

UCH-FC
MAG-B
P 521
C 1



**“Determinación del set mínimo de SNPs necesarios para la
caracterización del máximo de haplogrupos de ADN mitocondrial y
Cromosoma Y en poblaciones mestizas Latinoamericanas”**

Tesis

Entregada A La

Universidad De Chile

En Cumplimiento Parcial De Los Requisitos

Para Optar Al Grado De

Magíster en Ciencias Biológicas

Facultad de Ciencias

Por

Patricio Alejandro Pezo Valderrama

Agosto, 2014

Director de Tesis:

Dr. Mauricio Moraga

**FACULTAD DE CIENCIAS
UNIVERSIDAD DE CHILE
INFORME DE APROBACION
TESIS DE MAGÍSTER**

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Magíster presentada por el candidato.

Patricio Alejandro Pezo Valderama

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Magíster en Ciencias Biológicas, en el examen de Defensa Privada de Tesis rendido el día 31 de Julio, 2014.

Director de Tesis:

Dr.

Mauricio Moraga



Co-Director de Tesis

Dr.

.....

Comisión de Evaluación de la Tesis

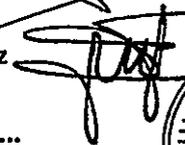
Dr.

Elie Poulin



Dr.

Marco Méndez



Dr.

.....

Dr.

.....



Dedicatorias

A Dios por otorgarme una vida de logros, oportunidades que nunca imaginé, de rodearme de personas únicas y valiosas, y por sobretodo, hacerme entender que los fracasos son oportunidades de crecimiento y superación.

A mis padres por haberme brindado valiosos años de su juventud en invertir tanto monetaria como psicológicamente en mi formación educativa y por sobretodo valórica. A mis hermanos por ser ese condimento especial en mi vida y ser motivo de orgullo para mí.

A mis compañeros del laboratorio (Adriana, Cony, Pame, Michael, Mario, Jacqueline, Paloma) por ayudarme en varias oportunidades y por crear un ambiente de trabajo del mejor nivel, y por su apoyo, soporte y esos excelentes momentos de conversaciones innecesarias y sin sentido que irónicamente son muy necesarias para el a veces estresante día a día.

Al profesor Mauricio Moraga por su apoyo (más monetario que emocional) y soportar mi presencia durante tres años, mis insistencias, bromitas, y más que otras salidas de contexto.

Quiero terminar dedicando estas últimas palabras a alguien demasiado importante en mi vida, alguien que llena mi corazón días tras día, alguien que se hizo tan parte de mí que no imagino un momento sin ella, alguien que es amiga, confidente y por sobretodo una hermosa mujer. Te amo Sandra.

“El éxito no se logra sólo con cualidades especiales. Es por sobretodo un trabajo de constancia, de método, y de organización”.

J.P Sergent

Agradecimientos

Al proyecto FONDEF 2011 D10I1007: "Genomics of the Chilean Populations: Genetic Characterization Necessary for Biomedical Research, Public Health and Forensic Medicine", del cual se encuentra inmerso este trabajo.

Al profesor Mauricio Moraga por permitirme ser parte del laboratorio y por el apoyo constante de nuevas ideas y también críticas al desarrollo de este estudio.

A Larry González por haberme asistido en el análisis bioinformático y haberme ayudado en la confección de nuevas programaciones en nuestro estudio.

A Ricardo Verdugo por su ayuda al principio del magíster y por sus constantes constructivas críticas durante el desarrollo del trabajo.

Y finalmente, a los miembros de la comisión: Elie Poulin y Marco Méndez por sus acertadas correcciones y críticas de este trabajo.

Contenido	Página
Introducción	1
I. Antecedentes Generales	3
1.- Origen de las Poblaciones Mestizas en Latino América	3
2.- Linajes Uniparentales: ADN mitocondrial y Cromosoma Y	4
3.- Caracterización de la Variabilidad Genética a nivel Genómico	5
II.- Antecedentes Específicos	6
1.- Población Mestiza Chilena	6
2.- Composición Genética de los Linajes Uniparentales en Latino América	7
a) ADN mitocondrial	7
b) Cromosoma Y	9
3.- Consideraciones Finales	11
Pregunta de Investigación	12
Objetivo General	12
Objetivos Específicos	12
Metodología	13
I.- Descripción de la Muestra	13
II.- Elaboración Tabla de Haplogrupos	13
1.- ADN mitocondrial	13
2.- Cromosoma Y	14
III.- Confección de un set mínimo de SNPs para el ADN mitocondrial y Cromosoma Y	16
1.- Confección de un set mínimo de SNPs	16
a) Método de Exclusión Independiente	17
b) Método de Árbol de Decisiones	19
2.- Ajuste del set mínimo de SNPs para Poblaciones Mestizas de Latinoamérica	19
IV.- Análisis Experimental	20
1.- ADN mitocondrial y Cromosoma Y	20
a) Composición Genética de la muestra mediante PCR-RFLP	20
b) Caracterización filogenética mediante secuenciación completa	20
c) Caracterización de la Región Control	24
d) Caracterización a través de un Chip comercial	24
d.1) Construcción de la matriz a través del Chip Axiom Latino	27
e) Asignación de haplogrupos mediante el set mínimo de SNPs	28
Resultados	29
I.- Confección Tabla de Haplogrupos	29
1.- ADN mitocondrial	29
2.- Cromosoma Y	29
II.- Confección de un set mínimo de SNPs para el ADN mitocondrial y Cromosoma Y	31

1.- Optimización de un set mínimo de SNPs	31
a) ADN mitocondrial	31
b) Cromosoma Y	33
2.- Ajuste del set de SNPs optimizado para poblaciones mestizas de Latinoamérica	35
IV.- Desarrollo Experimental	35
1.- ADN mitocondrial	35
a) Determinación de la composición genética de la muestra mediante PCR-RFLP	35
b) Caracterización de la región control	35
c) Caracterización filogenética mediante secuenciación completa	35
d) Microarreglo de alta densidad: Chip Axiom Latino	36
d.1) Descripción del linaje materno dentro del Chip	36
d.2) Caracterización de haplogrupos asociados al Chip	39
e) Asignación de haplogrupos por el set mínimo de SNPs	40
2.- Cromosoma Y	41
a) Composición Genética de la muestra mediante PCR-RFLP	41
b) Caracterización filogenética mediante Secuenciación Completa	42
c) Microarreglo de alta densidad: Chip Axiom Latino	42
c.1) Descripción del linaje paterno dentro del Chip	42
c.2) Caracterización de haplogrupos asociados al Chip	43
d) Asignación de Haplogrupos por el set de SNPs optimizados	44
3.- Comparación entre la Tabla de haplogrupos y el Microarreglo de alta Densidad	44
Discusión	46
I.- Elaboración Tabla de Haplogrupos	46
II.- Confección de un set mínimo de SNPs para el ADN mitocondrial y Cromosoma Y	49
III.- Desarrollo Experimental	50
1.- ADN mitocondrial	50
2.- Cromosoma Y	50
IV.- Comparación del set mínimo de SNPs con el microarreglo de alta densidad y la secuenciación genómica	51
Conclusiones	53
Bibliografía	54

Lista de Tablas

Tablas	Contenido	Páginas
Tabla I	ADN mitocondrial	22
Tabla II	Cromosoma Y	23
Tabla III	Representación parcial de la asignación de los distintos haplogrupos	36
Tabla IV	Matriz parcial de haplogrupos	37
Tabla V	Comparación con cinco métodos diferentes para la caracterización de linajes mitocondriales	41
Tabla VI	Comparación con cuatro métodos diferentes para la caracterización de linajes del Cromosoma Y	44

