the Waorani of eastern Ecuador engaged in an endless cycle of revenge killing in which men responded to the death of kin by attacking their enemies (Boster et al., 2003). This sociocultural scenario was predominantly characterized by inevitable conflicts, anger and hostility, so that an important fraction of the population mortality could be ascribed to internal fights, violence and warfare episodes (Yost, 1981). Mortality crises resulting from quarrels within the tribe would have triggered strong population bottlenecks, thus promoting recurrent genetic drift events.

Along these lines, some authors have suggested that the Waorani population size in the 1950s did not exceed 500 individuals (Yost, 1981; Holt et al., 2004), which dispersed over ~ 20,000 km² renders an extremely low population density (0.025 inhabitant/km²). Likewise, the demographic size of the Waorani community toward 1993 was estimated in around 1300 individuals (Smith, 1993). Nowadays, population density remains low at <1 inhabitant/km² (Beckerman et al., 2009). In such a context, the impact of the genetic drift might have been reinforced by: i) a marked population isolation, and therefore, a virtually nonexistent effect of the gene flow as a consequence of the behavioral and linguistic singularities of the Waorani, and their reluctance to any contact with other human groups, and ii) a historically low effective population size, on account of high mortality rates and low reproductive performances (Beckerman et al., 2009).

As we have mentioned, the main consequence of population bottlenecks and the genetic drift associated would have been a drastic decline of the Waorani’s genomic diversity. In this work, the reduction of the genetic diversity in the Waorani was corroborated by the substantial allelic

loss of the *Alu* loci examined: four (ACE, APO, HS4.65, and FXIIIIB) out of 12 *Alu* markers proved to be monomorphic, whereas other four *Alu* elements were very close to the fixation for the presence (D1, PV92) or the lack (A25, HS4.32) of the insertion. The markedly low heterozygosity of this Amerindian population is in all probability strengthened by the peculiar marital structure of the Waorani, with wide predominance of close consanguineous matings (Larrick et al., 1985).

In a study on the Asian origin of native American groups based on five polymorphic *Alu* insertions, Novick et al. (1998) found six out of 24 Central and South American populations featuring very low genetic diversity, with heterozygosity values below 0.20. Specifically, these populations were the Ngobe from Western Panama (0.19 ± 0.08), Arhuaco (0.17 ± 0.06) and Chimila (0.18 ± 0.10) from Northern Colombia, Guayabero (0.15 ± 0.09) from East-central Colombia, and Karitiana (0.16 ± 0.09) and Surui (0.19 ± 0.07) from the Brazilian Amazon. The cited authors pointed out that, bearing in mind that Aborigines from the New World are thought to derive from migrant groups with a limited number of individuals, genetic diversity is expected to be reduced in some Amerindian populations because of founder effects and population bottlenecks. The heterozygosity obtained for the same five *Alu* insertions (ACE, TPA25, APO, PV92, and FXIIIIB) was even lower in the Waorani collection (0.13 ± 0.07). This result corroborates the low diversity of the study population and seems to support the idea that genetic drift events other than those generated by the early settlement of the Americas could have had an impact on the loss of genetic diversity in the Waorani.

The isolated condition of the Waorani tribe is clearly perceptible in the results of the centroid analysis (see Fig. 3). Their remote position below the regression line (i.e., below the expected heterozygosity) gives credence to the notion of the Waorani being one of the Native American groups less impacted by gene flow processes. In contrast, the Mvskoke, which have been reported to have experienced a substantial degree of admixture with non-American native groups (Kasprisin et al., 1987; Novick et al., 1998), appeared with gene flow levels far above the average.

Previous analyses of the Waorani population based on Y-chromosome microsatellites and mitochondrial DNA have also found a significantly low genetic diversity for this interesting anthropological group. Results derived from the analysis of Y-chromosomal short tandem repeats (Y-STRs) revealed reduced haplotype variability within the population, and a large genetic distance between the Waorani and the major Ecuadorian ethnic groups (González-Andrade et al., 2009). Similarly, studies of the mitochondrial DNA D-loop region revealed the significant homogeneity of Waorani maternal lineages, with a wide predominance of a Waorani-specific variant of the haplogroup A2 (Cardoso et al., 2008; Baeta et al., 2009).

Other Native American samples such as Ache, Yanomami, Wai Wai, and Gavião plotted considerably below the expected heterozygosity values as well, in accordance with previously published data (Battilana et al., 2006). Among them, the Yanomami have been reported to possess quite similar habitat conditions and behavioral characteristics than those of the Waorani, including aggressiveness (Chagnon, 1968, 1988). Nevertheless, some peculiarities regarding the cultural rules that guide

these two hostile lowland South American peoples have led the specialists to consider the Waorani even more warlike than the Yanomami (see Beckerman et al., 2009 for a review). One major difference between the two societies is the spacing of revenge raids in the Yanomami. Sociocultural mores that guide the Yanomami violence dictate that an interval of respite follows an exchange of homicides, during which both sides usually stand down for about a generation. On the other hand, there is no evidence that the Waorani have any type of cultural regulation mandating a lapse of time (or limitation in the number of victims) in revenge exchanges. In other words, no tradition of standing down (even for a short period) has been observed in the Waorani.

Contrasting cultural patterns may also be involved in another crucial difference between the cited societies. Among the Yanomami, male deaths from warfare and/or homicide have been reported to be 5–10 times more numerous than female deaths. In contrast, among the Waorani, male deaths (at all ages) caused by internal quarrels and revenge killings have been reported to be only 1.4 times more numerous than female deaths from the same cause. This peculiar behavior of killing women and girls might be a key factor in explaining the historical failure of the population growth of Waorani community. Likewise, this demographic contrast between Yanomami and Waorani may account for differences in the intensity of population bottlenecks, and therefore, in the differential impact of genetic drift on the genetic diversity of these two Native American groups.

The prolonged isolation of the Waorani, along with a historically low effective population size (Yost, 1981; Boster et al., 2003; Beckerman et al., 2009) and high endogamy levels (Larrick et al., 1985), have resulted in the drastic reduction of their genetic diversity, because of the effects of strong genetic drift episodes. Recurrent population bottlenecks presumably determined by certain deep-rooted sociocultural practices of the Waorani (characterized by violence, internal quarrels and revenge killings until recent times) are likely responsible for their genetic singularity, and therefore, for their genetic micro-differentiation with respect to other Native American populations. In summary, the findings of this study illustrate the importance of sociocultural factors in the variation and evolutionary history of the human gene pool.

ACKNOWLEDGMENT

The authors are deeply indebted to the Waorani people who cooperated generously to the development of this study.

LITERATURE CITED

- Arcot SS, Adamson AW, Lamerdin JE, Kanagy B, Deininger PL, Carrano AV, Batzer MA. 1996. *Alu* fossil relics-distribution and insertion polymorphism. *Genome Res* 6:1084–1092.
- Baeta M, Núñez C, González-Andrade F, Sosa C, Casalod Y, Bolea M, Zweynert S, Vacas-Cruz OW, González-Solorzano J, Geppert M, Roewer L, Martínez-Jarreta B. 2009. Mitochondrial analysis revealed high homogeneity in the Waorani population—The last nomadic group of hunter-gatherers from Ecuador. *Forensic Sci Int Genet Suppl Series* 2:313–314.
- Battilana J, Bonatto SL, Freitas LB, Hutz MH, Weimer TA, Callegari-Jacques SM, Batzer MA, Hill K, Hurtado AM, Tsuneto LT, Petzl-Erler ML, Salzano FM. 2002. *Alu* insertions versus blood group plus protein genetic variability in four Amerindian populations. *Ann Hum Biol* 29:334–347.
- Battilana J, Fagundes NJ, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E, Munkhbat B, Munkhtuvshin N, Krylov M, Benevolenskaia L,

- Arnett FC, Batzer MA, Deininger PL, Salzano FM, Bonatto SL. 2006. *Alu* insertion polymorphisms in Native Americans and related Asian populations. *Ann Hum Biol* 33:142–160.
- Batzer MA, Deininger PL. 2002. *Alu* repeats and human genomic diversity. *Nat Rev Genet* 3:370–379.
- Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E. 1996. Standardized nomenclature for *Alu* repeats. *J Mol Evol* 42:3–6.
- Beckerman S, Erickson PL, Yost J, Regalado J, Jaramillo L, Sparks C, Iromenga M, Long K. 2009. Life histories, blood revenge, and reproductive success among the Waorani of Ecuador. *Proc Natl Acad Sci USA* 106:8134–8139.
- Boster JS, Yost J, Peeke C. 2003. Rage, Revenge, and Religion: honest signaling of aggression and nonaggression in Waorani coalitional violence. *Ethos* 31:471–494.
- Cardoso S, Alfonso-Sánchez MA, González-Andrade F, Valverde L, Odriozola A, Pérez-Miranda AM, Peña JA, Martínez-Jarreta B, de Pancorbo MM. 2008. Mitochondrial DNA in Huaorani (Ecuadorian amerindians): A new variant in haplogroup A2. *Forensic Sci Int Genet Suppl Se*; 1:269–270.
- Chagnon N. 1968. *Yanomamo: the fierce people*. New York: Holt, Rinehart and Winston.
- Chagnon N. 1988. Life histories. Blood revenge and warfare in a tribal population. *Science* 239:985–992.
- Cordaux R, Lee J, Dinoso L, Batzer MA. 2006. Recently integrated *Alu* retrotransposons are essentially neutral residents of the human genome. *Gene* 24:138–144.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567.
- Felsenstein J. 1989. PHYLIP: phylogeny inference package (version 3.2). *Cladistics* 5:164–166.
- Finer M, Jenkins CN, Pimm SL, Keane B, Ross C. 2008. Oil and gas projects in the Western Amazon: threats to wilderness, biodiversity, and indigenous peoples. *PLoS One* 3:e2932.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, Vidales C, Arroyo D, Peña JA. 2006. Genetic position of Valencia (Spain) in the Mediterranean basin according to *Alu* insertions. *Am J Hum Biol* 18:187–195.
- Gómez-Pérez L, Alfonso-Sánchez MA, Dipierri JE, Alfaro E, García. Obregón S, De Pancorbo MM, Bailliet G, Peña JA. 2011. Microevolutionary processes due to landscape features in the province of Jujuy (Argentina). *Am J Hum Biol* 23:177–184.
- González-Andrade F, Roewer L, Willuweit S, Sánchez D, Martínez-Jarreta B. 2009. Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis. *Forensic Sci Int Genet* 3:e83–91.
- Goudet J, Raymond M, de-Meeus T, Rousset F. 1996. Testing differentiation in diploid populations. *Genetics* 144:1933–1940.
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.
- Harpending HC, Ward RH. 1982. Chemical systematics and human populations. In: Nitecki M, editor. *Biochemical aspects of evolutionary biology*. Chicago: University of Chicago. p 213–256.
- Holt FL, Bilsborrow RE, Oña AI. 2004. Demography, household economics, and land and resource use of five indigenous populations in the northern Ecuadorian Amazon: a summary of ethnographic research. North Carolina: Carolina Population Center Occasional Paper.
- Kasprisin DO, Crow M, McClintock C, Lawson J. 1987. Blood types of the native Americans of Oklahoma. *Am J Phys Anthropol* 73:1–7.
- Kidd KK, Kidd JR, Pakstis AJ, Bonnè-Tamir B, Grigorenko E. 2000. Nuclear genetics variation of European population in a global context. In: Renfrew C, Boyle K, editors. *Archaeogenetics: DNA and the populations prehistory of Europe*. Cambridge: The McDonald Institute for Archaeological Research. p 109–149.
- Larrick JW, Yost J, Gourley C, Buckley CE 3rd, Plato CC, Pandey JP, Burck KB, Kaplan J. 1985. Markers of genetic variation among the Waorani Indians of the Ecuadorian Amazon headwaters. *Am J Phys Anthropol* 66:445–453.
- Lu FE. 2001. The common property regime of the Huaorani Indians of Ecuador: implications and challenges to conservation. *Hum Ecol* 29:425–447.
- Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ. 1998. Polymorphic *Alu* insertions and the Asian origin of Native American populations. *Hum Biol* 70:23–39.
- Orr C, Levinsohn SH, Peeke MC. 1991. Estudios gramaticales en Napo quichua y Huaorani. *Cuadernos Etnolingüísticos* 16. Quito: Instituto Lingüístico de Verano (ILV).
- Peeke MC. 1973. Preliminary grammar of Auca. Norman: Summer Institute of Linguistics and University of Oklahoma.
- Peeke MC. 1979. El idioma Huarao (Auca): gramática pedagógica. *Cuadernos Etnolingüísticos* 3. Quito: Instituto Lingüístico de Verano (ILV).
- Raymond M, Rousset F. 1995. An exact test for population differentiation. *Evolution* 49:1280–1283.
- Reynolds J, Weir BS, Cockerman CC. 1983. Estimation of the coancestry coefficient: bases for a short term genetic distance. *Genetics* 105:767–779.
- Rival L. 2002. *Trekking through history*. New York: Columbia University Press.
- Robarchek CA, Robarchek CJ. 1992. Cultures of war and peace: a comparative study of Waorani and Semai. In: Silverberg J, Gray P, editors. *Aggression and peacefulness in humans and other primates*. Oxford: Oxford University Press.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Smith R. 1993. *Drama bajo el manto amazónico: el turismo y otros problemas de los Huaorani en la actualidad*. Quito: Ediciones Abya-Yala.
- Terreros MC, Alfonso-Sánchez MA, Novick GE, Luis JR, Lacau H, Lowery RK, Regueiro M, Herrera RJ. 2009. Insights on human evolution: an analysis of *Alu* insertion polymorphisms. *J Hum Genet* 54:603–611.
- Wallis EE. 1973. *Aucas downriver*. New York: Harper and Row.
- Yost JA. 1981. Twenty years of contact: the mechanisms of change in Huao (Auca) culture. In: Whitten NA, editor. *Cultural transformations and ethnicity in modern Ecuador*. Urbana: University of Illinois Press. p 677–704.

Supplementary Table S1. Results (*P*-values) of population differentiation tests based on *Alu* allele frequencies between the Waorani community (Ecuadorian Amazon) and other Native American groups. Statistically significant differences are highlighted in bold. Population key: North American populations: Mvsko (MvSK), Alaskan Inuits (ALSK) and Greenland Inuits (GREL); Central American populations: Maya from Yucatan (MAYA); South American populations: Yanomamo from Venezuela (YNNM), Quechua from Arequipa (QARE) and from Tayacaja (QTAY) provinces in Peru, Aché from Paraguay (ACHE), and Cinta Larga (CNTL), Gavião (GAVI), Suruí (SURU), Wai-Wai (WWAI), Zoró (ZORO), Guarani (GUAR), Kaingang (KAIG) and Xavante (XAVT) from Brazil.

	ALSK	GREL	MVSK	MAYA	YNNM	QARE	QTAY	ACHE	CNTL	GAVI	SURU	WWAI	ZORO	GUAR	KAIG	XAVT
A25	0.1195	0.0466	0.0066	0.0070	0.0138	0.1652	0.5437	0.3283	0.0143	0.0142	1.0000	0.5361	0.1666	0.6131	0.5393	0.0023
ACE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0137	0.0002	0.4985	0.1235	0.0000	0.0000	0.0000
APO	0.0073	0.0290	0.1220	0.0284	1.0000	1.0000	0.4941	1.0000	0.1207	1.0000	1.0000	1.0000	0.4956	0.0145	0.2447	1.0000
B65	0.6700	0.0014	0.3896	0.0532	0.3169	N.D.D.	N.D.D.	N.D.D.	N.D.D.	N.D.D.	N.D.D.	N.D.D.	N.D.D.	N.D.D.	N.D.D.	N.D.D.
D1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	N.D.D.	0.0000	0.0000	0.0000	0.0000	0.0000	N.D.D.	N.D.D.	N.D.D.
FXIII	0.0069	0.0000	0.0000	0.0020	1.0000	1.0000	0.0006	0.0000	0.0298	1.0000	1.0000	0.0002	1.0000	0.0065	0.0003	1.0000
HS3.23	0.0006	0.0000	0.0000	0.4715	0.2440	0.5717	0.2574	0.0000	0.7735	0.0028	0.1966	0.3232	0.2039	0.0000	0.0668	0.0449
HS4.32	0.1184	0.0000	0.0014	0.0072	0.0236	0.0001	0.0020	0.1664	0.8327	0.5349	0.0855	1.0000	0.0583	0.4043	0.0140	0.0241
HS4.59	0.1160	0.0005	0.6568	0.8819	0.0001	0.0862	0.3565	0.0000	0.0000	0.0055	0.3106	0.2210	0.0875	0.5598	0.0023	0.2808
HS4.65	0.0149	0.2466	0.4948	0.4975	1.0000	1.0000	0.0035	1.0000	0.1215	0.0000	0.0032	0.2481	0.1215	0.2494	0.0000	0.0000
PV92	0.0000	0.0000	0.0000	0.0001	1.0000	0.0000	0.0000	0.0020	0.0000	0.0835	0.4934	0.0162	0.0014	0.0000	0.0001	0.0005
TPA25	0.0053	0.0208	1.0000	0.0447	0.0089	0.0630	0.0038	0.0000	0.4916	0.0001	0.2586	0.0000	0.0106	0.0031	0.0237	0.3986

Significant values ($p < 0.05$) are highlighted in bold.

N.D.D.: No datos disponibles

PRINCIPALES RESULTADOS

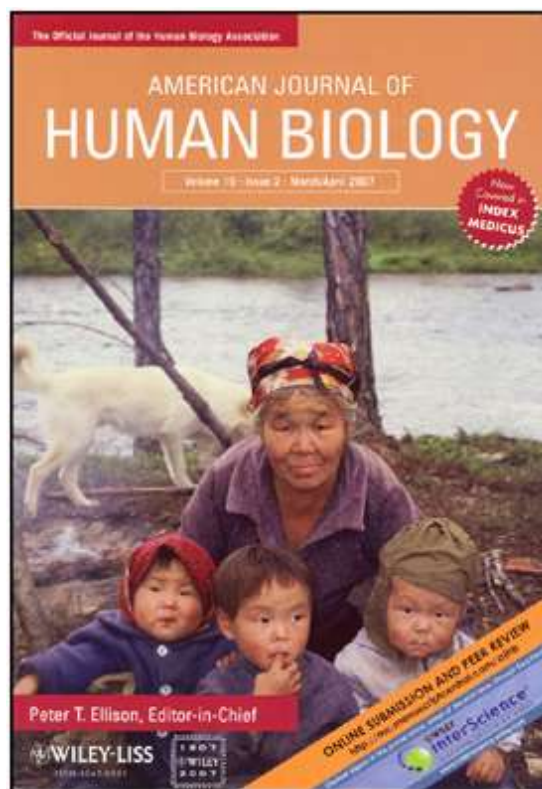
Como cabría esperar, se ha encontrado una escasa diversidad genética en esta población, hallándose varios de los marcadores analizados en estado monomórfico, así como unos valores de heterocigosidad comparativamente bajos.

El grado de diferenciación genética entre la población Huaorani y otras poblaciones amerindias es notable, encontrando diferencias significativas con todas ellas para al menos 5 loci (*Supplementary Table S1*). Es también perfectamente lógico que se haya encontrado un menor número de diferencias significativas con las poblaciones de Sudamérica que con las norteamericanas.

Estos resultados se han visto reflejados en el análisis de escalamiento multidimensional, donde la población Huaorani aparece situada en un extremo de la variabilidad del continente. No obstante, la dispersión de todas las poblaciones por el gráfico es alta, revelando un alto grado de heterogeneidad genética, particularmente entre las poblaciones de Norte y Sudamérica, que aparecen en extremos opuestos.

El aislamiento genético de la población Huaorani se pone de manifiesto en el análisis del centroide, donde se representa la heterocigosidad frente a la distancia al Centroides (*Figura 3, Manuscrito IV*). En él se refleja que esta población es una de las poblaciones más aisladas, junto con los Aché de Paraguay.

A partir de los resultados obtenidos se concluye que los factores culturales y las costumbres de la etnia Huaorani, que les han llevado a un aislamiento severo, han potenciado el efecto de un reducido tamaño poblacional, incrementado el efecto de la deriva genética, algo observado en otros trabajos (Baeta, 2011; Cardoso *et al.* 2012; Baeta *et al.* 2013). En definitiva, se observa que las características culturales de una etnia pueden amplificar de forma notable los procesos de microdiferenciación que ésta experimente.



MANUSCRITO V

Young *Alu* insertions within the MHC class I region in Native American populations: insights into the origin of the MHC-*Alu* repeats

Luis Gómez-Pérez, Miguel A. Alfonso-Sánchez, José E. Dipierri, Dora Sánchez, Ibone Espinosa, Marian M. de Pancorbo, José A. Peña. 2013.

American Journal of Human Biology 25, 359-365

INTRODUCCIÓN

En este trabajo se han analizado una serie de inserciones *Alu* localizadas en el MHC (6p21.31). Son los primeros resultados publicados de estos marcadores en poblaciones de origen amerindio y se incluyen los Huaorani y los Jujeños. La muestra procedente de Jujuy no incluye la totalidad de los individuos empleados en el artículo 3, sino que únicamente se utilizaron las de los individuos provenientes de las regiones más altas de la provincia (>2500 m.s.n.m.), dado que se ha observado en esta región un muy escaso flujo génico con poblaciones no amerindias (*Manuscrito III*).

Los principales objetivos del trabajo se han centrado en analizar la heterogeneidad genética de ambas poblaciones, desde el punto de vista de las inserciones *Alu* del MHC, pero también en evaluar la distribución y diversidad de los haplotipos compuestos por este grupo de marcadores, valorando en su caso la existencia de diferencias en la distribución de los haplotipos entre poblaciones europeas, africanas, asiáticas y americanas.

Además, se ha pretendido obtener nuevos datos sobre las relaciones de ligamiento existentes entre este grupo de inserciones y eventualmente añadir nueva información acerca de su origen.

Original Research Article

Young *Alu* Insertions Within the MHC Class I Region in Native American Populations: Insights into the Origin of the MHC-*Alu* Repeats

LUIS GÓMEZ-PÉREZ,¹ MIGUEL A. ALFONSO-SÁNCHEZ,¹ JOSÉ E. DIPIERRI,² DORA SÁNCHEZ,³ IBONE ESPINOSA,¹ MARIAN M. DE PANCORBO,⁴ AND JOSÉ A. PEÑA^{1*}

¹Departamento de Genética y Antropología Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), Bilbao, Spain

²Instituto de Biología de la Altura, Universidad Nacional de Jujuy, San Salvador de Jujuy, Argentina

³Hospital Metropolitano, Laboratorio de Genética Molecular, Quito, Ecuador

⁴BIOMiCs Research Group, Centro de Investigación "Lascaray" Ikerkunea, Universidad del País Vasco (UPV/EHU), Vitoria-Gasteiz, Spain

Objectives: Genetic heterogeneity of two Amerindian populations (Jujuy province, Argentina, and Waorani tribe, Ecuador) was characterized by analyzing data on polymorphic *Alu* insertions within the human major histocompatibility complex (MHC) class I region (6p21.31), which are completely nonexistent in Native Americans. We further evaluated the haplotype distribution and genetic diversity among continental ancestry groups and their potential implications for the dating of the origin of MHC-*Alus*.

Methods: Five MHC-*Alu* elements (*Alu*MicB, *Alu*TF, *Alu*HJ, *Alu*HG, and *Alu*HF) were typed in samples from Jujuy ($N = 108$) and Waorani ($N = 36$). Allele and haplotype frequency data on worldwide populations were compiled to explore spatial structuring of the MHC-*Alu* diversity through AMOVA tests. We utilized the median-joining network approach to illustrate the continental distribution of the MHC-*Alu* haplotypes and their phylogenetic relationships.

Results: Allele and haplotype distributions differed significantly between Jujuy and Waorani. The Waorani featured a low average heterozygosity attributable to strong population isolation. Overall, *Alu* markers showed great genetic heterogeneity both within and among populations. The haplotype distribution was distinctive of each continental ancestry group. Contrary to expectations, Africans showed the lowest MHC-*Alu* diversity.

Conclusions: Genetic drift mainly associated to population bottlenecks seems to be reflected in the low MHC-*Alu* diversity of the Amerindians, mainly in Waorani. Geographical structuring of the haplotype distribution supports the efficiency of the MHC-*Alu* loci as lineage (ancestry) markers. The markedly low *Alu* diversity of African populations relative to other continental clusters suggests that these MHC-*Alus* might have arisen after the anatomically modern humans expanded out of Africa. *Am. J. Hum. Biol.* 25:359–365, 2013. © 2013 Wiley Periodicals, Inc.

INTRODUCTION

Native South American groups are characterized by high levels of genetic heterogeneity among populations. This high genetic diversity, which is consistent with the linguistic and cultural diversity of South Amerindians, seems to be the consequence of contrasting regional patterns of genetic drift and gene flow owing to the presence of geographic barriers and a wide spectrum of biomes generated by a very complex orography, among other factors (Malhi et al., 2002; Schurr and Sherry, 2004; Tarazona-Santos et al., 2001).

In this work, we analyzed two South American communities who have experienced microevolutionary changes attributable mainly to a marked population isolation, albeit resulting from factors of different nature in each case: Native Americans from Jujuy province, northwestern Argentina (Gómez-Pérez et al., 2011a), and the Waorani community from the Ecuadorian Amazon (Cardoso et al., 2012; Gómez-Pérez et al., 2011b). Both populations show a notable interest from the anthropological viewpoint.

On account of the proximity of the Andes mountain range, the region of Jujuy can be divided in various geographical/ecological areas well differentiated by significant variations in altitude above sea level, which generate prominent differences in their local environmental conditions. For this work, Jujuy sample was collected in the highest regions of the province, namely, Pre Puna (~2,500 m above sea level), and Puna (>3,300 m). The introduc-

tion of African slaves by the Spanish colonists during colonial times (16th–19th centuries) was mainly concentrated in the subtropical forests of Jujuy lowlands, thus displacing the autochthonous inhabitants to higher-altitude regions (Martinez-Sarasola, 1992). Reflecting such a historical background, Native American individuals are proportionally more numerous in the highest population nuclei. In this way, some authors have reported very restricted amounts of gene flow and low admixture rates for the highland communities of Jujuy (Dipierrri et al., 1998, 2000). In fact, even nowadays the Amerindian genetic component is highly predominant in Jujuy, while the European and African contributions to the gene pool are practically negligible (Dipierrri et al., 2000; Gómez-Pérez et al., 2011a). Jujuy has been also characterized by low population densities over time (~1.3 inh/km²) (INDEC,

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Basque Country Government; Contract grant number: Grupo Consolidado IT-424-07; Contract grant sponsor: University of the Basque Country UPV/EHU; Contract grant number: ELDUNANO-TEK UFI 11/32; Contract grant sponsor: Saiotek; Contract grant number: SA-2010/00075; Contract grant sponsor: UPV/EHU (L.G.-P.).

*Correspondence to: José A. Peña, Departamento de Genética, Antropología Física y Fisiología Animal, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain. E-mail: joseangel.pena@ehu.es

Received 11 October 2012; Revision received 17 December 2012; Accepted 7 January 2013

DOI: 10.1002/ajhb.22377

Published online 6 April 2013 in Wiley Online Library (wileyonlinelibrary.com).

2001), notably in the areas of more difficult access because of the altitude.

The Waorani constitute one of the world's last uncontacted peoples that have lived in voluntary isolation (Finer et al., 2008). They inhabit much of the lowland Amazon tropical forests (Wallis, 1973) in the Yasuní National Park, located between the eastern Ecuadorian provinces of Orellana, Napo, and Pastaza. The Waorani are known to be descendants from hunter-gatherer groups and they continue to primarily subsist on hunting and horticulture, with a seminomadic lifestyle. The Waorani society has traditionally been highly inbred, owing to a marked prevalence of prearranged bilateral cross-cousin marriages (Beckerman et al., 2009; Larrick et al., 1985; Robarchek and Robarchek, 1992). The demographic size of the Waorani oscillates between 1,500 and 2,000 individuals at present, but some population data suggest that this demographic size could have been notably lower during the recent past due to strong population bottlenecks (Beckerman et al., 2009; Holt et al., 2004; Smith, 1993; Yost, 1981).

The striking isolation of the Waorani seems to have been essentially conditioned by linguistic and behavioral factors. On the one hand, they speak the Wao Tededo (the autonym is Huao Tededo), which is an unclassified and isolated language without any congener in the Amazon region (Lu, 2001; Peeke, 1973; Yost, 1981). On the other hand, the Waorani have been considered as one of the most warlike peoples, because of certain deeply ingrained sociocultural practices characterized by anger, aggressiveness, violence, internal quarrels, and revenge killings until relatively recent times (Beckerman et al., 2009; Robarchek and Robarchek, 1992; Yost, 1981). In fact, the first peaceful contact of the Waorani community with western cultures did not occur until the 1950s (Wallis, 1973). Afterward, the Waorani have had increasing contact with outsiders, though some familial clans still live totally isolated in the Amazonian forest (Boster et al., 2003; Yost, 1981).

Numerous previous publications have stressed the advantageous properties of the *Alu* polymorphisms for human evolutionary studies. These include selective neutrality of the *Alu* markers, the existence of a single common mutation for all chromosomes with the same *Alu* repeat (identity by descent not just by state), the possibility of knowing the ancestral state (absence of the insertion), and the nonexistence of a known mechanism for the specific removal of an *Alu* element (Batzer and Deininger, 2002; Terreros et al., 2009; among others). Likewise, haplotype combinations of *Alu* loci from autosomal chromosomes are considered to be useful lineage markers, as they have proved to be efficient for detecting heterogeneity among human groups with different genetic ancestry (Dunn et al., 2003, 2007; Kulski and Dunn, 2005; Kulski et al., 2010; Tian et al., 2008; Yao et al., 2010).

This work was intended to characterize the genetic diversity of the targeted Native American groups (Jujuy province and Waorani) by analyzing a set of polymorphic *Alu* insertions located in the major histocompatibility complex (MHC) class I genomic region (6p21.31). With this analysis, we sought to contribute new data on allele and haplotype frequencies of MHC-*Alu* repeats, which are completely nonexistent in Amerindian communities to date. We further compiled haplotype frequencies from earlier investigations to explore the genetic relationships of these two Native American populations in a wide geographical context and to analyze the differences in genetic diversity

among continental groups and their potential implications for the dating of the origin of the MHC-*Alu* insertions.

MATERIAL AND METHODS

Samples

Five polymorphic *Alu* insertions (*Alu*MicB, *Alu*TF, *Alu*HJ, *Alu*HG, and *Alu*HF) located in the MHC region from the human chromosome 6 were analyzed in two South Amerindian populations ($N = 144$). One of these samples consisted of 108 individuals from Jujuy province, northwestern Argentina. As mentioned above, individuals included in the sample lived in different population nuclei from the highest zones of the province, characterized by their altitude ($>2,500$ m above sea level). The sample of the Waorani tribe consisted of 36 individuals living in different communal longhouses or *nanicabos* (traditional domestic groups) settled at the western Amazon basin, in the Yasuní National Park (Ecuador). According to the surname composition of the sample examined, we estimated that the Waorani individuals included in the study belonged to at least 11 different familial clans. In this respect, results from a recent analysis of both X-chromosome microsatellites (X-STRs) and Y-chromosome microsatellites (Y-STRs) in the same Waorani population sample examined herein reported the presence of several different haplotypes, suggesting the existence of various paternal lineages, and therefore, the absence of close genetic kinship between the Wao individuals included in the sample (Cardoso et al., 2012). In this way, genetic data provided by sex chromosome-linked markers supported the hypothesis of the existence of a number of different familial clans based only on a cultural approach (surnames). Ethical guidelines for research with human beings were adhered to as stipulated by each of the institutions involved in the investigation. The study protocol was approved by the Institutional Review Board from Universidad del País Vasco (UPV/EHU).

Alu typing

In the case of Jujuy sample, genomic DNA was extracted from bloodstains using a Qiagen kit (QIAmp DNA Micro Kit, Qiagen, Hilden, Germany). In the Waorani collection, DNA was extracted from blood samples using the Wizard Genomic DNA Purification Kit System (Promega, Madison, WI). In both cases, DNA was stored at -20°C when not in use. Typing of the polymorphic MHC-*Alu* markers was performed by PCR amplification, according to procedures described in previous publications (Kulski and Dunn, 2005). PCR products were further electrophoresed in 1.5% agarose gels. The DNA bands obtained by electrophoresis were stained with ethidium bromide (0.5 $\mu\text{l/ml}$) for a better visualization, and documented by photography under UV light. Allele identification was performed using a Roche DNA Molecular Weight Marker V ladder. Quality control of both the PCR and the electrophoresis was carried out by including a positive control (homozygous for the insertion) as well as a negative control containing all the components of the PCR mix except the template DNA. Likewise, to minimize the risk of misgenotyping as a result of amplification errors, 20% random quality controls were performed.

Statistical analysis

Allele frequencies with standard errors, as well as observed and expected heterozygosity values for each of

TABLE 1. Allele frequencies (presence of insertion) with standard errors (\pm SE) for five MHC-*Alu* loci in Amerindian samples from Jujuy province (Argentina) and from the Waorani tribe (Ecuadorian Amazon)

Population		<i>Alu</i> marker					Avg ^b
		<i>Alu</i> MicB	<i>Alu</i> TF	<i>Alu</i> HJ	<i>Alu</i> HG	<i>Alu</i> HF	
Jujuy	N ^a	216	216	216	216	216	
	Freq	0.556	0.301	0.227	0.338	0.019	
	\pm SE	0.034	0.032	0.024	0.033	0.009	
	He	0.494	0.421	0.350	0.448	0.036	0.350
	Ho	0.426	0.343	0.296	0.380	0.037	0.296
	P ^c	0.168	0.063	1.000	0.124	1.000	
Waorani	N ^a	72	72	72	70	72	
	Freq	0.267	0.633	0.083	0.190	0.000	
	\pm SE	0.056	0.064	0.031	0.047	0.000	
	He	0.391	0.464	0.153	0.307	0.000	0.263
	Ho	0.333	0.333	0.167	0.310	0.000	0.229
	P ^c	0.643	0.104	1.000	1.000	-	

^aSample size in number of chromosomes analyzed.

^bAverage of heterozygosity values (including all *Alu* loci).

^cP values for the Fisher's exact test (Hardy-Weinberg equilibrium).

Freq, insertion frequency; He, expected heterozygosity; Ho, observed heterozygosity.

TABLE 2. Fixation indices (F_{ST} , F_{SC} , and F_{CT}) from a hierarchical analysis of molecular variance (AMOVA) for five polymorphic *Alu* insertions and the five-*Alu* haplotypes, considering population groups^a classed according to ancestry

<i>Alu</i> marker	Fixation index (%)		
	F_{ST}	F_{SC}	F_{CT}
MicB	49.77*	38.12***	18.83***
TF	55.34 n.s.	50.54***	9.69***
HJ	41.75*	26.87***	20.34***
HG	32.98**	16.51***	19.73***
HF	74.30 n.s.	70.08***	14.12***
Five- <i>Alu</i> haplotypes	11.99**	6.42***	5.96***

^aGroups are **Africa**: Sekele San, Kung San and SE African Bantu (Kulski and Dunn, 2005); **Asia**: NE Thais, Khalk population from Mongolia (Kulski and Dunn, 2005), Han Chinese from Guangdong, Han Chinese from Hunan, Han Chinese from Inner Mongolia and Mongols from Inner Mongolia (Tian et al., 2008), Maonan Chinese from Guangxi province, and Bulang Chinese, Dai Chinese, Han Chinese, Hani Chinese, Jinuo Chinese and Wa Chinese from Yunnan province (Yao et al., 2010). **Europe**: Navarre, Guipuzcoa (García-Obregón et al., 2012), Valencia (García-Obregón et al., 2011) and European Australian (Kulski and Dunn, 2005). **America**: Jujuy province and Waorani (study populations).

F_{ST} , genetic variation within populations; F_{SC} , genetic variation among populations within groups; F_{CT} , genetic variation among groups.

*Statistical significance for $P < 0.05$.

**Statistical significance for $P < 0.01$.

***Statistical significance for $P < 0.001$.

n.s., nonsignificant.

the *Alu* elements analyzed were calculated using the Power Marker software v3.0 (Liu and Muse, 2004). Hardy-Weinberg equilibrium (HWE) was assessed by an exact test (Guo and Thompson, 1992) in the Arlequin v3.5 program (Excoffier and Lischer, 2010).