Lista de Figuras

Figuras	Contenido	Página
Figura 1	Esquema con el proceso de selección de haplogrupos para el ADN mitocondrial y Cromosoma Y	16
Figura 2	Diagrama de Flujo del algoritmo de exclusión independiente	18
Figura 3	Esquema Comparación entre la matriz del Chip Axiom Latino con la matriz de los individuos	25
Figura 4	Diagrama de Flujo para los algoritmos de asignación de haplogrupos	26
Figura 5	Esquema procedimiento para el análisis de los SNPs del Chip Axiom Latino	28
Figura 6	Distribución y frecuencia de los haplogrupos para ADN mitocondrial y Cromosoma Y	30
Figura 7	Frecuencia de los distintos set de SNPs por el método de exclusión independiente	31
Figura 8	Gráfico con la frecuencia para cada SNP dentro de los haplogrupos mitocondriales con exclusión independiente	32
Figura 9	Gráfico con la frecuencia para cada SNP dentro de los haplogrupos del Cromosoma Y con exclusión independiente	33
Figura 10	Gráfico con la frecuencia para cada SNP dentro de los haplogrupos del Cromosoma Y con el árbol de decisiones	34
Figura 11	Caracterización macro-haplogrupos mitocondriales representados por el Chip Axiom Latino	38
Figura 12	Esquema con dos haplogrupos en el Chip Axiom Latino bajo el mismo set de SNPs	40
Figura 13	Filogenia del Cromosoma Y	43
Figura 14	Comparación entre la tabla de haplogrupos con el chip Axiom Latino para el ADN mitocondrial y el Cromosoma Y	45
Figura 15	Número de citas para el ADN mitocondrial y Cromosoma Y en los últimos 40 años en Genética de Poblaciones Humana	47

RESUMEN

Introducción. Desde la llegada de los españoles a América hace un poco más de 500 años atrás se inició un proceso de formación de un nuevo tipo de población: la mestiza. Esta población se generó por la mezcla asimétrica que hubo entre soldados españoles con mujeres nativas americanas. Sin embargo este proceso no ocurrió de la misma forma en todo el continente, por lo que cada población mestiza hoy en día contiene una historia diferente. Una forma de poder analizar el aporte de cada grupo en una población es mediante la herencia uniparental: ADN mitocondrial y Cromosoma Y.

Objetivo. El objetivo principal de este estudio es confeccionar un set de SNPs mínimos capaz de caracterizar filogenéticamente el linaje materno (ADN mitocondrial) y paterno (Cromosoma Y) de un individuo en poblaciones mestizas de Latinoamérica.

Metodología. En la fase teórica se elaboró una tabla de haplogrupo con la distribución de los principales linajes del ADN mitocondrial y Cromosoma Y en América, Europa, África, Asia y Oceanía. Luego, se redujo el número de SNPs que caracterizan a estos haplogrupos mediante un proceso de optimización con dos algoritmos en paralelo. En una fase experimental se caracterizó el linaje uniparental en 55 individuos mediante distintos métodos: PCR-RFLP, secuenciación de la región control, secuenciación genómica, tipificación con un microarreglo de alta densidad y a través del nuevo set de SNPs optimizado.

Resultados. Se describió un total de 351 haplogrupos mediante 724 SNPs para el ADN mitocondrial y 97 haplogrupos mediante 301 SNPs para el Cromosoma Y. Estos datos fueron analizados mediante los algoritmos de exclusión independiente y árbol de decisiones. Para el ADN mitocondrial se eligieron 196 SNPs, mientras que para el Cromosoma Y a 70 SNPs. La caracterización de haplogrupos a través de este nuevo set de SNPs muestra resultados similares a los de la secuenciación genómica.

Discusión. La elaboración de la tabla de haplogrupos es fundamental para poder describir la distribución en un elevado detalle de los principales linajes materno y paterno en poblaciones mestizas en Latinoamérica puesto que posee información de poblaciones nativas americanas, europeas y africanas. Debido a la gran cantidad de SNPs que definen los haplogrupos descritos se hace necesario poder optimizar el número de SNPs manteniendo la resolución en la

caracterización. Las mayores ventajas de este nuevo set de SNPs es que con el mínimo de AIMs posibles, se puede llegar casi a los mismos resultados que con la secuenciación genómica en la caracterización de linajes uniparentales sobre un individuo.

Conclusión. Los resultados de nuestros trabajos nos permiten generar un set de SNPs optimizados con el fin de caracterizar el componente genético de una población a través del ADN mitocondrial y Cromosoma Y, mediante: 1) la elaboración de una tabla de haplogrupos, y 2) el diseño e implementación de un algoritmo capaz de reducir y optimizar el número de SNPs dentro de un haplogrupo para poder caracterizar un individuo con la mínima información posible.

ABSTRACT

Introduction. The arrival of the Spanish people to America some 500 years ago started the formation process of a new population, the admixed one. This new population formed itself from the despaired association between Spanish soldiers and Native American women. None the less, this process did not occur in the same way throughout the continent, the result being that every admixture population to this day, has a different background story. One way to analyze the contribution of each parental group to a population is through uniparental inheritance: Mitochondrial DNA and Y chromosome.

Objective. The main goal of this study is to create a minimal SNP set capable of phylogenetically characterize the maternal (mitochondrial DNA) and paternal (Y chromosome) lineage of an individual from the admixture populations of America.

Methodology. In the theoretical phase an haplogroup table was made of the main lineages of mitochondrial DNA and Y Chromosome's distribution in America, Europe, Africa, Asia and Oceania. The number of SNPs that characterized this haplogroups was then reduced by an optimization process with two algorithms. In an experimental phase the uniparental lineage of 55 individuals was characterized using different methods: RFLP-PCR, control region sequencing, genomic sequencing, high density microarray typification and through the new optimized SNP set.

Results. A total of 351 haplogroups were described by 724 SNPs for mitochondrial DNA and 97 haplogroups were described by 301 SNPs for the Y chromosome. The data was analyzed through independent exclusion algorithm and decision tree method. For mitochondrial DNA 196 SNPs were chosen while for the Y chromosome there were 70. The characterization of haplogroups through this new set of SNPs shows similar results to those of genomic sequencing.

Discussion. Making this haplogroup table is fundamental to describe, in high detail, the distribution of the more prominent maternal and paternal lineages in admixed populations of Latin America since it has data from Native American, European and African populations. Since there is a great number of SNPs defining each haplogroup it has been made necessary to reduce the number of them without compromising the resolutive power of the

characterization. The biggest advantages with this new SNP set is that with the least number of AIMS possible, you are able to reach the same results as with genomic sequencing for the uniparental lineages involving on individual.

Conclusión. Our works results allowed us to generate an optimized set of SNPs with which to characterize the genetic component of a population trough mitochondrial DNA and Y chromosome studies by: 1) The creation of an haplogroup table, and 2) the design and implementation of an algorithm capable of reducing and optimizing the number of SNPs inside an haplogroup so that we are able to characterize an individual with the least amount of information possible.

INTRODUCCIÓN

Gran parte de las poblaciones que habitan actualmente el continente americano son producto de una mezcla de diferentes culturas, provenientes de diferentes regiones geográficas. Este proceso comenzó hace más de 500 años con la llegada de los españoles e involucro mayoritariamente a los grupos nativos americanos y los invasores europeos. Posteriormente, grupos africanos se unieron a esta nueva configuración poblacional en ciertas regiones del continente. El proceso de mezcla que se generó entre estos tres grupos constituye el inicio de las poblaciones mestizas en América, que actualmente mantienen aporte europeo, nativo americano y africano, pero que no necesariamente mantienen las mismas proporciones a lo largo del continente americano. Una propiedad adicional de las poblaciones mestizas en América corresponde al cruzamiento asimétrico original que existió, donde el aporte paterno fue principalmente en europeos (españoles) mientras que el aporte materno fue mayoritariamente de mujeres nativo americanas. Esta situación la podemos corroborar fácilmente si analizamos la composición genética de diversas poblaciones urbanas mediante herramientas de la genética molecular, las que revelan la existencia de una elevada frecuencia de linajes europeos para el Cromosoma Y, y una elevada frecuencia de linajes nativo americanos para el ADN mitocondrial. Desde la conquista de América hasta nuestros días las poblaciones mestizas han sufrido cambios abruptos en su demografía aumentando el número de individuos, que debido a diversos procesos culturales de estas poblaciones tienden a concentrarse en ciudades, de manera que podemos encontrar en América un importante número de ciudades con un elevado número de individuos. Esto es bastante importante porque podemos obtener, a través de herramientas genéticas la historia y procesos recientes en una población o región en particular, basados en el estudio de los linajes uniparentales.