Linkage disequilibrium (LD) values (Slatkin and Excoffier, 1996) were calculated for each loci pair using Arlequin v3.5 program. Bearing in mind that *Alu* markers are autosomal loci, we computed maximum-likelihood haplotype frequencies through an Expectation-Maximization (E-M) algorithm in the Arlequin software. This procedure is an iterative process aiming specifically at obtaining maximum-likelihood estimates of haplotype frequencies from multilocus genotype data when the gametic phase is unknown (phenotypic data).

To determine the fraction of genetic variability attributable to differences within and among populations, genetic variance was hierarchically apportioned through the analysis of molecular variance, AMOVA (Excoffier et al., 1992; Weir and Cockerham, 1984). AMOVA tests were carried out based on both allele and haplotype frequencies, and considering four population groups classed according to ancestry (African, Asian, European, and Native American). Data on the populations included in the comparative analyses such as population designation, site of collection, and ethnic affiliation (ancestry) are given in Supporting Information Table 1.

With the aim of illustrating the genetic heterogeneity of the Native American collections in a worldwide context, we utilized the median-joining network approach (Bandelt et al., 1995) available in the Network v4.5.0.0 software (Fluxus Technology), to simultaneously represent the five-*Alu* haplotype frequencies and their phylogenetic relationships. As we performed this analysis by considering continental (ancestry) clusters, we first estimated average haplotype frequencies by continental group (establishing the same sample size for each cluster). Obviously, depicting of the phylogenetic relationships among haplotypes is merely a global approach, because of the existence of recombination processes and the inability of

knowing the gametic phase of the haplotypes. But this limitation is largely compensated by the abovementioned advantageous properties of the *Alu* polymorphisms in evolutionary studies, as well as by the possibility of obtaining a joint vision (by considering both sexes) about the evolutionary processes that have contributed to the shaping of the human gene pool.

RESULTS

Allele frequencies of the *Alu* insertions in Jujuy province and the Waorani community are listed in Table 1. One of the *Alu* markers (*Alu*HF) proved to be monomorphic for the absence of the insertion in Waorani. The rest of the *Alu* loci examined were polymorphic in both populations. No significant departure from HWE expectations was detected for any of the *Alu* elements (Table 1). As expected according to the degree of population isolation, observed heterozygosity values were lower in Waorani than in Jujuy for all *Alu* elements. Accordingly, average heterozygosity was also lower in Waorani (0.229) than in Jujuy (0.296).

Three significant associations between the MHC-*Alu* markers were detected through LD analyses. Specifically, significant LD values were detected between *Alu*MicB and *Alu*TF ($P < 0.05$) and between *Alu*HJ and *Alu*HG ($P < 0.001$) in Jujuy, whereas *Alu*TF and *Alu*HG loci were found to be strongly associated ($P < 0.01$) in the Waorani group. Previous studies have reported statistically significant LD values between the *Alu* loci examined herein, notably between the *Alu* elements within the alpha block of the MHC (García-Obregón et al., 2012; Tian et al., 2008).

A total of 16 different haplotypes (*Alu*MicB-*Alu*TF-*Alu*HJ-*Alu*HG-*Alu*HF) were identified in the sample from Jujuy province, whereas only 10 haplotypic combinations were found among the Waorani (Supporting Information Table 2). The most frequent haplotype in Jujuy proved to be 2-1-1-1-1 (i.e. insertion only at the *Alu*MicB locus), with a frequency of 0.210. In Waorani, the most common

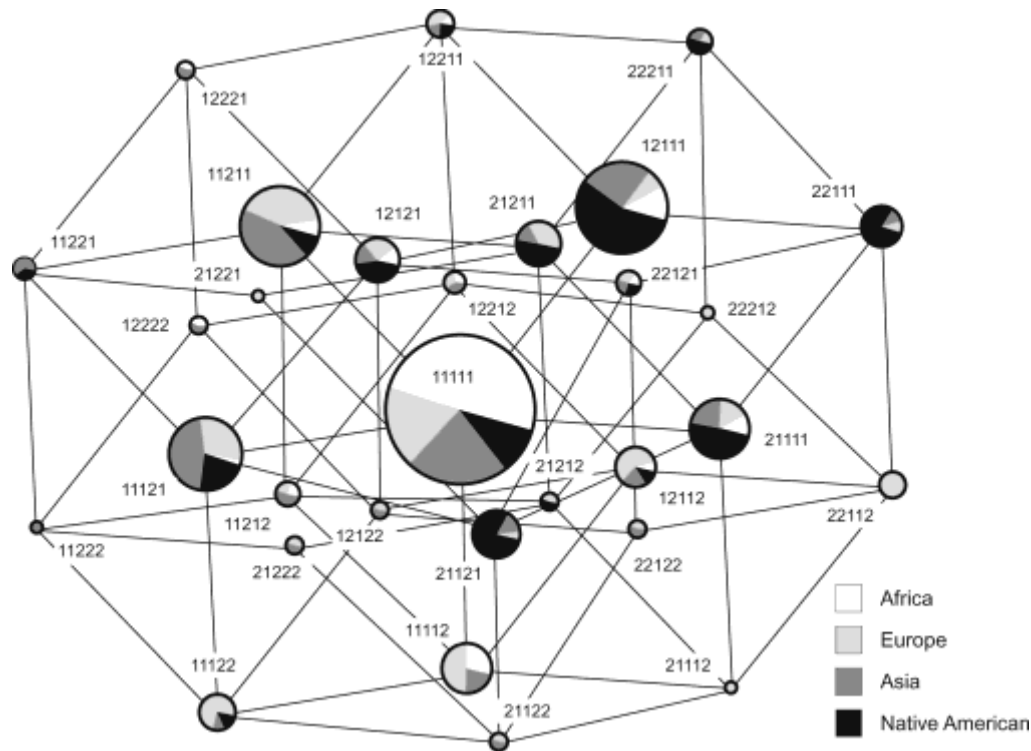


Fig. 1. Median-joining network of polymorphic *Alu* insertions (*Alu*MicB, *Alu*TF, *Alu*HJ, *Alu*HG, and *Alu*HF) within the human major histocompatibility complex (MHC). The network depicts the MHC-*Alu* haplotype distribution in four population clusters classed according to continental ancestry (Africans, Europeans, Asians and Native Americans), and their phylogenetic affinities. Native American cluster includes the populations examined (Jujuy province, NW Argentina, and Waorani from the Ecuadorian Amazon). Populations included in the other continental clusters can be consulted in Supporting Information Table 1. The size of each circle is proportional to the haplotype frequency (considering the whole set of populations). The grayscale illustrates the frequency of the MHC-*Alu* haplotypes in each continental group.

haplotype (frequency: 0.455) was 1-2-1-1-1, that is, presence of the insertion only at *Alu*TF. As can be noticed, the haplotype formed by the insertion of all *Alu* elements (2-2-2-2-2) was not found in any of the Amerindian samples. Results of an exact test of population differentiation (Goudet et al., 1996; Raymond and Rousset, 1995) revealed highly significant differences ($P < 0.0001$) in haplotype distribution between Jujuy and Waorani samples (data not shown).

Hierarchical apportionment of the genetic diversity of the *Alu* elements in worldwide populations (including Native Americans) was performed by AMOVA. To that end, data for those populations summarized in Supporting Information Table 1 were employed. AMOVA tests were performed based on both allele and five-*Alu* haplotype frequencies, and considering four populations groups classed according to geography (continent) and/or ancestry (Table 2). In general, all the MHC-*Alu* loci examined showed a great heterogeneity, as indicated by the fixation indices generated by the AMOVA. In this way, the fixation index F_{SC} (variation among populations within groups) revealed statistically significant differences ($P < 0.001$) for both allele and haplotype frequencies, thus reflecting the great genetic heterogeneity of the populations included in each cluster. As expected, the genetic variation among continental groups (fixation index F_{CT}) also rendered highly

significant differences for alleles and haplotypes, thus revealing a significant trend toward the spatial structuring of the *Alu* variation.

The continental distribution of the five-*Alu* haplotype frequencies and their phylogenetic affinities were depicted using the median-joining network approach (Fig. 1). As can be noticed, African populations featured the lowest haplotypic diversity. The Africans only showed an average frequency above 0.1 for the haplotype combining the absence of insertion at all loci (1-1-1-1-1), that is, the ancestral haplotype. In the Asian groups, the average frequency was higher than 0.1 not only for the ancestral haplotype, but also for haplotypes showing a single insertion at the *Alu*HG (1-1-1-2-1), *Alu*HJ (1-1-2-1-1), and *Alu*TF (1-2-1-1-1) loci. It is worth mentioning that the haplotype combining insertions at *Alu*HJ, *Alu*HG, and *Alu*HF (1-1-2-2-2) was solely identified in the Asian cluster, which suggests its potential autochthony. Regarding the European collections, both the ancestral haplotype and those with a single insertion at *Alu*HG (1-1-1-2-1) or at *Alu*HJ (1-1-2-1-1) also showed frequencies above 0.1. Likewise, two haplotypes seemed to be confined only to human groups of European ancestry: 2-2-1-1-2 (insertions at *Alu*MicB, *Alu*TF, and *Alu*HF), and 2-2-2-1-2 (insertions at all loci, excepting *Alu*HG). Two further haplotypic combinations were almost autochthonous to Europeans: 2-1-1-1-2

(insertions at *AluMicB* and *AluHF*), and 2-1-2-2-1 (insertions at *AluMicB*, *AluHJ*, and *AluHG*). In addition to the European-ancestry samples, the latter haplotypes were observed in some Asian populations although at very low frequencies. Finally, we identified four haplotypes with frequencies above 0.1 in Native Americans. These were the ancestral, no-insertion haplotype (1-1-1-1-1), the haplotype with a single insertion at *AluTF* (1-2-1-1-1), which is also common in Asian populations, and the haplotypes with a single insertion at *AluMicB* (2-1-1-1-1), and with insertions at *AluMicB* and *AluHG* (2-1-1-2-1), both of which proved to be rather infrequent in the other continental clusters. As previously mentioned, the most frequent haplotype in Jujuy was the one with a single insertion at *AluMicB* (2-1-1-1-1), whereas in Waorani the combination with a single insertion at *AluTF* (1-2-1-1-1) was predominant. On the whole, most of the African chromosomes featured the ancestral haplotype. The ancestral haplotypic combination was the most frequent in Asia and Europe as well, but in these continental clusters, we additionally identified a sizeable number of haplotypes showing at least one insertion. Interestingly, some of the haplotypes with insertions proved to be even more frequent than the ancestral haplotype among the Amerindian groups.

In accord with the haplotype frequency distribution depicted in the median-joining network, Africans featured the lowest haplotype diversity (0.470) followed, in ascending order, by Asians (0.764), Native Americans (0.832), and Europeans (0.878). Regarding the Amerindian groups, the haplotype diversity was visibly higher in Jujuy (0.894) than in Waorani (0.771), although both values fell within the range of non-African populations.

DISCUSSION

This study reports the first data on *Alu* insertion frequencies within the MHC class I genomic region for Native American populations, specifically from Argentina (Jujuy province) and Ecuador (Waorani tribe).

Native Americans experienced two major episodes of population constriction: one related to the peopling of the Americas and the other caused by the European contact (Mulligan et al., 2004; Schurr and Sherry, 2004). Both of them would have promoted genetic heterogeneity among the Amerindian populations by genetic drift episodes associated with founding events and population bottlenecks (Malhi et al., 2002, 2003). Some critical MHC-*Alu* frequency data observed in the Amerindian populations seem to reflect the effects of genetic drift. Thus, for instance, Jujuy exhibited the highest insertion frequency of *AluMicB* observed so far, as well as the highest frequency of the corresponding haplotype (2-1-1-1-1). Likewise, the Waorani showed the highest frequency for the *AluTF* locus and the corresponding haplotype (1-2-1-1-1). In addition, the Waorani community and the Wa from Yunnan province (China) were the only populations where the *AluHF* insertion was not identified. In the Native American groups analyzed, the effects of genetic drift and other stochastic processes might have been reinforced by their specific characteristics, notably their reduced demographic sizes and a strong, population isolation over time. In Jujuy, population isolation seems to be a consequence of the poor human habitability conditions inherent to high altitudes above sea level (Dipierrri et al., 2000; Gómez-

Pérez et al., 2011a). In Waorani, isolation has been argued to be mostly determined by their linguistic peculiarities (Lu, 2001; Peeke, 1973), and by some ethnopsychological factors leading to the prevalence of hostility, belligerence, fights within the tribe, and revenge homicides until recent times (Beckerman et al., 2009; Boster et al., 2003; Yost, 1981). Some recent studies on mitochondrial DNA (Cardoso et al., 2012) and polymorphic *Alu* insertions (Gómez-Pérez et al., 2011b) have both reported a strikingly reduced genetic diversity in Waorani. Recurrent genetic drift events mostly associated with dramatic population bottlenecks would also account for the great genetic heterogeneity detected between Jujuy and Waorani.

It is widely accepted that Africa was the source of the anatomically modern humans that dispersed all over the world during the past 100,000 years. Africa has been identified as the region of the world with the highest levels of human genetic diversity within and between populations (Jorde et al., 1998; Reed and Tishkoff, 2006; Tishkoff and Verrelli, 2003). In turn, the reduction of the genetic diversity in continents other than Africa is thought to be a consequence of the intensification of the genetic drift when the anatomically modern humans began their migration out of Africa, because of the intense sampling process that implied the limited demographic size of the groups that left Africa. Several authors have hypothesized an evolutionary scenario in which modern humans first evolved in Africa, thus accumulating genetic diversity. Then small subsets of the African population began to expand out of the continent and experienced bottlenecks and founder effects potentiated by the small population sizes of the migrant groups (Gonder et al., 2007; Ingman et al., 2000; Jorde and Wooding, 2004).

According to the above considerations, then the strikingly low haplotype diversity obtained for the African collections does not fit the expectations of the out-of-Africa model. A plausible explanation for the lower genetic diversity of the African population cluster, bearing in mind the physiological role of HLA genes and the physical proximity of the MHC-*Alu* markers with some HLA class I genes, could be the existence of selective pressures against some of these HLA genes that, by effect of genetic hitchhiking, would have also reduced the MHC-*Alu* diversity in Africans. Nevertheless, findings of a recently published work revealed no signal of genetic hitchhiking effects (Kulski et al., 2011).

A much more likely hypothesis suggests that this set of *Alu* elements, or at least some of them, might have a recent evolutionary origin, so that the event of insertion occurred after the anatomically modern humans left Africa. This situation has been deduced at least for *AluHG*, whose insertion seems to have arisen in a common ancestor of Asians and Europeans at a time after their migration from Africa, taking into account that its origin is later to that of the HLA-A2 allele (Dunn et al., 2002). The presence of the *AluHG* insertion in African populations might then be explained by human back migrations from West Asia to sub-Saharan Africa (Hammer et al., 1998). Indeed, the haplotype with insertion only at the *AluHG* locus (1-1-1-2-1) proved to be remarkably infrequent in Africa and totally nonexistent in !Kung, being more frequent the haplotype with insertions at *AluTF* and *AluHG* loci (1-2-1-2-1), which probably originated from a recombination between haplotypes 1-2-1-1-1 and 1-1-1-2-1. A similar scenario could be argued to

explain the continental distribution of *Alu*HJ, or alternatively, a moderate increase in the frequency generated by population bottlenecks in human groups out of Africa, bearing in mind that the haplotype 1-1-2-1-1 is clearly more frequent in Europe and Asia than in Africa, and also that other related haplotypes (excepting 1-1-2-1-2) are virtually nonexistent in the African populations. Interestingly, no previous study on MHC-*Alu* has reported the presence of the haplotype combining *Alu* insertions at all loci (2-2-2-2-2) in the worldwide populations surveyed to date, which also points towards the notion of a relatively recent origin, evolutionarily speaking, of the targeted MHC-*Alu* markers.

Summarizing, even when *Alu* elements are known to be conservative and stable markers from the evolutionary viewpoint, they permitted to detect substantial genetic heterogeneity between the Native American samples examined most likely conditioned by stochastic evolutionary forces. Genetic drift associated to founding events and population bottlenecks seems to be mostly reflected in the comparatively reduced average heterozygosity obtained for the Waorani, attributable to the strong population isolation of this Amerindian community as a consequence of a sociocultural uniqueness determined by a striking ethnography. Analyses of MHC-*Alu* including populations from different continents revealed a clear trend to the spatial structuring of the genetic heterogeneity, since haplotype diversity notably overlapped the geographical origins of the human groups. These findings indicate that the MHC-*Alu* polymorphisms might be useful as lineage (ancestry) markers in several types of investigations, as they showed an adequate resolving power to properly discriminate between the major ethnic groups. Finally, the conspicuously low genetic diversity of the African populations relative to the other continental clusters seems to suggest that the origin of these MHC-*Alu*s might have occurred after the anatomically modern humans migrated out of Africa. Yet, MHC-*Alu* frequency data for many other worldwide populations as well as other lines of evidence would be necessary to support such an attractive hypothesis.

ACKNOWLEDGMENTS

Technical and human support provided by SGIker (UPV/EHU, MICINN, GV/EJ, ERDF, and ESF) is gratefully acknowledged. The authors are particularly indebted to all voluntary donors from Jujuy (Argentina) and from the Waorani community (Ecuador) who generously cooperated to the development of this research.

LITERATURE CITED

- Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753.
- Batzler MA, Deininger PL. 2002. *Alu* repeats and human genomic diversity. *Nat Rev Genet* 3:370–379.
- Beckerman S, Erickson PI, Yost J, Regalado J, Jaramillo L, Sparks C, Iromenga M, Long K. 2009. Life histories, blood revenge, and reproductive success among the Waorani of Ecuador. *Proc Natl Acad Sci USA* 106:8134–8139.
- Boster JS, Yost J, Peeke C. 2003. Rage, revenge, and religion: honest signaling of aggression and nonaggression in Waorani coalitional violence. *Ethos* 31:471–494.
- Cardoso S, Alfonso-Sánchez MA, Valverde L, Sánchez D, Zarrabeitia MT, Odriozola A, Martínez-Jarreta B, de Pancorbo MM. 2012. Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. *Heredity* 108:609–615.
- Dipierrri JE, Alfaro E, Martínez-Marignac VL, Baillet G, Bravi CM, Cejas S, Bianchi NO. 1998. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol* 70:1001–1010.
- Dipierrri JE, Alfaro E, Peña JA, Constans J, Dugoujon JM. 2000. GM, KM immunoglobulin allotypes and other serum genetic markers (HP, GC, PI, and TF) among South American populations living at different altitudes (Jujuy Province, Argentina): admixture estimates. *Hum Biol* 72:305–319.
- Dunn DS, Choy MK, Phipps ME, Kulski JK. 2007. The distribution of major histocompatibility complex class I polymorphic *Alu* insertions and their associations with HLA alleles in a Chinese population from Malaysia. *Tissue Antigens* 70:136–143.
- Dunn DS, Naruse T, Inoko H, Kulski JK. 2002. The association between HLA-A alleles and young *Alu* dimorphisms near the HLA-J, -H, and -F genes in workshop cell lines and Japanese and Australian populations. *J Mol Evol* 55:718–726.
- Dunn DS, Ota M, Inoko H, Kulski JK. 2003. Association of MHC dimorphic *Alu* insertions with HLA class I and MIC genes in Japanese HLA-B48 haplotypes. *Tissue Antigens* 62:259–262.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567.
- Excoffier L, Smouse P, Quattro J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
- Finer M, Jenkins CN, Pimm SL, Keane B, Ross C. 2008. Oil and gas projects in the Western Amazon: threats to wilderness, biodiversity, and indigenous peoples. *PLoS One* 3:e2932.
- García-Obregón S, Alfonso-Sánchez MA, Gómez-Pérez L, Pérez-Miranda AM, Arroyo D, de Pancorbo MM, Peña JA. 2011. Microsatellites and *Alu* elements from the human MHC in Valencia (Spain): analysis of genetic relationships and linkage disequilibrium. *Int J Immunogenet* 38:483–491.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, Gómez-Pérez L, de Pancorbo MM, Peña JA. 2012. Ancestry markers from the human chromosome 6: *Alu* repeats from the MHC in autochthonous Basques. *Hum Immunol* 73:720–725.
- Gómez-Pérez L, Alfonso-Sánchez MA, Dipierrri JE, Alfaro E, García-Obregón S, de Pancorbo MM, Baillet G, Peña JA. 2011a. Microevolutionary processes due to landscape features in the province of Jujuy (Argentina) *Am J Hum Biol* 23:177–184.
- Gómez-Pérez L, Alfonso-Sánchez MA, Sánchez D, García-Obregón S, Espinosa I, Martínez-Jarreta B, de Pancorbo MM, Peña JA. 2011b. *Alu* polymorphisms in the Waorani tribe from the Ecuadorian Amazon reflect the effects of isolation and genetic drift. *Am J Hum Biol* 23:790–795.
- Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. 2007. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol Biol Evol* 24:757–768.
- Goudet J, Raymond M, de-Meeus T, Rousset F. 1996. Testing differentiation in diploid populations. *Genetics* 144:1933–1940.
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.
- Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL. 1998. Out of Africa and back again: recent clastic analysis of human Y chromosome variation. *Mol Biol Evol* 15:427–441.
- Holt FL, Bilsborrow RE, Oña AI. 2004. Demography, household economics, and land and resource use of five indigenous populations in the Northern Ecuadorian Amazon: a summary of ethnographic research. North Carolina: Carolina Population Center Occasional Paper, Chapel Hill.
- INDEC. 2001. Censo Nacional de Población y Vivienda 1991 y Censo Nacional de Población, Hogares y Viviendas. Available at: <http://www.indec.gov.ar/webcenso/index.asp>. Last accessed June, 2010.
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713.
- Jorde LB, Bamshad M, Rogers AR. 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* 20:126–136.
- Jorde LB, Wooding SP. 2004. Genetic variation, classification and 'race'. *Nat Genet* 36 (11 Suppl):S28–S33.
- Kulski JK, Dunn DS. 2005. Polymorphic *Alu* insertions within the major histocompatibility complex class I genomic region: a brief review. *Cytogenet Genome Res* 110:193–202.
- Kulski JK, Shigenari A, Inoko H. 2011. Genetic variation and hitchhiking between structurally polymorphic *Alu* insertions and HLA-A, -B, and -C alleles and other retroelements within the MHC class I region. *Tissue Antigens* 78:359–377.
- Kulski JK, Shigenari A, Shiina T, Inoko H. 2010. Polymorphic major histocompatibility complex class II *Alu* insertions at five loci and their association with HLA-DRB1 and -DQB1 in Japanese and Caucasians. *Tissue Antigens* 76:35–47.

- Larrick JW, Yost J, Gourley C, Buckley CE III, Plato CC, Pandey JP, Burck KB, Kaplan J. 1985. Markers of genetic variation among the Waorani Indians of the Ecuadorian Amazon headwaters. *Am J Phys Anthropol* 66: 445–53.
- Liu K, Muse SV. 2004. PowerMarker: new genetic data analysis software. Version 3.0. Available at: <http://www.powermarker.net>. Last accessed May, 2010.
- Lu FE. 2001. The common property regime of the Huaorani Indians of Ecuador: implications and challenges to conservation. *Hum Ecol* 29:425–447.
- Malhi RS, Eshleman JA, Greenberg JA, Weiss DA, Schultz Shook BA, Kaestle FA, Lorenz JG, Kemp BM, Johnson JR, Smith DG. 2002. The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America. *Am J Hum Genet* 70:905–919.
- Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gorodezky C, Smith DG. 2003. Native American mtDNA prehistory in the American Southwest. *Am J Phys Anthropol* 20:108–124.
- Martinez-Sarasola C. 1992. Nuestros paisanos los indios. Buenos Aires: EMECE.
- Mulligan CJ, Hunley K, Cole S, Long JC. 2004. Population genetics, history, and health patterns in native Americans. *Annu Rev Genomics Hum Genet* 5:295–315.
- Peeke C. 1973. Preliminary grammar of Auca. Norman: the Summer Institute of Linguistics and University of Oklahoma.
- Raymond M, Rousset F. 1995. An exact test for population differentiation. *Evolution* 49:1280–1283.
- Reed FA, Tishkoff SA. 2006. African human diversity, origins and migrations. *Curr Opin Genet Dev* 16:597–605.
- Robarchek CA, Robarchek CJ. 1992. Cultures of war and peace: a comparative study of Waorani and Semai. In: Silverberg J, Gray P, editors. *Aggression and peacefulness in humans and other primates*. Oxford: Oxford University Press. p 189–213.
- Schurr TG, Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol* 16:420–439.
- Slatkin M, Excoffier L. 1996. Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity* 76:377–383.
- Smith R. 1993. Drama bajo el manto amazónico: el turismo y otros problemas de los Huaorani en la actualidad. Quito: Ediciones Abya-Yala.
- Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* 68:1485–1496.
- Terreros MC, Alfonso-Sánchez MA, Novick GE, Luis JR, Lacau H, Lowery RK, Regueiro M, Herrera RJ. 2009. Insights on human evolution: an analysis of Alu insertion polymorphisms. *J Hum Genet* 54:603–611.
- Tian W, Wang F, Cai JH, Li LX. 2008. Polymorphic insertions in 5 Alu loci within the major histocompatibility complex class I region and their linkage disequilibria with HLA alleles in four distinct populations in mainland China. *Tissue Antigens* 72:559–567.
- Tishkoff SA, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet* 4:293–340.
- Wallis EE. 1973. *Aucas downriver*. New York: Harper and Row.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Yao Y, Shi L, Shi L, Kulski JK, Chen J, Liu S, Yu L, Lin K, Huang X, Tao Y, Tokunaga K, Chu J. 2010. The association and differentiation of MHC class I polymorphic Alu insertions and HLA-B/Cw alleles in seven Chinese populations *Tissue Antigens* 76:194–207.
- Yost JA. 1981. Twenty years of contact: the mechanisms of change in Huao (Auca) culture. In: Whitten NA, editor. *Cultural transformations and ethnicity in modern Ecuador*, Urbana: University of Illinois Press. p 677–704.

Supplementary Table 1. Worldwide populations included in the comparative analyses

Continent/Population	Ancestry	Reference
Africa		
South-East Bantu	African	Kulski and Dunn (2005)
!Kung San	African	"
Sekele San	African	"
Asia		
North-East Thais	Asian	Kulski and Dunn (2005)
Khalk Mongolian	Asian	"
Han Chinese (Guangdong province)	Asian	Tian <i>et al.</i> (2008)
Han Chinese (Hunan province)	Asian	"
Han Chinese (Inner Mongolia)	Asian	"
Mongols (Inner Mongolia)	Asian	"
Maonan (Guangxi province)	Asian	Yao <i>et al.</i> (2010)
Bulang (Yunnan province)	Asian	"
Dai (Yunnan province)	Asian	"
Han (Yunnan province)	Asian	"
Hani (Yunnan province)	Asian	"
Jinuo (Yunnan province)	Asian	"
Wa (Yunnan province)	Asian	"
Oceania		
Australian	European	Kulski and Dunn (2005)
Europe		
Navarre	European	García-Obregón <i>et al.</i> (2012)
Guipuzcoa	European	"
Valencia	European	García-Obregón <i>et al.</i> (2011)

Supplementary Table 1 References:

- García-Obregón S, Alfonso-Sánchez MA, Gómez-Pérez L, Pérez-Miranda AM, Arroyo D, de Pancorbo MM, Peña JA. 2011. Microsatellites and *Alu* elements from the human MHC in Valencia (Spain): analysis of genetic relationships and linkage disequilibrium. *Int J Immunogenet* 38, 483-491.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, Gómez-Pérez L, de Pancorbo MM, Peña JA. 2012. Ancestry markers from the human chromosome 6: *Alu* repeats from the MHC in autochthonous Basques. *Hum Immunol* 73, 720-725.
- Kulski JK, Dunn DS, Gaudieri S, Shiina T, Inoko H. 2001. Genomic and phylogenetic analysis of the human CD1 and HLA class I multicopy genes. *J Mol Evol* 53, 642-650.
- Tian W, Wang F, Cai JH, Li LX. 2008. Polymorphic insertions in 5 *Alu* loci within the major histocompatibility complex class I region and their linkage disequilibria with HLA alleles in four distinct populations in mainland China. *Tissue Antigens* 72, 559-567.
- Yao Y, Shi L, Shi L, Kulski JK, Chen J, Liu S, Yu L, Lin K, Huang X, Tao Y, Tokunaga K., Chu J. 2010. The association and differentiation of MHC class I polymorphic *Alu* insertions and HLA-B/Cw alleles in seven Chinese populations. *Tissue Antigens* 76, 194-207.

Supplementary Table 2. Five-*Alu* (*Alu*MicB-*Alu*TF-*Alu*HJ-*Alu*HG-*Alu*HF) haplotype frequencies in Native American samples from Jujuy province (Argentina) and Waorani (Ecuadorian Amazon)

Haplotype ^a	Frequencies (\pm SE)			
	Jujuy		Waorani	
1-1-1-1-1	0.128 \pm 0.023	28	0.147 \pm 0.047	9
2-1-1-1-1	0.210 \pm 0.028	45	0.023 \pm 0.020	1
1-2-1-1-1	0.092 \pm 0.020	20	0.455 \pm 0.066	27
1-1-2-1-1	0.029 \pm 0.011	6	0.040 \pm 0.026	2
1-1-1-2-1	0.090 \pm 0.020	20	0.057 \pm 0.031	3
2-2-1-1-1	0.056 \pm 0.016	12	0.099 \pm 0.040	6
2-1-2-1-1	0.104 \pm 0.021	22	-	-
2-1-1-2-1	0.121 \pm 0.022	26	0.094 \pm 0.039	6
1-2-2-1-1	0.008 \pm 0.006	2	0.021 \pm 0.019	1
1-2-1-2-1	0.076 \pm 0.018	16	0.038 \pm 0.025	2
1-2-1-1-2	0.008 \pm 0.006	2	-	-
1-1-2-2-1	0.007 \pm 0.006	2	-	-
1-1-1-2-2	0.006 \pm 0.005	1	-	-
2-2-2-1-1	0.022 \pm 0.010	5	0.025 \pm 0.021	1
2-2-1-2-1	0.038 \pm 0.013	8	-	-
2-1-2-1-2	0.005 \pm 0.005	1	-	-

^a leyenda: 1 = absence of the *Alu* element, 2 = insertion of the *Alu* element

SE, standard error

Supplementary Table 3. Matrix of *p*-values generated by the exact test of population differentiation (Raymond *et al.* 1995; Goudet *et al.* 1996), among populations from different continents and/or ancestry. Populations abbreviations in Supplementary Table 1, except the analyzed Native American populations Jujuy (JUJY) and Waorani (WAO). Highlighted figures are nonsignificant *p*-values

	JUJY	WAO	BANT	KUNG	SEKE	NTHA	KHLK	HMON	IMON	BLNG	WA	DAI	MAON	HANI	JNUO	GUDG	HNAN	FYDH	NNAV	GUIP	AUST	
WAO	0.0001	*																				
BANT	0.0001	0.0001	*																			
KUNG	0.0001	0.0001	0.2576	*																		
SEKE	0.0001	0.0001	0.7571	0.0048	*																	
NTHA	0.0001	0.0001	0.0001	0.0001	0.0001	*																
KHLK	0.0010	0.0001	0.0001	0.0001	0.0001	0.0004	*															
HMON	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	*														
IMON	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.6209	*													
BLNG	0.0001	0.0001	0.0001	0.0001	0.0001	0.0009	0.0001	0.0001	0.0001	*												
WA	0.0001	0.0001	0.0560	0.0006	0.0018	0.0001	0.0001	0.0001	0.0001	0.0001	*											
DAI	0.0001	0.0001	0.0001	0.0001	0.0001	0.1205	0.0001	0.0001	0.0001	0.0001	0.0001	*										
MAON	0.0001	0.0001	0.0001	0.0001	0.0001	0.0159	0.0001	0.0001	0.0001	0.0001	0.0001	0.0084	*									
HANI	0.0001	0.0001	0.0001	0.0001	0.0001	0.1711	0.0010	0.0001	0.0001	0.0001	0.0001	0.0009	0.0142	*								
JNUO	0.0001	0.0001	0.0001	0.0001	0.0001	0.0010	0.0001	0.0001	0.0001	0.0001	0.0001	0.1417	0.1587	0.0002	*							
GUDG	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	*						
HNAN	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0130	0.0086	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.4521	*					
FYDH	0.0001	0.0001	0.0001	0.0001	0.0001	0.2204	0.0001	0.0001	0.0001	0.0002	0.0001	0.0102	0.0001	0.0060	0.0001	0.0001	0.0001	*				
NNAV	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	*			
GUIP	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0091	*	
AUST	0.0001	0.0001	0.0001	0.0001	0.0001	0.0301	0.0001	0.0001	0.0001	0.0001	0.0001	0.0004	0.0001	0.0011	0.0001	0.0001	0.0001	0.0205	0.0003	0.0043	*	
VALN	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0729	0.0702	0.0473	

PRINCIPALES RESULTADOS Y MATERIAL COMPLEMENTARIO

A partir de los resultados obtenidos en este trabajo se pudo confirmar que la deriva genética ha tenido una importancia notable en el modelado del patrimonio genético de las dos poblaciones analizadas y probablemente de todo el continente, como se ha mencionado ya en otros trabajos para diferentes marcadores (Baeta *et al.* 2009; Cardoso *et al.* 2012) y en anteriores resultados de esta Tesis (*Manuscrito III y IV*)

En la muestra Huaorani se detecta una muy baja diversidad alélica y haplotípica. Son, junto con los Wa de la provincia de Yunnan (China), las dos únicas poblaciones analizadas donde no se ha encontrado la inserción *AluHF*. Además, los Huaorani muestran la mayor frecuencia de inserción en el locus *AluTF* y de su correspondiente haplotipo (1-2-1-1-1) que está presente en más del 45% de sus cromosomas. Asimismo, entre otros resultados relevantes se ha encontrado un muy bajo número de haplotipos. Todo ello apunta a la existencia de un proceso de homogeneización experimentado por esta población. Esta misma conclusión ha sido subrayada en *el Manuscrito III* de esta memoria, así como por otros autores (Cardoso *et al.* 2008, 2012; Baeta *et al.* 2009, 2011).

Por su parte, Jujuy muestra unos niveles más bajos de homogeneización poblacional, producto de un menor aislamiento. A pesar de ello, se ha encontrado la mayor frecuencia registrada de inserción de *AluMicB*, así como la frecuencia más alta del haplotipo correspondiente a la presencia de dicha inserción (2-1-1-1-1), portado por el 21% de sus cromosomas.

La deriva ha originado en gran medida la heterogeneidad genética entre las poblaciones amerindias (Malhi *et al.* 2002, 2003). En este trabajo se detecta dicha heterogeneidad mediante el test exacto de diferenciación poblacional, mediante el cual se ha observado una diferenciación altamente significativa entre Huaorani y Jujeños, las únicas poblaciones americanas analizadas.

Mediante un test AMOVA se ha observado la existencia de una estructuración geográfica en la distribución de los haplotipos, lo que justificaría la consideración de las inserciones *Alu* del MHC como AIMs (*Tabla 2, Manuscrito V*).

A partir del análisis de las frecuencias haplotípicas en poblaciones de diferentes continentes, se han observado unos resultados que entran en conflicto con los esperados de acuerdo al modelo *Out of Africa*, ya que la diversidad de estas inserciones *Alu* es particularmente baja en poblaciones africanas. Una posible explicación se basa en que algunos de los marcadores *Alu* del MHC podrían haber surgido recientemente, concretamente después de que los humanos modernos se expandieran fuera de África. Esta hipótesis se apoya, entre otros datos, en las frecuencias de *AluHG*, inserción que parece haber surgido en un ancestro común de asiáticos y europeos en algún momento después de su migración fuera de África (Dunn *et al.* 2002). Este marcador es muy poco frecuente en África e inexistente en los Kung (*Figura 1, Manuscrito V*). Un argumento similar podría utilizarse para explicar la distribución continental de frecuencias de *AluHJ*.

La ausencia del haplotipo compuesto por la presencia de inserción *Alu* en los cinco loci (2-2-2-2-2) en todas las poblaciones analizadas, es otro indicio que sugiere un origen muy reciente, evolutivamente hablando, de los marcadores MHC-*Alu*.