Los individuos provenientes de la primera generación del proceso de mestizaje tendrán para cada par de autosomas un cromosoma entero con un origen particular (europeo, americano o africano), mientras que la segunda generación tendrá cromosomas conteniendo bloques de diferente ancestría. A medida que el proceso continúe en el tiempo, estos bloques de diferentes orígenes se harán más pequeños. Consecuentemente, los cromosomas de los individuos en las

poblaciones actuales muestran una compleja mezcla de ancestría dependiendo del número de generaciones de mezcla, patrones de unión, y las características de la población. Por lo tanto, la necesidad de conocer la composición genética de estas poblaciones a través de los linajes uniparentales se hace fundamental, permitiéndonos conocer como fue el proceso de mezcla en la primera generación; cuantos grupos iniciales se vieron involucrados, y en qué proporción aportó cada grupo en una población en particular. Además, es posible comparar estos resultados entre distintas poblaciones y evaluar si los procesos de formación de cada población fueron similares. Por último, podemos determinar patrones de distribución de linajes específicos en una región.

El ADN mitocondrial y Cromosoma Y son buenas herramientas para el estudio de la genética de poblaciones, debido básicamente a la ausencia de recombinación y la tasa de mutación. Estos reflejan los patrones demográficos y de migración en una o varias poblaciones desde que el hombre moderno salió de África. De esta manera a medida que las poblaciones se fueron dispersando y separando entre sí fueron acumulando mutaciones propias en sus genomas generando haplotipos específicos a esas poblaciones y regiones geográficas. De esta manera al estudiar los linajes uniparentales sobre poblaciones mestizas, podemos decir con bastante exactitud el origen geográfico del linaje en un individuo. La cantidad de información concerniente a la variabilidad genética dentro del ADN mitocondrial y del Cromosoma Y es bastante, de hecho hay diferentes plataformas informáticas con grandes bases de datos.

Actualmente, existen múltiples herramientas genético-moleculares para caracterizar los linajes uniparentales en poblaciones humanas, donde se encuentran desde las más tradicionales como PCR-RFLP y la secuenciación de la región control, como los métodos de secuenciación masiva (NGS) sobre el genoma completo, y plataformas de genotipificación llamados microarreglos. Este último método se ha perfeccionado para estimar dentro de una población el aporte genético de cada grupo (europeos, nativos americanos y africanos) en el genoma de cada individuo, sin embargo se desconoce si pueden caracterizar el linaje uniparental de un individuo. En base a los mencionados métodos de caracterización de linajes uniparentales, así como también el aumento del volumen de información para los diferentes haplogrupos alrededor del mundo nace nuestro objetivo principal, el cual busca analizar la distribución de los principales haplogrupos en América, Europa, África, Asia y Oceanía y confeccionar el set de SNPs mínimo e

ideal para que se pueda representar con una máxima resolución la variabilidad genética en diversas poblaciones mestizas de Latinoamérica, considerando que estas poblaciones son el fruto de una mezcla inicial poblaciones de diferentes continentes.

I.- Antecedentes Generales

1.- Origen de las Poblaciones Mestizas en Latino América

Las gran mayoría de las poblaciones en América Latina son mestizas, y se generaron producto de la interacción inicial de tres grupos poblacionales distintos (Wall et al., 2011). Este proceso empezó en el noreste de Asia con la llegada de pequeños grupos hacia América entre 15000 a 18000 años atrás (Schurr 2004; Tamm et al., 2007; Achilli et al., 2008; Fagundes et al., 2008 a; b). Luego de la llegada de los europeos y africanos hace un poco más de cinco siglos atrás, un complejo proceso de mezcla tomó lugar entre estos tres grupos. Estas poblaciones que entraron en contacto durante este tiempo fueron muy diversos entre sí; las poblaciones nativas americanas, europeos (principalmente Españoles y Portugueses), y africanos, quienes fueron inicialmente traídos a la región como esclavos (Salzano & Sans, 2014; Verdu & Rosenberg, 2011). Wang et al., (2008) estimaron que el tiempo promedio desde el primer evento de mezcla se encuentra entre seis a 14 generaciones, excluyendo las poblaciones del Caribe donde el proceso empezó, por lo que este número podría ser mayor. Mientras que el mestizaje en un principio involucró básicamente a hombres españoles y mujeres Nativas Americanas, se expandió corta y tempranamente hacia los hombres europeos o “criollos” (descendientes europeos nacidos en América) y mujeres mestizas o hombres y mujeres mestizos. Este proceso ya complicado de por sí, se vio intervenido nuevamente por la introducción de los esclavos africanos hacia América a principios del siglo XVI (Salzano & Sans 2014). Además, las características sociales, culturales y económicas de las poblaciones nativas, aspectos de la conquista europea y colonización, y por sobre todo la naturaleza de los europeos y africanos al arribar al continente influenciaron el proceso de mestizaje (Salzano & Sans 2014). Es por esto que las poblaciones Latino Americanas contemporáneas varían mucho entre sí, dado que presentan diversas variables históricas y específicas para cada región (Bryc et al., 2010; Via et al., 2011; Wall et al., 2011), como la densidad

poblacional de las poblaciones Nativas Americanas que ocupaban los espacios donde actualmente hay poblaciones urbanas (Wang et al., 2008), el grado y tasa a la cual los asentamientos europeos desplazaron a las poblaciones nativas, si la esclavitud fue introducida o no en una dada región, y, si es así, el tamaño y tiempo de la introducción de la población esclava africana (Bryc et al., 2010; Choudhry et al., 2006; Via et al., 2011; Wang et al., 2008).

2.- Linajes Uniparentales: ADN mitocondrial y Cromosoma Y

Los escenarios de mezcla han sido abundantes sobre el curso de la humanidad, y por cada evento de mezcla que haya ocurrido recientemente, se puede obtener una marca detectable en los genomas de los individuos mestizos (Bocca & Rosenberg, 2011). Un acercamiento para poder analizar estos escenarios de mezcla, específicamente en poblaciones mestizas, es separar la contribución materna y paterna a través de marcadores uniparentales (M Sans, 2000). Estos marcadores como el ADN mitocondrial y Cromosoma Y son eficaces al momento de estudiar los eventos demográficos o migratorios dentro de una población en un pasado reciente (Halder et al., 2008; Wilkins & Marlowe, 2006) debido a las propiedades que poseen (Brown et al., 1979; Giles et al., 1980; Jobling & Tyler-Smith, 2003). Una de estas características corresponde a la ausencia de recombinación, lo que genera haplotipos que pasan intactos de una generación a la siguiente, viéndose afectados únicamente por eventos de mutación, y por lo tanto, preservan un registro de alta resolución en la historia de la especie (Mark & Jobling & Tyler-Smith, 2003; Wilkins & Marlowe, 2006).

Diversos estudios señalan que existe una asimetría en el flujo genético de las poblaciones parentales hacia las poblaciones mestizas actuales (Salazar-Flores et al., 2010), donde la diversidad genética encontrada en el Cromosoma Y pertenece mayoritariamente a grupos europeos, mientras que los linajes del ADN mitocondrial son en su mayoría de mujeres amerindias y/o africanas (Bryc et al., 2010; Sandoval et al., 2012; Via et al., 2011), coincidiendo con los registros y datos históricos (Simms et al., 2012). Y, a pesar que fue un proceso relativamente rápido, se plasmó como huella imborrable en aquellos patrones de herencia uniparental (ADN mitocondrial y Cromosoma Y). Este proceso ha mostrado ser extraordinariamente consistente entre ciudades y poblaciones incluyendo Argentina (Bobillo et al., 2010; Catelli et al., 2011; Corach et al., 2010; Seldin et al., 2007; Toscanini et al., 2007), Brasil

(Carvalho et al., 2008; Marrero et al., 2005, 2007), Colombia (Bedoya et al., 2006; Carvajal-Carmona et al., 2000; 2003), Costa Rica (Carvajal-Carmona et al., 2003), Ecuador (Baeta et al., 2012; González-Andrade et al., 2003), Guatemala (Martínez-González et al., 2012), Guyana Francesa (Mazières et al., 2008), Haití y Jamaica (Simms et al., 2012), México (Rangel-Villalobos et al., 2008, 2009; Salazar-Flores et al., 2010; Sandoval et al., 2012; Silva-Zolezzi et al., 2009).

3.- Caracterización de la Variabilidad Genética a nivel Genómico.

Indudablemente, la huella de esta mezcla asimétrica no sólo se ve reflejada en linajes uniparentales, sino que también se puede rastrear a lo largo de todo el genoma en un individuo, encontrando que los mayores aportes hacia estas poblaciones mestizas han sido de grupos nativos americanos y europeos (Wang et al., 2008). Ahora bien, la estructuración genética que muestran estas poblaciones señala que existe una variación importante en torno al porcentaje de cada grupo parental relacionado con diversas variables como el asentamiento de los grupos indígenas, clases sociales, etc (Wang et al., 2008; Via et al., 2011). La principal estrategia para detectar la variación entre estos componentes es a través de los AIMs (*ancestry informative marker*) que se definen como polimorfismos (SNPs) que exhiben grandes diferencias en las frecuencias alélicas entre poblaciones y pueden ser usadas para inferir el origen geográfico de los individuos (Galanter et al., 2012). Actualmente, existe un afán de poder describir el grado de ancestría para cada individuo ha llevado a la construcción de un set de marcadores AIMs en todo el genoma para poder cuantificar la ancestría asociada a grupos nativos americanos, europeos y africanos sobre poblaciones mestizas en Latinoamérica (Galanter et al., 2012). Via et al., 2011 estudió mediante 93 AIMs el proceso de mestizaje asociado a la distribución geográfica en poblaciones de Puerto Rico. Avena et al., (2012) analiza a través de 99 marcadores AIMs la variabilidad genética en diferentes poblaciones de Argentina, encontrando una importante variación en el aporte de cada grupo nativo americano/europeo entre distintas regiones.

Hoy en día, existen múltiples plataformas con elevadas densidades de genotipificación que permiten caracterizar cientos y miles de SNPs distribuidos tanto en cromosomas autosómicos como en uniparentales (Hoheisel, 2006; Syvänen, 2005). Estas plataformas, también llamados microarreglos radican básicamente en la elevada afinidad y especificidad que cada hebra de ADN posee para reconocer secuencias complementarias a través del apareamiento de sus bases

(Brown & Botstein, 1999). Las principales aplicaciones que poseen los microarreglos van desde el análisis de expresión genética y factores de transcripción hasta la genotipificación masiva (Bumgarner, 2013; Lipshutz et al., 1999). Últimamente, los estudios de asociación genética son la razón principal para la construcción de estas plataformas, los que conllevan necesariamente la selección de un set de SNPs que representen de mejor manera la variabilidad dentro de una población (Hoffmann et al., 2011a). Sin embargo, muchas poblaciones actuales son étnicamente heterogéneas, debido a su proceso de formación o migraciones posteriores, por lo tanto al aplicar una plataforma de microarreglo con un set de SNPs universales genera una pérdida en la cobertura de variantes que son específicas de un grupo étnico (Hoffmann et al., 2011b). Bryc et al (2010) tomó en cuenta este problema y analizó aquellos SNPs comunes en dos plataformas: Affymetrix 500K e Illumina 610-Quad y que mostraban frecuencias alélicas que podían discriminar entre poblaciones nativo americanas, africanas y europeas en diferentes poblaciones latinas. Más tarde, Moreno-Estrada et al., (2014) analiza 20 poblaciones indígenas y 11 mestizas en México a través del análisis de un microarreglo (Axiom Latino, Affymetrix) enfocado a poblaciones mestizas en Latinoamérica, encontrando fuertes diferencias entre las poblaciones nativas.

II.- Antecedentes Específicos

1.- Población Mestiza Chilena

La población Chilena es considerada el producto de una mezcla asimétrica entre soldados españoles con mujeres nativas (Encina 1983; Cifuentes et al., 2004; Barrai et al., 2012). La población Africana no parece haber tenido en Chile un rol muy importante como en otros países de Latino América. Mediante el uso de marcadores tradicionales (grupos sanguíneos) Acuña estima entre un 5 a 10% el aporte africano en poblaciones mixtas del norte del país (Acuña et al., 2000), mientras que en el resto del país, la mano de obra principal eran los mismos indígenas nativos, quienes realizaban servicios de encomienda. A nivel genómico Wang et al., (2008) señala a través del estudio de dos localidades chilenas que la composición es básicamente dihíbrida (europea/americana). Diversos estudios genéticos realizados a poblaciones mixtas de nuestro país pueden detectar esta huella de asociación asimétrica, encontrando una baja frecuencia de

cromosomas Y con origen amerindio en la ciudad de Santiago (22% aproximadamente) (Cifuentes et al., 2004), así como un alto aporte (sobre el 80%) de linajes mitocondriales Amerindios (Rocco et al., 2001; Henríquez et al., 2004; García et al., 2006; Leiva, 2010; Pezo et al., 2014 *en revisión*). Por otro lado, investigaciones basadas en marcadores localizados en cromosomas autosómicos de poblaciones chilenas muestran un 40% de ancestría Amerindia (*fide* Cifuentes et al., 2004), la cual varía de acuerdo al estrato socioeconómico de la población: una baja frecuencia para estratos altos, y elevada para estratos más bajos (*fide* Cifuentes et al., 2004). Posterior a la llegada de la República, pequeños grupos de inmigrantes europeos se asentaron en nuestro país, por ejemplo, colonos alemanes arribaron al sur de Chile a mediados del siglo XX, mientras que inmigrantes croatas se instalaron en el extremo austral de Chile a principios del siglo XX, entre otros. La contribución genética de estos grupos, no obstante, no ha alterado de manera importante la mezcla asimétrica original (Cifuentes et al., 2004).

2.- Composición Genética de los Linajes Uniparentales en Latino América

a) ADN mitocondrial

La presencia y frecuencia de los linajes de ADN mitocondrial en las poblaciones mestizas de Latino América puede estar asociada a tres grupos parentales principales: nativos americanos, europeos y africanos. Los grupos Nativos Americanos se originaron a partir de grupos asiáticos que se mantuvieron aislados en Beringia, pudiendo colonizar América sólo hace 19000 a 15000 años atrás a finales del último máximo glacial (Schurr 2004; Tamm et al., 2007; Achilli et al., 2008; Fagundes et al., 2008 a; b). Los linajes principales que llevaban estas poblaciones fundadoras eran principalmente cuatro (A2, B2, C1 y D1) analizados mediante marcadores de RFLP, región control y genomas completos (Bailliet et al., 1994; Merriwether & Ferrell, 1996; Schurr et al., 1990; Torroni et al., 1992). Se detectó después la presencia de un quinto haplogrupo fundador, X2a únicamente en Norteamérica (Dornelles et al., 2005; Malhi et al., 2002). Dentro de estos haplogrupos principales, se han identificado recientemente nuevos linajes fundadores: C1b, C1c y C1d los cuales se encuentran distribuidos a lo largo de todo el nuevo mundo, D4h3a presente desde Alaska a Tierra de Fuego y C4c en dos poblaciones nativas de Colombia (Tamm et al., 2007; Achilli et al., 2008; O'Rourke & Raff, 2010). En diversas poblaciones al Norte de Norte América se

han identificado otros haplogrupos, tales como A2a, D2a, D3 los cuales están restringidos a esa región y dan cuenta de llegadas tardías al continente americano (D2a, D3), así como migraciones inversas hacia Siberia (A2a) (Achilli et al., 2008; Kemp et al., 2007; Perego et al., 2009; Tamm et al., 2007; Volodko et al., 2008). En Venezuela se encontraron dos nuevos clados del haplogrupo B2: B2j y B2k, ambos presentes en dos poblaciones urbanas (Venezuela y Pueblo Llano) (Gómez-Carballa et al., 2012). En la población amazónica Waorani del Ecuador se encontró el linaje A2y, el cual es específico de ese grupo y es reflejo de la deriva génica causada por efectos fundadores, aislamiento y pequeños tamaños poblacionales (Cardoso et al., 2012). En Uruguay, se encontró un nuevo linaje para el haplogrupo C1, caracterizado por variantes en la HVI (16051G, 16140C, 16288C) y en la región codificante (12378T), asociado a los indígenas precolombinos Charrúas que habitaban la zona (Sans et al., 2012).