MANUSCRITO VI

Jujuy province (NW Argentina): STR loci unveil genetic footprints of microevolutionary responses to a steep orography

Luis Gómez-Pérez, Miguel A. Alfonso-Sánchez, José E. Dipierri, José A. Peña

American Journal of Human Biology. Enviado.

INTRODUCCIÓN

Con el fin de profundizar en el análisis del patrimonio genético de la población de Jujuy, se ha procedido a examinar con otros marcadores la región más alta de la provincia, ampliando así los resultados obtenidos en el *Manuscrito III*. Este análisis se ha realizado a mediante un grupo de marcadores STRs, ya que poseen un mayor poder de resolución que las inserciones *Alu* en eventos recientes

La región analizada, que se ha denominado el Alto Jujuy (*Jujuy highlands*), incluye las comarcas de Quebrada Baja, Quebrada Alta y La Puna, todas ellas por encima de los 2500 m.s.n.m. Se han seleccionado aquellas regiones que presentan un componente casi exclusivamente amerindio en su patrimonio genético. El objetivo ha sido analizar los procesos de microdiferenciación de la población autóctona amerindia debido a la Deriva Genética, por lo que se ha procurado minimizar el efecto del mestizaje provocado por la llegada de poblaciones de otros continentes.

Esta muestra, no expuesta al mestizaje, debería verse afectada por tanto tan sólo por los condicionantes a la movilidad derivados de una complicada orografía y por una muy baja densidad poblacional.

Los marcadores STRs han sido seleccionados por su alto grado de polimorfismo. Se han elegido cuatro STRs del CMH, con un cierto desequilibrio de ligamiento entre ellos, con el fin de poder utilizar en los análisis estadísticos las correspondientes frecuencias haplotípicas, que lógicamente deberían mostrar una mayor resolución.

Jujuy province (NW Argentina): STR loci unveil genetic footprints of microevolutionary responses to a steep orography

Luis Gómez-Pérez ¹, Miguel A. Alfonso-Sánchez ¹, José E. Dipierri ², José A. Peña ^{1*}

1. Departamento de Genética y Antropología Física, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), 48080 Bilbao, Spain

2. Instituto de Biología de la Altura, Universidad Nacional de Jujuy, 4600 San Salvador de Jujuy, Argentina

Text pages: 14. Tables, figures and supporting information tables: 7

Abbreviated title (running headline): Genetic footprints of microevolution in Jujuy

Key words: Jujuy, STR, microevolution, landscape genetics, genetic drift, gene flow

* Address for correspondence:

Dr. José A. Peña

Departamento de Genética, Antropología Física y Fisiología Animal

Universidad del País Vasco (UPV/EHU)

Apartado 644, 48080 Bilbao, Spain

e-mail: joseangel.pena@ehu.es

Grant sponsors:

L.G.-P has a doctoral fellowship from the University of the Basque Country (UPV/EHU).

Grant sponsors: 'Programa de Subvención a Grupos Consolidados, Departamento de Educación, Universidades e Investigación' of the Basque Government (Contract grant number IT-424-07), and Saiotek (Contract grant number SA-2010/00075).

ABSTRACT

Objectives: Assessing the effects of genetic drift on the microevolutionary changes and genetic substructuring of three high-altitude populations from Jujuy province (NW Argentina). These human communities are characterized by extreme living conditions and low population densities, associated with a prominent height above sea level and a steep orography.

Methods: Four autosomal STRs (D6S2792, D6S265, D6S105 and D6S2666) were typed in 102 individuals from Quebrada Baja (~ 2,500 m), Quebrada Alta (~ 3,300 m), and Puna (> 3,500 m). Genetic diversity was measured by computing the observed (H_o) and expected (H_e) heterozygosities, as well as haplotype diversity. Population differentiation tests based on allele and haplotype frequencies were performed to assess genetic heterogeneity among subpopulations.

Results: No deviation from HWE was detected for each subpopulation analyzed separately. Yet, significant departures from HWE (D6S2792 and D6S105) were detected in an analysis considering the whole study area. Genetic diversity showed a decreasing trend from Puna to Quebrada Baja. Accordingly, both allele and haplotype frequencies differed mostly between these subpopulations, sited at the edges of the variation in altitude.

Conclusions: This study corroborates the effectiveness of the resolving power of the STRs to identify microevolutionary changes at a microgeographic level. The main evolutionary consequence of the altitude and, especially, of the steep landscape of Jujuy would be an imbalance in the interplay between gene flow and genetic drift favoring the latter. The combined effect of a restricted gene flow with an intense genetic drift has promoted local genetic differentiation between Jujuy highlands' subpopulations.

INTRODUCTION

Evolutionary changes can take place on different scales. Specifically, the term ‘microevolution’ has been generally equated with events at or below the species level. Accordingly, a microevolutionary process is basically the variation in allele frequencies within a population over time. Evolution at this scale can be observed over relatively short time periods, which is particularly interesting from a population genetics viewpoint. Changes in allele frequencies might come about because: *i*) natural selection favored some gene or set of genes, *ii*) the population received new individuals (immigrants) from regions with distinct allele frequencies, or even carrying new allelic variants, *iii*) some alleles might have mutated, appearing new private gene variants, and *iv*) the impact of random genetic drift on small, isolated populations could have substantially altered the gene frequencies (Hendry and Kinnison, 1999; Carroll, 2001).

Ancient demographic events frequently leave their genetic signature in human DNA. As a consequence, we can obtain inferences about the structure and demographic history of present-day populations by interpreting genetic diversity in terms of the evolutionary forces. Studies of population genetic structure have generally concentrated on how linguistic differences, cultural contacts, population movements (migrations), and geographical factors have affected the genetic diversity of human populations (Jorde, 1980; Jorde et al., 1982; Wood et al., 1982; Relethford, 1996; Gómez-Pérez et al., 2011a, b; Capocasa et al., 2014). Sociocultural and linguistic differences between human groups are known to affect rates of gene flow. In the same way, a sharp decline in population size due to geographical and ecological constraints could promote genetic drift (Relethford, 2012).

Genomic markers of quite different nature have been used to identify microevolutionary changes in human populations. Among them, polymorphisms from non-coding DNA regions are assumed to be selectively neutral, so their evolution in populations can be mostly explained by the interplay between gene flow and genetic drift. Because of their hypervariable nature, microsatellite or short tandem repeat (STR) polymorphisms have been extensively analyzed in several fields of population genetics. Microsatellites feature comparatively high mutation rates that range between 3.3×10^{-4} per locus per generation (Forster et al., 2000) and 15.2×10^{-4} per locus per generation

(Zhivotovsky et al., 2004), considering generations of 25 years. The high mutation rate of human STRs may therefore provide valuable genetic information for the elucidation of human population history and for studying genetic substructuring among local populations (Agrawal and Khan, 2005; Shepard et al., 2005).

The province of Jujuy, NW Argentina, have proved to be a suitable model for exploring the influence of landscape features on human microevolution. On account of the significant orography of the region, Jujuy is predominantly characterized by conspicuous variations in height above sea level among its population centers (Fig. 1). This phenomenon implies the existence of numerous intricate places, and therefore, the emergence of partially isolated villages. Such features most likely reduce the homogenizing effects of the gene flow (Dipierri et al., 1998, 2000; Gómez-Pérez et al., 2011a), thus promoting genetic substructuring in the area by effects of random genetic drift in small-size populations.

In a previous work based on polymorphic *Alu* elements, Gómez-Pérez et al. (2011a) identified a clear influence of the altitude on the spatial distribution of the *Alu* insertion frequencies. Consequently, the five study regions of Jujuy showed disparate levels of admixture between Native Americans and individuals of European or African ancestry depending on height above sea level. High-altitude populations (> 2500 m) showed almost 100% of Native American ancestry, with a negligible genetic contribution of the other ethnic groups. Genetic diversity therefore showed a decreasing trend from the high-altitude regions (Quebrada Baja, Quebrada Alta and Puna) to those at lower altitudes (Valle and Selva). Virtually no genetic heterogeneity was observed within the high-altitude populations, as expected according to their similarity in admixture levels and ethnic ancestry.

The present study was designed to perform a refined analysis of the genetic diversity in the three high-altitude populations of Jujuy, whose Native American genetic background has been preserved almost intact. To that end, we take advantage of the high level of polymorphism of the STR markers, whose hypervariable nature provides a higher resolution than *Alu* markers for the detection of recent demographic phenomena that might have affected the gene pool of the aforementioned populations or, in other words, might have promoted microevolutionary processes. The main focus of this work was the assessment of the potential impact of genetic drift in the genetic substructuring of the high-

altitude populations of Jujuy, which are characterized by their extreme living conditions, and consequently, by their remarkably low population densities.

MATERIAL AND METHODS

Four autosomal microsatellite loci (D6S2792, D6S265, D6S105 and D6S2666) located at chromosome 6 were screened in a population sample of 102 Native American individuals from Jujuy province, NW Argentina. All individuals included in the sample lived in different population centers from the province highlands, which are characterized by their prominent height above sea level, a.s.l. (> 2500 m a.s.l.) (Fig. 1). To refine the approach of the genetic structuring of human populations settled in the highlands of Jujuy, samples were divided into three groups according to altitude: *i*) Puna (N = 41; mean altitude > 3,500 m a.s.l.), *ii*) Quebrada Alta (N = 31; mean altitude ~ 3,300 m a.s.l.), and *iii*) Quebrada Baja (N = 30; mean altitude ~ 2,500 m a.s.l.). For the sake of simplicity, the population cluster formed by these three subregions will be called hereafter 'Jujuy highlands'.

Voluntary donors were asked to provide personal and family information helpful to our study, such as surnames and geographical origin of their parents and grandparents, among others. From these interviews, three-generation pedigree charts were constructed to ensure the collection from native individuals from Jujuy highlands. Ethical guidelines for research with human beings were adhered to as stipulated by the institutions implicated in the research project: Universidad del País Vasco (UPV/EHU), Spain, and Universidad Nacional de Jujuy, Argentina. Blood donors gave their informed consent before inclusion in the sample. The study protocol was approved by the Institutional Review Board from the UPV/EHU.

PCR amplification and STR typing

Genomic DNA was extracted from bloodstains using the standard phenol:chloroform procedure (Sambrook et al., 1989). Typing of the polymorphic STR markers was performed by PCR amplification, according to methods previously described (García-Obregón et al., 2010). Positive and negative PCR controls were run to assess the quality of the amplification process. Amplified STR fragments were further detected and separated by capillary electrophoresis, using an ABI PRISM 310

Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). To estimate the size of the amplified products, an internal size standard (GeneScan TAMRA-500, Applied Biosystems) was included. Fragment sizes were determined automatically using GeneMapper Analysis software v3.7, and by comparison with the supplied allelic ladders.

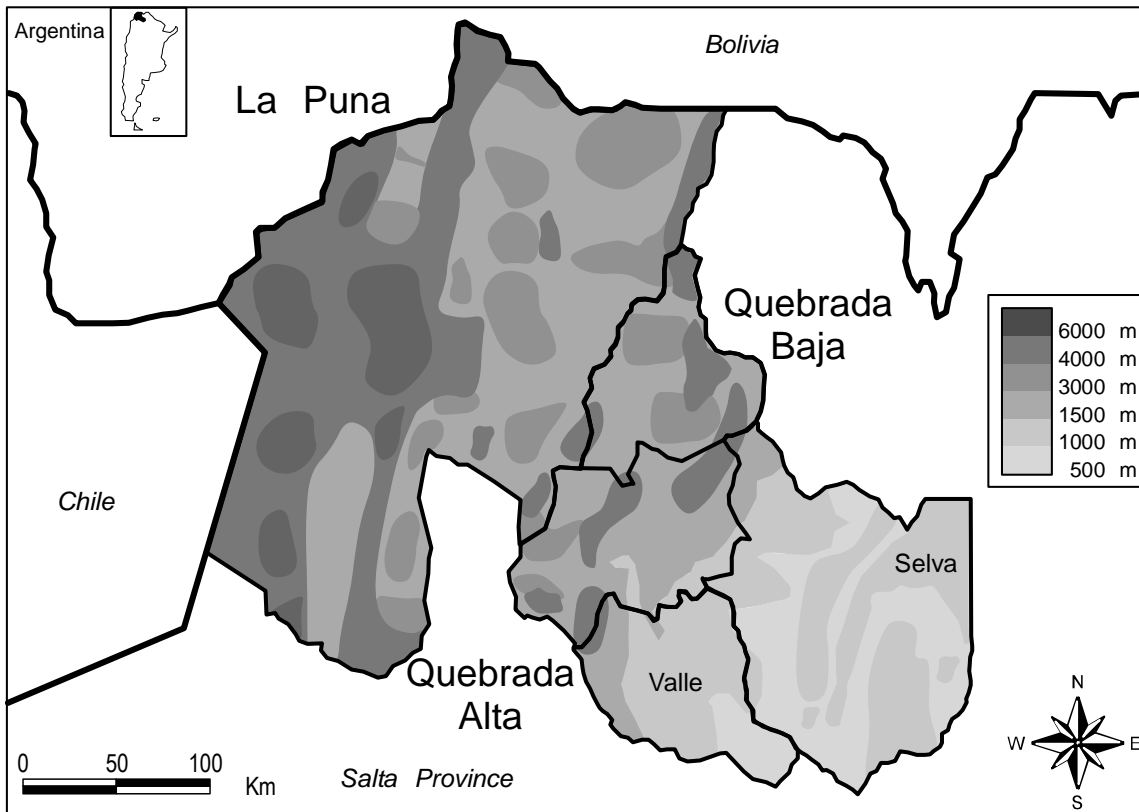


Figure 1. Map of the Jujuy province (NW Argentina) showing the geographic position of the subpopulations involved in the study (Puna, Quebrada Alta and Quebrada Baja). Altitude contour lines illustrate the height above sea level of the different population nuclei.

Statistical analysis

Allele frequencies were calculated by direct counting. To test for Hardy-Weinberg equilibrium (HWE), a Fisher's exact probability test using the Markov-chain Monte Carlo method was conducted to estimate *P*-values (Guo and Thompson, 1992). Bearing in mind that all microsatellite markers are located in chromosome 6, linkage disequilibrium values (Slatkin and Excoffier, 1996) were computed for each loci pair.

To measure the genetic variability of the STRs examined, observed (H_o) and expected (H_e) heterozygosities, as well as the haplotype diversity (Nei and Tajima, 1981) were computed. Haplotypes were constructed using genotype data for each individual (D6S2792-D6S265-D6S105-D6S2666). Haplotypes were inferred, and their frequencies were then calculated. Both allele and haplotype frequencies were utilized to perform F_{ST} pairwise comparisons (Reynolds et al., 1983) to thoroughly explore the genetic heterogeneity between the different subregions of Jujuy highlands. Statistical analyses were carried out by means of the Arlequin v3.5 software (Excoffier and Lischer 2010).

RESULTS

Allele frequencies of the STRs, both for the three subpopulations (Quebrada Baja, Quebrada Alta and Puna) and for the whole region (Jujuy highlands) are listed in Table 1. A total of 10 alleles for the D6S2792 locus, 14 alleles for D6S105, 7 alleles for D6S265 and 6 for D6S2666 were identified in the sample surveyed. Overall, the more frequent alleles per loci coincided in the three subpopulations. These were specifically the cases of allele 115 of D6S2792, allele 120 of D6S105, and allele 124 of D6S265. Exception to the above was the D6S2666 locus, with allele 170 being the most frequent in Quebrada Baja and Quebrada Alta and allele 144 predominating in Puna.

The Fisher's exact probability test revealed no significant deviation from the expected allele frequencies when HWE was checked for each single subpopulation (Table 2). In this analysis, results indicated that only D6S105 in Quebrada Alta was very close to the statistical significance after applying the Bonferroni correction for multiple testing. However, when the same analysis was performed considering the whole study area (Jujuy highlands) we found significant departures from HWE in two out of four microsatellites, namely D6S2792 and D6S105.

Linkage disequilibrium analyses (Supporting Information Table 1) among the STRs examined revealed several associations. Specifically, significant linkage disequilibrium values were detected between D6S2792 and D6S105 ($p = 0.013$), D6S2792 and D6S2666 ($p = 0.006$), and mainly between D6S265 and D6S105 ($p < 0.001$).

Table 1. Allele frequencies and standard error (\pm SE) for autosomal microsatellites from chromosome 6 in three populations from Jujuy province (NW Argentina)

Locus	Allele (bp)	Frequency \pm SE			
		Quebrada Baja (2N = 60)	Quebrada Alta (2N = 62)	Puna (2N = 82)	Total ^a (2N = 204)
D6S2792	97	0.117 \pm 0.037	0.097 \pm 0.040	0.122 \pm 0.036	0.113 \pm 0.022
	99	0.100 \pm 0.035	0.065 \pm 0.036	0.098 \pm 0.034	0.088 \pm 0.021
	103	0.017 \pm 0.016	-	-	0.005 \pm 0.005
	105	0.017 \pm 0.016	0.048 \pm 0.025	0.012 \pm 0.012	0.025 \pm 0.011
	107	0.250 \pm 0.049	0.210 \pm 0.047	0.134 \pm 0.037	0.191 \pm 0.026
	109	0.150 \pm 0.040	0.065 \pm 0.029	0.024 \pm 0.016	0.074 \pm 0.017
	111	-	-	0.024 \pm 0.016	0.010 \pm 0.007
	115	0.333 \pm 0.060	0.468 \pm 0.064	0.561 \pm 0.057	0.466 \pm 0.037
	117	0.017 \pm 0.016	0.016 \pm 0.015	0.012 \pm 0.012	0.015 \pm 0.008
	123	-	0.032 \pm 0.031	0.012 \pm 0.012	0.015 \pm 0.011
D6S105	108	0.233 \pm 0.062	0.210 \pm 0.052	0.073 \pm 0.032	0.162 \pm 0.029
	110	-	0.032 \pm 0.031	0.024 \pm 0.016	0.020 \pm 0.012
	112	0.017 \pm 0.016	-	-	0.005 \pm 0.005
	114	-	0.048 \pm 0.034	-	0.015 \pm 0.011
	116	0.133 \pm 0.045	0.226 \pm 0.060	0.256 \pm 0.050	0.211 \pm 0.031
	118	0.067 \pm 0.029	0.032 \pm 0.021	0.073 \pm 0.026	0.059 \pm 0.015
	120	0.350 \pm 0.056	0.355 \pm 0.058	0.427 \pm 0.053	0.382 \pm 0.033
	122	0.033 \pm 0.022	0.032 \pm 0.021	-	0.020 \pm 0.009
	124	0.033 \pm 0.022	-	0.024 \pm 0.016	0.020 \pm 0.009
	126	0.017 \pm 0.016	-	0.024 \pm 0.016	0.015 \pm 0.008
	128	0.067 \pm 0.030	0.016 \pm 0.015	0.024 \pm 0.023	0.034 \pm 0.014
	130	0.050 \pm 0.026	0.016 \pm 0.015	0.061 \pm 0.030	0.044 \pm 0.015
	132	-	0.032 \pm 0.021	-	0.010 \pm 0.007
	138	-	-	0.012 \pm 0.012	0.005 \pm 0.005

Table 1. (Cont.)

Locus	Allele (bp)	Frequency \pm SE			
		Quebrada Baja (2N = 60)	Quebrada Alta (2N = 62)	Puna (2N = 82)	Total ^a (2N = 204)
D6S265	116	0.083 \pm 0.040	0.065 \pm 0.029	0.024 \pm 0.016	0.054 \pm 0.017
	120	0.150 \pm 0.051	0.210 \pm 0.060	0.159 \pm 0.034	0.172 \pm 0.027
	122	0.050 \pm 0.026	0.048 \pm 0.026	0.037 \pm 0.026	0.044 \pm 0.015
	124	0.500 \pm 0.069	0.483 \pm 0.066	0.512 \pm 0.057	0.500 \pm 0.038
	126	0.217 \pm 0.057	0.194 \pm 0.051	0.244 \pm 0.047	0.221 \pm 0.030
	128	-	-	0.012 \pm 0.012	0.005 \pm 0.005
	132	-	-	0.012 \pm 0.012	0.005 \pm 0.005
D6S2666	144	0.233 \pm 0.057	0.387 \pm 0.062	0.439 \pm 0.050	0.363 \pm 0.034
	148	0.017 \pm 0.016	-	-	0.005 \pm 0.005
	164	0.083 \pm 0.032	0.065 \pm 0.036	0.195 \pm 0.043	0.123 \pm 0.024
	168	0.033 \pm 0.022	0.016 \pm 0.015	0.024 \pm 0.016	0.025 \pm 0.015
	170	0.600 \pm 0.059	0.516 \pm 0.065	0.342 \pm 0.047	0.471 \pm 0.035
	172	0.033 \pm 0.032	0.016 \pm 0.015	-	0.015 \pm 0.011

2N, sample size in number of chromosomes analyzed

^a Jujuy highlands (Quebrada Baja, Quebrada Alta and Puna)

Genetic diversity parameters computed for the three subpopulations analyzed herein are shown in Supporting Information Table 2. In general terms, the lowest values of genetic diversity in Jujuy highlands were observed in Puna (> 3,500 m a.s.l.). Between Quebrada Baja (~ 2,500 m a.s.l.) and Quebrada Alta (~ 3,300 m a.s.l.), differences in genetic diversity were less apparent, although with a perceptible decreasing trend towards Quebrada Alta.

To assess genetic heterogeneity among Jujuy highlands' subpopulations, F_{ST} pairwise tests were performed from both the allele and the haplotype frequencies (Table 3). Several statistical significant differences were observed between subpopulations. The heterogeneity in frequencies for both the individual STR loci and the STR haplotypes were mostly detected between Puna and Quebrada Baja (D6S2792, D6S105, D6S2666 and the four-loci haplotype), which are those subpopulations occupying

the extremes of the variation in altitude above sea level. In addition, two microsatellite loci (D6S105 and D6S2666) were found to show significant frequency differences between Puna and Quebrada Alta, whereas D6S2666 differed between Quebrada Baja and Quebrada Alta.

Table 2. Hardy-Weimberg equilibrium (HWE) for autosomal STRs from chromosome 6 region in three population samples from Jujuy province (NW Argentina)

Locus	HWE by population ^a			Total ^c
	Quebrada Baja	Quebrada Alta	Puna	
D6S2792	0.290	0.073	0.138	0.011
D6S105	0.148	0.037	0.121	0.002
D6S265	0.052	0.328	0.094	0.058
D6S2666	0.234	0.180	0.583	0.034
2N ^b	60	62	82	204

^a Figures are *P*-values generated by the Fisher's exact probability test (Guo and Thompson, 1992)

^b 2N, sample size in number of chromosomes analyzed

^c Jujuy highlands (Quebrada Baja, Quebrada Alta and Puna)

In bold, significant *P*-values after Bonferroni correction for multiple testing (threshold value, $p < 0.0125$)

Table 3. Pairwise F_{ST} comparisons^a based on autosomal STR markers to test for genetic differentiation among three populations of Jujuy province (NW Argentina)

STR marker	QB / QA	QB / Puna	QA / Puna
D6S2792	0.063 ± 0.023	<0.001 ± <0.001	0.126 ± 0.043
D6S105	0.180 ± 0.027	<0.001 ± <0.001	0.045 ± 0.020
D6S265	0.847 ± 0.028	0.766 ± 0.058	0.559 ± 0.052
D6S2666	0.036 ± 0.015	<0.001 ± <0.001	0.009 ± 0.009
Haplotype	0.676 ± 0.043	<0.001 ± <0.001	0.207 ± 0.036

^a Figures are *P*-values ± standard errors for pairwise F_{ST} tests

QB, Quebrada Baja; QA, Quebrada Alta

Statistically significant F_{ST} values between populations are highlighted in bold

DISCUSSION

Geographic distance is considered a crucial factor in human population structure, because of its limiting effects on migratory movements (see Jorde, 1980). However, such a simple model certainly excludes a number of parameters that could have an impact in the shaping of the genetic structure of a population. The developing field of landscape genetics takes also into account other ecological

determinants influencing gene flow and, therefore, genetic variation, mostly associated with geographic landforms like rivers, lagoons and mountains, among others (Storfer et al., 2007). Recent studies on this issue indicate that the spatial heterogeneity of the environment is key in explaining gene flow patterns and population differentiation processes even at a microgeographic scale (Manel et al., 2003; McRae and Beier, 2007; Manel and Holderegger, 2013; Duforet-Frebourg and Blum, 2014).

An earlier study on the genetic diversity of Jujuy province based on polymorphic *Alu* insertions found that Native American ancestry increased with altitude above sea level (Gómez-Pérez et al., 2011a) since immigrants tended to settle at lower altitudes, leading to more genetic admixture. Populations at the highest altitude (Jujuy highlands) showed almost 100% of Native American ancestry, whereas the gene pool of the lowest population nuclei showed both European and African contribution. Accordingly, these authors also identified a decreasing trend of genetic heterozygosity from Jujuy highlands' to the population nuclei sited at lower-altitude regions (Valle and Selva). Yet, *Alu* markers failed to detect significant genetic differentiation among the Jujuy highlands' subpopulations (see Supporting Information Table 3).

Results of the present study confirm the effectiveness of the resolving power of the STR markers in identifying population differentiation processes even at a microgeographic level, as can be inferred from the comparative data presented Table 3 (STR markers) and Supporting Information Table 3 (*Alu* markers). Conflicting with the seeming lack of genetic heterogeneity in Jujuy highlands derived from the analysis of *Alu* insertions, the geographic distribution of the STR allele frequencies revealed spatial patterning, suggesting genetic substructuring among the three high-altitude human groups.

Genotype frequencies in populations with non-random mating (i.e., non-panmictic populations) frequently diverge from the expected proportions under HWE assumptions. Along these lines, results of the HWE test for each one of the Jujuy highlands' subpopulations classed according to altitude (Quebrada Baja, Quebrada Alta and Puna) showed no deviation from the HWE expectations. However, when we considered the three subgroups as a single population, 50% of the STR loci examined did not meet HWE (see Table 2). These findings contribute robust evidence on the existence of geographical barriers to random mating between the analyzed subpopulations. The remarkable

limitation of gene flow in Jujuy highlands could be explained in terms of landscape characteristics, bearing in mind the steep orography across the region.

Microsatellites are one of the most informative PCR-based genomic markers (Budowle et al., 1999; Rowold and Herrera, 2005). STR loci data have been argued to be useful for a variety of reasons, including high degree of heterozygosity, large number of codominant alleles, widespread distribution within the human genome and, last but not least, ease in genotyping and scoring. All these features make STRs one of the most powerful tools in human population genetics, and therefore, highly suitable for investigations with evolutionary, biomedical or forensic perspectives.

Genetic heterogeneity in Jujuy highlands was also reflected by the statistically significant F_{ST} values generated by the population differentiation tests (Table 3), thus contributing further evidence on the microevolutionary processes that have occurred in the study region. As expected, genetic microdifferentiation was particularly detectable between those populations located at the edges of the altitude range (Quebrada Baja: ~ 2500 m; Puna: > 3500 m). Puna is the more remote and intricate region of Jujuy highlands. These factors, in conjunction with the hostile living conditions associated with extreme altitudes have undoubtedly reinforced its degree of isolation. In fact, population density in Puna is 1.34 inh/km² (INDEC, 2010), the lowest in Jujuy highlands (Quebrada Baja: 5.64 inh/km²; Quebrada Alta: 2.95 inh/km²). Obviously, such a significantly low population size would have propitiated the reduction of the genetic diversity by genetic drift effect.

Taking into account the relatively short linear geographic distances between the population centers of Quebrada Baja and Quebrada Alta (minimum: 5.5 km, maximum: 43.2 km), genetic heterogeneity could be basically explained in terms of landscape characteristics, notably the steep orography. Specifically in Puna, orographic features would have substantially reinforced the partial population isolation generated, in principle, by geographic distance. A similar pattern of microevolutionary changes by lack of panmictic mating has been observed in several model and non-model organisms as well as in human populations, suggesting that the influence of landscape on local dispersal is a leading evolutionary force (Ramachandran et al., 2005).

Summarizing, the main evolutionary consequence of the altitude and, especially, of the steep landscape of Jujuy highlands would be an imbalance in the interplay between gene flow and genetic

drift favoring the impact of this latter evolutionary force. On the one hand, extreme living conditions typical of prominent altitudes limit the immigrants arrival, thus reducing the gene flow effects and the genetic diversity within each subpopulation. On the other hand, population isolation owing to a very steep local orography, together with markedly low population densities would have promoted intense genetic drift processes. The main consequence of the combined effect of a restricted gene flow with a significant impact of genetic drift would therefore be local genetic differentiation between Jujuy highlands' subpopulations (Slatkin, 1994). Both the loss of genetic diversity and the population differentiation are detectable mainly in Puna, coinciding with the highest population and the one exhibiting the lowest admixture levels (Gómez-Pérez et al., 2011a). Furthermore, the hypervariable nature of microsatellite markers has proved to be highly advantageous for the detection of recent microevolutionary signatures in the gene pool of human populations, as postulated in previous works (Shepard and Herrera, 2006; García-Obregón et al., 2007, among others). In contrast, *Alu* repeats are unique evolutionary events (Batzler et al., 1994) and thus much more conservative, since they are not exposed to the fluctuations of recurrent mutational events. In this way, polymorphic *Alu* elements could better detect the divergent ancestry of a set of geographically related populations (Gómez-Pérez et al., 2011a). The combined examination of *Alu* and STR loci may therefore provide complementary perspectives in the reconstruction of the demographic and evolutionary histories of partially isolated human groups.

ACKNOWLEDGMENTS

The authors are particularly indebted to all voluntary donors from Jujuy province (Argentina), who generously cooperated to the development of this research.

LITERATURE CITED

Agrawal S, Khan F. 2005. Reconstructing recent human phylogenies with forensic STR loci: a statistical approach. *BMC Genet* 6: 47.

- Batzler MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL. 1994. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci USA* 91:12288–12292.
- Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. 1999. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *J Forensic Sci* 44: 1277-1286.
- Capocasa M, Taglioli L, Anagnostou P, Paoli G, Danubio ME. 2014. Determinants of marital behaviour in five Apennine communities of Central Italy inferred by surname analysis, repeated pairs and kinship estimates. *Homo* 65: 64-74.
- Carroll SB. 2001. The big picture. *Nature* 409: 669.
- Dipierri JE, Alfaro E, Martínez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO. 1998. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol* 70: 1001-1010.
- Dipierri JE, Alfaro E, Peña JA, Constans J, Dugoujon JM. 2000. GM, KM immunoglobulin allotypes and other serum Genetic markers (HP, GC, PI and TF) among South American populations living at different altitudes (Jujuy Province, Argentina): Admixture estimates. *Hum Biol* 72: 305–319.
- Duforet-Frebourg N, Blum MG. 2014. Nonstationary patterns of isolation-by-distance: inferring measures of local genetic differentiation with Bayesian kriging. *Evolution* 68: 1110-1123.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population Genetics analyses under Linux and Windows. *Mol Ecol Resour* 10: 564-567.
- Forster P, Röhl A, Lünemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. *Am J Hum Genet* 67: 182-196.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, de Pancorbo MM, Peña JA. 2007. Polymorphic Alu insertions and the genetic structure of Iberian Basques. *J Hum Genet* 52: 317-327.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, Gómez-Pérez L, de Pancorbo MM, Peña JA. 2010. Genetic variability in autochthonous Basques from Guipuzcoa: a view from MHC microsatellites. *Int J Immunogenet* 37: 279-287.
- Gómez-Pérez L, Alfonso-Sánchez MA, Dipierri JE, Alfaro E, García-Obregón S, De Pancorbo MM, Bailliet G, Peña JA. 2011a. Microevolutionary processes due to landscape features in the province of Jujuy (Argentina). *Am J Hum Biol* 23: 177-184.
- Gómez-Pérez L, Alfonso-Sánchez MA, Sánchez D, García-Obregón S, Espinosa I, Martínez-Jarreta B, De Pancorbo MM, Peña JA. 2011b. Alu polymorphisms in the Waorani tribe from the Ecuadorian Amazon reflect the effects of isolation and genetic drift. *Am J Hum Biol*. 23: 790-795.
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48: 361–372.

-
- Hendry AP, Kinnison MT. 1999. The pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53: 1637-1653.
- INDEC. 2001. Censo Nacional de Población y Vivienda 1991 y Censo Nacional de Población, Hogares y Viviendas. Available at: <http://www.indec.gov.ar/webcenso/index.asp>.
- Jorde LB, Eriksson AW, Morgan K, Workman PL. 1982. The genetic structure of Iceland. *Hum Hered.* 32: 1-7.
- Jorde LB. 1980. The Genetic Structure of Subdivided Human Populations, a review. En: Mielke J, Crawford MD, editors. *Current developments in Anthropological Genetics, Vol 1. Theory and Methods.* New York, USA: Plenum Press. p 135-208.
- Manel S, Holderegger R. 2013. Ten years of landscape genetics. *Trends Ecol Evol* 28: 614-621.
- Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol* 18: 189-197.
- McRae B, Beier P. 2007. Circuit theory predicts gene flow in plant and animal populations. *Proc. Natl. Acad. Sci. USA* 104: 19885-19890.
- Nei M, Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 97: 145-163.
- Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci U S A.* 102: 15942-15947.
- Relethford JH. 1996. Genetic drift can obscure population history: problem and solution. *Hum Biol* 68: 29-44.
- Relethford JH. 2012. *Human Population Genetics.* Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term Genetic distance. *Genetics* 105: 767-779.
- Rowold DJ1, Herrera RJ. 2005. On human STR sub-population structure. *Forensic Sci Int* 151: 59-69.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual.* 2nd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Shepard EM, Chow RA, Suafo'a E, Addison D, Pérez-Miranda AM, Garcia-Bertrand RL, Herrera RJ. 2005. Autosomal STR variation in five Austronesian populations. *Hum Biol* 77: 825-851.
- Shepard EM, Herrera RJ. 2006. Iranian STR variation at the fringes of biogeographical demarcation. *Forensic Sci Int* 158: 140-148.
- Slatkin M, Excoffier L. 1996. Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity* 76: 377-383.
- Slatkin M. 1994. Gene flow and population structure. In: Real LA, (Ed). *Ecological Genetics.* New Jersey, USA: Princeton University Press. pp 3-18.
- Storfer A1, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP. 2007. Putting the "landscape" in landscape genetics. *Heredity (Edinb).* 98:128-142.
-

- Wood JW, Johnson PL, Kirk RL, McLoughlin K, Blake NM, Matheson FA. 1982. The genetic demography of the Gainj of Papua New Guinea. I. Local differentiation of blood group, red cell enzyme, and serum protein allele frequencies. *Am J Phys Anthropol* 57: 15-25.
- Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 74: 50-61.

Supporting information Table 1: p-values linkage disequilibrium^a for autosomal STRs from chromosome 6 region in three population samples from Jujuy province (NW Argentina)

Locus	D6S2792	D6S265	D6S105	D6S2666
D6S2792	-			
D6S265	0.204 ±0.013	-		
D6S105	0.013 ±0.004	0.000 ±0.000	-	
D6S2666	0.006 ±0.002	0.324 ±0.014	0.884 ±0.008	-

^a Lewontin and Kojima, 1960 and Slatkin and Excoffier, 1996.

Statistically significant p-values linkage disequilibrium between autosomal STR markers are highlighted in bold

Supporting information Table 2. Genetic diversity parameters for autosomal STRs (D6S2792, D6S105, D6S265, D6S2666) in three population samples from Jujuy province (NW Argentina)

Diversity parameter	Quebrada Baja	Quebrada Alta	Puna
Average H_o	0.633	0.556	0.616
Average H_e	0.716	0.699	0.679
HD	0.997	0.993	0.973

H_o , observed heterozygosity; H_e , expected heterozygosity; HD, unbiased haplotype diversity

Average values were calculated considering the four loci examined

Supporting information Table 3. Pairwise F_{ST} comparisons^a based on polymorphic *Alu* insertions^b to test for genetic differentiation among three populations of Jujuy province (NW Argentina).

<i>Alu</i> marker	QB / QA	QB / Puna	QA / Puna
A25	0.991 ± 0.003	0.676 ± 0.037	0.658 ± 0.047
APO	0.991 ± 0.003	0.784 ± 0.043	0.757 ± 0.039
ACE	0.991 ± 0.003	0.991 ± 0.003	0.432 ± 0.045
B65	0.658 ± 0.039	0.036 ± 0.020	0.234 ± 0.034
D1	0.865 ± 0.034	0.072 ± 0.023	0.333 ± 0.039
FXIIIB	0.288 ± 0.042	0.279 ± 0.034	0.991 ± 0.003
PV92	0.991 ± 0.003	0.072 ± 0.023	0.090 ± 0.024
TPA25	0.667 ± 0.051	0.108 ± 0.030	0.460 ± 0.029

^a Figures are *P*-values ± standard errors for pairwise F_{ST} tests

^b Data source: Gomez-Pérez et al. (2011a)

QB, Quebrada Baja; QA, Quebrada Alta

Statistically significant F_{ST} values between populations are highlighted in bold

PRINCIPALES RESULTADOS Y MATERIAL COMPLEMENTARIO

A través del análisis de los marcadores STR, se ha detectado una subdivisión significativa en el Alto Jujuy. Este fenómeno no fue apreciado mediante inserciones *Alu*, en un estudio previo (*Manuscrito III*). Ello, probablemente, sea debido al alto grado de polimorfismo de los marcadores STRs, más indicados para detectar evidencias de microevolución reciente en poblaciones humanas (Shepard y Herrera, 2006; García-Obregón *et al.* 2007).

En este trabajo se ha detectado heterogeneidad entre las subpoblaciones a partir de varios análisis. En primer lugar, se detectó desviación del equilibrio Hardy-Weinberg cuando se analizó la población de Jujuy Highlands como una única población, sin embargo estas desviaciones prácticamente desaparecieron una vez que cada subpoblación era considerada como una unidad independiente (*Tabla 2, Manuscrito VI*). Asimismo, el test F_{ST} (*Tabla 3, Manuscrito VI*), revela una diferenciación poblacional significativa entre regiones en varias comparaciones (7 de 15) y especialmente entre Puna y Quebrada Baja (4 de 5), las subpoblaciones que se encuentran separadas por una mayor diferencia de altitud. Sin embargo, puede observarse que en sólo 7 de las 80 comparaciones realizadas entre las comarcas de Jujuy para 8 inserciones *Alu* (*Tabla suplementaria S1, Manuscrito III*) el test F_{ST} mostró significación. En sólo un caso, además, implicó a poblaciones del Alto Jujuy (Puna Vs Quebrada Baja, para B65, $p=0,036$).

En todo caso, como se observaba en el *Manuscrito III*, se ha encontrado un cierto gradiente de reducción de la diversidad génica en relación con la altitud mediante varios índices (*Supporting information Table 2, Manuscrito V*). Entre las causas más probables para explicar este hecho, se proponen el aislamiento, debido a una escarpada orografía, y la deriva genética, por una baja densidad poblacional.

Tanto la pérdida de la diversidad genética como la microdiferenciación poblacional son más evidentes en la región de La Puna, coincidiendo con la población localizada a mayor altitud, con menor densidad poblacional y que exhibe los niveles más bajos de mestizaje (*Manuscrito III*).

En resumen, la principal consecuencia evolutiva del escarpado paisaje del alto Jujuy sería un desequilibrio en la interacción entre el Flujo Génico y la Deriva Genética, favoreciendo la acción de esta última. Por un lado, las condiciones de vida extremas típicas de altitudes importantes limitan la llegada inmigrantes, lo que reduce los efectos del Flujo Génico. Por otro lado, el aislamiento de la población debido a la orografía local, junto con densidades bajas de población marcadamente habría promovido intensos procesos de Deriva Genética, reduciendo la diversidad genética dentro de cada subpoblación. La principal consecuencia del efecto combinado de un flujo de genes restringido con un impacto significativo de la deriva genética sería por tanto la diferenciación genética local entre subpoblaciones del alto Jujuy (Slatkin, 1994b). Por lo tanto, a partir de estos resultados se ha llegado a la conclusión de que existe una tendencia a la microdiferenciación dentro de la región del Alto Jujuy, y que es independiente de los efectos del mestizaje con poblaciones provenientes de otros continentes que habitan en Jujuy.

Tabla 6. Promedio del índice de diversidad genética Garza-Williamson (Garza y Williamson, 2001) evaluado a para cada uno de los 4 STRs del MHC en la región de Jujuy

	Puna	Quebrada Alta	Quebrada Baja
Avg G-W Index	0.304	0.331	0.369

Con el fin de completar la información acerca de los procesos que han ocurrido en las Además se realizó el test para calcular el índice Garza-Williamson (*Tabla 6*). Gracias a éste test se puede comprobar la estabilidad que han tenido las poblaciones en su tamaño poblacional a lo largo del tiempo, es decir los posibles cuellos de botella que han sufrido las poblaciones. Este índice fue calculado a través del software ARLEQUIN v3.5 (Excoffier and Lischer 2010).

Los resultados del test nos mostraron valores bajos (cerca de 0) para la población general de Jujuy, indicando que ha sufrido variaciones recientes en el tamaño poblacional, siendo una población poco estable (Garza y Williamson, 2001). Sin embargo, cabe destacar que los valores obtenidos no son excesivamente bajos; esto sugiere que las reducciones en el tamaño poblacional no ha sido muy drásticas. Es posible que estas poblaciones hayan mantenido tamaños poblacionales bajos durante los últimos siglos y que en determinados momentos, hayan sufrido ligeros cuellos de botella, que han acentuado la acción de la Deriva Genética.

Al observar los valores encontrados en las diferentes subregiones, se puede apreciar que existe un gradiente descendente de Quebrada Baja hacia Puna; por lo tanto, las reducciones en el tamaño poblacional que han sufrido Quebrada Alta y sobre todo Puna, han sido poblaciones menos estables y más afectadas por los cuellos de botella, por lo tanto más susceptibles de haber sufrido los efectos de la Deriva Genética. Esto se ve reflejado a lo largo de los demás trabajos realizados en la Tesis Doctoral en los índices de diversidad genética realizados (*Manuscrito III, Manuscrito V y Manuscrito VI*).

Desequilibrio de Ligamiento

Para completar el estudio de ligamiento realizado en esta región del genoma se han incorporado los datos relativos a las inserciones *Alu* analizadas (*Manuscrito V*). El análisis de ligamiento entre los 9 marcadores (5 inserciones *Alu* y 4 STRs) se muestra en la *Tabla 7*. A partir de este análisis, se observan 13 asociaciones significativas, siendo 7 de ellas mixtas, entre STR e inserción *Alu*. Se observa, por tanto, un notable grado de desequilibrio de ligamiento entre los diferentes marcadores analizados.

Es destacable que en la realización del test con las combinaciones de STR y *Alu* se observe una comparación significativa entre inserciones *Alu* (MICB-HJ), que en otros test realizados anteriormente (*Manuscrito V*) no dio significativa.

El hecho de trabajar con marcadores heterogéneos en una región con desequilibrio de ligamiento abre nuevas líneas potenciales de trabajo, ya que los STRs podrían contribuir a detectar haplotipos específicos de una población o un continente, en base a su alto grado de polimorfismo, en tanto que las inserciones *Alu* permitirán diferenciar los haplotipos derivados de los ancestrales, ya que se conoce el sentido de la mutación, que es siempre de no inserción hacia inserción.

Tabla 7. Desequilibrio de ligamiento entre 9 marcadores localizados en la región del MHC realizado mediante el Programa Arlequin V3.5 Se muestran los valores de p

	<i>Alu</i> MICB	<i>Alu</i> HJ	<i>Alu</i> HG	<i>Alu</i> TF	<i>Alu</i> HF	D6S2792	D6S265	D6S105	D6S2666
<i>Alu</i> MICB	/								
<i>Alu</i> HJ	0.014	/							
<i>Alu</i> HG	0.242	0.000	/						
<i>Alu</i> TF	0.011	0.249	0.677	/					
<i>Alu</i> HF	0.863	0.894	1.000	0.829	/				
D6S2792	0.000	0.172	0.026	0.016	0.000	/			
D6S265	0.009	0.000	0.000	0.506	0.194	0.391	/		
D6S105	0.288	0.657	0.492	0.716	0.155	0.025	0.000	/	
D6S2666	0.097	0.062	0.143	0.058	0.869	0.008	0.622	0.885	/

Aparecen destacados en negrita los valores de *P* significativos ($p < 0.05$)

Aparecen destacadas en cursiva un par marcadores *Alu* cuya comparación no dio significativa en el análisis realizado en el artículo

DISCUSIÓN

DISCUSIÓN GENERAL

El estudio de la microevolución en poblaciones humanas es un proceso complejo debido a los numerosos factores que pueden afectarla, otorgándole así un carácter multidisciplinar.

Las interacciones entre las presiones microevolutivas, principalmente el flujo génico y la deriva genética y eventualmente los procesos de adaptación locales, pueden verse afectadas por factores orográficos, demográficos y ecológicos, que pueden actuar como barreras a la panmixia. De este modo, estos factores contribuyen a modelar el patrimonio genético de las poblaciones en la forma de procesos de microevolución. Entre los factores demográficos, es interesante tener en cuenta el tamaño efectivo de la población reproductora o el tamaño censal en su caso, su densidad, su estructura por edad y sexo, tasas de natalidad y mortalidad y patrones de búsqueda del cónyuge, considerando la distribución geográfica de todos ellos (Relethford, 2012).

Los factores demográficos y ecológicos pueden a su vez verse condicionados por determinantes culturales. Si bien la especie humana no es la única que posee cultura (Whiten *et al.* 1999), es, en todo caso, la más compleja, habiendo llegado en ocasiones a afectar notablemente a la densidad, distribución geográfica e incluso la supervivencia de las poblaciones. Por lo tanto, es evidente que puede afectar de forma perceptible a su estructura genética (Relethford, 2012).

En consecuencia, un análisis exhaustivo de la distribución espacial de la heterogeneidad genética en una población requerirá de un enfoque multidisciplinar para interpretar con precisión los procesos poblacionales que hayan podido afectar a su acervo genético.

Esta Tesis se ha centrado en una parte de los factores mencionados, incluyendo el tamaño y la densidad poblacional, la orografía y otros factores del entorno, así como algunos aspectos socioculturales, haciendo hincapié en la incidencia que han mostrado todos ellos sobre la deriva y el flujo génico.

UTILIDAD DE LOS MARCADORES *Alu* Y STR

Para el estudio de los procesos microevolutivos en poblaciones sudamericanas se han empleado 21 marcadores genéticos con diferentes características, en varias poblaciones también significativamente heterogéneas. Se han utilizado dos tipos de marcadores (inserciones *Alu* y Microsatélites) por su capacidad para evaluar acontecimientos a diferente escala temporal. Ello es debido a las características específicas que poseen. Las inserciones *Alu*, con bajas tasas de mutación ($2,3 \times 10^{-9}$ por locus y por generación) (Relethford, 2012), son útiles para informar acerca de eventos ancestrales (Bamshad *et al.* 2003; Relethford, 2012). Por otro lado, los microsatélites tienen una elevada tasa de mutación (entre $1,52 \times 10^{-3}$ y $8,5 \times 10^{-4}$ por locus y por generación) (Zhivotovsky *et al.* 2000, 2003) por lo que muestran mayor variabilidad y serán más adecuados para la descripción de procesos recientes (Mesa *et al.* 2000).

Las propiedades de las inserciones *Alu* polimórficas, mencionadas en capítulos anteriores de esta memoria, hacen que sean marcadores altamente interesantes para la investigación de la estructura de las poblaciones humanas (Raaum *et al.* 2010). En varios trabajos disponibles en la bibliografía se han utilizado para la discriminación de individuos y poblaciones (Bamshad *et al.* 2003; Watkins *et al.* 2003; Witherspoon *et al.* 2006), con buenos resultados.

En el *Manuscrito I* de esta memoria, un grupo de 8 *Alu* autosómicas se ha revelado apto para la correcta clasificación de las poblaciones de diferentes subcontinentes, siendo por tanto útiles para la estima de tasas de mestizaje y de diferenciación entre poblaciones. Además, como se describe en el *Manuscrito II*, la mayor parte de estas inserciones (ACE, APO, FXIII_B, PV92 y TPA 25) han superado los valores de δ requeridos para su utilización como *AIMs*. De hecho, algunas ya han sido utilizadas con una cierta frecuencia como tales en diferentes trabajos (Bonilla *et al.* 2004; Luizon *et al.* 2008; Terreros *et al.* 2009).

En otra línea de trabajo, elementos *Alu* localizados en una misma región y por lo tanto próximos entre sí, han mostrado que su utilización en forma de haplotipos proporciona una

notable capacidad de discriminación entre poblaciones (*Manuscrito V*). Lamentablemente, su uso se ha visto limitado a causa de la ausencia de datos disponibles para un número amplio de poblaciones. No obstante, los análisis realizados muestran claros indicios de su utilidad, particularmente a partir de los resultados obtenidos mediante AMOVAs entre poblaciones de diferentes continentes. A pesar del bajo grado de polimorfismo de las inserciones *Alu*, que limita el número de combinaciones haplotípicas posibles, han aparecido 3 haplotipos específicos de continente, dos de ellos de distribución europea y uno de distribución asiática.

Por otra parte, los microsatélites han mostrado una mayor efectividad a la hora de revelar procesos de microdiferenciación locales (*Manuscrito VI*), procesos que mediante el análisis de inserciones *Alu* no se encontraron significativos.

PROCESOS DE MICRODIFERENCIACIÓN EN POBLACIONES DE SUDAMÉRICA

En los trabajos incluidos en esta Tesis se han observado los efectos de diferentes factores sobre los procesos microevolutivos. Entre estos factores se incluye algunos intrínsecos o inherentes a las poblaciones (características demográficas, culturales, étnicas, sociales, etc.) y otros externos a las propias poblaciones (el entorno, la orografía, etc.). Se ha constatado que todos estos factores pueden modificar la intensidad y distribución del flujo génico, y modular la incidencia de la deriva genética sobre las poblaciones; afectando, por tanto, a la estructura de las poblaciones y contribuyendo a dar forma a su patrimonio genético.

El flujo génico es un componente esencial de la estructura poblacional, dado que su intensidad caracteriza el grado de independencia de cada subpoblación desde un punto de vista microevolutivo. Un notable flujo génico entre subpoblaciones generalmente determinará una tendencia a su homogeneización; si por el contrario existen barreras al flujo génico y por tanto éste es escaso, las subpoblaciones evolucionarán independientemente, con una tendencia a la

microdiferenciación, lo que implicará aislamiento reproductivo y eventualmente el establecimiento de linajes evolutivamente independientes (Slatkin, 1994b).

En el contexto de las poblaciones sudamericanas analizadas, se han observado unos bajos niveles de flujo génico principalmente en la etnia Huaorani, en la región del Alto Jujuy y en la población afrocolombiana; reflejando un cierto aislamiento reproductor con un probablemente alto grado de endogamia. Esta situación prolongada en el tiempo y unida a un escaso tamaño poblacional, puede desencadenar una serie de consecuencias:

- I. Pérdida de diversidad genética y una progresiva homogeneización entre los individuos de una población (Alfaro y Dipierri, 1996; Cardoso *et al.* 2008, 2012; Baeta *et al.* 2009, 2013; González-Andrade *et al.* 2009; Baeta, 2011), ya que se intensifican los efectos de la deriva genética (*Manuscritos III, IV y VI*).
- II. Diferenciación paulatina de la población respecto de otras poblaciones próximas, elevando así la heterogeneidad interpoblacional y promoviendo la microevolución (Slatkin, 1994b). Es por ello que la magnitud del flujo génico que mantienen las poblaciones actúa como un importante factor regulador, provocando grados variables de diferenciación (*Manuscritos III, IV, V y VI*).
- III. Finalmente, unos elevados niveles de endogamia generarán a su vez un aumento de la consanguinidad. En efecto, un habitante de una población aislada verá limitados sus cónyuges potenciales a los existentes dentro de su población; en consecuencia, aumentará la consanguinidad aleatoria, simplemente por la dificultad de encontrar un cónyuge no emparentado (Relethford, 2012; Capocasa *et al.* 2014).

Situaciones de este tipo se han observado tanto en la etnia Huaorani como en la región de Jujuy, y en menor medida entre los afrocolombianos. En cada caso la reducción del flujo génico presenta unas características diferenciadas para cada población.

En el caso de la población Huaorani, se ha encontrado una importante reducción de la diversidad genética, así como bajos valores de heterocigosidad. El severo aislamiento genético que han sufrido ha sido el resultado tanto de factores intrínsecos a la población como de factores externos (*Manuscrito IV*).

Entre los factores endógenos, destacan los socioculturales. Los Huaorani se caracterizan por constituir sociedades agresivas y belicosas, lo que evidentemente ha condicionado sus contactos con otras poblaciones, incluyendo a menudo otros grupos (*Huaomoni*) de su misma etnia. También pueden incluirse en este grupo los factores demográficos, ya que muestran un reducido tamaño poblacional, que ha potenciado la pérdida de diversidad debido al efecto de la deriva genética. El modo de vida seminómada, que aún practican una parte de ellos, limita considerablemente el tamaño de sus poblaciones, haciendo que el efecto de la deriva sea notable. Por otra parte, entre los Huaorani los emparejamientos se limitan al grupo familiar, con preferencia por los matrimonios entre primos, promoviéndose de esta forma la consanguinidad (Larrick *et al.* 1985; Robarchek y Robarchek, 1992; Beckerman *et al.* 2009).

Entre los factores externos, es preciso considerar el entorno de pluvisilva, un hábitat poco accesible, que puede considerarse como un entorno hostil para la mayor parte de los humanos, lo que ha facilitado el aislamiento de los Huaorani tanto entre ellos como con otras poblaciones nativas americanas y que ha retrasado su contacto con los pobladores europeos o africanos que llegaron a partir del siglo XVI al continente americano.

Este conjunto de factores ha provocado una diferenciación genética muy importante con otras poblaciones amerindias y una considerable pérdida de diversidad, con altas tasas de homocigosidad (Baeta *et al.* 2009, 2013; González-Andrade *et al.* 2009; Baeta, 2011; Cardoso *et al.* 2012).

En Jujuy, el aislamiento reproductor ha sido causado principalmente por variables exógenas. La orografía ha tenido en su poblamiento una importancia fundamental, condicionando su

estructura genética y en buena medida modelando su acervo genético; algo destacado en otras poblaciones de montaña (Capocasa *et al.* 2014) y reportado para esta población a través de análisis de ADN antiguo (Mendisco *et al.* 2014). Sin embargo no ha sufrido un aislamiento tan severo como la etnia Huaorani, por lo que sus consecuencias son menos evidentes (*Manuscritos III, IV y VI*).

En todo caso, no pueden desecharse por completo los factores endógenos como causa de la microevolución en Jujuy. En efecto, algunas variables demográficas han podido condicionar la composición del patrimonio genético de la población de Jujuy, fundamentalmente debido a unos tamaños poblacionales pequeños, básicamente en su parte más alta, habiéndose observado efectos de una notable acción de la deriva genética en el Alto Jujuy (*Manuscritos VI*). Aunque a su vez la escasa densidad de población de este territorio se habrá visto determinada por una evidente escasez de recursos del entorno.

La imponente orografía de Jujuy ha afectado claramente al flujo génico entre las poblaciones de esta región. Así, existe un proceso de microdiferenciación entre el Alto Jujuy y las zonas más bajas de la provincia (*Manuscrito III*). De hecho, y gracias a la diferente capacidad de resolución de los marcadores empleados, se ha encontrado una incipiente microdiferenciación también entre las regiones más altas (Alto Jujuy), las cuales, mediante el análisis de inserciones *Alu* aparecían como una población homogénea y con escasos indicios de microdiferenciación (*Manuscrito VI*).

En la provincia de Jujuy, y como consecuencia de la convivencia entre individuos de diferentes orígenes, existe un mestizaje entre las diferentes etnias. Este mestizaje ha afectado a la estructura genética de la población. Sin embargo, la orografía ha dificultado e incluso impedido el mestizaje entre los diferentes grupos humanos en algunas comarcas (Dipierri *et al.* 1998, 2000; Bejarano *et al.* 2009). Por ello, se ha observado que el mestizaje no ocurre de una manera regular entre las subpoblaciones de Jujuy, sino que se ha impuesto un patrón en relación a la altitud (*Manuscritos III*), que se explica por la dificultad para la movilidad de los individuos entre los

diferentes núcleos poblacionales, así como por el interés económico diferencial en los recursos susceptibles de ser explotados.

El patrón de asentamiento ha sido diferencial desde la época colonial. Las poblaciones inmigrantes han estado históricamente asentadas en las zonas más bajas de la provincia, desplazando a las amerindias a las zonas localizadas a mayor altitud. De este modo, las poblaciones localizadas a mayor altitud quedaron más aisladas y vieron reducido su contacto con poblaciones no amerindias. Es por ello que la composición del patrimonio genético del Alto Jujuy es casi exclusivamente amerindio (Dipierri *et al.* 1998, 2000; Bejarano *et al.* 2009; *Manuscritos III, V y VI*).

En cuanto a la población Afrocolombiana, a partir de las estimas sobre su grado de mestizaje se ha deducido que esta población ha sufrido un aislamiento notable en relación a otras etnias de Antioquia, principalmente respecto a la de origen caucasioide (*Manuscrito II*). Esta circunstancia ha sido descrita asimismo por otros autores, que han sugerido que el componente europeo en el acervo genético de las poblaciones afroamericanas procedentes de palenques tiende a ser escasa, debido a la hostilidad tradicional de sus habitantes, los cimarrones, hacia los colonos europeos (Castro de Guerra 1993; Bortolini *et al.* 1995; Martínez-Labarga *et al.* 1999; Cotrim *et al.* 2004). El comportamiento de otras comunidades afroamericanas no procedentes de los palenques o quilombos es muy variable (Basu *et al.* 1976; Crawford *et al.* 1981; Schneider *et al.* 1987; Biondi *et al.* 1988; Bortolini *et al.* 1992, 1995, 1997, 1998, 1999; Castro de Guerra *et al.* 1996; Martínez-Labarga *et al.* 1999; Da Silva *et al.* 1999; Sans *et al.* 2002; Zabala Fernández *et al.* 2005; Salzano y Sans, 2014) aunque generalmente con mayores valores de mestizaje (*Tabla suplementaria S1, Manuscrito III*).

Con toda probabilidad han sido las circunstancias ocurridas durante el proceso de formación de sus poblaciones las que han condicionado dicha hostilidad y el consecuente aislamiento, por lo que podrían servir como paradigma del aislamiento genético debido a causas intrínsecas a la población. La esclavitud, la persecución de los esclavos huidos y el origen de los palenques como su lugar de refugio han sido presumiblemente los factores esenciales de una barrera social a la

panmixia en estas poblaciones. En todo caso, se ha observado una gran heterogeneidad de comportamientos, de modo que en otras poblaciones afroamericanas con un origen similar se han encontrado grados menores de aislamiento (Arpini-Sampaio *et al.* 1999; Bortolini *et al.* 1999; Da Silva *et al.* 1999; Sans *et al.* 2002), similares (Castro de Guerra *et al.* 1996) y también hay registrados algunos casos con un aislamiento mayor (Zabala Fernández *et al.* 2005).

Sin embargo, el grado de aislamiento no aparece necesariamente reflejado mediante las medidas habituales de diversidad genética. Así, los afrocolombianos son una población cuyo origen se encuentra en África, continente que presenta la mayor diversidad genética. Las poblaciones afroamericanas actuales son, además, el resultado de una amalgama de individuos provenientes de diferentes etnias y regiones del África subsahariana (Curtin 1969; Mello y Souza, 2006; Gontijo *et al.* 2014), lo que sin duda ha contribuido al mantenimiento de una elevada diversidad genética en esta población a pesar de su presumiblemente elevada endogamia. Y es que dicha endogamia habría ocurrido entre individuos provenientes de diferentes etnias africanas que convivían en un mismo espacio debido a una situación peculiar derivada de la trata de esclavos.

Por otra parte, la población mestiza de Antioquia habrá sido una población abierta al mestizaje, siendo el resultado una población con una composición de origen mayoritariamente europeo, pero con aportes relevantes de amerindios y afroamericanos (*Manuscrito II*). Este resultado se encuentra en consonancia con los observados en otros estudios realizados sobre poblaciones sudamericanas próximas (Sandoval *et al.* 1993; Zabala-Fernández *et al.* 2005; Bedoya *et al.* 2006; Wang *et al.* 2008). Es preciso mencionar, no obstante, el amplio rango de comportamientos observado en las poblaciones mestizas a este respecto en diferentes regiones de Sudamérica (*Supplementary table S1, Manuscrito II*).

En definitiva, si desde un punto de vista microevolutivo, a partir de los siglos XVIII y XIX, la situación generalizada en Europa y amplias regiones de otros continentes, es la derivada de la rotura de los aislados, con una exogamia generalizada y una tendencia a la homogeneización de las

poblaciones, en Sudamérica puede todavía hoy encontrarse un amplio rango de situaciones, que oscilan desde la exogamia al aislamiento más extremo, existiendo poblaciones que aún hoy en día tienden a la microdiferenciación, siendo la principal consecuencia de los desequilibrios en la interacción entre las fuerzas microevolutivas del Flujo Génico y de la Deriva Genética.

CONCLUSIONES

CONCLUSIONES

Teniendo en cuenta los resultados presentados en esta memoria, se han extraído las siguientes conclusiones:

- I. Se ha constatado la eficiencia de un pequeño número de inserciones *Alu* para la clasificación de poblaciones de orígenes diversos y consecuentemente para la estima de tasas de mestizaje.
- II. Mediante la utilización de diferentes marcadores genéticos, como inserciones *Alu* y STRs, pueden obtenerse niveles de resolución variables en la identificación de procesos microevolutivos, de modo que puede resultar adecuado su uso de forma conjunta cuando se pretenda analizar una población con variabilidad en los patrones espaciales de microdiferenciación.
- III. Las poblaciones sudamericanas muestran un rango de variación muy amplio en la intensidad y el sentido (es decir, tendencia a la microdiferenciación o a la homogeneización) de los procesos microevolutivos.
- IV. Los mestizos de Antioquia, que pueden considerarse representativos de la mayor parte de la población colombiana, se caracterizan por tener una proporción importante de su patrimonio genético con un origen europeo y una fracción menor de origen amerindio y africano.
- V. Los afrocolombianos de Antioquia analizados en esta Tesis Doctoral son un caso extremo entre los afroamericanos por su escaso grado de mestizaje. Han mostrado muy poco flujo génico con poblaciones de amerindios y ninguno con poblaciones de origen europeo, debido probablemente a su origen como descendientes de cimarrones refugiados en palenques.

- VI. La provincia de Jujuy ha revelado un patrón de diversidad genética marcado por procesos de microdiferenciación. Dichos procesos han sido principalmente determinados por la orografía del territorio, mostrando notables variaciones en la intensidad de la acción de la deriva y el flujo génico en las diferentes comarcas.
- VII. Entre los Huaorani se ha observado una reducción drástica de su diversidad genética, debida a una intensa acción de la deriva genética. Las causas hay que buscarlas en este caso preferentemente entre los factores culturales, ya que su prolongado aislamiento se ha visto favorecido por una tradición de enfrentamientos con otros grupos vecinos, que han ocasionado los consiguientes procesos de cuello de botella, así como por una marcada tendencia a la búsqueda preferente de un cónyuge emparentado.
- VIII. La orografía del entorno, el medio ambiente y las características socioculturales de las poblaciones pueden condicionar en gran medida la acción de la deriva genética y el flujo génico, determinando la intensidad y el sentido de los procesos microevolutivos en las poblaciones humanas.

REFERENCIAS BIBLIOGRÁFICAS

- Abraham LJ, French MA, Dawkins RL. 1993. Polymorphic MHC ancestral haplotypes affect the activity of tumour necrosis factor-alpha. *Clin Exp Immunol* 92, 14-18.
- Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ. 2008. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 12, 3 e1764.
- Ade C, Roy-Engel AM, Deininger PL. 2013. Alu elements: an intrinsic source of human genome instability. *Curr Opin Virol* 3, 639-645.
- Adovasio JM, Pedler DR. 2004. Pre-Clovis sites and their implications for human occupation before the Last Glacial Maximum. En: Madsen DM (Ed). *Entering America: Northeast Asia and Beringia before the last Glacial Maximum. Salt Lake City, Utah, USA: Univ of Utah Press.* pp 139-158.
- Ahmed W, Malik M, Saeed I, Khan AA, Sadeque A, Kaleem U, Ahmed N, Ajmal M, Azam M, Qamar R. 2011. Role of tissue plasminogen activator and plasminogen activator inhibitor polymorphism in myocardial infarction. *Mol Biol Rep* 38, 2541-2548.
- Alfaro E, Dipierri JE. 1996. Isonimia, endogamia, exogamia y distancia marital en la provincia de Jujuy. *Rev Arg Antropol Biol* 1, 41-56.
- Alfaro E. 2010. Dinámica antroponímica y estructura demogenética en Casabindo: siglos XVI al XXI. Tesis doctoral. Universidad Nacional de la Plata.
- Alper CA, Awdeh Z, Yunis EJ. 1992. Conserved, extended MHC haplotypes. *Exp Clin Immunogenet* 9, 58-71.
- Alvarez VM. 1996. Poblamiento y población en el valle de Aburrá y Medellín, 1541-1951. En: Melo JO (Ed). *Historia de Medellín. Medellín, Colombia: Suramericana.* pp 23-27.
- Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67, 444-461.
- Ammerman AJ, Biagi P. 2003. *The widening harvest. Boston, USA: Archaeological Institute of America.*
- Ammerman AJ, Cavalli-Sforza LL. 1971. Measuring the rate of spread of early farming in Europe. *Man* 6, 674-688.
- Ammerman AJ, Cavalli-Sforza LL. 1973. A population model for the diffusion of early farming in Europe. En: Renfrew C (Ed). *The explanation of culture change. London, UK: Duckworth.* pp 343-357.
- Ammerman AJ, Cavalli-Sforza LL. 1984. *The Neolithic Transition and the Genetics of Populations in Europe. Princeton, New Jersey, USA: Princeton University Press.*
- Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP, Herrera RJ. 2002. Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis* 23, 3346-3356.
- Aranguren Méndez JA, Román Bravo R, Isea W, Villasmil Y, Jordana J. 2005. Los microsatélites (STR's), marcadores moleculares de ADN por excelencia para programas de conservación: una revisión. Facultad de Ciencias Veterinarias, La Universidad de Zulia. *Arch Latinoam Prod Anim* 13, 1-6.
- Arcot SS, Adamson AW, Lamerdin JE, Kanagy B, Deininger PL, Carrano AV, Batzer MA. 1996. Alu fossil relics—Distribution and insertion polymorphism. *Genome Res* 6, 1084-1092.
- Arcot SS, Fontius JJ, Deininger PL, Batzer MA. 1995a. Identification and analysis of a "young" polymorphic Alu element. *Biochim Biophys Acta* 1263, 99-102.
- Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA. 1995b. Alu repeats: a source for the Genesis of primate microsatellites. *Genomics* 29, 136-144.
- Armour JAL, Neumann R, Gobert S, Jeffreys AJ. 1994. Isolation of human simple repeat loci by hybridisation selection. *Hum Mol Genet* 3, 599-605.
- Arnold SJ, Pfrender ME, Jones AG. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* 112-113, 9-32.

- Arpini-Sampaio Z, Costa MCB, Melo AA, Carvalho MFVA, Deus MSM Simões AL. 1999. Genetic polymorphisms and ethnic admixture in African-derived black communities of northeastern Brazil. *Hum Biol* 71:69–85.
- Asari M, Omura T, Oka K, Maseda C, Tasaki Y, Shiono H, Matsubara K, Matsuda M, Shimizu K. 2012. Multiplex PCR-based *Alu* insertion polymorphisms genotyping for identifying individuals of Japanese ethnicity. *Genomics* 99, 227–232.
- Atkinson QD, Gray RD, Drummond AJ. 2008. mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Mol Biol Evol* 25, 468–474.
- Baeta M, Núñez C, Aznar JM, Sosa C, Casalod Y, Bolea M, González-Andrade F, de Pancorbo MM, Martínez-Jarreta B. 2013. Analysis of 10 X-STRs in three population groups from Ecuador. *Forensic Sci Int Genet* 7, e19–20.
- Baeta M, Núñez C, González-Andrade F, Sosa C, Casalod Y, Bolea M, Zweynert S, Vacas-Cruz OW, González-Solorzano J, Geppert M, Roewer L, Martínez-Jarreta B. 2009. Mitochondrial analysis revealed high homogeneity in the Waorani population—The last nomadic group of huntergatherers from Ecuador. *Forensic Sci Int Genet Suppl Series 2*, 313–314.
- Baeta M. 2011. Estudio de la variabilidad del genoma mitocondrial y de marcadores sexuales en grupos étnicos de Ecuador. Tesis Doctoral. Universidad de Zaragoza.
- Bahri R, Ben Halima A, Ayadi I, Esteban E, Alfidhli SM, Rebai A, Moral P, Chaabani H. 2013. Genetic position of Bahrain natives among wider Middle East populations according to *Alu* insertion polymorphisms. *Ann Hum Biol* 40, 35–40.
- Bailey SL, Carpentier PA, McMahon EJ, Begolka WS, Miller SD. 2006. Innate and adaptive immune responses of the central nervous system. *Crit Rev Immunol* 26, 149–188. Review.
- Bamshad M, Wooding S, Salisbury BA, Stephens JC. 2004. Deconstructing the relationship between Genetics and race. *Nat Rev Genet* 5, 598–609.
- Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB. 2003. Human population genetic structure and inference of group membership. *Am J Hum Genet* 72, 578–589.
- Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evo* 16, 37–48.
- Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang YP. 2003. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Ann Hum Genet* 67, 512–524.
- Barbujani G, Colonna V. 2010. Human genome diversity: frequently asked questions. *Trends Genet* 26, 285–295.
- Barnette T, Constantin A, Gourraud PA, Abbal M, Garnier JG, Cantagrel A, Cambon-Thomsen A. 2006. Microsatellite typing of the human leucocyte antigen region: analytical approach and contribution to rheumatoid arthritis ImmunoGenetic studies. *Tissue Antigens* 68, 390–398.
- Basu A, Namboodiri KK, Weitkamp LR, Brown WH, Pollitzer WS, Spivey MA. 1976. Morphology, serology, dermatoglyphics, and microevolution of some village populations in Haiti, West Indies. *Hum Biol* 48, 245–269.
- Battaglia V, Grugni V, Perego UA, Angerhofer N, Gomez-Palmieri JE, Woodward SR, Achilli A, Myres N, Torroni A, Semino O. 2013. The first peopling of South America: new evidence from Y-chromosome haplogroup Q. *PLoS One* 8, e71390.
- Battilana J, Fagundes NJ, Heller AH, Goldani A, Freitas LB, Tarazona-Santos E, Munkhbat B, Munkhtuvshin N, Krylov M, Benevolenskaia L, Arnett FC, Batzer MA, Deininger PL, Salzano FM, Bonatto SL. 2006. *Alu* insertion polymorphisms in Native Americans and related Asian populations. *Ann Hum Biol* 33, 142–160.
- Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M. 1996a. Genetic variation of recent *Alu* insertions in human populations. *J Mol Evol* 42, 22–29.
- Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E. 1996b. Standardized nomenclature for *Alu* repeats. *J Mol Evol* 42, 3–6.

- Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. *Nat Rev Genet* 3, 370-379. Review.
- Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL. 1990. Structure and variability of recently inserted Alu family members. *Nucleic Acids Res* 18, 6793-6798.
- Batzer MA, Rubin CM, Hellmann-Blumberg U, Alegria-Hartman M, Leeflang EP, Stern JD, Bazan HA, Shaikh TH, Deininger PL, Schmid CW. 1995. Dispersion and insertion polymorphism in two small subfamilies of recently amplified human Alu repeats. *J Mol Biol* 31, 247, 418-427.
- Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ. 1994. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci USA* 91, 12288-12292.
- Beckerman S, Erickson PI, Yost J, Regalado J, Jaramillo L, Sparks C, Iromenga M, Long K. 2009. Life histories, blood revenge, and reproductive success among the Waorani of Ecuador. *Proc Natl Acad Sci USA* 106, 8134-8139.
- Beckman JS, Weber JL. 1992. Survey of human and rat microsatellites. *Genomics* 12, 627-631.
- Bedoya G, Montoya P, Garcia J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, Ruiz-Linares A. 2006. Admixture dynamics in Hispanics: A shift in the nuclear Genetic ancestry of a South American population isolate. *Proc Natl Acad Sci USA* 103, 7234-7239.
- Bejarano I, Dipierri JE, Alfaro E. 2006. Datos censales y evolución del flujo génico en la provincia de Jujuy, Argentina (1869-1991). *Antropo* 13, 19-33. www.didac.ehu.es/antropo.
- Bejarano IF, Dipierri JE, Andrade A, Alfaro EL. 2009. Geographic altitude, surnames, and height variation of Jujuy (Argentina) conscripts. *Am J Phys Anthropol* 138, 158-163.
- Benham FJ, Spurr N, Povey S, Brinton BT, Goodfellow PN, Solomon E, Harris TJ. 1984. Assignment of tissue-type plasminogen activator to chromosome 8 in man and identification of a common restriction length polymorphism within the gene. *Mol Biol Med* 2, 251-259.
- Bercovitch S. 1975. The puritan origins of the American self. *New Haven, Connecticut, USA: Yale University Press*.
- Berkman CC, Dinc H, Sekeryapan C, Togan I. 2008. Alu insertion polymorphisms and an assessment of the genetic contribution of Central Asia to Anatolia with respect to the Balkans. *Am J Phys Anthropol* 136, 11-18.
- Berovides-Alvarez V, Alfonso-Sánchez MA. 1995. Biología Evolutiva. *La Habana, Cuba: Editorial Pueblo y Educación*.
- Bielejec F, Lemey P, Baele G, Rambaut A, Suchard MA. 2014. Inferring heterogeneous evolutionary processes through time: from sequence substitution to phylogeography. *Syst Biol* 63, 493-504.
- Biondi G, Battistuzzi G, Rickards O, Carli A, De Stefano GF, Santachiara-Benerecetti SA, Ranzani GN, Beretta M, Astolfi P, Santolamazza C. 1988. Migration pattern and genetic marker distribution of the Afro-American population of Bluefields, Nicaragua. *Ann Hum Biol* 15, 399-412.
- Biral AC, Magalhaes RF, Wastowski IJ, Simoes R, Donadi EA, Simoes AL, Mendes-Junior CT, Tanaka AM, Kraemer MH. 2006. Association of HLA-A, -B, -C Genes and TNF microsatellite polymorphism with psoriasis vulgaris: a study of Genetic risk in Brazilian patients. *Eur J Dermatol* 16, 523-529.
- Bodner M, Perego UA, Huber G, Fendt L, Röck AW, Zimmermann B, Olivieri A, Gómez-Carballa A, Lancioni H, Angerhofer N, Bobillo MC, Corach D, Woodward SR, Salas A, Achilli A, Torroni A, Bandelt HJ, Parson W. 2012. Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Res* 22, 811-820.
- Bolnick DA, Smith DG. 2003. Unexpected patterns of mitochondrial DNA variation among Native Americans from the southeastern United States. *Am J Phys Anthropol* 122, 336-354.
- Bonato SL, Salzano FM. 1997a. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci USA* 94, 1866-1871.

- Bonatto SL, Salzano FM. 1997b. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *Am J Hum Genet* 61, 1413-1423.
- Bonilla C, Parra EJ, Pfaff CL, Dios S, Marshall JA, Hamman RF, Ferrell RE, Hoggart CL, McKeigue PM, Shriver MD. 2004. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann Hum Genet* 68, 139-153.
- Bortolini MC, Da Silva WA Junior W, De Guerra DC, Remonato G, Mirandola R, Hutz MH, Weimer TA, Silva MC, Zago MA, Salzano FM. 1999. African-derived South American populations: A history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. *Am J Hum Biol* 11, 551-563.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73, 524-539.
- Bortolini MC, Silva-Junior WA, Weimer TA, Zago MA, Castro de Guerra D, Schneider MPC, Layrisse Z, Mendez Castellano H, Salzano FM. 1998. Protein and hypervariable tandem repeat diversity in eight African-derived South American populations: Inferred relationships do not coincide. *Hum Biol* 70, 443-461.
- Bortolini MC, Weimer TA, Franco MH, Salzano FM, Layrisse Z, Schneider H, Schneider MP, Harada ML. 1992. Genetic studies in three South American black populations. *Gene Geogr* 6, 1-16.
- Bortolini MC, Weimer TA, Salzano FM, Callegari-Jacques SM, Schneider H, Layrisse Z, Bonatto SL. 1995. Evolutionary relationships between black South American and African populations. *Hum Biol* 67, 547-559.
- Bortolini MC, Weimer TA, Salzano FM, Moura LB, Silva MCBO. 1997. Genetic structure of two urban afro-brazilian populations. *Int J Anthropol* 12: 5-16.
- Boster JS, Yost J, Peeke C. 2003. Rage, Revenge, and Religion: honest signaling of aggression and nonaggression in Waorani coalitional violence. *Ethos* 31, 471-494.
- Bottenus RE, Ichinose A, Davie EW. 1990. Nucleotide sequence of the *Gene* for the b subunit of human factor XIII. *Biochemistry* 29, 11195-11209.
- Boulanger LM, Shatz CJ. 2004. Immune signalling in neural development, synaptic plasticity and disease. *Nat Rev Neurosci* 5, 521-531.
- Bourgeois S, Yotova V, Wang S, Bourtoumieu S, Moreau C, Michalski R, Moisan JP, Hill K, Hurtado AM, Ruiz-Linares A, Labuda D. 2009. X-chromosome lineages and the settlement of the Americas. *Am J Phys Anthropol* 140, 417-428.
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368, 455-457
- Boyd-Bowman P. 1976. Patterns of Spanish emigration to the Indies until 1600. *Hispanic Am Hist Rev* 66, 580-604.
- Bradley B, Stanford D. 2004. The North Atlantic ice-edge corridor: a possible Palaeolithic route to the New World. *World Archaeology* 36, 459-478.
- Breslow JL, Ross D, McPherson J, Williams H, Kurnit D, Nussbaum AL, Karathanasis SK, Zannis VI. 1982. Isolation and characterization of cDNA clones for human apolipoprotein A-I. *Proc Natl Acad Sci USA* 79, 6861-6865.
- Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B. 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* 62, 1408-1415.
- Brinkmann B, Sajantila A, Goedde HW, Matsumoto H, Nishi K, Wiegand P. 1996. Population Genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci. *Eur J Hum Genet* 4, 175-182.
- Brinkmann B. 1996. The STR approach. En: Carracedo A, Brinkmann B, Bär W (Eds). *Advances in forensic haemagenetics Vol 6. Berlin, Heidelberg, New York: Springer.* pp 41-51.

- Britten RJ. 1994. Evidence that most human Alu sequences were inserted in a process that ceased about 30 million years ago. *Proc Natl Acad Sci USA* 91, 6148-6150.
- Britten RJ. 1997. Mobile elements inserted in the distant past have taken on important functions. *Gene* 205, 177-182.
- Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani F, Wallace DC. 1998. MtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *Am J Hum Genet* 63, 1852-1861.
- Brubaker LB, Anderson PM, Edwards ME, Lozhkin AV. 2005. Beringia as a Glacial Refugium for Boreal Trees and shrubs: New perspectives from mapped pollen data. *J Biogeogr* 32, 833-848.
- Burkholder MA, Johnson LL. 2003. Colonial Latin America. *Oxford, UK: Oxford University Press.*
- Cabodevilla MA. 1994. Los Huaorani en la historia de los pueblos del oriente. *Coca, España: CICAME.*
- Cabodevilla MA. 2008. Zona intangible: peligro de muerte! *Quito, Ecuador: CICAME.*
- Cagliani R, Fumagalli M, Riva S, Pozzoli U, Comi GP, Bresolin N, Sironi M. 2010. Genetic variability in the ACE Gene region surrounding the Alu I/D polymorphism is maintained by balancing selection in human populations. *Pharmacogenet Genomics* 20, 131-134.
- Callaway E. 2015. South America settled in one go. *Nature* 20, 598-599
- Capocasa M, Taglioli L, Anagnostou P, Paoli G, Danubio ME. 2014. Determinants of marital behaviour in five Apennine communities of Central Italy inferred by surname analysis, repeated pairs and kinship estimates. *Homo* 65, 64-74.
- Cardoso S, Alfonso-Sánchez MA, González-Andrade F, Valverde L, Odriozola A, Pérez-Miranda AM, Peña JA, Martínez-Jarreta B, de Pancorbo MM. 2008. Mitochondrial DNA in Huaorani (Ecuadorian amerindians): A new variant in haplogroup A2. *Forensic Sci Int Genet Suppl Series* 1, 269-270.
- Cardoso S, Alfonso-Sánchez MA, Valverde L, Sánchez D, Zarrabeitia MT, Odriozola A, Martínez-Jarreta B, de Pancorbo MM. 2012. Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. *Heredity* 108, 609-615.
- Cardoso S, Palencia-Madrid L, Valverde L, Alfonso-Sánchez MA, Gómez-Pérez L, Alfaro E, Bravi CM, Dipierri JE, Peña JA, de Pancorbo MM. 2013a. Mitochondrial DNA control region data reveal high prevalence of Native American lineages in Jujuy province, NW Argentina. *Forensic Sci Int Genet* 7, e52-55.
- Cardoso S, Valverde L, Alfonso-Sánchez MA, Palencia-Madrid L, Elcoroaristizabal X, Algorta J, Catarino S, Arteta D, Herrera RJ, Zarrabeitia MT, Peña JA, de Pancorbo MM. 2013b. The expanded mtDNA phylogeny of the Franco-Cantabrian region upholds the pre-neolithic genetic substrate of Basques. *PLoS One* 8, e67835.
- Carnese FR, Mendisco F, Keyser C, Dejean CB, Dugoujon JM, Bravi CM, Ludes B, Crubézy E. 2010. PaleoGenetical study of pre-Columbian samples from Pampa Grande (Salta, Argentina). *Am J Phys Anthropol* 141, 452-462.
- Carroll MC, Katzman P, Alicot EM, Koller BH, Geraghty DE, Orr HT, Strominger JL, Spies T. 1987. Linkage map of the human major histocompatibility complex including the tumor necrosis factor Genes. *Proc Natl Acad Sci USA* 84, 8535-8539.
- Carroll SB. 2001. The big picture. *Nature* 409, 669.
- Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, Ruiz-Linares A. 2003. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum Genet* 112, 534-541.
- Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, Ruiz-Linares A. 2000. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet* 67, 1287-1295.

- Castro de Guerra D, Arvelo H, Rodríguez-Larralde A, Salzano FM. 1996. Genetic study in Panaquire, a Venezuelan population. *Hum Her* 46, 323-328.
- Castro F, Acevedo E, Ciusani E, Angulo JA, Wollheim FA, Sandberg-Wollheim M. 2001. Tumour necrosis factor microsatellites and HLA-DRB1*, HLA-DQA1*, and HLA-DQB1* alleles in Peruvian patients with rheumatoid arthritis. *Ann Rheum Dis* 60, 791-795.
- Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human Genes. *Princeton*: Princeton University Press.
- Cerna M, Falco M, Friedman H, Raimondi E, Maccagno A, Fernandez-Viña M, Stastny P. 1993. Differences in HLA class II alleles of isolated South American Indian populations from Brazil and Argentina. *Hum Immunol* 37, 213-220.
- Chakrabarti CS, Roy M, Sengupta NK, Lalthantluanga R, Majumder PP. 2002. Genetic relationships among some tribal groups inhabiting the north-eastern, eastern and sub-Himalayan regions of India. *Ann Hum Genet* 66, 361-368.
- Chakraborty R. 1985. Gene identity in racial hybrids and estimation of admixture rates. En: Neel JV, Ahuja Y (Eds). Genetic microdifferentiation in man and other animals. *Delhi, India*: Indian Anthropological Association, Delhi University. pp 171-180.
- Charron D. 1997. Genetic diversity of HLA. Functional and medical implication. *París, Francia*: EDK international publisher
- Chatters JC, Kennett DJ, Asmerom Y, Kemp BM, Polyak V, Blank AN, Beddows PA, Reinhardt E, Arroyo-Cabrales J, Bolnick DA, Malhi RS, Culleton BJ, Erreguerena PL, Rissolo D, Morell-Hart S, Stafford TW Jr. 2014. Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans. *Science* 344, 750-754
- Choo SY. 2007. The HLA system: Genetics, immunology, clinical testing, and clinical implications. *Yonsei Med J* 48, 11-23.
- Cinnioğlu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114, 127-148.
- Claas FH, Duquesnoy RJ. 2008. The polymorphic alloimmune response in clinical transplantation. *Curr Opin Immunol* 20, 566-567.
- Clark PU, Mix AC. 2002. Ice Sheets and Sea Level of the Last Glacial Maximum. *Quaternary Science Reviews* 21, 1-7.
- Cohly HH, Panja A. 2005. Immunological findings in autism. *Int Rev Neurobiol* 71, 317-341.
- Colleoni P, Proaño J. 2010. Caminantes en la selva. Los Pueblos en Aislamiento de la Amazonia Ecuatoriana. Informe 7 IWGIA. Grupo Internacional de Trabajo sobre Asuntos Indígenas (IWGIA). Available: http://www.iwgia.org/publicaciones/buscar-publicaciones?publication_id = 275.
- Colmenares G. 1979. Historia Económica y social de Colombia. Popayán: una sociedad esclavista. 1680-1800. *Bogotá, Colombia*: La Carrera Inéditos Ltda.
- Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A. 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong Genetic boundary through the Gibraltar Straits. *Hum Genet* 107, 312-319.
- Cook SF, Borah W. 1971. Essays in population history: Mexico and the Caribbean. *Berkeley, California, USA*: University of California Press.
- Cordaux R, Lee J, Dinoso L, Batzer MA. 2006. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. *Gene* 373, 138-144. .
- Cotrim NH, Auricchio MT, Vicente JP, Otto PA, Mingroni-Netto RC. 2004. Polymorphic Alu insertions in six Brazilian African-derived populations. *Am J Hum Biol* 16, 264-277.

- Crawford MH, Gonzalez NL, Schanfield MS, Dykes DD, Skradski K, Polesky HF. 1981. The Black Caribs (Garifuna) of Livingston, Guatemala: genetic markers and admixture estimates. *Hum Biol* 53, 87-103.
- Crawford MH. 1998. The Origins of Native Americans. Anthropological Genetic Perspective. *Cambridge, UK: Cambridge University Press.*
- Creasy Rk, Resnik R, Iams J. 2004. Maternal-fetal medicine. *Philadelphia, Pennsylvania, USA: EDK Saunders.*
- Crosby AW. 2003. The Columbian exchange: biological and cultural consequences of 1492. *Westport, CT, USA: Greenwood Press.*
- Crosby AW. 2004. Ecological imperialism: the biological expansion of Europe. *New York, USA: Cambridge University Press.*
- Crouau-Roy B, Briant L, Bouissou C, Stravropoulos C, Pociot F, Cambon-Thomsen A, Clayton J. 1993. Tumor necrosis factor microsatellites in four European populations. *Hum Immunol* 38, 213-216.
- Cullheim S, Thams S. 2007. The microglial networks of the brain and their role in neuronal network plasticity after lesion. *Brain Res Rev* 55, 89-96.
- Curtin PD. 1969. The Atlantic Slave Trade: A Census. *Madison, Milwaukee, USA; University Wisconsin Press.*
- Da Silva WA Jr, Bortolini MC, Meyer D, Salzano FM, Elion J, Krishnamoorthy R, Schneider MP, De Guerra DC, Layrisse Z, Castellano HM, Weimer TD, Zago MA. 1999. Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. *Am J Phys Anthropol* 109, 425-437.
- DANE. Censo General. 2005. *Departamento Administrativo Nacional de Estadística.* 499 pp. (<http://www.dane.gov.co>).
- Danser AH, Schunkert H. 2000. Renin-angiotensin system Gene polymorphisms: potential mechanisms for their association with cardiovascular diseases. *Eur J Pharmacol* 410, 303-316.
- Dawkins RL, Zhang WJ, Degli-Esposti MA, Abraham L, McCann V, Christiansen FT. 1991. Genetics of diabetes: Studies of MHC haplotypes by pulsed field gel electrophoresis. *Baillieres Clin Endocrinol Metab* 5, 285-297.
- de Pancorbo MM, López-Martínez M, Martínez-Bouzas C, Castro A, Fernández-Fernández I, de Mayolo GA, de Mayolo AA, de Mayolo PA, Rowold DJ, Herrera RJ. 2001. The Basques according to polymorphic Alu insertions. *Hum Genet* 109, 224-233.
- Degen SJ, Rajput B, Reich E. 1986. The human tissue plasminogen activator gene. *J Biol Chem* 261, 6972-6985.
- DeGiorgio M, Degnan JH, Rosenberg NA. 2011. Coalescence-time distributions in a serial founder model of human evolutionary history. *Genetics* 189, 579-593.
- Degler CC. 1984. Out of our past: the forces that shaped modern America. *New York: Harpen Colophon Books.*
- Degler CN. 1986. Historia de Estados Unidos: La formación de una potencia (1600-1860). *Barcelona, España: Ariel.*
- Deininger PL, Batzer MA. 1999. Alu repeats and human disease. *Mol Genet Metab* 67, 183-193.
- Deininger PL, Moran JV, Batzer MA, Kazazian HH Jr. 2003. Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* 13, 651-658.
- Deininger PL, Sherry ST, Risch g, Donaldson C, Robichaux MB, Soodyall H, Jenkins T, Sheen F, Swergold G, Stoneking M, Batzer MA. 1999. Interspersed repeat insertion polymorphisms for studies of human molecular anthropology. En: Papiha SS, Deka R, Chakraborty R (Eds). *Genomic Diversity: Applications in Human Population Genetics.* New York, usa: Kluwer Academic/Plenum Publishers. pp 201-212.
- Dejean CB, Crouau-Roy B, Goicoechea AS, Avena SA, Carnese FR. 2004. Genetic variability in Amerindian populations of Northern Argentina. *Genet and Mol Biol* 27, 489-495.
- Del Castillo N. 1981. La llave de las Indias. *Bogotá, Colombia: El Tiempo.*

- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, Dorzhu C, Luzina F, Lee HK, Vanacek T, Villems R, Zakharov I. 2007. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 81, 1025-1041.
- Derenko MV, Grzybowski T, Malyarchuk BA, Czarny J, Miścicka-Sliwka D, Zakharov IA. 2001. The presence of mitochondrial haplogroup x in Altaians from South Siberia. *Am J Hum Genet* 69, 237-241.
- Dillehay TD, Ramírez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD. 2008. Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* 320, 784-786.
- Dillehay TD. 1997. *Monte Verde: A Late Pleistocene Settlement in Chile*. Washington DC, USA: Smithsonian Institution Press.
- Dipierri JE, Alfaro E, Martínez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO. 1998. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol* 70, 1001-1010.
- Dipierri JE, Alfaro E, Peña JA, Constans J, Dugoujon JM. 2000. GM, KM immunoglobulin allotypes and other serum Genetic markers (HP, GC, PI and TF) among South American populations living at different altitudes (Jujuy Province, Argentina): Admixture estimates. *Hum Biol* 72, 305-319.
- Dixon EJ. 2001. Human colonization of the Americas: Timing, technology and process. *Quaternary Science Reviews* 20, 277-299.
- Dobyns HF. 1993. Disease transfer at contact. *Annu Rev Anthropol* 22, 273-291.
- Duforet-Frebourg N, Blum MG. 2014. Nonstationary patterns of isolation-by-distance: inferring measures of local genetic differentiation with Bayesian kriging. *Evolution* 68, 1110-11123.
- Dunn DS, Choy MK, Phipps ME, Kulski JK. 2007. The distribution of major histocompatibility complex class I polymorphic Alu insertions and their associations with HLA alleles in a Chinese population from Malaysia. *Tissue Antigens* 70, 136-143.
- Dunn DS, Inoko H, Kulski JK. 2003. Dimorphic Alu element located between the TFIID and CDSN Genes within the major histocompatibility complex. *Electrophoresis* 24, 2740-2748.
- Dunn DS, Naruse T, Inoko H, Kulski JK. 2002. The association between HLA-A alleles and young Alu dimorphisms near the HLA-J, -H, and -F Genes in workshop cell lines and Japanese and Australian populations. *J Mol Evol* 55, 718-726.
- Dunn PP, Day S, Williams S, Bendukidze N. 2005. HLA-DQB1 sequencing-based typing using newly identified conserved nucleotide sequences in introns 1 and 2. *Tissue Antigens* 66, 99-106.
- Dyke AS. 2004. An outline of North American deglaciation with emphasis on central and northern Canada. En: Ehlers J, Gibbard PL (Eds). *Quaternary Glaciations—Extent and Chronology, Part II: North America*. Amsterdam, Netherlands: Elsevier.
- Economou EP, Bergen AW, Warren AC, Antonarakis SE. 1990. The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc Natl Acad Sci USA* 87, 2951-2954.
- Edwards A, Civitello A, Hammond HA, Caskey CT. 1991. DNA typing and Genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 49, 746-756.
- Edwards MC, Gibbs RA. 1992. A human dimorphism resulting from loss of an Alu. *Genomics* 14, 590-597.
- Efron B. 1979. Bootstrap Methods: Another look at the Jackknife. *Annals of Statistics* 7, 1-26.
- El Moncer W, Esteban E, Bahri R, Gayà-Vidal M, Carreras-Torres R, Athanasiadis G, Moral P, Chaabani H. 2010. Mixed origin of the current Tunisian population from the analysis of Alu and Alu/STR compound systems. *J Hum Genet* 55, 827-833.
- Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calò C, De Montis A, Atzori M, Marini M, Tofanelli S, Francalacci P, Pagani L, Tyler-Smith C, Xue Y, Cucca F, Schurr TG, Gaieski JB, Melendez C, Vilar MG, Owings AC, Gómez R, Fujita R, Santos FR, Comas D, Balanovsky O, Balanovska E, Zalloua P, Soodyall H, Pitchappan R, Ganeshprasad A, Hammer M, Matisoo-Smith L, Wells RS; Genographic Consortium. 2014.

- Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun* 29, 3513.
- Estrada-Mena B, Estrada FJ, Ulloa-Arvizu R, Guido M, Méndez R, Coral R, Canto T, Granados J, Rubí-Castellanos R, Rangel-Villalobos H, García-Carrancá A. 2010. Blood group O alleles in Native Americans: implications in the peopling of the Americas. *Am J Phys Anthropol* 142, 85-94.
- Ewens WJ. 1972. The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3, 87-112.
- Excoffier L, Laval G, Shneider S. 2005. Arlequin ver 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1, 47-50.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population Genetics analyses under Linux and Windows. *Mol Ecol Resour* 10, 564-567.
- Excoffier L, Slatkin M. 1998. Incorporating genotypes of relatives into a test of linkage disequilibrium. *Am J Hum Genet* 62, 171-180.
- Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479-491.
- Excoffier L. 2007. Analysis of population subdivision. En: Balding DJ, Bishop M, Cannings C (Eds). Handbook of statistical Genetics, 3ª edición. Chichester, UK: John Wiley & Sons, Ltd., pp 980-1020.
- Fagan BM. 2004. People of the earth. 11º Ed. New York, USA: Prentice-Hall.
- Fagundes NJ, Kanitz R, Bonatto SL. 2008a. A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS One* 3, e3157.
- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM, Smith DG, Silva WA Jr, Zago MA, Ribeiro-dos-Santos AK, Santos SE, Petzl-Erler ML, Bonatto SL. 2008b. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82, 583-592.
- Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci USA*. 104, 17614-17619.
- Fay JC, Wyckoff GJ, Wu CI. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* 415, 1024-1026.
- Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, Zhang K, Gumbs C, Castagna A, Cossarizza A, Cozzi-Lepri A, De Luca A, Easterbrook P, Francioli P, Mallal S, Martinez-Picado J, Miro JM, Obel N, Smith JP, Wyniger J, Descombes P, Antonarakis SE, Letvin NL, McMichael AJ, Haynes BF, Telenti A, Goldstein DB. 2007. A whole-genome association study of major determinants for host control of HIV-1. *Science* 317, 944-947.
- Felsenstein, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164-166.
- Fernández-Gómez AA, Velasco-Ortiz A. 2011. Los primeros humanos en America. *Estrat Critic* 5, 379-387.
- Fernández-Mestre M, Alfonso-Sánchez MA, Gendzekhadze K, Layrisse Z, Peña JA. 2006. Genetic polymorphisms at four STR loci from the HLA region in a Venezuelan population. *J Forensic Sci* 51, 703-704.
- Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD. 2008. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet* 4, e1000024.
- Ferree PM, Barbash DA. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol* 10, e1000234.
- Ferree PM, Gomez K, Rominger P, Howard D, Kornfeld H, Barbash DA. 2014. Heterochromatin position effects on circularized sex chromosomes cause filicidal embryonic lethality in *Drosophila melanogaster*. *Genetics* 4, 1001-1005.
- Finer M, Jenkins CN, Pimm SL, Keane B, Ross C. 2008. Oil and gas projects in the Western Amazon: threats to wilderness, biodiversity, and indigenous peoples. *PLoS One* 3, e2932.

- Fischer A, Wiebe V, Paabo S, Przeworski M. 2004. Evidence for a complex demographic history of chimpanzees. *Mol Biol Evol* 5, 799–808.
- Fisher, RA. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Ann Eugen* 7, 179–188.
- Foissac A, Crouau-Roy B, Fauré S, Thomsen M, Cambon-Thomsen A. 1997. Microsatellites in the HLA region: on overview. *Tissue Antigens* 49, 197-214.
- Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59, 935-945.
- Friedemann NS. 1993. La saga del negro: presencia africana en Colombia. *Bogotá, Colombia*: Pontificia Universidad Javeriana.
- Fuentes BC. 1997. Huaomoni, huarani, cowudi: una aproximación a los Huaorani en la práctica política multi-étnica ecuatoriana. Series Pueblos de Ecuador. *Quito, Ecuador*: CEDIME/Abya Yala.
- Gamboa-Balbin CL. 2007. Régimen especial transectorial de protección a favor de pueblos indígenas en aislamiento y en contacto inicial. *Lima, Peru*: Bellido Ediciones EIRL.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, de Pancorbo MM, Peña JA. 2007. Polymorphic Alu insertions and the genetic structure of Iberian Basques. *J Hum Genet* 52: 317-327.
- García-Obregón S, Alfonso-Sánchez MA, Gómez-Pérez L, Pérez-Miranda AM, Arroyo D, de Pancorbo MM, Peña JA. 2011. Microsatellites and Alu elements from the human MHC in Valencia (Spain): analysis of Genetic relationships and linkage disequilibrium. *Int J InmunoGenet* 38, 483–491.
- García-Obregón S, Alfonso-Sánchez MA, Pérez-Miranda AM, Vidales C, Arroyo D, Peña JA. 2006. Genetic position of Valencia (Spain) in the Mediterranean basin according to Alu insertions. *Am J Hum Biol* 18, 187-195.
- Garza JC, Williamson EG. 2001. Detection of reduction in population size using data from microsatellite loci. *Mol Ecol* 10, 305-318.
- Gaudieri S, Leelayuwat C, Townend DC, Kulski JK, Dawkins RL. 1997. Genomic characterization of the region between HLA-B and TNF: implications for the evolution of multicopy *Gene* families. *J Mol Evol* 44, Suppl 1, 147-154.
- Gayà-Vidal M, Dugoujon JM, Esteban E, Athanasiadis G, Rodríguez A, Villena M, Vasquez R, Moral P. 2010. Autosomal and X chromosome Alu insertions in Bolivian Aymaras and Quechuas: two languages and one genetic pool. *Am J Hum Biol* 22, 154-162.
- Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. 2010. The NCBI BioSystems database. *Nucleic Acids Res* 38, D492-D496.
- Gil R. 2000-2001. Género y etnicidad en la "Puna de Jujuy". *Población & Sociedad* 8/9, 41-70.
- Gilbert MT, Jenkins DL, Götherstrom A, Naveran N, Sanchez JJ, Hofreiter M, Thomsen PF, Binladen J, Higham TF, Yohe RM 2nd, Parr R, Cummings LS, Willerslev E. 2008a. DNA from pre-Clovis human coprolites in Oregon, North America. *Science* 320, 786-789.
- Gilbert MT, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Götherström A, Campos PF, Rasmussen M, Metspalu M, Higham TF, Schwenninger JL, Nathan R, De Hoog CJ, Koch A, Møller LN, Andreasen C, Meldgaard M, Villems R, Bendixen C, Willerslev E. 2008b. Paleo- Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* 320, 1787–1789.
- Gill P, Kimpton C, D'Aloja E, Andersen JF, Bar W, Brinkmann B, Holgersson S, Johnsson V, Kloosterman AD, Lareui MV, Nellemann L, Pfitzinger H, Phillips CP, Schmitter H, Schneidern PM, Stenersen M. 1994. Report of the European DNA profiling group (EDNAP)-towards standardisation of short tandem repeat (STR) loci. *Forensic Sci Int* 65, 51-59.
- Goddard CA, Butts DA, Shatz CJ. 2007. Regulation of CNS synapses by neuronal MHC class I. *Proc Natl Acad Sci USA* 104, 6828-6833.
- Goebel T, Waters MR, Dikova M. 2003. The archaeology of Ushki Lake, Kamchatka, and the Pleistocene peopling of the Americas. *Science* 301, 501-505.

- Goebel T, Waters MR, O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. *Science* 319, 1497–1502.
- Goldstein D, Schlötterer C. 1999. Microsatellites: Evolution and applications. *Oxford, UK: Oxford University Press*.
- Gomes MV. 2006a. El escamoteo de la entidad y de la identidad afro-negra. *Cuadernos del Instituto de Pensamiento Latinoamericano* 1, 159-174.
- Gomes MV. 2006b. La presencia negroafricana en la Argentina: pasado y permanencia. *Boletín digital de la Biblioteca del Congreso*, Nº 9.
- Gontijo CC, Guerra Amorim CE, Godinho NM, Toledo RC, Nunes A, Silva W, Da Fonseca Moura MM, De Oliveira JC, Pagotto RC, De Nazaré Klautau-Guimarães M, De Oliveira SF. 2014. Brazilian quilombos: A repository of Amerindian alleles. *Am J Hum Biol* 26, 142-150.
- González-Andrade F, Roewer L, Willuweit S, Sánchez D, Martínez-Jarreta B. 2009. Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis. *Forensic Sci Int Genet* 3, e83–91.
- González-José R, Bortolini MC, Santos FR, Bonatto SL. 2008. The peopling of America: Craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am J Phys Anthropol* 137, 175–187.
- González-Pérez E, Esteban E, Via M, Gayà-Vidal M, Athanasiadis G, Dugoujon JM, Luna F, Mesa MS, Fuster V, Kandil M, Harich N, Bissar-Tadmouri N, Saetta A, Moral P. 2010. Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems). *Am J Phys Anthropol* 141, 430-439.
- González-Pérez E, Moral P, Via M, Vona G, Varesi L, Santamaria J, Gaya-Vidal M, Esteban E. 2007. The ins and outs of population relationships in west-Mediterranean islands: data from autosomal Alu polymorphisms and Alu/STR compound systems. *J Hum Genet* 52, 999-1010.
- Goudet J, Raymond M, de-Meeus T, Rousset F. 1996. Testing differentiation in diploid populations. *Genetics* 144, 1933–1940.
- Gourraud PA, Cambon-Thomsen A, Dauber EM, Feolo M, Hansen J, Mickelson E, Single RM, Thomsen M, Mayr WR. 2007. Nomenclature for HLA microsatellites. *Tissue Antigens* 69, Suppl 1, 210-213. Review.
- Graham RR, Ortmann W, Rodine P, Espe K, Langefeld C, Lange E, Williams A, Beck S, Kyogoku C, Moser K, Gaffney P, Gregersen PK, Criswell LA, Harley JB, Behrens TW. 2007. Specific combinations of HLA-DR2 and DR3 class II haplotypes contribute graded risk for disease susceptibility and autoantibodies in human SLE. *Eur J Hum Genet* 15, 823-830.
- Green LD, Derr JN, Knight A. 2000. mtDNA affinities of the peoples of North-Central Mexico. *Am J Hum Genet* 66, 989-998.
- Greenberg JH, Turner CG, Zegura SL. 1986. The settlement of the Americas: a comparison of the linguistic, dental and Genetic evidence. *Current Anthropol* 27, 477-495.
- Grover D, Kannan K, Brahmachari SK, Mukerji M. 2005. ALU-ring elements in the primate genomes. *Genetica* 124, 273-289.
- Grubić Z, Stingl K, Zunec R, Car H, Cecuk-Jelčić E, Brkljčić-Kerhin V. 2007. Linkage disequilibria between human leucocyte antigen-B and closely linked microsatellites in the Croatian population. *Tissue Antigens* 69, 86-94
- Grundmann U, Nerlich C, Rein T, Zettlmeissl G. 1990. Complete cDNA sequence encoding the B subunit of human factor XIII. *Nucleic Acids Res* 18, 2817-2818.
- Guo SW, Thompson EA. 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48, 361–372.
- Guthrie RD. 1990. Frozen Fauna of the Mammoth Steppe: The story of Blue Babe. *Chicago y London: The University of Chicago Press*.

- Hajeer AH, Hutchinson IV. 2000. TNF-alpha *Gene* polymorphism: clinical and biological implications. *Microsc Res Tech* 50, 216-228.
- Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Annu Rev Ecol Evol System* 40, 151-172.
- Halverson MS, Bolnick DA. 2008. An ancient DNA test of a founder effect in Native American ABO blood group frequencies. *Am J Phys Anthropol* 137, 342-347.
- Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4, 9. http://palaeo-electronica.org/2001_1/past/issue1_01.htm
- Hancock JM. 1999. Microsatellites and other simple sequences: Genomic context and mutational mechanisms. En: Goldstein DB, Schlötterer C (Eds). *Microsatellites: Evolution and Applications*. New York: Oxford University Press. pp 1-9.
- Harpending H, Rogers AR. 2000. Genetic perspectives on human origins and differentiation. *Annu Rev Genom Hum Genet* 1, 361-385.
- Harpending H, Ward R. 1982. Chemical systematic and human evolution. En: Nitecki M (Ed). *Biochemical of Evolutionary Biology*. Chicago, USA: University of Chicago Press. pp 213-256.
- Harpending HC, Jenkins T. 1973. Genetic distance among southern African populations. En: Crawford M, Workman P (Eds). *Method and Theory in Anthropological Genetics*. Albuquerque, USA: University of New Mexico Press. pp 177-199.
- Hartl DL, Clark AG. 1997. Principles of population Genetics, Third edition. Massachusetts, USA: Sinauer Associates, Inc.
- Häsler J, Strub K. 2006. Alu elements as regulators of *Gene* expression. *Nucleic Acids Res* 34, 5491-5497.
- Hawks J, Hunley K, Lee S, Wolpoff M. 2000. Population Bottlenecks and Pleistocene Human Evolution. *Mol Biol Evol* 17, 2-22.
- Haynes G. 2002. The early settlement of North America: The Clovis Era. Cambridge, UK: Cambridge University Press.
- Hearne CM, Ghosh S, Todd JA. 1992. Microsatellites for linkage analysis of Genetic traits. *Trends Genet* 8, 288-294. Review.
- Helgason A, Pálsson G, Pedersen HS, Angulalik E, Gunnarsdóttir ED, Yngvado' ttir B, Stefánsson K. 2006. mtDNA variation in Inuit populations of Greenland and Canada: Migration history and population structure. *Am J Phys Anthropol* 130, 123-134.
- Hendry AP, Kinnison MT. 2001. An introduction to microevolution: rate, pattern, process. *Genetica* 112-113, 1-8.
- Herrera RJ, Rojas DP, Terreros MC. 2007. Polymorphic Alu insertions among Mayan. *J Hum Genet* 52, 129-142.
- Hingorani AD, Jia H, Stevens PA, Hopper R, Dickerson JE, Brown MJ. 1995. Renin-angiotensin system *Gene* polymorphisms influence blood pressure and the response to angiotensin converting enzyme inhibition. *J Hypertens* 13, 1602-1609.
- Ho SY, Endicott P. 2008. The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am J Hum Genet* 83, 142-146.
- Hoffecker JF, Elias SA, O'Rourke DH. 2014. Out of Beringia? *Science* 343, 979-980.
- Hoffecker JF, Powers WR, Goebel T. 1993. The colonization of beringia and the peopling of the new world. *Science* 259, 46-53.
- Holden C. 1999. Were Spaniards among the first Americans? *Science* 286, 1467-1468.
- Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K. 1993. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol Biol Evol* 10, 23-47.
- Houck CM, Rinehart FP, Schmid CW. 1979. A ubiquitous family of repeated DNA sequences in the human genome. *J Mol Biol* 132, 289-306.