Continuando más al sur, se observó una transición C16187T en poblaciones Mapuches y Huiliches de Chile y Argentina (Ginther et al., 1993; Horai, 1994; Moraga et al., 2000). Bodner et al., (2012) describe esta variante como D1g, la cual junto a D1j están presentes únicamente en el cono sur de Sudamérica. D1g se encuentra tanto en poblaciones nativas de Chile y Argentina, mientras que D1j parece estar más restringido al noreste de Argentina (García et al., 2012). Se han descrito también dos nuevos linajes con una distribución similar a D1g: B2i2 el cual posee una edad estimada de ~15000 años (semejante a D1g) y se originó probablemente en poblaciones cazadoras-recolectoras, y C1b13 que tuvo una aparición más tardía que B2i2 y D1g (~3000-4000 años) en grupos en crecimiento poblacional, posiblemente agricultoras (de Saint Pierre et al., 2012). Estos últimos autores también describen la presencia del haplogrupo D4h3a5, el cual se encuentra sólo en poblaciones de Tierra del Fuego, descrito inicialmente por Perego et al., (2009).

Dentro de la composición genética europea, el haplogrupo más frecuente en Europa corresponde a H (Achilli et al., 2004; Loogväli et al., 2004; Richards et al., 2000), el cual se ha sugerido que se originó en el Medio Este aproximadamente entre 30 a 25 mil años atrás, expandiéndose hacia Europa contemporáneo con la difusión de la cultura Gravettiana entre 25 a 20 mil años atrás, y estuvo envuelta en numerosos episodios de flujo génico en el oeste de Eurasia pero particularmente durante la re-población post-glacial de Europa (Torroni et al., 1998; Richards et al., 2000; Loogvali et al., 2004). Dentro de este clado podemos encontrar otros sub-linajes frecuentes como H1, H2, H3, H4, H5a, H6a, H8, H11 distribuidos en Europa Central, Oeste

de Europa y algunos en la Península Ibérica como H1, H3, H5a y H6a (Achilli et al., 2004; Brandstätter et al., 2008; Maca-Meyer et al., 2003; Malyarchuk et al., 2008; Rose et al., 2008). También podemos encontrar los haplogrupos V, K, T, J, I y U (U5, U5b) principalmente en Francia, Italia y España (Dubut et al., 2004; Ingman & Gyllensten, 2007; Maca-Meyer et al., 2003; Rose et al., 2008), así como también R0, W y X principalmente en el Oeste de Europa (Achilli et al., 2011; Brisighelli et al., 2009).

La diáspora de África fue el resultado de la esclavitud en América de probablemente 11 millones de individuos, principalmente traídos del oeste de África. Los linajes mitocondriales de estos grupos que llegaron a América corresponden principalmente a L0, L1b, L1c, L2, L3b, L3d, L3e (Ely et al., 2006; Salas et al., 2004; 2005). Existen diferencias sin embargo en las frecuencias del componente africano a través del continente americano debido a variaciones en la contribución materna (Stefflova et al., 2011).

b) Cromosoma Y

Del lado paterno, podemos encontrar que la composición genética de estos linajes proviene de grupos parentales europeos, nativos americanos y africanos. Dentro del set de linajes que ingresaron a América por parte de los europeos, se espera que el linaje R1-M173 sea el más frecuente en las poblaciones mestizas, puesto que es el linaje más común en Europa (Francalacci & Sanna, 2008). También es probable encontrar en menor frecuencia macrohaplogrupos como el F, relacionado con la segunda migración "fuera de África", el G asociado con la expansión neolítica (O. Semino, 2000), el haplogrupo H presente exclusivamente en los Balcanes (Pericic et al., 2005), el haplogrupo I que pudo haber llegado a Europa en el Paleolítico Superior (Semino, 2000) y el haplogrupo J con un gradiente Este-Oeste indicando su arribo desde el cercano Este.

La península Ibérica se caracteriza por presentar en su pool génico diferentes linajes, propios de la historia reciente, así como las fuentes de eventos pasados de recolonización de Europa. Esta región fue un refugio durante el último máximo glacial, y su estructura genética refleja naturalmente el aislamiento, diferenciación y subsecuente expansión, siendo la fuente para la recolonización del Mesolítico hacia el norte y este de Europa. Entre los haplogrupos que acá se originaron, R1b1b2-M269 (antes R1b1c) es particularmente importante, puesto que cubre la mayor parte del Oeste de Europa (Francalacci et al., 2010; Francalacci & Sanna, 2008), también

el haplogrupo I2a1-M26 –ahora I2a1a (ISOGG 2012) – presente en varias regiones de España, pero rara en otras poblaciones del Oeste de Europa, incluso ausente en el resto del continente, así como también I1a-M253 –ahora I1 (ISOGG 2012) – el cual tuvo su origen en el refugio de Iberia (Francalacci & Sanna, 2008; López-Parra et al., 2009; Martínez-Cruz et al., 2012). También es posible encontrar esporádicamente haplogrupos como E3b1b-M81 – ahora Eb1b1b1a (ISOGG, 2012) – perteneciente a grupos Bereberes del Norte de África en representación de la influencia musulmana, además de los haplogrupos J2 y G los cuales son bastante frecuentes en los judíos del mediterráneo (Adams et al., 2008; Francalacci et al., 2010).

Los haplogrupos que componen la diversidad genética de los grupos nativos americanos derivaron genéticamente de poblaciones asiáticas. Estos linajes que entraron se clasifican sin ningún tipo de ambigüedad en dos haplogrupos principales, C3-M217 y Q-M242 (Jota et al., 2011). En Norteamérica se identificó el SNP P39, asociado al haplogrupo C3b, encontrado en bajas proporciones en Norteamérica (Zegura et al., 2004). De acuerdo a Santos et al (2007) (*vide* Jota et al (2011)), el linaje Q1a3a1-M3 representa la gran mayoría de los cromosomas Y nativos en América, situación que se corrobora en trabajos posteriores (Bailliet et al., 2009; Blanco-Verea et al., 2010; Geppert et al., 2011; Toscanini et al., 2011). A partir de este linaje se han descrito dos nuevos subhaplogrupos presentes en las poblaciones amerindias, Q1a3a1-M19 (Bortolini et al., 2003) –ahora Q1a3a1a (ISOGG, 2012) – y Q1a3a4-SA01 (Jota et al., 2011).

En lo que respecta a las poblaciones africanas, tenemos que los linajes más frecuentes son el clado A, el cual está presente en las regiones de Khoisan, Sudán y Etiopía. Además del linaje B que se encuentra en la región del Sub-Sahara. Sin embargo, hay un clado que es el más frecuente y diverso en África: el haplogrupo E, específicamente el haplogrupo E1b1a que se encuentra cubriendo en elevadas frecuencias la mayor parte del continente (Karafet et al., 2008). Abe-Sandes et al., (2004) señala en un trabajo de la composición genética en poblaciones Afro-Brasileñas que la mayoría de los Cromosomas Y pertenecieron al sub-Sahara de África.

3.- Consideraciones Finales

Finalmente, la caracterización genética de una población, vía paterna y materna es de mucha importancia, puesto que nos puede dar una visión histórica de los procesos de formación u otros eventos recientes (migraciones) sobre una población mestiza. Estos eventos pueden además ser distintos en cada población mestiza, ya sea a nivel macro-espacial (continental), así como a nivel local (dentro de un mismo país). Además, la cantidad de información que se puede obtener de las poblaciones mestizas en Latinoamérica es elevada dada su condición inicial híbrida por lo que es posible describir linajes de origen nativo americano, europeo y/o africano. Es por esto que la confección de un set ideal de SNPs que sea capaz de capturar esta información uniparental con el máximo de resolución es de vital importancia para poder generar una nueva herramienta de caracterización molecular.

PREGUNTA DE INVESTIGACIÓN

¿Cuál es el set ideal de SNPs para representar con máxima resolución la variabilidad de los haplogrupos del ADN mitocondrial y Cromosoma Y en poblaciones mestizas latinoamericanas?

OBJETIVO GENERAL

Confeccionar un set de SNPs mínimo capaz de caracterizar filogenéticamente el linaje materno (ADN mitocondrial) y paterno (Cromosoma Y) de casi cualquier individuo en poblaciones mestizas de Latinoamérica.