- Hrdlicka, A. 1923. The origin and antiquity of the American Indian. En: Annual Report of the Board of Regents of the Smithsonian Institution. *Washington DC, USA*: Government Printing Office. pp 481–493.
- Hubert C, Houot AM, Corvol P, Soubrier F. 1991. Structure of the angiotensin I-converting enzyme *gene*. Two alternate promoters correspond to evolutionary steps of a duplicated *Gene*. *J Biol Chem* 266, 15377-15383.
- Huh GS, Boulanger LM, Du H, Riquelme PA, Brotz TM, Shatz CJ. 2000. Functional requirement for class I MHC in CNS development and plasticity. *Science* 290, 2155-2159.
- Ibarra A, Freire-Aradas A, Martínez M, Fondevila M, Burgos G, Camacho M, Ostos H, Suarez Z, Carracedo A, Santos S, Gusmão L. 2014. Comparison of the genetic background of different Colombian populations using the SNPforID 52plex identification panel. *Int J Legal Med* 128, 19-25.
- INDEC. 2010. Censo Nacional de Población y Vivienda 1991 y Censo Nacional de Población, Hogares y Viviendas. www.indec.gov.ar/webcenso/index.asp.
- Jazwinska EC, Pyper WR, Burt MJ, Francis JL, Goldwurm S, Webb SI, Lee SC, Halliday JW, Powell LW. 1995. Haplotype analysis in Australian hemochromatosis patients: evidence for a predominant ancestral haplotype exclusively associated with hemochromatosis. *Am J Hum Genet* 56, 428-433.
- Jenkins DL. 2007. Distribution and Dating of Cultural and Paleontological Remains in the Paisley Five Mile Point Caves in the Northern Great Basin: An Early Assessment. En: Graft KE, Schmitt DN (Eds). Paleoindian or Paleoarchaic? Great basin human ecology at the Pleistocene/Holocene transition. *Salt Lake City, Utah, USA*: University of Utah Press. pp 57-81.
- Jones MA. 1996. Historia de Estados Unidos. 1607-1992. *Madrid, España*: Cátedra.
- Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sunq S, Kere J, Harpending HC. 1997. Microsatellite diversity and the demographic history of modern humans. *Proc Natl Acad Sci USA* 94, 3100-3103.
- Jorde LB, Workman PL, Erikson AW. 1982. Genetic microevolution in the Åland Islands, Finland. En: Crawford MD, Mielke JH (Eds). Current developments in Anthropological Genetics, Vol. 2. Ecology and Populations Structure. *New York, USA*: Plenum Press. pp 333-366.
- Jorde LB. 1980. The Genetic Structure of Subdivided Human Populations, a review. En: Mielke J, Crawford MD (Eds). Current developments in Anthropological Genetics. Vol 1. Theory and Methods. *New York, USA*: Plenum Press. pp 135-208.
- Joyce DJ. 2006. Chronology and new research on the Schaefer mammoth (*mammuthus primigenius*) site, Kenosha County. *Wisconsin, USA. Quaternary International* 142-143, 44-57.
- Jurka J, Zuckerkandl E. 1991. Free left arms as precursor molecules in the evolution of Alu sequences. *J Mol Evol* 33, 49-56.
- Kapitonov V, Jurka J. 1996. The age of Alu subfamilies. *J Mol Evol* 42, 59-65.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18, 830–838.
- Karafet TM, Zegura SL, Hammer MF. 2006. Y chromosomes. En: Ubelaker DH (Ed). Handbook of North American Indians, vol. 3. Environment, Origins and Population. *Washington DC, USA*: Smithsonian Institution Press. pp 831–839.
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF. 1999. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64, 817-831.
- Karathanasis SK. 1985. Apolipoprotein multiGene family: tandem organization of human apolipoprotein AI, CIII, and AIV Genes. *Proc Natl Acad Sci USA* 82, 6374-6378.

- Kass DH, Aleman C, Batzer MA, Deininger PL. 1994. Identification of a human specific Alu insertion in the factor XIII^B gene. *Genetica* 94, 1-8.
- Kaufman T, Golla V. 2000. Language groupings in the new world: their reliability and usability in cross-disciplinary studies. En: Renfrew C (ed). *America past, America present: Genes and languages in the Americas and beyond*. Cambridge, UK: The McDonald Institute for Archaeological Research. pp 47-57.
- Kehoe PG, Russ C, McIlroy S, Williams H, Holmans P, Holmes C, Liolitsa D, Vahidassr D, Powell J, McGleenon B, Liddell M, Plomin R, Dynan K, Williams N, Neal J, Cairns NJ, Wilcock G, Passmore P, Lovestone S, Williams J, Owen MJ. 1999. Variation in DCP1, encoding ACE, is associated with susceptibility to Alzheimer disease. *Nat Genet* 21, 71-72.
- Kemp BM, Malhi RS, McDonough J, Bolnick DA, Eshleman JA, Rickards O, Martinez-Labarga C, Johnson JR, Lorenz JG, Dixon EJ, Fifield TE, Heaton TH, Worl R, Smith DG. 2007. Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas. *Am J Phys Anthropol*. 132, 605-621.
- Kitchen A, Miyamoto MM, Mulligan CJ. 2008. A three-stage colonization model for the peopling of the Americas. *PLoS One* 3, e1596.
- Kluft C. 1994. Constitutive synthesis of tissue type plasminogen activator and plasminogen activator inhibitor type 1: conditions and therapeutic targets. *Fibrinolysis* 8, 1-7.
- Knight A, Batzer MA, Stoneking M, Tiwari HK, Scheer WD, Herrera RJ, Deininger PL. 1996. DNA sequences of Alu elements indicate a recent replacement of the human autosomal Genetic complement. *Proc Natl Acad Sci USA* 93, 4360-4364.
- Kolman CJ, Sambuughin N, Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142, 1321-1334.
- Konkel MK, Batzer MA. 2010. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Semin Cancer Biol* 20, 211-221.
- Kuehnen P, Krude H. 2012. Alu elements and human common diseases like obesity. *Mob Genet Elements* 2, 197-201.
- Kulski JK, Dunn DS, Hui J, Martinez P, Romphruk AV, Leelayuwat C, Tay GK, Oka A, Inoko H. 2002. Alu polymorphism within the MICB Gene and association with HLA-B alleles. *Immunogenetics* 53, 975-979. Erratum in: *Immunogenetics* 54, 365.
- Kulski JK, Dunn DS. 2005. Polymorphic Alu insertions within the Major Histocompatibility Complex class I genomic region: a brief review. *Cytogenet Genome Res* 110, 193-202.
- Kulski JK, Gaudieri S, Bellgard M, Balmer L, Giles K, Inoko H, Dawkins RL. 1997. The evolution of MHC diversity by segmental duplication and transposition of retroelements. *J Mol Evol* 45, 599-609. Erratum in: *J Mol Evol* 46, 734.
- Kulski JK, Gaudieri S, Dawkins RL. 2000. Using Alu J elements as molecular clocks to trace the evolutionary relationships between duplicated HLA class I genomic segments. *J Mol Evol* 50, 510-519.
- Kulski JK, Gaudieri S, Martin A, Dawkins RL. 1999. Coevolution of PERB11 (MIC) and HLA class I Genes with HERV-16 and retroelements by extended genomic duplication. *J Mol Evol* 49, 84-97.
- Kulski JK, Martinez P, Longman-Jacobsen N, Wang W, Williamson J, Dawkins RL, Shiina T, Naruse T, Inoko H. 2001. The association between HLA-A alleles and an Alu dimorphism near HLA-G. *J Mol Evol* 53, 114-123.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla

- AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglu S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Lareu MV, Barral S, Salas A, Pestoni C, Carracedo A. 1998. Sequence variation of a hypervariable short tandem repeat at the D1S1656 locus. *Int J Legal Med* 111, 244–247.
- Lareu MV, Pestoni C, Schürenkamp M, Rand S, Brinkmann B, Carracedo A. 1996a. A highly variable STR at the D12S391 locus. *Int J Legal Med* 109, 134–138.
- Lareu MV, Pestoni MC, Barros F, Salas A, Carracedo A. 1996b. Sequence variation of a hypervariable short tandem repeat at the D12S391 locus. *Gene* 182, 151–153.
- Larrick JW, Yost J, Gourley C, Buckley CE 3rd, Plato CC, Pandey JP, Burck KB, Kaplan J. 1985. Markers of genetic variation among the Waorani Indians of the Ecuadorian Amazon headwaters. *Am J Phys Anthropol* 66, 445–453.
- Lee HS, Li W, Lee A, Rodine P, Graham RR, Ortmann WA, Batliwalla F, Lee KW, Bae SC, Behrens TW, Gregersen PK. 2006. Microsatellite typing for DRB1 alleles: application to the analysis of HLA associations with rheumatoid arthritis. *Genes Immun* 7, 533–543.
- Lee J, Kim YJ, Mun S, Kim HS, Han K. 2015. Identification of human-specific AluS elements through comparative genomics. *Gene* 555, 208–216.
- León-S FE, Ariza-Deleon A, Leon-S ME, Ariza-C A. 1996. Peopling the Americas. *Science* 273, 723–725.
- Levinson G, Gutman GA. 1987. Slipped-strand mispairing: major mechanism for DNA sequence evolution. *Mol Biol Evol* 4, 203–221.
- Lewis, MP, Simons GF, Fennig CD. 2014. Ethnologue: Languages of the World, 17th edition. Dallas, Texas, USA: SIL International.
- Li C, Jia J, Wei Y, Wan L, Hu L, Ye J. 2014. The selection of 30 ancestry informative markers and its application in ancestry inference. *Yi Chuan* 36, 779–785.
- Li WH, Sadler LA. 1991. Low nucleotide diversity in man. *Genetics* 129, 513–523.
- Litt M, Luty JA. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44, 397–401.
- Livi-Bacci. 2006. The Depopulation of Hispanic America after the Conquest. *Popul dev rev* 32, 199–232.
- Locksley RM, Killeen N, Lenardo MJ. 2001. The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* 104, 487–501.

- Long JC, Williams RC, McAuley JE, Medis R, Partel R, Tregellas WM, South SF, Rea AE, McCormick SB, Iwaniec U. 1991. Genetic variation in Arizona Mexican Americans: Estimation and interpretation of admixture proportions. *Am J Phys Anthropol* 84, 141–157.
- Lovejoy PE. 1982. The volumen of the Atlantic slave trade: A synthesis. *Journal of African History* 22, 473–501.
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605.
- Luizon MR, Mendes-Junior CT, De Oliveira SF, Simões AL. 2008. Ancestry informative markers in Amerindians from Brazilian Amazon. *Am J Hum Biol* 20, 86–90.
- Majumder PP, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N, Roy M, Thakurta PG, Sil SK. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur J Hum Genet* 7, 435–446.
- Malhi RS, Cybulski JS, Tito RY, Johnson J, Harry H, Dan C. 2010. Brief communication: mitochondrial haplotype C4c confirmed as a founding genome in the Americas. *Am J Phys Anthropol* 141, 494–497.
- Malhi RS, Eshleman JA, Greenberg JA, Weiss DA, Schultz Shook BA, Kaestle FA, Lorenz JG, Kemp BM, Johnson JR, Smith DG. 2002. The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America. *Am J Hum Genet* 70, 905–919.
- Malhi RS, Gonzalez-Oliver A, Schroeder KB, Kemp BM, Greenberg JA, Dobrowski SZ, Smith DG, Resendez A, Karafet T, Hammer M, Zegura S, Brovko T. 2008. Distribution of Y chromosomes among native North Americans: a study of Athapaskan population history. *Am J Phys Anthropol* 137, 412–424.
- Malhi RS, Mortensen HM, Eshleman JA, Kemp BM, Lorenz JG, Kaestle FA, Johnson JR, Gorodezky C, Smith DG. 2003. Native American mtDNA prehistory in the American Southwest. *Am J Phys Anthropol* 120, 108–124.
- Mandryk CAS. 2004. Invented traditions and the ultimate origin myth: in the beginning... there was an ice free-corridor. En: Barton CM, Clark GA, Yesner DR, Pearson GA (eds). *The settlement of the American continents. Tucson, USA: University of Arizona Press.* pp 113–122.
- Manel S, Holderegger R. 2013. Ten years of landscape genetics. *Trends Ecol Evol* 28, 614–621.
- Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol* 18, 189–197.
- Mann CC. 2006. 1491. Una nueva historia de las Américas antes de Colón. *Madrid, España: Taurus.*
- Manta FS, Pereira R, Caiafa A, Silva DA, Gusmão L, Carvalho EF. 2013. Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. *Ann Hum Biol* 40, 94–98.
- Marcheco-Teruel B, Parra EJ, Fuentes-Smith E, Salas A, Buttenschøn HN, Demontis D, Torres-Español M, Marín-Padrón LC, Gómez-Cabezas EJ, Alvarez-Iglesias V, Mosquera-Miguel A, Martínez-Fuentes A, Carracedo A, Børglum AD, Mors O. 2014. Cuba: exploring the history of admixture and the genetic basis of pigmentation using autosomal and uniparental markers. *PLoS Genet* 10, e1004488.
- Mardini AC, Rodenbusch R, Schumacher S, Chula FG, Michelon CT, Gastaldo AZ, Maciel LP, de Matos Almeida SE, da Silva CM. 2013. Mutation rate estimates for 13 STR loci in a large population from Rio Grande do Sul, Southern Brazil. *Int J Legal Med* 127, 45–47.
- Martin MP, Carrington M. 2005. Immunogenetics of viral infections. *Curr Opin Immunol* 17, 510–516.
- Martin PS. 1973. The discovery of America: The First Americans may have swept the Western Hemisphere and decimated its fauna within 1000 years. *Science* 179, 969–974.
- Martin TR. 1992. Ancient Greece: From Prehistoric to Hellenistic Times. *Princeton, USA: Yale University Press.*
- Martinez-Labarga C, Rickards O, Scacchi R, Corbo RM, Biondi G, Pena JA, De Vieira CV, Guevara AEB, Santurino MSM, De Stefano GF. 1999. Genetic population structure of two African- Ecuadorian communities of Esmeraldas. *Am J Phys Anthropol* 109, 159–174.
- Martínez-Sarasola C. 1992. Nuestros paisanos los indios. *Buenos Aires, Argentina: EMECE.*

- Matera AG, Hellmann U, Hintz MF, Schmid CW. 1990. Recently transposed Alu repeats result from multiple source genes. *Nucleic Acids Res* 18, 6019-6023.
- Mathias S, Dressler KA, Kolesnick RN. 1991. Characterization of a ceramide-activated protein kinase: stimulation by tumor necrosis factor alpha. *Proc Natl Acad Sci USA* 88, 10009-10013.
- Matsuo R, Asada A, Fujitani K, Inokuchi K. 2001. LIRF, a gene induced during hippocampal long-term potentiation as an immediate-early gene, encodes a novel RING finger protein. *Biochem Biophys Res Commun* 289, 479-484.
- Mattei MG, Hubert C, Alhenc-Gelas E, Roeckel N, Corvol P, Soubrier F. 1989. Angiotensin converting enzyme gene is on chromosome 17. *Cytogenet Cell Genet* 51, 1041.
- McAulay KA, Higgins CD, Macsween KF, Lake A, Jarrett RF, Robertson FL, Williams H, Crawford DH. 2007. HLA class I polymorphisms are associated with development of infectious mononucleosis upon primary EBV infection. *J Clin Invest* 117, 3042-3048.
- McCue AD, Nuthikattu S, Slotkin RK. 2013. Genome-wide identification of Genes regulated in trans by transposable element small interfering RNAs. *RNA Biol* 10, 1379-1395.
- McElroy JP, Oksenberg JR. 2008. Multiple sclerosis Genetics. *Curr Top Microbiol Immunol* 318, 45-72.
- McManus R, Wilson AG, Mansfield J, Weir DG, Duff GW, Kelleher D. 1996. TNF2, a polymorphism of the tumour necrosis-alpha Gene promoter, is a component of the celiac disease major histocompatibility complex haplotype. *Eur J Immunol* 26, 2113-2118.
- McRae B, Beier P. 2007. Circuit theory predicts gene flow in plant and animal populations. *Proc Natl Acad Sci USA* 104, 19885-19890.
- Mello E, Souza EM. 2006. Reis Negros no Brasil escravista: historia da festa de Coroacao de Rei Congo, 1 edn. *Belo Horizonte, Brazil*: Editora UFMG
- Mendisco F, Keyser C, Seldes V, Rivolta C, Mercolli P, Cruz P, Nielsen AE, Crubezy E, Ludes B. 2014. Genetic diversity of a late prehispanic group of the Quebrada de Humahuaca, northwestern Argentina. *Ann Hum Genet* 78, 367-380.
- Merriwether DA, Ferrell RE. 1996a. The four founding lineage hypothesis for the New World: a critical reevaluation. *Mol PhyloGenet Evol* 5, 241-246.
- Merriwether DA, Hall WW, Vahlne A, Ferrell RE. 1996b. mtDNA variation indicates Mongolia may have been the source for the founding population for the New World. *Am J Hum Genet* 59, 204-212.
- Merriwether DA, Huston S, Iyengar S, Hamman R, Norris JM, Shetterly SM, Kamboh MI, Ferrell RE. 1997. Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. *Am J Phys Anthropol* 102, 153-159.
- Merriwether DA, Rothhammer F, Ferrell RE. 1995. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am J Phys Anthropol* 98, 411-430.
- Merriwether DA. 2006. Mitochondrial DNA. En: Ubelaker DH (Ed). Handbook of North American Indians, vol. 3. Environment, Origins and Population. *Washington DC, USA*: Smithsonian Institution Press. pp 817-830.
- Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortiz-Barrientos D, García LF, Velez ID, Bravo ML, Múnera JG, Bedoya G, Bortolini MC, Ruiz-Linares A. 2000. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of Gene flow in South America. *Am J Hum Genet* 67, 1277-1286.
- Meyer E, Wiegand P, Brinkmann B. 1995. Phenotype differences of STRs in 7 human populations. *Int J Legal Med* 107, 314-322.
- Mighell AJ, Markham AF, Robinson PA. 1997. Alu sequences. *FEBS Lett* 417, 1-5.
- Mills RE, Bennett EA, Iskow RC, Luttig CT, Tsui C, Pittard WS, Devine SE. 2006. Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet* 78, 671 - 679.

- Mix AC, Bard E, Schneider R. 2001. Environmental processes of the ice age: land, oceans, glaciers (EPILOG). *Quaternary Science Reviews* 20, 627 - 657
- Mnuková-Fajdelová M, Satta Y, O'hUigin C, Mayer WE, Figueroa F, Klein J. 1994. Alu elements of the primate major histocompatibility complex. *Mamm Genome* 5, 405-415.
- Montgomery HE, Kiernan LA, Whitworth CE, Fleming S, Unger T, Gohlke P, Mullins JJ, McEwan JR. 1998. Inhibition of tissue angiotensin converting enzyme activity prevents malignant hypertension in TGR (mREN2)27. *J Hypertens* 16, 635-643.
- Morales ME, White TB, Strevva VA, DeFreece CB, Hedges DJ, Deininger PL. 2015. The contribution of alu elements to mutagenic DNA double-strand break repair. *PLoS Genet* 11, e1005016.
- Mulligan CJ, Hunley K, Cole S, Long JC. 2004. Population Genetics, history, and health patterns in native americans. *Annu Rev Genomics Hum Genet* 5, 295-315.
- Mulligan CJ, Kitchen A, Miyamoto MM. 2008. Updated three-stage model for the peopling of the Americas. *PLoS One* 3, e3199.
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51, 263-273
- Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, Stoneking M. 2001. Alu insertion polymorphisms and the Genetic structure of human populations from the Caucasus. *Eur J Hum Genet* 9, 267-272.
- Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. *Oxford, UK: University Press, Oxford*
- Nei M, Li WH. 1979. Mathematical model for studying Genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76, 5269-5273.
- Nei M. 1987. Molecular Evolutionary Genetics. *New York, USA: Columbia University Press.*
- Novick GE, Batzer MA, Deininger PL, Herrera RJ. 1996. *The mobile genetic element Alu in the human genome.* *BioScience* 46, 32-41.
- Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ. 1998. Polymorphic Alu insertions and the Asian origin of Native American populations. *Hum Biol* 70, 23-39.
- O'Fallon BD, Fehren-Schmitz L. 2011. Native Americans experienced a strong population bottleneck coincident with European contact. *Proc Natl Acad Sci USA* 108, 20444-20448.
- Ohtsuka M, Inoko H, Kulski JK, Yoshimura S. 2008. Major histocompatibility complex (MHC) class Ib gene duplications, organization and expression patterns in mouse strain C57BL/6. *BMC Genomics* 9, 178.
- O'Keefe GM, Nguyen VT, Benveniste EN. 2002. Regulation and function of class II major histocompatibility complex, CD40, and B7 expression in macrophages and microglia: Implications in neurological diseases. *J Neurovirol* 8, 496-512.
- Olson KW. 1980. Reseña de la historia de los Estados Unidos. *Washington, USA: Servicio Informativo y Cultural de los Estados Unidos de América.*
- O'Rourke DH, Raff JA. 2010. The human Genetic history of the Americas: the final frontier. *Curr Biol* 20, R202-207.
- Overstreet DF. 2005. Late-Glacial Ice-Marginal Adaptation in Southeastern Wisconsin. En: Bonnicksen R, Lepper BT, Stanford D, Waters MR (Eds). *Paleoamerican Origins: Beyond Clovis. College station TX: Center for the study of the first Americans, department of Anthropology, Texas, USA: A&M University Press.* pp 183-195.
- Palacio LG, Rivera D, Builes JJ, Jiménez ME, Salgar M, Anaya JM, Jiménez I, Camargo M, Arcos-Burgos M, Sánchez JL. 2002. Multiple sclerosis in the tropics: Genetic association to STR's loci spanning the HLA and TNF. *Mult Scler* 8, 249-255.
- Pappalardo SE, De Marchi M, Ferrarese F. 2013. Uncontacted Waorani in the Yasuní Biosphere Reserve: Geographical Validation of the Zona Intangible Tagaeri Taromenane (ZITT). *PLoS One* 8, e66293.

- Papworth S, Milner-Gulland EJ, Slocombe K. 2013. The natural place to begin: the ethnoprimateology of the Waorani. *Am J Primatol* 75, 1117-1128.
- Park JK, Kim W, Kim SW, Koh GY, Park SK. 1999. Gene-polymorphisms of angiotensin converting enzyme and endothelial nitric oxide synthase in patients with erectile dysfunction. *Int J Impot Res* 11, 273-276.
- Patiño-Lopez G, Hevezi P, Lee J, Willhite D, Verge GM, Lechner SM, Ortiz-Navarrete V, Zlotnik A. 2006. Human class-I restricted T cell associated molecule is highly expressed in the cerebellum and is a marker for activated NKT and CD8+ T lymphocytes. *J Neuroimmunol* 171, 145-155.
- Peeke MC. 1973. Preliminary grammar of Auca. *Norman, Oklahoma, USA*: Summer Institute of Linguistics and University of Oklahoma.
- Peña JA, Calderon R, Perez-Miranda A, Vidales C, Dugoujon JM, Carrion M, Crouau-Roy B. 2002. Microsatellite DNA markers from HLA region (D6S105, D6S265 and TNFa) in autochthonous Basques from Northern Navarre (Spain). *Ann Hum Biol* 29, 176-191.
- Peña JA, Alfonso-Sánchez MA, Pérez-Miranda AM, García-Obregón S, Gómez-Pérez L. 2009. GeDis: Un programa para análisis de datos en Antropogenética. *Antropo* 20, 49-56. www.didac.ehu.es/antropo.
- Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Hooshiar Kashani B, Ritchie KH, Scozzari R, Kong QP, Myres NM, Salas A, Semino O, Bandelt HJ, Woodward SR, Torroni A. 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19, 1-8.
- Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, Carossa V, Ekins JE, Gómez-Carballa A, Huber G, Zimmermann B, Corach D, Babudri N, Panara F, Myres NM, Parson W, Semino O, Salas A, Woodward SR, Achilli A, Torroni A. 2010. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20, 1174-1179.
- Pérez-Miranda AM, Alfonso-Sánchez MA, Peña JA, Calderón R. 2003. HLA-DQA1 polymorphism in autochthonous Basques from Navarre (Spain): Genetic position within European and Mediterranean scopes. *Tissue Antigens* 61, 465-474.
- Pérez-Miranda AM, Alfonso-Sánchez MA, Vidales MC, Calderón R, Peña JA. 2004. Genetic polymorphism and linkage disequilibrium of the HLA-DP region in Basques from Navarre (Spain). *Tissue Antigens* 64, 264-275.
- Perna NT, Batzer MA, Deininger PL, Stoneking M. 1992. Alu insertion polymorphism: a new type of marker for human population studies. *Hum Biol.* 64, 641-648.
- Phillips C, Rodriguez A, Mosquera-Miguel A, Fondevila M, Porras-Hurtado L, Rondon F, Salas A, Carracedo A, Lareu MV. 2008. D9S1120, a simple STR with a common Native American-specific allele: forensic optimization, locus characterization and allele frequency studies. *Forensic Sci Int Genet* 3, 7-13.
- Pickrell JK, Reich D. 2014. Toward a new history and geography of human genes informed by ancient DNA. *Trends Genet* 30, 377-389.
- Powell JF, Neves WA. 1999. Craniofacial morphology of the first Americans: Pattern and process in the peopling of the New World. *Am J Phys Anthropol Suppl* 29, 153-188.
- Powell JF. 2005. *The First Americans: Race, Evolution, and the Origin of Native Americans*. Cambridge: Cambridge University Press.
- Prendergast JG, Chambers EV, Semple CA. 2014. Sequence-level mechanisms of human epigenome evolution. *Genome Biol Evol* 6, 1758-1771.
- Presidencia de la Republica de Ecuador. 1999. Decreto Ejecutivo No. 552. Quito, Ecuador.
- Presidencia de la Republica. 2007. Decreto Ejecutivo No. 2187. Quito, Ecuador. Available: http://saveamericasforests.org/Yasuni/News/ZonaIntangible/DECRETO_ZONA_INTANGIBLE.pdf.
- Pyper W, Burt M, Powell L, Webb S, Adès L, Halliday J, Jazwinska E. 1995. A region of primer binding variation at the D6S265 locus associated with HLA-A25 and HLA-A26 antigens. *Hum Genet* 96, 490-492.
- Quentin Y. 1992a. Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. *Nucleic Acids Res* 20, 487-493.

- Quentin Y. 1992b. Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic Acids Res* 20, 3397-3401.
- Raaum RL, Wang AB, Al-Meerri AM, Mulligan CJ. 2010. Efficient population assignment and outlier detection in human populations using biallelic markers chosen by principal component-based rankings. *Biotechniques* 48, 449-454.
- Raff JA, Bolnick DA, Tackney J, O'Rourke DH. 2011. Ancient DNA perspectives on American colonization and population history. *Am J Phys Anthropol* 146, 503-514.
- Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliusson TS, Grønnow B, Appelt M, Gulløv HC, Friesen TM, Fitzhugh W, Malmström H, Rasmussen S, Olsen J, Melchior L, Fuller BT, Fahrni SM, Stafford T Jr, Grimes V, Renouf MA, Cybulski J, Lynnerup N, Lahr MM, Britton K, Knecht R, Arneborg J, Metspalu M, Cornejo OE, Malaspina AS, Wang Y, Rasmussen M, Raghavan V, Hansen TV, Khusnutdinova E, Pierre T, Dneprovsky K, Andreasen C, Lange H, Hayes MG, Coltrain J, Spitsyn VA, Götherström A, Orlando L, Kivisild T, Villems R, Crawford MH, Nielsen FC, Dissing J, Heinemeier J, Meldgaard M, Bustamante C, O'Rourke DH, Jakobsson M, Gilbert MT, Nielsen R, Willerslev E. 2014. The genetic prehistory of the New World Arctic. *Science* 345, 1255-1259.
- Ragoussis J, Bloemer K, Weiss EH, Ziegler A. 1988. Localization of the Genes for tumor necrosis factor and lymphotoxin between the HLA class I and III regions by field inversion gel electrophoresis. *Immunogenetics* 27, 66-69.
- Raha-Chowdhury R, Andrews SR, Gruen JR. 2005. CAT 53: a protein phosphatase 1 nuclear targeting subunit encoded in the MHC Class I region strongly expressed in regions of the brain involved in memory, learning, and Alzheimer's disease. *Brain Res Mol Brain Res* 138, 70-83.
- Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK. 2012. ALFRED: an allele frequency resource for research and teaching. *Nucleic Acids Res* 40, D1010- D1015.
- Ramenenovsky AF. 1987. Vectors of death: the archeology of European contact. *Albuquerque, New Mexico, USA*: University of New Mexico Press.
- Rao RC. 1948. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, Series B* 10, 159-203.
- Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M5, Stafford TW Jr, Rasmussen S, Moltke I, Albrechtsen A, Doyle SM, Poznik GD, Gudmundsdottir V, Yadav R, Malaspina AS, White SS 5th, Allentoft ME, Cornejo OE, Tambets K, Eriksson A, Heintzman PD, Karmin M, Korneliusson TS, Meltzer DJ18, Pierre TL, Stenderup J, Saag L, Warmuth VM, Lopes MC, Malhi RS, Brunak S, Sicheritz-Ponten T, Barnes I, Collins M, Orlando L, Balloux F, Manica A, Gupta R, Metspalu M, Bustamante CD, Jakobsson M, Nielsen R, Willerslev E. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506, 225-229.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MT, Wang Y, Raghavan M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorenzen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grønnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TF, Ramsey CB, Hansen TV, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén T, Villems R, Nielsen R, Krogh A, Wang J, Willerslev E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463, 757-762.
- Raymond M, Rousset F. 1995. An exact test for population differentiation. *Evolution* 49, 1280-1283.
- Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, García LF, Triana O, Blair S, Maestre A, Dib JC, Bravi CM, Bailliet G, Corach D, Hünemeier T, Bortolini MC, Salzano FM, Petzl-Erler ML, Acuña-Alonzo V, Aguilar-Salinas C, Canizales-Quinteros S, Tusié-Luna T, Riba L, Rodríguez-Cruz M, Lopez-Alarcón M, Coral-Vazquez R, Canto-Cetina T, Silva-Zolezzi I, Fernandez-Lopez JC, Contreras AV, Jimenez-Sanchez G, Gómez-Vázquez MJ, Molina J, Carracedo A, Salas A, Gallo C, Poletti G, Witonsky DB, Alkorta-Aranburu G, Sukernik RI, Osipova L, Fedorova SA, Vasquez R, Villena M, Moreau C, Barrantes R, Pauls D, Excoffier L, Bedoya G, Rothhammer F, Dugoujon JM, Larrouy G, Klitz W, Labuda D, Kidd J, Kidd K, Di Rienzo A, Freimer NB, Price AL, Ruiz-Linares A. 2012. Reconstructing Native American population history. *Nature* 488, 370-374.

- Relethford JH. 1996. Genetic drift can obscure population history: problem and solution. *Hum Biol* 68, 29-44.
- Relethford JH. 2012. Human population Genetics. *Hoboken, New Jersey, USA: Wiley-Blackwell*.
- Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term Genetic distance. *Genetics* 105, 767-779.
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, Bandelt HJ. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 62, 241-260.
- Richards M. 2003. The Neolithic invasion of Europe. *Annu Rev Anthropol* 32, 135-162.
- Ridker PM, Baker MT, Hennekens CH, Stampfer MJ, Vaughan DE. 1997. Alu-repeat polymorphism in the gene coding for tissue-type plasminogen activator (t-PA) and risks of myocardial infarction among middle-aged men. *Arterioscler Thromb Vasc Biol* 17, 1687-1690.
- Rival L. 2002. Trekking through history. *New York, USA: Columbia University Press*.
- Robarchek CA, Robarchek CJ. 1992. Cultures of War and Peace: A Comparative Study of Waorani and Semai. En: Silverberg J, Gray JP (Eds). Aggression and Peacefulness in Humans and Other Primates. *New York, USA: Oxford University Press*. pp 189-213.
- Rodas C, Gelvez N, Keyeux G. 2003. Mitochondrial DNA studies show asymmetrical Amerindian admixture in Afro-Colombian and Mestizo populations. *Hum Biol* 75, 13-30.
- Rodríguez J, Vives L, Jordà M, Morales C, Muñoz M, Vendrell E, Peinado MA. 2008. Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells. *Nucleic Acids Res* 36, 770-784.
- Roewer L, Nothnagel M, Gusmão L, Gomes V, González M, Corach D, Sala A, Alechine E, Palha T, Santos N, Ribeiro-Dos-Santos A, Geppert M, Willuweit S, Nagy M, Zweynert S, Baeta M, Núñez C, Martínez-Jarreta B, González-Andrade F, Fagundes de Carvalho E, da Silva DA, Builes JJ, Turbón D, Lopez Parra AM, Arroyo-Pardo E, Toscanini U, Borjas L, Barletta C, Ewart E, Santos S, Krawczak M. 2013. Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. *PLoS Genet* 9, e1003460.
- Rojas W, Parra MV, Campo O, Caro MA, Lopera JG, Arias W, Duque C, Naranjo A, García J, Vergara C, Lopera J, Hernandez E, Valencia A, Caicedo Y, Cuartas M, Gutiérrez J, López S, Ruiz-Linares A, Bedoya G. 2010. Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers. *Am J Phys Anthropol* 143, 13-20.
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. *Science* 298, 2381-2385.
- Rothhammer F, Cocilovo JA, Quevedo S. 1984. El poblamiento temprano de Sudamérica. *Chungará* 13, 99-108.
- Rothhammer F, Dillehay TD. 2009. The late Pleistocene Colonization of South America: An interdisciplinary Perspective. *Ann Hum Genet* 73, 540-549.
- Rothhammer F, Silva C. 1992. Gene Geography of South America: Testing models of populations displacement based on archaeological evidence. *Am J Phy Anthropol* 89, 441-446.
- Rowold DJ, Herrera RJ. 2000. Alu elements and the human genome. *Genetica* 108, 57-72.
- Rowold DJ, Herrera RJ. 2005. On human STR sub-population structure. *Forensic Sci Int* 151, 59-69.
- Roy-Engel AM, Batzer MA, Deininger P. 2008. Eolution of human retrosequences: Alu. En: Kehrer-Sawatzki H (Ed). Encyclopedia of life sciences. *Chichester, UK: John Wiley & Sons, Ltd*. DOI: 10.1002/9780470015902.a0005131.pub2
- Roy-Engel AM, Carroll ML, El-Sawy M, Salem AH, Garber RK, Nguyen SV, Deininger PL, Batzer MA. 2002. Non-traditional Alu evolution and primate genomic diversity. *J Mol Biol* 316, 1033-1040.

- Ruiz-Linares A. 2014. How Genes Have Illuminated the History of Early Americans and Latino Americans. *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a008557. [Epub ahead of print]
- Rupert JL, Devine DV, Monsalve MV, Hochachka PW. 1999. Angiotensin-converting enzyme (ACE) alleles in the Quechua, a high altitude South American native population. *Ann Hum Biol* 26, 375-380.
- Rupert JL, Kidd KK, Norman LE, Monsalve MV, Hochachka PW, Devine DV. 2003. Genetic polymorphisms in the Renin-Angiotensin system in high-altitude and low-altitude Native American populations. *Ann Hum Genet* 67, 17-25.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-425.
- Salzano FM, Bortolini MC. 2002. The Evolution and Genetics of Latin American Populations. *Cambridge, UK: Cambridge University Press.*
- Salzano FM, Sans M. 2014. Interethnic admixture and the evolution of Latin American populations. *Genet Mol Biol* 37, 151-170.
- Sánchez-Albornoz N. 1974. The Population of Latin America: A History. *Berkeley, California, USA: University California Press.*
- Sánchez-Albornoz N. 1977. La población de América Latina: desde los tiempos precolombinos al año 2000. 2ª ed. *Madrid, España: Editorial Alianza.*
- Sánchez-Domingo R. 2012. Las Leyes de Burgos de 1512 y la doctrina jurídica de la conquista. *Revista Jurídica de Castilla y León* 28, ISSN 2254-3805. Historia del derecho.
- Sánchez-Mazas A, Meyer D. 2014. The relevance of HLA sequencing in population genetics studies. *J Immunol Res* 2014, 971818. doi: 10.1155/2014/971818
- Sandoval C, De la Hoz A, Yunis EJ. 1993. Estructura genética de la población colombiana. Análisis de mestizaje. *Rev Fac Med UN Colombia* 43, 3-14.
- Sans M, Salzano FM, Chakraborty R. 1997. Historical Genetics in Uruguay: estimates of biological origins and their problems. *Hum Biol* 69, 161-170.
- Sans M, Weimer TA, Franco MHL, Salzano FM, Bentancor N, Alvarez I, Bianchi NO, Chakraborty R. 2002. Unequal contributions of male and female *Gene* pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am J Phys Anthropol* 118, 33-44.
- Sato H, Okada Y, Seiki M. 1997. Membrane-type matrix metalloproteinases (MT-MMPs) in cell invasion. *Thromb Haemost* 78, 497-500.
- Schneider H, Guerreiro JF, Santos SE, Weimer TA, Schneider MPC, Salzano FM. 1987. Isolate breakdown in Amazonia – The Blacks of the Trombetas River. *Rev Bras Genet* 10: 565-74
- Schroeder KB, Jakobsson M, Crawford MH, Schurr TG, Boca SM, Conrad DF, Tito RY, Osipova LP, Tarskaia LA, Zhadanov SI, Wall JD, Pritchard JK, Malhi RS, Smith DG, Rosenberg NA. 2009. Haplotypic background of a private allele at high frequency in the Americas. *Mol Biol Evol* 26, 995-1016.
- Schurr TG, Sherry ST. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: Evolutionary and demographic evidence. *Am J Hum Biol* 16, 420-439.
- Schurr TG. 2004. The peopling of the New World: Perspectives from molecular anthropology. *Annu Rev Anthropol* 33, 551-583.
- Schwartz ML, Pizzo SV, Hill RL, McKee PA. 1973. Human Factor XIII from plasma and platelets. Molecular weights, subunit structures, proteolytic activation, and cross-linking of fibrinogen and fibrin. *J Biol Chem* 248, 1395-1407.
- Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. 2007. Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biology* 8, R127.
- Sens-Abuázar C, Santos PS, Bicalho MG, Petzl-Erler ML, Sperandio-Roxo V. 2009. MHC microsatellites in a Southern Brazilian population. *Int J Immunogenet* 36, 269-274.

- Seymour RM, Allan MJ, Pomiankowski A, Gustafsson K. 2004. Evolution of the human ABO polymorphism by two complementary selective pressures. *Proc Biol Sci* 271, 1065-1072.
- Shen MR, Batzer MA, Deininger PL. 1991. Evolution of the master Alu gene(s). *J Mol Evol* 33, 311-320.
- Shepard EM, Herrera RJ. 2006. Iranian STR variation at the fringes of biogeographical demarcation. *Forensic Sci Int* 158: 140-148.
- Shiina T, Hosomichi K, Inoko H, Kulski JK. 2009. The HLA genomic loci map: expression, interaction, diversity and disease. *J Hum Gen* 54, 15-39.
- Shook BA, Smith DG. 2008. Using ancient mtDNA to reconstruct the population history of northeastern North America. *Am J Phys Anthropol* 137, 14-29.
- Shriver MD, Kittles RA. 2004. Genetic ancestry and the search for personalized Genetic histories. *Nature Rev Genet* 5, 611-618.
- Silva WA Jr, Bonatto SL, Holanda AJ, Ribeiro-Dos-Santos AK, Paixão BM, Goldman GH, Abe-Sandes K, Rodriguez-Delfin L, Barbosa M, Paçó-Larson ML, Petzl-Erler ML, Valente V, Santos SE, Zago MA. 2002. Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. *Am J Hum Genet* 71, 187-192.
- Simon M, Bourel M, Fauchet R, Genetet B. 1976. Association of HLA-A3 and HLA-B14 antigens with idiopathic haemochromatosis. *Gut* 17, 332-334.
- Simons AM. 2002. The continuity of microevolution and macroevolution. *Journal of Evolutionary Biology* 15, 688-701.
- Simpson GG. 1944. *Tempo and Mode in Evolution*. New York, USA: Columbia University Press.
- Simpson GG. 1953. *The Major Features of Evolution*. New York, USA: Columbia University Press.
- Sjödín P, François O. 2011. Wave-of-advance models of the diffusion of the Y chromosome haplogroup R1b1b2 in Europe. *PLoS One* 6, e21592.
- Slatkin M, Excoffier L. 1996. Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity* 76, 377-383.
- Slatkin M, Excoffier L. 2012. Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics* 191, 171-181.
- Slatkin M. 1994a. Linkage disequilibrium in growing and stable populations. *Genetics* 137, 331-336.
- Slatkin M. 1994b. Gene flow and population structure. En: Real LA, (Ed). *Ecological Genetics*. New Jersey, USA: Princeton University Press. pp 3-18.
- Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139, 457-462.
- Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8, 272-285.
- Smith DG, Malhi RS, Eshleman JA, Kaestle FA, Kemp BM. 2005. En: Paleoamerican Origins: Beyond Clovis. Bonnicksen R, Lepper BT, Stanford D, Waters MR (Eds). *Paleoamerican Origins: Beyond clovis. College station TX: Center for the study of the first Americans, department of Anthropology, Texas: A&M University Press*. pp. 243-254.
- Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the Mantel Test of matrix correspondence. *Systematic Zoology* 35, 627-632.
- Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt HJ, Torroni A, Richards M. 2010. The archaeogenetics of Europe. *Curr Biol* 20, R174-R183.
- Sousa V, Peischl S, Excoffier L. 2014. Impact of range expansions on current human genomic diversity. *Curr Opin Genet Dev* 29, 22-30.

- Stanford D, Bradley B. 2002. Ocean Trails and Prairie Paths? Thoughts About Clovis Origins. En: Jablonski NG (ed). *The First Americans: The Pleistocene Colonization of the New World (Memoirs of the California Academy of Sciences, No. 27)*. San Francisco, USA: California Academy of Sciences. pp 255–271.
- Starikovskaya EB, Sukernik RI, Derbeneva OA, Volodko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, Wallace DC. 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet* 69, 67–89.
- Stewart FM. 1977. Computer algorithm for obtaining a random set of allele frequencies for a locus in an equilibrium population. *Genetics* 86, 482–483.
- Stone AC, Stoneking M. 1998. mtDNA Analysis of a Prehistoric Oneota Population: Implications for the Peopling of the New World. *Am J Hum Genet* 62, 1153–1170.
- Stone C, Pointon JJ, Jazwinska EC, Halliday JW, Powell LW, Robson KJ, Monaco AP, Weatherall DJ. 1994. Isolation of CA dinucleotide repeats close to D6S105; linkage disequilibrium with haemochromatosis. *Hum Mol Genet* 3, 2043–2046.
- Stoneking M, Deininger PL, Batzer MA. 2001. Alu insertion polymorphisms in humans: a review. En: Donnelly P, Foley R (eds). *Genes, Fossils and Behaviour. Amsterdam, Netherlands: IOS Press*. pp. 111–121.
- Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* 7, 1061–1071.
- Straus LG. 2000. Solutrean Settlement of North America? A Review of Reality. *American Antiquity (Society for American Archaeology)* 65, 219–226.
- Su M, Han D, Boyd-Kirkup J, Yu X, Han JD. 2014. Evolution of Alu elements toward enhancers. *Cell Rep* 7, 376–385.
- Sun H, Liu S, Zhang Y, Whittle MR. 2014. Comparison of southern Chinese Han and Brazilian Caucasian mutation rates at autosomal short tandem repeat loci used in human forensic genetics. *Int J Legal Med* 128, 1–9.
- Szathmary EJE. 1993. Genetics of aboriginal North Americans. *Evol Anthropol* 1, 202–220.
- Takeuchi K, Majima Y, Sakakura Y. 2000. Tumor necrosis factor gene polymorphism in chronic sinusitis. *Laryngoscope* 110, 1711–1714.
- Tamaki K, Jeffreys AJ. 2005. Human tandem repeat sequences in forensic DNA typing. *Leg Med (Tokyo)* 7, 244–250.
- Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Vilems R, Malhi RS. 2007. Beringian standstill and spread of Native American founders. *PLoS One* 2, 1–6.
- Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* 68, 1485–1496.
- Tarragó MN. 2000. Chacras y pukara. Desarrollos Sociales Tardíos. En: Tarragó MN (Dir). *Nueva historia argentina. Vol. 1, Los pueblos originarios y la conquista: 257–300. Buenos Aires Argentina: Sudamericana*.
- Tautz D, Schlötterer C. 1994. Simple sequences. *Curr Opin Genet Dev* 4, 832–837.
- Terreros MC, Alfonso-Sánchez MA, Novick GE, Luis JR, Lacau H, Lowery RK, Regueiro M, Herrera RJ. 2009. Insights on human evolution: an analysis of Alu insertion polymorphisms. *J Hum Genet* 54, 603–611.
- Thalmann O, Fischer A, Lankester F, Pääbo S, Vigilant L. 2007. The complex evolutionary history of gorillas: insights from genomic data. *Mol Biol Evol* 24, 146–158.

- The MHC sequencing consortium. 1999. Complete sequence and *Gene* map of a human major histocompatibility complex. *Nature* 401, 921-923.
- Thomas H. 1997. *The Slave Trade*. New York, USA: Simon and Schuster.
- Tian W, Wang F, Cai JH, Li LX. 2008. Polymorphic insertions in 5 Alu loci within the major histocompatibility complex class I region and their linkage disequilibria with HLA alleles in four distinct populations in mainland China. *Tissue Antigens* 72, 559-567.
- Tiret L, Ducimetière P, Bonnardeaux A, Soubrier F, Poirier O, Ricard S, Cambien F, Marques-Vidal P, Evans A, Kee F, Arveiler D, Luc G. 1994. Synergistic effects of angiotensin-converting enzyme and angiotensin-II type 1 receptor gene polymorphisms on risk of myocardial infarction. *Lancet* 344, 910-913.
- Tiret L, Rigat B, Visvikis S, Breda C, Corvol P, Cambien F, Soubrier F. 1992. Evidence, from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. *Am J Hum Genet* 51, 197-205.
- Tishkoff SA, Pakstis AJ, Stoneking M, Kidd JR, Destro-Bisol G, Sanjantila A, Lu RB, Deinard AS, Sirugo G, Jenkins T, Kidd KK, Clark AG. 2000. Short tandem-repeat polymorphism/Alu haplotype variation at the PLAT locus: implications for modern human origins. *Am J Hum Genet* 67, 901-925.
- Tishkoff SA, Ruano G, Kidd JR, Kidd KK. 1996. Distribution and frequency of a polymorphic Alu insertion at the plasminogen activator locus in humans. *Hum Genet* 97, 759-764.
- Tishkoff SA, Verrelli BC. 2003. Pattern of human Genetic diversity: Implications for Human Evolutionary History and Disease. *Annu Rev Genomics Hum Genet* 4, 293-340.
- Tonelli LH, Postolache TT. 2005. Tumor necrosis factor alpha, interleukin-1 beta, interleukin-6 and major histocompatibility complex molecules in the normal brain and after peripheral immune challenge. *Neurol Res* 27, 679-684.
- Törn C, Hillman M, Sanjeevi CB, Landin-Olsson M. 2006. Polymorphisms of TNF microsatellite marker and HLA-DR-DQ in diabetes mellitus-a study in 609 Swedish subjects. *Hum Immunol* 67, 527-534.
- Torrioni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk HV, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus ML, Huoponen K, Laitinen V, Koivumäki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O, Scozzari R. 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69, 844-852.
- Torrioni A, Schurr TG, Yang CC, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130, 153-162.
- Torrioni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC. 1993. mtDNA variation of aboriginal Siberians reveals distinct Genetic affinities with Native Americans. *Am J Hum Genet* 53, 591-608.
- Trent JM, Ziegler A. 1993. Report of the First International Workshop on Human Chromosome 6 Mapping. Ann Arbor, Michigan, June 7-9, 1992. *Cytogenet Cell Genet* 62, 67-87.
- Trowsdale J. 1995. "Both man & bird & beast": comparative organization of MHC genes. *Immunogenetics* 41, 1-17.
- Uçer B, Kayıkçıoğlu O, Seymenoğlu G, Var A, Cam S. 2011. The relationship between angiotensin converting enzyme insertion/deletion polymorphism and age-related macular degeneration. *Ophthalmic Genet* 32, 158-161.
- Ullu E, Tschudi C. 1984. Alu sequences are processed 7SL RNA genes. *Nature* 312, 171-172.
- Urquhart A, Kimpton CP, Downes TJ, Gill P. 1994. Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers. *Int J Legal Med* 107, 13-20.

- Van den Eijnden-Schrauwen Y, Lakenberg N, Emeis JJ, de Knijff P. 1995. Alu-repeat polymorphism in the tissue type plasminogen activator *Gene* does not affect basal endothelial t-PA synthesis. *Thromb Haemost* 74, 1202.
- Van der Bom JG, de Knijff P, Haverkate F, Bots ML, Meijer P, de Jong PT, Hofman A, Kluit C, Grobbee DE. 1997. Tissue plasminogen activator and risk of myocardial infarction. The Rotterdam Study. *Circulation* 95, 2623-2627.
- Vilborg A, Bersani C, Wilhelm MT, Wiman KG. 2011. The p53 target Wig-1: a regulator of mRNA stability and stem cell fate? *Cell Death Differ* 18, 1434-1440.
- Volodko NV, Starikovskaya EB, Mazunin IO, Eltsov NP, Naidenko PV, Wallace DC, Sukernik RI. 2008. Mitochondrial Genome Diversity in Arctic Siberians, with Particular Reference to the Evolutionary History of Beringia and Pleistocenic Peopling of the Americas. *Am J Hum Genet* 82, 1084-1100.
- Volz A, Boyle JM, Cann HM, Cottingham RW, Orr HT, Ziegler A. 1994. Report of the Second International Workshop on Human Chromosome 6. *Genomics* 21, 464-472.
- Wallace DC, Torroni A. 1992. American Indian prehistory as written in the mitochondrial DNA: a review. *Hum Biol* 64, 403-416.
- Wallis EE. 1973. Aucas downriver. *New York, USA: Harper and Row.*
- Wang J. 2003. Maximum-Likelihood estimation of admixture proportions from genetic data. *Genetics* 164, 747-765.
- Wang N, Zhou X, Tan FK, Foster MW, Arnett FC, Chakraborty R. 2004. Genetic signatures of pre-expansion bottleneck in the Choctaw population of Oklahoma. *Am J Phys Anthropol* 124, 373-379.
- Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A. 2007. Genetic variation and population structure in native Americans. *PLoS Genet* 3, e185.
- Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L, Ruiz-Linares A. 2008. Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet* 21, 3.
- Waters MR, Stafford TW Jr, McDonald HG, Gustafson C, Rasmussen M, Cappellini E, Olsen JV, Szklarczyk D, Jensen LJ, Gilbert MT, Willerslev E. 2011. Pre-Clovis mastodon hunting 13,800 years ago at the Manis site, Washington. *Science* 334, 351-353.
- Waters MR, Stafford TW Jr. 2007. Redefining the age of Clovis: implications for the peopling of the Americas. *Science* 315, 1122-1126.
- Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB. 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am J Hum Genet* 68, 738-752.
- Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV, Reddy PG, Das PK, Batzer MA, Jorde LB. 2003. Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. *Genome Res* 13, 1607-1618.
- Watkins WS, Xing J, Huff C, Witherspoon DJ, Zhang Y, Perego UA, Woodward SR, Jorde LB. 2012. Genetic analysis of ancestry, admixture and selection in Bolivian and Totonac populations of the New World. *BMC Genet* 20, 13-39.
- Webb SD. 2005. First Floridians and Last Mastodons: The Page-Ladson Site in the Aucilla River. *Dordrecht, Netherlands: Springer.*
- Weber JL, May PE. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44, 388-396.
- Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW, Boesch. 1999. Culture in chimpanzees. *Nature* 399, 628-685.

- Willard C, Nguyen HT, Schmid CW. 1987. Existence of at least three distinct Alu subfamilies. *J Mol Evol* 26, 180–186.
- Witherspoon DJ, Marchani EE, Watkins WS, Ostler CT, Wooding SP, Anders BA, Fowlkes JD, Boissinot S, Furano AV, Ray DA, Rogers AR, Batzer MA, Jorde LB. 2006. Human population genetic structure and diversity inferred from polymorphic L1 (LINE-1) and *Alu* insertions. *Hum. Hered* 62, 30-46.
- Witherspoon DJ, Watkins WS, Zhang Y, Xing J, Tolpinrud WL, Hedges DJ, Batzer MA, Jorde LB. 2009. Alu repeats increase local recombination rates. *BMC Genomics* 10:530-541.
- Witherspoon DJ, Zhang Y, Xing J, Watkins WS, Ha H, Batzer MA, Jorde LB. 2013. Mobile element scanning (ME-Scan) identifies thousands of novel Alu insertions in diverse human populations. *Genome Res* 23, 1170-1181.
- Wood JW, Johnson PL, Kirk RL, McLoughlin K, Blake NM, Matheson FA. 1982. The genetic demography of the Gainj of Papua New Guinea. I. Local differentiation of blood group, red cell enzyme, and serum protein allele frequencies. *Am J Phys Anthropol* 57, 15-25.
- Worwood M, Raha-Chowdhury R, Darke C. 1994a. Distribution of alleles at D6S105 and D6S265 with possible HLA haplotype associations. *Tissue Antigens* 44, 322-325.
- Worwood M, Raha-Chowdhury R, Dorak MT, Darke C, Bowen DJ, Burnett AK. 1994b. Alleles at D6S265 and D6S105 define a haemochromatosis-specific genotype. *Br J Haematol* 86, 863-866.
- Wright S. 1950. Genetical structure of populations. *Nature* 166, 247-249.
- Wright S. 1951. The Genetic structure of populations. *Ann Eugen* 1, 323-334.
- Wright S. 1965. The interpretation of population structure by F-statistics with special regards to systems of mating. *Evolution* 19, 395-420.
- Wu J, Lee WP, Ward A, Walker JA, Konkel MK, Batzer MA, Marth GT. 2014. Tangram: a comprehensive toolbox for mobile element insertion detection. *BMC Genomics* 15, 795 [Epub ahead of print].
- Xiao BG, Link H. 1998. Immune regulation within the central nervous system. *J Neurol Sci* 157, 1-12.
- Xiao FX, Yang JF, Cassiman JJ, Decorte R. 2002. Diversity at eight polymorphic Alu insertion loci in Chinese populations shows evidence for European admixture in an ethnic minority population from northwest China. *Hum Biol* 74, 555-568.
- Xie Y, You C, Chen J. 2014. An updated meta-analysis on association between angiotensin I-converting enzyme gene insertion/deletion polymorphism and cancer risk. *Tumour Biol* 35, 6567-6579.
- Yacobaccio HD, Catá MP, Solá P, Alonso MS. 2008. Estudio arqueológico y fisicoquímico de pinturas rupestres en Hornillos 2 (Puna de Jujuy). *Estudios atacameños* 36, 5-28.
- Yang-Feng TL, Opdenakker G, Volckaert G, Francke U. 1986. Human tissue-type plasminogen activator *Gene* located near chromosomal breakpoint in myeloproliferative disorder. *Am J Hum Genet* 39, 79-87.
- Yao Y, Shi L, Shi L, Kulski JK, Chen J, Liu S, Yu L, Lin K, Huang X, Tao Y, Tokunaga K, Chu J. 2010. The association and differentiation of MHC class I polymorphic Alu insertions and HLA-B/Cw alleles in seven Chinese populations. *Tissue Antigens* 76, 194-207.
- Yao Y, Shi L, Shi L, Lin K, Tao Y, Yu L, Sun H, Huang X, Li Y, Chu J. 2009a. Polymorphic Alu insertions and their associations with MHC class I alleles and haplotypes in Han and Jinuo populations in Yunnan Province, southwest of China. *J Genet Genomics* 36, 51-58.
- Yao Y, Shi L, Shi L, Lin K, Yu L, Sun H, Huang X, Tao Y, Yi W, Liu S, Chu J. 2009b. The association between HLA-A, -B alleles and major histocompatibility complex class I polymorphic Alu insertions in four populations in China. *Tissue Antigens* 73, 575-581.
- Yost JA. 1981. Twenty years of contact: the mechanisms of change in Huaor (Auca) culture. En: Whitten NA (Ed). Cultural transformations and ethnicity in modern Ecuador. *Urbana, Illinois, USA*: University of Illinois Press. pp 677–704.

- Zabala-Fernandez WM, Borjas-Fajardo L, Fernandez Salgado E, Castillo C, Socca L, Portillo MG, Sanchez MA, Delgado W, Morales-Machin A, Layrisse Z, Pineda Bernal L. 2005. Use of short tandem repeats loci to study the Genetic structure of several populations from Zulia State, Venezuela. *Am J Hum Biol* 17, 451–459.
- Zazula GD, Froese DG, Schweger CE, Mathewes RW, Beaudoin AB, Telka AM, Harington CR, Westgate JA. 2003. Palaeobotany: Ice-age steppe vegetation in east Beringia. *Nature* 423, 603.
- Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol* 21, 164-175.
- Zhivotovsky LA, Bennett L, Bowcock AM, Feldman MW. 2000. Human population expansion and microsatellite variation. *Mol Biol Evol* 17, 757-767.
- Zhivotovsky LA, Rosenberg NA, Feldman MW. 2003. Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am J Hum Genet* 72, 1171-1186.
- Zhivotovsky LA, Underhill PA, Cinnioğlu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 74, 50-61.
- Zhu X, Bouzekri N, Southam L, Cooper RS, Adeyemo A, McKenzie CA, Luke A, Chen G, Elston RC, Ward R. 2001. Linkage and association analysis of angiotensin I-converting enzyme (ACE)-Gene polymorphisms with ACE concentration and blood pressure. *Am J Hum Genet* 68, 1139-1148.

MATERIAL ANEXO

Tabla Anexo 1: Departamentos pertenecientes a Colombia junto con su capital. La clave vincula los departamentos con su localización en el mapa a través de la *Figura 6*.

Clave	Departamento	Capital
1.	La Guajira	Riohacha
2.	San Andrés, Providencia y Santa Catalina	San Andrés
3.	Magdalena	Santa Marta
4.	Atlántico	Barranquilla
5.	Cesar	Valledupar
6.	Bolívar	Cartagena de Indias
7.	Sucre	Sincelejo
8.	Córdoba	Montería
9.	Norte de Santander	San José de Cúcuta
10.	Antioquia	Medellín
11.	Chocó	Quibdó
12.	Santander	Bucaramanga
13.	Arauca	Arauca
14.	Boyacá	Tunja
15.	Casanare	Yopal
16.	Vichada	Puerto Carreño
17.	Cundinamarca	Bogotá
18.	Caldas	Manizales
19.	Risaralda	Pereira
20.	Bogotá Distrito Capital	-
21.	Tolima	Ibagué
22.	Valle del Cauca	Cali
23.	Quindío	Armenia
24.	Meta	Villavicencio
25.	Guainía	Inírida
26.	Huila	Neiva
27.	Cauca	Popayán
28.	Guaviare	San José del Guaviare
29.	Caquetá	Florencia
30.	Nariño	San Juan de Pasto
31.	Vaupés	Mitú
32.	Putumayo	Mocoa

Tabla Anexo 2: Departamentos pertenecientes a la provincia de Jujuy detallados en la *Figura 9*. Se indica su localidad cabecera y la región a la que pertenecen.

Clave	Departamento	Localidad cabecera	Región
a	Santa Barbara	Santa Clara	Selva
b	Ledesma	Libertador General San Martín	Selva
c	San Pedro	San Pedro de Jujuy	Selva
d	Valle Grande	Valle Grande	Selva
e	El Cármen	El Cármen	Valle
f	Palpalá	Palpalá	Valle
g	San Antonio	San Antonio	Valle
h	Dr. Manuel Belgrano	San Salvador de Jujuy	Valle
i	Tumbaya	Tumbaya	Quebrada
j	Tilcara	Tilcara	Quebrada
k	Humahuaca	Humahuaca	Quebrada
l	Cochinoca	Abra Pampa	La Puna
m	Yavi	La Quiaca	La Puna
n	Rinconada	Rinconada	La Puna
o	Santa Catalina	Santa Catalina	La Puna
p	Susques	Susques	La Puna

Tabla Anexo 3: Poblaciones muestreadas en la provincia de Jujuy especificando la región donde se encuentran, la altitud y el tamaño muestral (N) obtenido. Dichas poblaciones se encuentran ubicadas en el mapa de la provincia mediante la *Figura 10*.

Clave	Población	Región	Altitud (m.s.n.m)	N
1	San Salvador de Jujuy	Valle	1200	61
2	Palpalá	Valle	1125	10
3	El Carmen	Valle	1140	2
4	Perico	Valle	895	2
5	La Mendieta	Selva	700	1
6	San Pedro de Jujuy	Selva	553	9
7	La Esperanza	Selva	519	3
8	Fraile Pintado	Selva	415	1
9	Libertador General San Martín	Selva	472	11
10	San lucas	Selva	500	1
11	El talar	Selva	318	3
12	Maimara	Quebrada	2500	37
13	El Moreno	Quebrada	3500	21
14	Humahuaca	Quebrada	3012	6
15	San Miguel de los Colorados-Tumbaya	Quebrada	3500	6
16	Tilcara	Quebrada	3139	3
17	Mina El Aguilar	La Puna	4260	1
18	Mina Pirquitas	La Puna	4147	3
19	Cochinoca	La Puna	3552	1
20	Abra Pampa	La Puna	3484	1
21	Paicone	La Puna	3600	26
22	Pucará de Rinconada	La Puna	3870	1
23	Rinconada	La Puna	3953	1
24	Chocoite-Yavi	La Puna	3440	1
25	Mina Pan de Azúcar	La Puna	3646	1
26	Pumahuasi-Yavi	La Puna	3440	1
27	La Quiaca	La Puna	3442	6
28	Casira	La Puna	3572	1
29	Yavi	La Puna	3440	1
30	Susques	La Puna	3896	1
31	Coranzulí	Puna	4028	1
NA	Sin población conocida	Selva	1200	17

Tabla Anexo 4: Provincias de Ecuador junto con su capital. La clave vincula a las provincias con su localización en el mapa a través de la *Figura 11*.

Clave	Provincia	Capital
1.	Azuay	Cuenca
2.	Bolívar	Guaranda
3.	Cañar	Azogues
4.	Cachi	Tulcán
5.	Chimborazo	Riobamba
6.	Cotopaxi	Latacunga
7.	El Oro	Machala
8.	Esmeraldas	Esmeraldas
9.	Galápagos	Puerto Baquerizo Moreno
10.	Guayas	Guayaquil
11.	Imbabura	Ibarra
12.	Loja	Loja
13.	Los Rios	Babahoyo
14.	Manabí	Portoviejo
15.	Morona-Santiago	Macas
16.	Napo	Tena
17.	orellana	Puerto Francisco de Orellana
18.	Zamora Chinchipe	Zamora
19.	Pichincha	Quito
20.	Santa Elena	Santa Elena
21.	Santo Domingos Tsáchilas	Santo Domingo
22.	Sucumbios	Nueva Loja
23.	Tungurahua	Ambato
S/N.	Pastaza	Puyo

Tabla Anexo 5: Se muestran los reactivos, su composición y concentración utilizados en la reacción PCR en un volumen total de 10 μ l.

Reactivo	Componentes	Marca	Stock
PCR Buffer	1,5 ml/tubo: 500 mM KCl, 1,5 mM MgCl ₂ , 100 mM tris-HCl (pH 9.0)	GE Helathcare	10X
Cloruro de Magnesio	1,2 ml/tubo: 25 mM MgCl ₂	GE Helathcare	25 mM
dATPs	250 μ l/Tubo: 100 mM de C ₁₀ H ₁₂ N ₅ O ₁₂ P ₃ Na ₄ disuelto en Sal Sódica	GE Helathcare	100 mM
dCTPs	250 μ l/Tubo: 100 mM de C ₉ H ₁₂ N ₃ O ₁₃ P ₃ Na ₄ disuelto en Sal Sódica	GE Helathcare	100 mM
dGTPs	250 μ l/Tubo: 100 mM de C ₁₀ H ₁₂ N ₅ O ₁₃ P ₃ Na ₄ disuelto en Sal Sódica	GE Helathcare	100 mM
dTTPs	250 μ l/Tubo: 100 mM de C ₁₀ H ₁₃ N ₂ O ₁₄ P ₃ Na ₄ disuelto en Sal Sódica	GE Helathcare	100 mM
Glicerol	100 ml/tubo de C ₃ H ₈ O ₃	Sigma	99%
Taq pol	50 μ l/Tubo: 250 U Taq DNA polimerasa, 50 mM tris-HCl (pH 7,5), 5 mM DTT, 0,1mM EDTA, 50% glicerol	GE Helathcare	5 U/ μ l

Tabla Anexo 6: Se muestra información acerca de las reacciones PCR realizadas para analizar los marcadores genéticos. Las concentraciones finales empleadas en los diferentes protocolos PCR seguidos para realizar el análisis de los marcadores empleados. Todos los reactivos empleados tenían las concentraciones originales descritas en la *Tabla Anexo 5*. La cantidad de ADN empleado varía dependiendo de cada muestra, por lo que es imposible detallarlo con total precisión.

Marcador analizado	Alu A25	Alu ACE	Alu APO	Alu B65	Alu D1	Alu FXIIIB	Alu HF	Alu HG	Alu HJ
Reactivo									
Buffer	1X	1X	1X	1X	1,3X	1X	1X	1X	1X
MgCl ₂	2 mM	1,5 mM	1,5 mM	1,75 mM	2,25 mM	1,5 mM	2mM	1,5 mM	2,5mM
dNTPs	1,5 mM	1 mM	1,5 mM	1,2 mM	1,5 mM	1,4 mM	2 mM	2 mM	2 mM
Glicerol	5%	5%	5%	5%	5%	5%	5%	5%	5%
Primer F/Primer R	0,6 µM	0,45 µM	0,6 µM	0,5 µM	0,5 µM	0,6 µM	0,5µM	0,5µM	1µM
Taq pol	0,06 U/µl	0,05 U/µl	0,05 U/µl	0,05 U/µl	0,06 U/µl	0,045 U/µl	0,05 U/µl	0,05 U/µl	0,05 U/µl

(Cont.)

Marcador analizado	Alu HS3.23	Alu HS4.32	Alu HS4.59	Alu HS4.65	Alu MicB	Alu PV92	Alu TF	Alu TPA25
Reactivo								
Buffer	1X	1X	1X	1X	1X	1X	1X	1X
MgCl ₂	3 mM	2,5 mM	2,5 mM	1,5 mM	2,5mM	1,5 mM	2 mM	1,25 mM
dNTPs	2 mM	1,5 mM	1,5 mM	1 mM	2mM	1 mM	2 mM	1 mM
Glycerol	5%	5%	5%	5%	5%	5%	5%	5%
Primer F/Primer R	0,5 µM	1 µM	1 µM	1 µM	0,5µM	0,45 µM	1µM	0,5 µM
Taq pol	0,25 U/uL	0,1U/µl	0,06 U/µl	0,025 U/µl	0,05 U/µl	0,05 U/µl	0,05 U/µl	0,04 U/µl

(Cont.)

Marcador analizado	D6S105	D6S265	D6S2666	D6S2792
Reactivo				
Buffer	1X	1X	1X	1X
MgCl ₂	1 mM	1,5 mM	0,375 mM	1,5 mM
dNTPs	0,125 mM	0,25 mM	1 mM	0,2 mM
Glicerol	N.A.	N.A.	N.A.	N.A.
Primer F/Primer R	5 µM	1 µM	0,2 µM	0,025 µM
Taq pol	0,025 U/ul	0,075 U/ul	0,025 U/ul	0,025 U/ul
Fluorocromo Utilizado	6-FAM	6-FAM	HEX	6-FAM

Tabla Anexa 7: Información acerca de los fluorocromos empleados en el análisis de STRs mediante Analizador genético (ABI prism 310, Applied Biosystem).

Fluorocromo	Marca	Color	Marcador asignado
6-FAM	Applied Biosystem (Life)	Azul	D6S105
6-FAM	Applied Biosystem (Life)	Azul	D6S265
HEX	Applied Biosystem (Life)	Amarillo	D6S2666
6-FAM	Applied Biosystem (Life)	Azul	D6S2792

Tabla Anexa 8. Información acerca de los programas utilizados para llevar a cabo la reacción PCR de los diferentes protocolos para cada marcador. La temperatura de anillamiento viene detallada en la Tabla 3 junto con la secuencia de los primers y el tamaño de los amplificados.

Marcador analizado	Alu A25	Alu ACE	Alu APO	Alu B65	Alu D1	Alu FXIII B	Alu HF	Alu HG	Alu HJ
Programa PCR									
Desnaturalización Inicial (94°C)	10 min	10 min	10 min	10 min	10 min	10 min	10 min	10 min	10 min
Ciclos previos	Desnaturalización (94°C)	1 min	/	1 min	0,5 min	1 min	/	/	/
	Anneling	1,5 min (60,2°C)	/	1,5 min (60°C)	0,75 min (59,5°C)	1,5 min (60°C)	/	/	/
	Extension (72°C)	1,5 min	/	1,5 min	1,5 min	1,5 min	/	/	/
	Número de ciclos	7	/	7	7	7	/	/	/
Desnaturalización (94°C)	1 min	1 min	1 min	1 min	1 min	1 min	0,5 min	0,5 min	0,5 min
Anneling (Tª en Tabla 3)	1 min	1,5 min	1,5 min	1 min	1,5 min	1 min	0,75 min	0,75 min	0,75 min
Extension (72°C)	1 min	1,5 min	1,5 min	1 min	2 min	1 min	0,75 min	0,75 min	0,75 min
Número de ciclos	25	30	30	28	25	25	35	35	35
Extension final (72°C)	10 min	10 min	10 min	10 min	10 min	10 min	10 min	10 min	10 min

(Cont.)

Marcador analizado	Alu HS3.23	Alu HS4.32	Alu HS4.59	Alu HS4.65	Alu MicB	Alu PV92	Alu TF	Alu TPA25
Programa PCR								
Desnaturalización Inicial (94°C)	10 min	10 min	10 min	10 min	10 min	10 min	10 min	10 min
Desnaturalización (94°C)	/	/	/	/	/	/	0,5 min	1 min
Anneling	/	/	/	/	/	/	0,75 min (67°C)	1,5 min (65°C)
Extension (72°C)	/	/	/	/	/	/	0,75 min	1,5 min
Número de ciclos	/	/	/	/	/	/	10	7
Desnaturalización (94°C)	1 min	1 min	1 min	1 min	0,5 min	1 min	0,5 min	1 min
Anneling (Tª en <i>Tabla 3</i>)	1,5 min	1,5 min	1,5 min	1,5 min	0,75 min	1,5 min	0,75 min	1 min
Extension (72°C)	1,5 min	1,5 min	1,5 min	1,5 min	0,75 min	1,5 min	0,75 min	1 min
Número de ciclos	35	35	35	35	35	30	25	25
Extension final (72°C)	10 min	10 min	10 min	10 min	10 min	10 min	10 min	10 min

Ciclos previos

(Cont.)

Marcador analizado	D6S105	D6S265	D6S2666	D6S2792
Programa PCR				
Desnaturalización Inicial (94°C)	3 min	3 min	3 min	5 min
Desnaturalización (94°C)	/	/	/	/
Anneling	/	/	/	/
Extensión (72°C)	/	/	/	/
Número de ciclos	/	/	/	/
Desnaturalización (94°C)	1 min	1 min	0,5 min	0,5 min
Anneling (Tª en Tabla 4)	1 min	1 min	0,5 min	0,5 min
Extensión (72°C)	1 min	1 min	1 min	0,16 min
Número de ciclos	23	23	35	37
Extensión final (72°C)	10 min	10 min	5 min	8 min

Ciclos previos

SOFTWARE EMPLEADO

A lo largo de la Tesis Doctoral, se han realizado numerosos análisis estadísticos (sección 2.3). Para realizarlos además de usar los programas informáticos típicos, se ha necesitado hacer uso de diferentes programas y paquetes estadísticos específicos de Biología que realizan los cálculos para las pruebas más específicas.