OBJETIVOS ESPECÍFICOS

- Establecer un mapa de distribución con los haplogrupos más representativos en América, Europa, África, Asia y Oceanía para el ADN mitocondrial y Cromosoma Y, a través de una tabla de haplogrupo con el fin de obtener una amplia y buena cobertura del linaje materno y paterno alrededor del mundo.
- Optimizar el número de SNPs de los haplogrupos descritos para el ADN mitocondrial y Cromosoma Y, a través de la implementación y evaluación de dos algoritmos.
- Aplicar el set de SNPs optimizado para la caracterización en 55 individuos y comparar su nivel de resolución respecto a diferentes técnicas para la asignación de haplogrupos: PCR-RFLP, Secuenciación Región Control, Secuenciación Genómica y tipificación con un microarreglo comercial de alta densidad.



METODOLOGÍA

I.- Descripción de la Muestra

El presente trabajo se encuentra enmarcado bajo el proyecto FONDEF 2011 D10I1007: "Genomics of the Chilean Populations: Genetic Characterization Necessary for Biomedical Research, Public Health and Forensic Medicine", que tiene como objetivo: (1) Identificar un panel de marcadores informativos de ancestría que permitan distinguir entre la ancestría amerindia y europea en todo el genoma de los chilenos, y (2) la caracterización de la composición Amerindia/Europea de poblaciones chilenas por locaciones geográficas y variables socioeconómicas. Se seleccionó en una primera instancia del proyecto una muestra de 55 individuos que fueron genotipificados con el Chip de Axiom Latino, de los cuales a 18 se les secuenció el genoma. Estas muestras se colectaron en Arica y Santiago durante el 2013. Las muestras se tomaron como parte de la ejecución del proyecto FONDEF 2011 D10I1007, y los consentimientos informados fueron visados por el comité de ética de la facultad de Medicina, Universidad de Chile y por los comités de ética locales en algunos casos. De esta manera en Arica se lograron recolectar 19 individuos de origen Aymara, y en Santiago a 36 individuos de origen Mapuche.

II.- Elaboración Tabla de Haplogrupos

1.- ADN mitocondrial

La selección de haplogrupos siguió un proceso ordenado, manteniendo como propósito principal que cada haplogrupo analizado debe estar asociado a una región específica, con esto buscamos obtener un número óptimo de haplogrupos que nos permita alcanzar una representación global para este marcador uniparental. Para esto, la confección de esta tabla tuvo que seguir un proceso estructurado y jerárquico, en el cuál se establecieron varias etapas de selección y filtrado como muestra la Figura 1.

La primera etapa se centró la búsqueda de información, tanto bibliográfica como de la base de datos online http://phylotree.org/mtDNA_seqs.htm (van Over & Kaiser, 2009). En esta etapa se realizó un proceso de filtración, seleccionando aquellos trabajos que describieran linajes

mitocondriales con un patrón de distribución geográfica específica. El proceso de selección de haplogrupos se realizó de manera separada para cada continente, con el propósito de organizar de manera más fácil y óptima los trabajos encontrados en la base de datos. Una vez elegidos estos haplogrupos, se registraron los SNPs que los definen en una tabla de acuerdo a lo descrito por los autores de cada trabajo.

En una segunda etapa de revisión y corrección, se identificaron aquellos haplogrupos que se repetían entre los distintos continentes (América, Europa, África, Asia, Oceanía) y se eliminaron. Luego de esta revisión, se compararon los haplogrupos con el árbol filogenético de la base de datos <http://phylotree.org/main/tree.htm> (van Oven & Kayser, 2009) para ver si hubo algún cambio en el nombre del haplogrupo. Este árbol se actualiza constantemente, agregando nuevos linajes mitocondriales, así como renombrando ya existentes.

En la tercera y final de resultados, se obtuvo una lista para cada continente con los haplogrupos, la región donde se encuentran mayormente frecuentes y los SNPs que lo identifican. Para América se aumentó el esfuerzo de caracterización, por lo que se mantuvieron todos los haplogrupos encontrados. Para el continente africano se puso especial énfasis en buscar aquellos linajes que llegaron a América, producto del comercio de esclavos.

2.- Cromosoma Y

La selección de haplogrupos para el cromosoma Y se rige bajo los mismos principios del ADN mitocondrial. Los haplogrupos seleccionados tienen que cumplir con la propiedad de representar regiones o áreas geográficas en particular, para así obtener información geográfica del origen ancestral del individuo. Se otorgó especial énfasis en aquellos haplogrupos presentes en el continente americano, europeo y en cierto grado al africano, dado el rol que jugó cada continente a la formación de las poblaciones mestizas en América, principalmente Chile. El proceso de selección de haplogrupos se encuentra descrito en la Figura 1.

En la primera etapa se recopiló información a través de revisión bibliográfica, así como de la plataforma de base de datos del Cromosoma Y (www.isogg.org). Se describió en una primera instancia la distribución global de los principales macrohaplogrupos (A, B, C, D, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T) (Karafet et al., 2008). Luego se seleccionaron los diferentes haplogrupos de acuerdo a los criterios antes mencionados. Cada análisis se llevó a cabo de

manera separada por cada continente (América, Europa, Asia, África, Oceanía). En esta etapa se registró además del haplogrupo, su mutación correspondiente.

En la segunda etapa se procedió a revisar y corregir cada haplogrupo seleccionado en la etapa anterior. Se corrigió y se estableció la nomenclatura a usar en este estudio. El nombre de cada haplogrupo va variando año tras año, por lo que se hace muy común que varios trabajos se refieran a un mismo haplogrupo, pero con distintos nombres a través del tiempo. Aquello que no cambia es la mutación terminal que define a cada haplogrupo, por lo que usaremos una combinación de los dos sistemas de nomenclaturas descritas por Karafet et al., (2008), en donde nombraremos el haplogrupo de manera “jerárquica”, seguido por la mutación que lo caracteriza. La referencia para la asignación “jerárquica” de este estudio de cada haplogrupo se tomará del árbol filogenético del Cromosoma Y 2012 (*International Society of Genetic Genealogy, 2012*). Cabe mencionar que a la fecha se encuentra la filogenia del Cromosoma Y actualizada. En la última etapa, se obtuvo una lista de diversos haplogrupos y sus respectivas mutaciones con diversos rangos de distribución para cada continente.

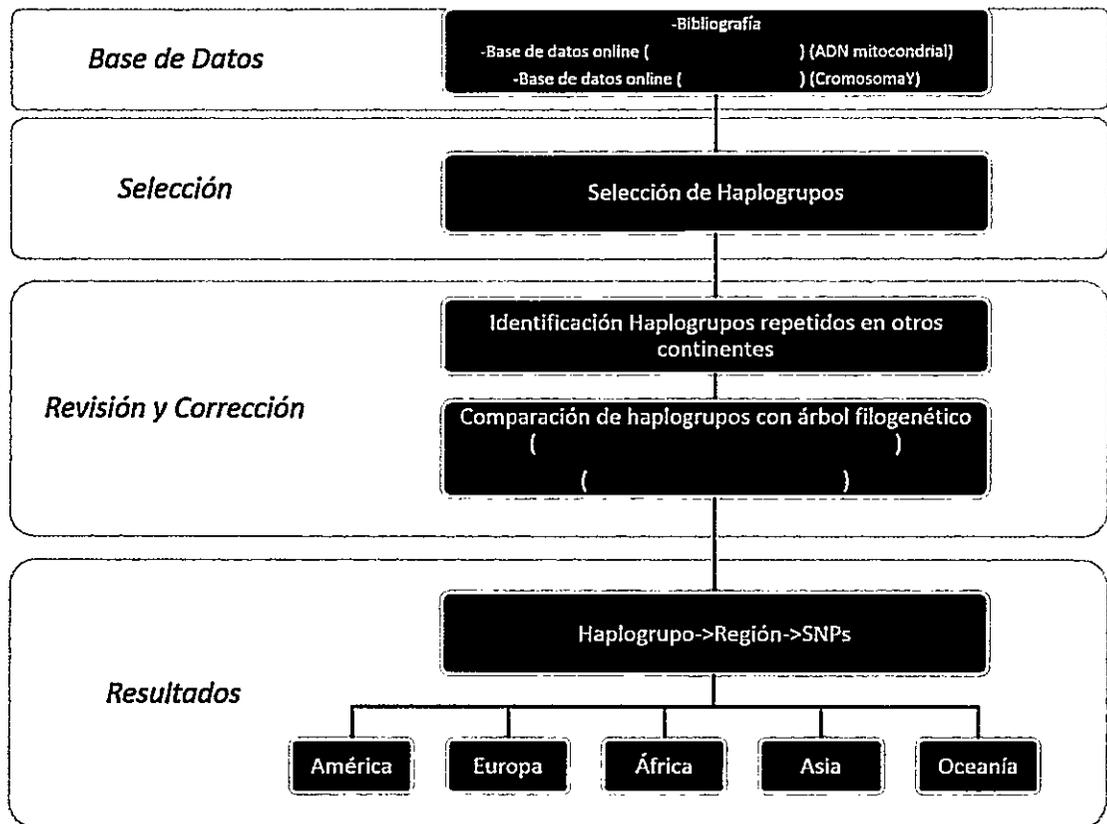


Figura 1. Esquema general con el proceso de selección de haplogrupos para el ADN mitocondrial y Cromosoma Y.