A continuación se detallan dichos programas:

ADMIX.PAS: Es un software empleado para calcular el mestizaje mediante el método de mínimos cuadrados según (Long *et al.* 1991).

ADMIX95: Es un programa utilizado para calcular el mestizaje mediante el método de *Gene Identity* de Chakraborty (1985). Se basa en los programas Admix2 y Admix3, que fueron desarrollados por el propio Chakraborty.

ARLEQUIN: Es un paquete de métodos básicos y de test estadísticos para análisis de datos de genética de poblaciones. Fue desarrollado por Excoffier *et al.* 2005 y Excoffier y Lischer (2010). En el desarrollo de la Tesis Doctoral ha sido utilizado para llevar a cabo pruebas y test como: el Equilibrio H-W, el AMOVA, Tests de Diferenciación poblacional (pairwise F_{ST} , Exact test of population differentiation, Desequilibrio de Ligamiento entre loci, inferencia de haplotipos, test de neutralidad de Ewens-Watterson).

GEDIS: (*Genetic and geographic DIStance*) Es un software libre diseñado para el análisis de datos poblacionales a partir de frecuencias génicas o haplotípicas. Ha sido desarrollado por Peña *et al.* (2009). Durante los trabajos que conforman la Tesis Doctoral ha sido empleado para realizar el cálculo de las Clinas y obtener matrices de distancia genética de Harpending y Jenkins.

NETWORK v4.6 (fluxus-engineering.com): programa diseñado por Bandelt *et al.* (1999) empleado para el tratamiento y representación de las relaciones entre haplotipos o individuos mediante

construcciones gráficas. Se ha empleado para construir las redes filogenéticas, inferir los haplotipos ancestrales.

PAST: (*PA*laeontological *S*Tatistics) Es un software libre diseñado para el análisis de datos originalmente dirigidos a la paleontología, pero también es útil en otras disciplinas. Fue desarrollado por Hammer *et al.* (2001). En este caso, se ha utilizado para realizar los Análisis Discriminantes, el Test de Mantel, el gráfico *ternary-plot* y la prueba K-S.

PHYLIP: (*the PHY*logeny *I*nference *P*ackage) Es un paquete de programas libre que se utiliza básicamente para inferir filogenias (árboles evolutivos). Desarrollado por Felsenstein *et al.* (1989). Durante los trabajos de la Tesis Doctoral se utilizó para obtener las matrices de distancia F_{ST} de Reynolds, dendrogramas Neighbor-joining, Bootstrap.

POWERMARKER: Es un paquete de herramientas básicas para el tratamiento de los datos genéticos. En este caso se ha utilizado para el cálculo de frecuencias alélicas, así como para los diferentes parámetros que se han obtenido de ellas: Heterocigosidad (observada y esperada), error estándar, PIC,...

SPSS: (*S*tatistical *P*ackage for the *S*ocial *S*ciences) Es un programa estadístico muy utilizado, desarrollado en principio para las ciencias sociales pero también en otras disciplinas por la capacidad para gestionar grandes bases de datos. En esta Tesis Doctoral se utilizó para realizar el cálculo del mestizaje mediante el método de Long *et al.* (1991), para la representación del MDS en dos dimensiones, así como para la obtención de sus coeficientes de estrés y de RSQ.

STR: Software desarrollado por José Ángel Peña., Es un paquete básico para el cálculo de frecuencias alélicas, así como para algunos parámetros que se han obtenido de ellas.

TREEVIEW: Es un programa para mostrar en forma de árbol las filogenias obtenidas mediante otros software a partir de matrices de distancia genética. Por lo tanto el uso que se le ha dado ha sido

construir los árboles filogenéticos obtenidos a partir de los cálculos realizados con Phylip, principalmente.

PUBLICACIONES RELACIONADAS

Se añaden aquí dos publicaciones en las que ha participado el autor de esta memoria de Tesis Doctoral, que han sido citadas en diferentes apartados de la misma. En ellas se presentan los análisis realizados sobre ADNmt de la población de Jujuy primero y de sus diferentes subpoblaciones después. En ellas se confirma que hay un origen nativo americano mayoritario en los pobladores de Jujuy para el linaje femenino. Lo que junto a los resultados expuesto en esta memoria (*Manuscrito III*) confirma que Jujuy es una población con un altísimo componente amerindio en su acervo genético.

En estas publicaciones se detalla la presencia de los haplogrupos sudamericanos principales (A2, B, C1 y D), destacando la elevada presencia del haplogrupo B (0,59) y en especial el B4b, con una frecuencia del 39% en la región del Alto Jujuy.

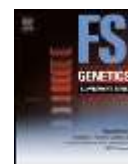
Las frecuencias de los diferentes haplogrupos varían ligeramente para cada una de las diferentes regiones; sin embargo el mayoritario en todas ellas es el haplogrupo B. Excepcionalmente, en la región de Valle se ha identificado un individuo con un haplogrupo de origen europeo (T) y dos con uno de origen africano (L). Este resultado apoya las hipótesis de un cierto mestizaje en las regiones más bajas de la provincia con individuos de origen africano y europeo durante la colonización del continente, mientras que en las regiones más altas el mestizaje es escaso o nulo. En estos trabajos, realizados sobre mtDNA, se muestra el panorama para los linajes femeninos, los cuales, lógicamente, no deberían haber experimentado una mayor tasa de mestizaje debido al fenómeno del mestizaje diferencial.

Asimismo, se ha detectado heterogeneidad genética entre las poblaciones de Valle y Puna, apoyando la hipótesis descrita en esta Tesis Doctoral acerca de la microdiferenciación genética dentro de Jujuy y los procesos microevolutivos implicados.



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS

Research article

Mitochondrial DNA control region in native population from the province of Jujuy (northwestern Argentina)

Sergio Cardoso^a, Maite Merino^a, Luis Gómez^b, Miguel A. Alfonso-Sánchez^a, José E. Dipierri^c, Emma Alfaro^c, Laura Valverde^a, Adrian Odriozola^a, Claudio M. Bravi^d, José A. Peña^b, Marian M. de Pancorbo^{a,*}

^aBIOMICS Research Group, Dpto. de Z. y Biología Celular A., Facultad de Farmacia, Universidad del País Vasco UPV/EHU, Paseo de la Universidad 7, 01006, Vitoria-Gasteiz, Spain

^bDpto. Genética, Antropología Física y Fisiología Animal, Facultad de Ciencia y Tecnología, Universidad del País Vasco, Leioa, Spain

^cInstituto de Biología de la Altura, Universidad Nacional de Jujuy, Avda. Bolívar 1661, (4600) San Salvador de Jujuy, Argentina

^dLaboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (IMBICE), CCT- CONICET-La Plata, Argentina

ARTICLE INFO

Article history:
Received 25 August 2009
Accepted 27 August 2009

Keywords:
Mitochondrial DNA
Control region
Haplotype
Native American
Argentina

ABSTRACT

The province of Jujuy is located in northwestern Argentina. We analyzed mitochondrial DNA (mtDNA) haplogroup composition in 100 autochthonous individuals living at altitudes between 1200 and 3500 m above sea level, in San Salvador de Jujuy, Quebrada de Humahuaca and Puna. The purpose was to explore the influence of non-Native American maternal lineages. Additionally, we evaluated the efficacy of analyzing the entire mtDNA control region to increase its power of discrimination with forensic purposes. Results showed that the population sample was entirely composed by Native American haplogroups, with haplotypes belonging to haplogroup B as the most common lineages. Analysis of the entire mtDNA D-loop region proved to be useful to increase the power of discrimination provided by the analysis of HVS-I and HVS-II segments and to refine haplogroup assignment.

© 2009 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Jujuy is a northwestern province from Argentina. It is divided in three different geographical areas, depending on altitude. These three areas are well differentiated by environmental conditions: the Valleys (1200 m.a.s.l.), which include the capital city of the province (San Salvador de Jujuy), the Quebrada de Humahuaca (2000–4000 m.a.s.l.) and the Puna region (>3800 m.a.s.l.). Several studies on American populations demonstrate that human migratory movements occurred since colonial times have generated an admixture pattern characterized by predominance of the ancestral Native American substrate in the maternal lineages and a high contribution of European-derived populations to the paternal lineages [1–3]. This phenomenon would have been caused by a strong sexual bias in the admixture processes occurred in Latin America during colonial times. Our study was aimed at determining mitochondrial DNA (mtDNA) composition of the Jujuy population to ascertain the proportion of non-native maternal lineages, as well as investigating the suitability of analyzing the entire mtDNA control region to increase the power of discrimination of this locus for forensic applications.

2. Materials and methods

Peripheral blood was taken from 100 healthy, maternally unrelated individuals from the province of Jujuy (northwestern Argentina): 22 from San Salvador de Jujuy, 40 from Quebrada de Humahuaca and 38 from the Puna region. All donors gave their informed consent prior to inclusion in the sample. DNA extraction was performed using the standard phenol:chloroform procedure [4]. Analysis of the entire mtDNA control region was carried out as described elsewhere [5].

3. Results and discussion

Sequence diversity was first calculated for each hypervariable segment separately, and then for the HVS-I–HVS-II and HVS-I–HVS-II–HVS-III complexes and the entire control region (Table 1). The sequence diversity observed for HVS-I–HVS-II haplotypes (0.9905 ± 0.0033) improved when we considered HVS-III (0.9922 ± 0.0030), and this value was even slightly higher when the entire control region was included in the calculation (0.9924 ± 0.0030), thereby resulting in higher power of discrimination. Concerning random match probability, we observed that analysis of the entire mtDNA control region constitutes a better approach for forensic applications when compared to the more largely used HVS-I and HVS-II regions.

* Corresponding author. Tel.: +34 945013045; fax: +34 945013014.
E-mail address: marianpancorbo@gmail.com (M.M. de Pancorbo).

Table 1
Diversity and forensic parameters in the population from the province of Jujuy (northwestern Argentina).

	HVS-I	HVS-II	HVS-III	HVS-I+HVS-II	HVS-I+HVS-II+HVS-III	Control region
Polymorphic sites	61	53	20	114	134	161
Haplotypes	50	53	17	72	75	76
Sequence diversity	0.9615 ± 0.0092	0.9835 ± 0.0048	0.7726 ± 0.0343	0.9905 ± 0.0033	0.9922 ± 0.0030	0.9924 ± 0.0030
Random match probability	4.83%	2.64%	23.52%	1.95%	1.79%	1.77%

Table 2
Haplogroup distribution derived from entire mtDNA control region haplotypes in the population sample from the province of Jujuy. Frequencies are shown as percentage.

Haplogroup	Frequency (%)
A2	16.0
B	59.0
B2	10.0
B2a	1.0
B2c	9.0
B4b	39.0
C	16.0
C1	1.0
C1b	15.0
D	9.0
D1	7.0
D4h3	2.0

As for haplogroup composition, all the maternal lineages identified in our sample were Native American. Most of the haplotypes belonged to haplogroup B (59.0%; Table 2). Analysis of the mtDNA control region permitted, in some instances, to confirm (np 489 for haplogroup B) or even refine (np 499 for haplogroup C1b) haplogroup assignment. Haplogroup distribution in the province of Jujuy was similar to other neighboring Native American populations, such as Quechua from Bolivia and Atacama from Chile [6]. It can be postulated that the relative isolation of the area under study may have contributed to the preservation of the Amerindian mtDNA pool of the Jujuy

population by preventing admixture with maternal lineages introduced by human migrations.

Conflict of interest

None.

Acknowledgements

Funds granted by the University of the Basque Country (GIU 05/51) and the Basque Government (IT-424).

References

- [1] S. Wang, N. Ray, W. Rojas, M.V. Parra, G. Bedoya, C. Gallo, G. Poletti, G. Mazzotti, K. Hill, A.M. Hurtado, B. Camrena, H. Nicolini, W. Klitz, R. Barrantes, J.A. Molina, N.B. Freimer, M.C. Bortolini, F.M. Salzano, M.L. Petzl-Erler, L.T. Tsuneto, J.E. Dipierri, E.L. Alfaro, G. Bailliet, N.O. Bianchi, E. Llop, F. Rothhammer, L. Excoffier, A. Ruiz-Linares, Geographic patterns of genome admixture in Latin American Mestizos, *PLoS Genet.* 4 (2008) 3.
- [2] M.A. Acosta, A. Blanco-Verea, M.V. Lareu, M. Brion, A. Carracedo, The genetic male component of two South-Western Colombian populations, *Forensic Sci. Int. Genet.* 3 (2009) e59–e61.
- [3] A. Salas, J.C. Jaime, V. Alvarez-Iglesias, A. Carracedo, Gender bias in the multiethnic genetic composition of central Argentina, *J. Hum. Genet.* 53 (2008) 662–674.
- [4] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989.
- [5] S. Cardoso, M.T. Zarrabeitia, L. Valverde, A. Odriozola, M.A. Alfonso-Sánchez, M.M. de Pancorbo, Variability of the entire mitochondrial DNA control region in a human isolate from the Pas Valley (northern Spain), *J. Forensic Sci.* 55 (5) (2010).
- [6] C.M. Lewis Jr., R.Y. Tito, B. Lizarraga, A.C. Stone, Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru, *Am. J. Phys. Anthropol.* 127 (2004) 351–360.



Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Forensic Population Genetics – Short Communication

Mitochondrial DNA control region data reveal high prevalence of Native American lineages in Jujuy province, NW Argentina

Sergio Cardoso^a, Leire Palencia-Madrid^a, Laura Valverde^a, Miguel A. Alfonso-Sánchez^b, Luis Gómez-Pérez^b, Emma Alfaro^c, Claudio M. Bravi^d, José E. Dipierri^c, José A. Peña^b, Marian M. de Pancorbo^{a,*}

^a BIOMICS Research Group, Centro de Investigación "Lascaray" Ikerkunea, Universidad del País Vasco (UPV/EHU), Vitoria-Gasteiz, Spain

^b Dpto. Genética, Antropología Física y Fisiología Animal, Facultad de Ciencia y Tecnología, Universidad del País Vasco (UPV/EHU), Leioa, Spain

^c Instituto de Biología de la Altura, Universidad Nacional de Jujuy, Avda. Bolivia 1661, 4600 San Salvador de Jujuy, Argentina

^d Laboratorio de Genética Molecular Poblacional, Instituto Multidisciplinario de Biología Celular (IMBICE), CCT-CONICET-La Plata, Argentina

ARTICLE INFO

Article history:

Received 15 September 2012

Received in revised form 29 December 2012

Accepted 29 January 2013

Keywords:

Mitochondrial DNA

Control region

Argentina

Jujuy

EMPOP

ABSTRACT

Mitochondrial control region (16024–576) sequences were generated from 180 individuals of four population nuclei from the province of Jujuy (NW Argentina), located at different altitudes above sea level. The frequency at which a randomly selected mtDNA profile would be expected to occur in the general population (random match probability) was estimated at 0.011, indicating a relatively high diversity. Analysis of the haplogroup distribution revealed that Native American lineages A2 (13.9%), B (56.7%), C1 (17.8%), D1 (8.9%) and D4h3a (1.1%) accounted for more than 98% of the total mtDNA haplogroup diversity in the sample examined. We detected a certain degree of genetic heterogeneity between two subpopulations located at different points along the altitudinal gradient (Valles and Puna), suggesting that altitude above sea level cannot be ruled out as a factor promoting divergences in mtDNA haplogroup frequencies, since altitude is closely associated with human living conditions, and consequently, with low demographic sizes and the occurrence of genetic drift processes in human communities. In all, mitochondrial DNA database obtained for Jujuy province strongly points to the need for creating local mtDNA databases, to avoid bias in forensic estimations caused by genetic substructuring of the populations.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Population

The entire control region of the mitochondrial DNA (mtDNA) remains virtually unexplored in many human populations from South America. This obviously hinders the construction of high-quality mtDNA databases, which in turn amplifies the limitations inherent to this molecular marker for forensic applications. The lacking of extensive mtDNA databases might be especially problematic in large countries such as Argentina, characterized by the presence of many different human groups with disparate levels of genetic diversity. A recent study on the mtDNA diversity in Argentina revealed that the country harbors a comprehensive admixture of worldwide mtDNA lineages [1]. Some of these lineages

show regional specificity, thus conditioning the construction of a nationwide mtDNA database for forensic and anthropological purposes. In this regard, Catelli et al. [2] demonstrated the existence of a clear-cut genetic substructuring in Argentina and the need for in-depth analyses of the different regions of the country to gradually disentangle its stratification. Here we present the first database of the complete mtDNA control region in a sample from the province of Jujuy (northwestern Argentina). Jujuy is characterized by remarkable differences in altitude along the different geographic regions of the province (see Fig. 1), all of them well differentiated by their typical environmental conditions: Selva or Yungas, located between 500 and 1200 m above sea level (m.a.s.l.), Valles (~1200 m.a.s.l.), which include the capital city of the province (San Salvador de Jujuy), Quebrada de Humahuaca (2000–3500 m.a.s.l.), and Puna (>3500 m.a.s.l.). Because of historico-geographical reasons, Native American individuals are proportionally more numerous in the highest regions of Jujuy, where even at present the Amerindian genetic component is highly predominant [3]. Furthermore, the region of Jujuy has been traditionally characterized by markedly low population densities over time notably in those areas of more difficult access (highlands),

* Corresponding author at: BIOMICS Research Group, Centro de Investigación "Lascaray" Ikerkunea, Universidad del País Vasco (UPV/EHU), Avda. Miguel de Unamuno 3, 01006 Vitoria-Gasteiz, Spain. Tel.: +34 945014528; fax: +34 945014458.

E-mail address: marian.mdepancorbo@ehu.es (M.M. de Pancorbo).



Fig. 1. Map showing the location of the province of Jujuy in northwestern Argentina. The subdivision of Jujuy in different regions (Selva, Valles, Quebrada and Puna) is also displayed.

which might have promoted genetic drift events, and therefore, genetic microdifferentiation processes between the diverse population nuclei of the zone.

2. Subjects, materials and methods

Peripheral blood was taken from 180 healthy and maternally unrelated individuals from the province of Jujuy (northwestern Argentina). Autochthony was verified by biographical information traced back at least three generations (mainly birthplaces of the donor's parents and grandparents). The total sample was divided into four different subgroups according to the sampling region: 59 from San Salvador de Jujuy (Valles), 41 from Selva, 41 from Quebrada de Humahuaca, and 39 from Puna. Samples from Valles and Selva were collected in urban and semi-urban areas. In these communities, admixture rates (~23%) of the Native American genetic background estimated from autosomal markers (polymorphic *Alu* insertions) indicated a major contribution of European ancestry populations, and a much more weak African component (Valle: 6.5%). Sampling locations in Quebrada de Humahuaca were rural communities with admixture rates between 8 and 13%, but in this case with a predominantly African contribution. Puna donors lived in rural, isolated localities with virtually no signals of gene flow in autosomal *Alu* elements [4]. All donors gave their informed consent prior to inclusion in the sample. DNA extraction was performed using the standard phenol:chloroform procedure [5]. DNA was quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA) in a DTX880 Multimode Detector (Beckman Coulter, Fullerton, CA). Full mitochondrial control region (CR) was amplified, sequenced and interpreted as reported in Cardoso et al. [6] using rCRS as reference sequence, instead of the recently released RSRS, as proposed by Salas et al. [7] for forensic DNA studies and casework.

3. Analysis of data

Diversity parameters and haplotype frequencies were calculated using the Arlequin Software v3.5 [8]. The probability of two randomly selected individuals from a population having identical mtDNA types was calculated as $p = \sum X_i^2$. The length polymorphisms of the poly-C stretches at HVS-I, HVS-II and HVS-III were disregarded in the analyses.

The mtDNA haplogroup assignment was performed according to the mtDNA tree Build 15 of the Phylotree [9]. Statistical differences in haplogroup distribution among subpopulations

were calculated using PAST software [10]. The conservative Bonferroni correction for multiple testing was used when necessary [11].

All the sequences have been deposited into EMPPOP [12] under accession number EMP00512. Likewise, sequences are available online at GenBank under accession numbers JQ736825–JQ737004.

4. Results and discussion

Summary statistics are shown in Table 1, whereas haplotypes are listed in Supplementary Table S1. A total of 188 polymorphic sites were identified, which defined 127 haplotypes. Of them, 96 were unique. This sequence variability rendered a high gene diversity value (0.9945 ± 0.0014) considering the entire control region. This figure is clearly higher than the diversity values estimated in other South American populations where the Native American maternal component is prevalent, such as Kichwa (0.8029 ± 0.0373) from Ecuador [13] and Mapuche (0.9725 ± 0.0107) from Argentina (calculated from de Saint Pierre et al. [14]). Interestingly, the sequence variability observed in Jujuy was comparable to the high values obtained for other South American populations with a long history of admixture, such as Santa Catarina in Brazil (0.9930) [15] or North Central Venezuela (0.9939) [16]. The great diversity of the maternal lineages was also reflected in the probability of a random match (RMP), with a value of 0.011.

Disregarding length variations in the poly-C stretches, the most frequent haplotypes in Jujuy, with 5 occurrences (2.8%) each, were (i) 263G-315.1C-499A-16183C-16188T-16189C-16217C-16519C, (ii) 73G-146C-249DEL-263G-290DEL-291DEL-315.1C-489C-493G-523DEL-524DEL-16223T-16298C-16325C-16327T-16519C, (iii) 63C-64T-66A-73G-185A-263G-315.1C-499A-16183C-16188T-16189C-16217C-16519C, and (iv) 46G-59A-61T-62T-73G-195C-249DEL-263G-290DEL-291DEL-315.1C-489C-493G-523DEL-524DEL-16092C-16223T-16249C-16298C-16325C-16327T-16400T-16463G-16519C. Haplotypes (i), (ii) and (iii) were identified in different Jujuy subregions; however, the distribution of haplotype (iv) was exclusively restricted to Puna, suggesting genetic drift effects in the area most likely promoted by prolonged population isolation, and accordingly, a traditional small community size.

Phylogeographic comparisons of the haplotypes identified in our study with those of other South American groups were hindered by the lack of high quality, complete control region sequences of relevant populations. Notwithstanding these limitations, we used a clade-defining approach to search for (near)-matches in an in-house database of published sequences of Native American origin derived from both indigenous and admixed South American populations.

Overall, most of the (near)-matches available for Jujuy samples pointed to a shared biological (maternal) history of southern Central Andean populations (Supplementary Table S2). The most frequent clade in Jujuy ($N = 28$) was the B2 haplotype carrying 16188T. Even though this clade has been previously described in different South American populations, the more

Table 1
Summary statistics for 180 control region (16024–576) sequences in a sample from Jujuy province (NW Argentina). Insertions at nucleotide positions 16193, 309, and 573 were ignored for all calculations.

Population statistics	Jujuy (N=180)
Random match probability (sum of squares)	0.011
Haplotypes	127 (31 shared)
Polymorphic positions	188
Mean pairwise differences	14.7932 ± 6.6503
Genetic diversity	0.9945 ± 0.0014

Table 2
Mitochondrial DNA haplogroup distribution in four subpopulations from Jujuy province (NW Argentina). Relative frequencies (as percentages) in parentheses.

	N	Haplogroup and frequency					P value ^a			
		A	B	C	D	L	T	Puna	Quebrada	Valles
Puna	39	8 (20.5)	17 (43.6)	8 (20.5)	6 (15.4)	–	–	–	–	–
Quebrada	41	7 (17.1)	25 (61.0)	7 (17.1)	2 (4.8)	–	–	0.3739	–	–
Valles	59	3 (5.1)	39 (66.1)	9 (15.3)	5 (8.5)	2 (3.4)	1 (1.7)	0.0396	0.4096	–
Selva	41	7 (17.1)	21 (51.2)	8 (19.5)	5 (12.2)	–	–	0.9147	0.7907	0.3025

^a Significance values of likelihood ratio tests (G-tests) to compare haplogroup frequencies between subpopulations. Significance *p* value 0.0125 after Bonferroni correction.

relevant frequencies appear in the southern Central Andes, peaking at 22–29% in North Chile, Bolivia and Peru [14,17,18].

Matches with haplotypes derived from some aDNA studies were also observed. In this regard, five ancient individuals from Nasca-Paracas in coastal southern Peru [19] showed a common pronounced motif with three mtDNA sequences from Puna.

Some lineages, such as those attributable to D1j and B2i2b show affinities with Central-Western Argentina and Patagonia most likely determined by gene flow processes. On the other hand, available mtDNA data clearly indicated the seclusion of two clusters in Jujuy region. One of them was a C1b clade, the second most frequent in our sample and signposted by the co-occurrence of 16092C-16400T, a combination with high incidence in Puna and, to the best of our knowledge, not yet reported. The other is a B2 clade with the motif 16145A-16156A-16157C present in five individuals, four of them from Quebrada.

As for haplogroup composition, the population from Jujuy showed a distribution almost entirely composed of Native American lineages (98.3%). Only three non-American haplogroups were observed: two African paragroup L sequences and a T2c lineage of European ancestry. These three exotic lineages were identified in Valles, the region of Jujuy showing the highest level of genetic admixture according to previous studies based on Y-chromosome markers [20] or on polymorphic *Alu* insertions [4]. These findings suggest that, despite the population admixture processes occurred in Jujuy since the 16th century between the indigenous population, the European settlers and the African slaves, there is a high predominance of the ancestral Native American substrate in the maternal lineages, probably mirroring the strong sexual bias of the immigrants settled in this sub-Andean zone from Argentina during colonial times. Asymmetry in the ancestry composition of maternal and paternal lineages has been reported in previous genetic studies carried out in other Ibero-American human groups from Brazil [21,22] and from other ancient Spanish colonies [23–25]. In Jujuy, as expected, the weak effect of the gene flow was detectable only in samples from the lowest geographic regions, since immigrants concentrated mainly in these zones [26].

With regard to the Native American lineages, we identified haplogroups A2 (13.9%), B4b (56.7%), C1 (17.8%), D1 (8.9%) and D4h3a (1.1%) in the study population. The prevalence of haplogroup B is commonplace among indigenous human groups from Central Andean South America, as corroborated by findings in other Amerindian populations such as Aymara (81%) and Quechua (61%) from Bolivia [27], Atacama (72%) from Chile [28], individuals from the Department of Ancash (52%), Peru [29], or Peruvian Amerindians from the Lake Titicaca basin (69%) [17], among others. Results of a likelihood-ratio test (G-test) indicated no statistically significant differences between the haplogroup distribution of the four Jujuy subpopulations considered (Table 2), at least at the macrohaplogroup level (A, B, C and D). Nevertheless, such results were indicative of a certain degree of genetic heterogeneity between two subpopulations sited at different points of the altitude range (Valles and Puna). As can be noted, the comparison

between the haplogroup frequencies of these two subgroups rendered a significance value ($P=0.0396$) very close to the threshold value obtained after applying the Bonferroni correction for multiple testing (0.0125), thus suggesting that differences in altitude cannot be ruled out as a potential factor in determining divergences in mtDNA haplogroup frequencies. In all probability, statistically significant differences in haplogroup distribution among Jujuy subpopulations might also be detected at a more refined level of the mtDNA phylogeny (subhaplogroups), or perhaps by increasing the sample size of the communities examined.

The existence of genetic heterogeneity even at a microgeographic scale strongly supports the creation of local databases such as the one presented herein, to avoid underestimation or overestimation of a given haplotype in routine forensic casework. These local databases can also be helpful in biomedical investigations, including case-control studies.

The findings of our study together with the information on mtDNA reported so far for Argentina both substantiate the notion that, in countries with quite isolated populations distributed over large territories, the creation of local mtDNA databases is essential.

5. Quality control

All data generated in this study are in compliance with ISFG and EDNAP guidelines for mitochondrial DNA analysis [30,31]. This publication follows ISFG guidelines for the publication of population genetic data [32].

Conflict of interest

Authors declare no competing interest in the content of this manuscript.

Acknowledgements

This project was supported by the Government of the Basque Country (Grupo Consolidado IT-424-07) and a grant from the University of the Basque Country UPV/EHU (ELDUNANOTEK UFI11/32). Technical and human support provided by SGiker (UPV/EHU, MICINN, GV/EJ, ESF) is gratefully acknowledged. We are also grateful to the EMPPOP team and especially to Prof. Walther Parson and Alexander W. Röck for assistance with database confirmation. Authors are deeply indebted to all voluntary donors who agreed to participate in this study. LP is supported by a grant from the Basque Government (Dpto. de Educación, Universidades e Investigación).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fsigen.2013.01.007>.

References

- [1] M.C. Bobillo, B. Zimmermann, A. Sala, G. Huber, A. Röck, H.J. Bandelt, D. Corach, W. Parson, Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database, *Int. J. Legal Med.* 124 (2010) 263–268.
- [2] M.L. Catelli, V. Alvarez-Iglesias, A. Gómez-Carballa, A. Mosquera-Miguel, C. Romanini, A. Borosky, J. Amigo, A. Carracedo, C. Vullo, A. Salas, The impact of modern migrations on present-day multi-ethnic Argentina as recorded on the mitochondrial DNA genome, *BMC Genet.* 12 (2011) 77.
- [3] J.E. Dipierrí, E. Alfaro, J.A. Peña, J. Constans, J.M. Dugoujon, G.M. KM immunoglobulin allotypes and other serum genetic markers (HP, GC, PI, and TF) among South American populations living at different altitudes (Jujuy Province, Argentina): admixture estimates, *Hum. Biol.* 72 (2000) 305–319.
- [4] L. Gómez-Pérez, M.A. Alfonso-Sánchez, J.E. Dipierrí, E. Alfaro, S. García-Obregón, M.M. de Pancorbo, G. Bailliet, J.A. Peña, Microevolutionary processes due to landscape features in the province of Jujuy (Argentina), *Am. J. Hum. Biol.* 23 (2011) 177–184.
- [5] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989.
- [6] S. Cardoso, M.J. Villanueva-Millán, L. Valverde, A. Odriozola, J.M. Aznar, S. Piñeiro-Hermida, M.M. de Pancorbo, Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country, *Forensic Sci. Int. Genet.* 6 (2012) e106–e108.
- [7] A. Salas, M. Coble, S. Desmyter, T. Grzybowski, L. Gusmão, C. Hohoff, M.M. Holland, J.A. Irwin, T. Kupiec, H.Y. Lee, B. Ludes, S. Lutz-Bonengel, T. Melton, T.J. Parsons, H. Pfeiffer, L. Prieto, A. Tagliabracci, W. Parson, A cautionary note on switching mitochondrial DNA reference sequences in forensic genetics, *Forensic Sci. Int. Genet.* 6 (2012) e182–e184.
- [8] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [9] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 30 (2) (2009) E386–E394.
- [10] Ø. Hammer, D.A.T. Harper, P.D. Ryan, PAST: paleontological statistics software package for educational and data analysis, *Paleontol. Electron.* 4 (2001) 9.
- [11] J.M. Bland, D.G. Altman, Statistics notes: multiple significance tests: the Bonferroni method, *BMJ* 310 (1995) 170.
- [12] W. Parson, A. Dür, EMPOP – a forensic mtDNA database, *Forensic Sci. Int. Genet.* 1 (2007) 88–92.
- [13] M. Baeta, C. Núñez, C. Sosa, M. Bolea, Y. Casalod, F. González-Andrade, L. Roewer, B. Martínez-Jarreta, Mitochondrial diversity in Amerindian Kichwa and Mestizo populations from Ecuador, *Int. J. Legal Med.* 126 (2012) 299–302.
- [14] M. de Saint Pierre, C.M. Bravi, J.M. Motti, N. Fuku, M. Tanaka, E. Llop, S.L. Bonatto, M. Moraga, An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups, *PLoS One* 7 (2012) e43486.
- [15] L. Palencia, L. Valverde, A. Alvarez, L.M. Cainé, S. Cardoso, M.A. Alfonso-Sánchez, M.F. Pinheiro, M.M. de Pancorbo, Mitochondrial DNA diversity in a population from Santa Catarina (Brazil): predominance of the European input, *Int. J. Legal Med.* 124 (2010) 331–336.
- [16] D. Castro de Guerra, C. Figueroa Pérez, C.M. Bravi, J. Saunier, M. Scheible, J. Irwin, M.D. Coble, A. Rodríguez-Laralde, Sequence variation of mitochondrial DNA control region in North Central Venezuela, *Forensic Sci. Int. Genet.* 6 (2012) e131–e133.
- [17] C. Barbieri, P. Heggarty, L. Castrí, D. Luiselli, D. Pettener, Mitochondrial DNA variability in the Titiaca basin: matches and mismatches with linguistics and ethnohistory, *Am. J. Hum. Biol.* 23 (2011) 89–99.
- [18] H.H. Afonso-Costa, M. Carvalho, V. Lopes, F. Balsa, A.M. Bento, A. Serra, L. Andrade, M.J. Anjos, M.C. Vide, S. Pantoja, D.N. Vieira, F. Corte-Real, Mitochondrial DNA sequence analysis of a native Bolivian population, *J. Forensic Leg. Med.* 17 (2010) 247–253.
- [19] L. Fehren-Schmitz, M. Reindel, E.T. Cagigao, S. Hummel, B. Herrmann, Pre-Columbian population dynamics in coastal southern Peru: a diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis, *Am. J. Phys. Anthropol.* 141 (2010) 208–221.
- [20] J.E. Dipierrí, E. Alfaro, V.L. Martínez-Marignac, G. Bailliet, C.M. Bravi, S. Cejas, N.O. Bianchi, Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina, *Hum. Biol.* 70 (1998) 1001–1010.
- [21] A.K. Ribeiro-dos-Santos, J.M. Pereira, M.R. Lobato, B.M. Carvalho, J.F. Guerreiro, S.E. Batista Dos Santos, Dissimilarities in the process of formation of Curiaú, a semi-isolated Afro-Brazilian population of the Amazon region, *Am. J. Hum. Biol.* 14 (2002) 440–447.
- [22] M.C. Bortolini, W.A. Da Silva Jr., D.C. De Guerra, G. Remonato, R. Mirandola, M.H. Hutz, T.A. Weimer, M.C. Silva, M.A. Zago, F.M. Salzano, African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers, *Am. J. Hum. Biol.* 11 (1999) 551–563.
- [23] I. Mendizabal, K. Sandoval, G. Berniell-Lee, F. Calafell, A. Salas, A. Martínez-Fuentes, D. Comas, Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba, *BMC Evol. Biol.* 8 (2008) 213.
- [24] L.G. Carvajal-Carmona, I.D. Soto, N. Pineda, D. Ortiz-Barrios, C. Duque, J. Ospina-Duque, M. McCarthy, P. Montoya, V.M. Alvarez, G. Bedoya, A.A. Ruiz-Linares, Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia, *Am. J. Hum. Genet.* 67 (2000) 1287–1295.
- [25] G. Bedoya, P. Montoya, J. García, I. Soto, S. Bourgeois, L. Carvajal, D. Labuda, V. Alvarez, J. Ospina, P.W. Hedrick, A. Ruiz-Linares, Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 7234–7239.
- [26] C. Martínez-Sarasola, Nuestros paisanos los indios, EMECE, Buenos Aires, 1992.
- [27] M. Gayà-Vidal, P. Moral, N. Saenz-Ruales, P. Gerbault, L. Tonasso, M. Villena, R. Vasquez, C.M. Bravi, J.M. Dugoujon, mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: different stories and special genetic traits of the Andean Altiplano populations, *Am. J. Phys. Anthropol.* 145 (2011) 215–230.
- [28] D.A. Merrifether, F. Rothhammer, R.E. Ferrell, Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World, *Am. J. Phys. Anthropol.* 98 (1995) 411–430.
- [29] C.M. Lewis Jr., R.Y. Tito, B. Lizárraga, A.C. Stone, Land, language, and loci: mtDNA in Native Americans and the genetic history of Peru, *Am. J. Phys. Anthropol.* 127 (2005) 351–360.
- [30] A. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tully, M. Wilson, DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing, *Forensic Sci. Int.* 110 (2000) 79–85.
- [31] G. Tully, W. Bär, B. Brinkmann, A. Carracedo, P. Gill, N. Morling, W. Parson, P. Schneider, Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles, *Forensic Sci. Int.* 124 (2001) 83–91.
- [32] A. Carracedo, J. Butler, L. Gusmão, W. Parson, L. Roewer, P. Schneider, Publication of population data for forensic purposes, *Forensic Sci. Int. Genet.* 4 (2010) 145–147.