III.- Confección de un set mínimo de SNPs para el ADN mitocondrial y Cromosoma Y

1.- Confección del set mínimo de SNPs

Se construyó una matriz general con la información (SNPs) de todos los haplogrupos descritos en la elaboración de la tabla de haplogrupos para el ADN mitocondrial y Cromosoma Y en la etapa anterior. La reducción del número de SNPs se llevó a cabo mediante una etapa de optimización realizada con el lenguaje estadístico de R (R Core Team 2013) y bajo dos condiciones que deben cumplirse bajo esta etapa de optimización: 1) no debe quedar ningún linaje con el

mismo set de SNPs que otro, y 2) no debe quedar ningún linaje sin información (SNP). En esta etapa se implementaron dos algoritmos diferentes hacia la matriz de haplogrupos: método de exclusión independiente y árbol de decisiones.

a) Método de Exclusión Independiente

El diseño de este algoritmo fue creado en este estudio y tiene como objetivo principal la reducción del número de SNPs para un grupo de haplogrupos. El algoritmo fue desarrollado a través del lenguaje estadístico de R (R Core Team, 2013) y consta principalmente del análisis por separado de cada columna (SNP) de la matriz. Primero se remueve la primera columna de la matriz y luego se evalúa si tal efecto de remoción produce incompatibilidad con las dos condiciones planteadas previamente. Si se cumplen estas condiciones, se confirma la remoción del SNP y se sigue con la columna siguiente, en cambio, si cualquiera de las dos condiciones o ambas no se cumplen, se reincorpora la columna eliminada y se continúa con el mismo análisis en la columna siguiente. Asociado a esto, se realizaron 1000 permutaciones en el orden de las columnas bajo este algoritmo para evaluar la robustez (frecuencias absolutas) de cada variante (SNP) dentro de un haplogrupo a ser el elegido por el algoritmo. En la Figura 2 podemos ver el diagrama de flujo de este algoritmo de manera más detallada.

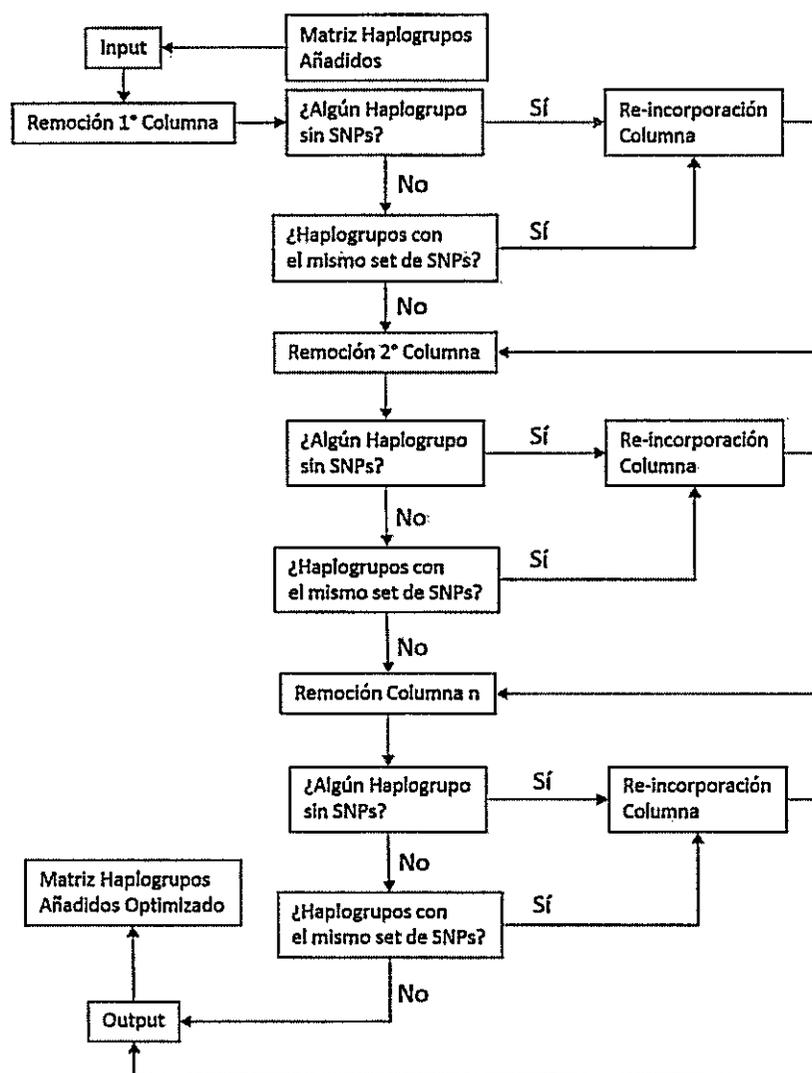


Figura 2. Diagrama de Flujo del algoritmo de exclusión independiente. Como datos de entrada se necesita la matriz de haplogrupos añadidos. En cada etapa de remoción de una columna (SNP) se evalúa el cumplimiento de las dos condiciones impuestas, si alguna de estas (o ambas) no se cumple, se reincorpora la columna. Como dato de salida (Output) se obtuvo una matriz optimizada basada en una reducción del número de SNPs para cada haplogrupo.

b) Método de Árbol de Decisión

Esto se origina a partir de la capacidad que tiene la muestra para dividirse en pequeños subconjuntos de grupos u observaciones relacionadas. La manera de crear estos grupos es buscar o preguntar sobre el set de datos totales, donde cada pregunta traería de vuelta un subconjunto de datos. En nuestro caso, se busca obtener una separación dentro de la matriz para todos los linajes descritos en relación a la presencia o no de un SNP en la matriz. Para esto, el árbol de decisiones va a tomar cada SNP para generar una respuesta dicotómica en función de aquellos linajes que poseen el SNP, de aquellos que no. Luego se analizan estos nuevos subconjuntos generados en el paso anterior y se analizan de la misma forma con un nuevo SNP a analizar, formando nuevos subconjuntos hasta llegar a un linaje único. Como resultado de salida, se entregó aquella combinación de SNPs (de varios resultados posibles) que permitió diferenciar a todos los linajes dentro de la matriz. Este método viene diseñado en el paquete rpart (Therneau et al., 2014) para su uso en R (R Core Team 2013). De la misma forma que en el método anterior, se procedió a permutar las columnas dentro de la matriz 1000 veces con el propósito de reportar las frecuencias absolutas de cada SNP dentro de cada haplogrupo.

Una vez hecho este análisis, se eligieron tanto para el ADN mitocondrial como para el Cromosoma Y un set mínimo de SNPs que nos permita caracterizar a cada linaje de la tabla de haplogrupos. La elección de cada SNPs tendrá como criterio ideal que esté presente en un clado común (sinapomorfía) y que no corresponda a un SNP de carácter homoplásico.

2.- Ajuste del set de SNPs optimizado para poblaciones mestizas de Latinoamérica

Una vez optimizado el set de SNPs con la información de los haplogrupos para todo el mundo se procedió a crear un sub-set de SNPs, enfocado principalmente en capturar la información de los linajes uniparentales en poblaciones mestizas de Latinoamérica. Para esto se filtraron los datos y se seleccionaron haplogrupos de origen americano, europeo y africano. Para el continente europeo se seleccionaron los haplogrupos más frecuentes y asociados principalmente a la península ibérica. Para el continente africano se incluyeron haplogrupos de la costa oeste y asociados fuertemente a América producto del comercio de esclavos.

IV.- Análisis Experimental

1.- ADN mitocondrial y Cromosoma Y

a) Composición Genética de la muestra mediante PCR-RFLP

Se caracterizó la composición genética del linaje materno, a través del ADN mitocondrial, presente en los 55 individuos de la muestra. Para esto se determinaron los haplogrupos A, B, C y D mediante PCR-RFLP, aquellas muestras que no hayan sido clasificadas a ninguno de estos haplogrupos, fueron consideradas como linajes No Amerindios. Para el Cromosoma Y se caracterizaron 11 marcadores bialélicos de la porción no recombinante a través de PCR-RFLP, y PCR alelo específico (PCR-ASP) sobre 28 individuos. Estos marcadores pertenecen a los haplogrupos DE-M1, E-M40, F-M89, H-M69, I-M258, J-M304, K-M9, P-M45, R-M207, Q-M242 y Q1a3a1-M3. La tipificación se determinó de manera jerárquica, caracterizando primero los haplogrupos P-M45, R-M207 y Q1a3a1-M3, debido a que a priori se espera que estos sean los más frecuentes en la muestra. Las condiciones del PCR, partidores y enzimas de restricción utilizadas en este estudio para el ADN mitocondrial y el Cromosoma Y se encuentran descrito en las Tabla I y II respectivamente. Los amplificados y los productos de digestión fueron analizados con electroforesis en geles de Agarosa al 2,5%.

b) Caracterización de haplogrupos mediante secuenciación completa

Se analizaron 18 individuos provenientes de Arica (n=9) y Santiago (n=9). Se secuenció el genoma mediante ligación con SOLiD Sequencer de Applied Biosystem en la empresa OMICS Solutions, Chile como parte de los objetivos del proyecto FONDEF 2011 D10I1007. La información del genoma mitocondrial y del Cromosoma Y requeridos para este estudio se extrajeron mediante los fragmentos de secuenciación (reads) y una planilla con los resultados de la secuenciación genómica para cada individuo. Se utilizó el GRCh37 (IHGSC, 2004) como genoma de referencia estándar para el alineamiento de las secuencias genómicas en el proyecto.

A partir de la información entregada para las secuenciaciones genómicas del proyecto para ADN mitocondrial y Cromosoma Y se registraron en una planilla, donde se anotaron varios datos importantes como la identidad y ubicación física de los SNPs polimórficos en las muestras. Luego, de acuerdo a las variantes que presentaba cada individuo se procedió a clasificarlos filogenéticamente a su haplogrupo correspondiente. Para hacer más óptimo el proceso de

clasificación, tanto para el ADN mitocondrial como para el Cromosoma Y, nos dirigimos directamente a los clados a los cuales correspondía la muestra de acuerdo a la información obtenida previamente por PCR-RFLP. El árbol filogenético mitocondrial utilizado como referencia para este análisis correspondió al *mtDNA tree Build 15 (30 Sep 2012)* de la plataforma online <http://phylotree.org> (van Oven y Kaiser, 2009). Mientras que para el Cromosoma Y, se utilizó la filogenia señalada por la “International Society of Genetic Genealogy” (ISOGG) de la plataforma online <http://isogg.org>.

El criterio para la asignación de haplogrupos, consiste en una lectura jerárquica del clado filogenético. Se procedió a evaluar los sub-linajes del clado en cuestión hasta llegar al punto máximo de resolución en donde no se podían reconocer más variantes en común entre el clado filogenético estudiado con la información genómica para cada individuo.

Tabla I. Primers utilizados para los linajes mitocondriales, con sus respectivas temperaturas de alineamiento y enzimas de restricción.

Mitochondrial Lineage	Marker	Primers	Primer Position (5'→3')	Aligning Temperature	Primer Reference
Haplogroup A	HaeIII 663 (+)	M13 (F)	583-604 GTAGCTTACCTCCTCAAAGCAA	55°C	Moraga et al., 2000
		M14 (R)	727-708 AGGGTGAACCTCACTGGAACG		
Haplogroup B	deletion 9 base pairs COII/tARNlys	M5 (F)	8195-8214 CACAGTTTCATGCCCATCGT	55°C	Moraga et al., 2000
		M6 (R)	8316-8294 ATGCTAAGTTAGCTTTACAGTGG		
Haplogroup C	HincII 13259 (-)	M7 (F)	13209-13232 CGCCCTTACACAAAAATGACATCAA	55°C	Moraga et al., 2000
		M8 (R)	13365-13344 GGAGCACATAAATAGTATGGC		
Haplogroup D	AluI 5176 (-)	M11 (F)	5099-5120 CCTAACTACTACCGAATTCCTA	55°C	Moraga et al., 2000
		M12 (R)	5277-5255 ATTCTCGATAATGGCCCATTTG		
Control Region (D-loop)	Sequences	M1 (F)	15978-15997 CACCATTAGCACCCAAAGCT	60°C	Moraga et al., 2000
		M14 (R)	727-708 AGGGTGAACCTCACTGGAACG		



Tabla II. Primers y Marcadores utilizados para la caracterización del Cromosoma Y, con sus respectivas enzimas y temperaturas de alineamiento.

Filogenia	Marcador	Técnica	Marcador	Primer (5'→3')		Alineamiento	Referencia
				Forward	Reverse		
D	M1	PCR- inserción Alu	150pb/450pb	cagggaagataaagaaa ta	actgctaagaaggagggat	51°C	Hammer y Horai 1995
E	M40	PCR-RFLP	MbiI(BsrBI)	acatctcagatcgtgttgg gt	tagaaggctcggagatgca	58°C	de Saint Pierre, en preparación
C	M89	PCR-RFLP	Nla III	Acagaagatgctgctcag ctt	gcaactcaggcaaaagtgagac at	56°C	Bailliet et al., 2008
F	M69	PCR-ASP		M69F	M69RT/M69RA	69°C	de Saint Pierre, en preparación
G	M258	PCR-RFLP	BsrI	cacagtcctcagggaatct	cctgagaacaaggtaactg	58°C	de Saint Pierre, en preparación
H	M304	PCR-RFLP	Tsp45I	Gctgttaatcactgtttact aat	ttctacgttcatctgcatgta	58°C	de Saint Pierre, en preparación
I	M9	PCR-RFLP	Hinf I	gcagcataataaaacttca gg	aaaacctaaacttgcacaagc	54°C	de Saint Pierre, en preparación
J	M45	PCR-RFLP	Bfa I	Attggcagtgaaaaaattat agcta	tgcccttggctacaactccta	58°C	Bailliet et al., 2008
K	M242	PCR-RFLP	BsiHKAI	Tcagatggcaagattttaa gtaca	ttcatgcccctttatactgag	56°C	de Saint Pierre, en preparación
L	M304	PCR-RFLP	Mfe I	TAATCAGTCTCTCC AGCA	TAGGTACCAGCTCTCC AATT	61°C	Underhill et al., 2001
M	M207	PCR-RFLP	Dra I	ggggcaaatgtaagtaag c	tttctagctgttgcgtct	54°C	Bailliet et al., 2008

c) Caracterización de la región control

Se amplificó y se secuenció la región control ADN mitocondrial humano desde la posición 16024 hasta la 576, siguiendo las recomendaciones dadas en detalle por (Parson & Bandelt, 2007) para los 55 individuos que conforman la muestra. Las condiciones de amplificación están descritas en la Tabla I. La secuenciación y purificación de los productos del PCR se realizaron en Macrogen. Inc, Corea del Sur. Con el fin de evitar mutaciones originados por errores en la lectura, se aplicó un proceso de revisión, siguiendo algunos parámetros descritos por Bandelt et al., (2003). Las secuencias fueron alineadas de acuerdo a la *secuencia de referencia revisada de Cambridge* (rCRS) sujetas a los parámetros de control estándares (Bandelt & Parson, 2008), los polimorfismos fueron confirmados y las secuencias editadas directamente usando Geneious (Geneious versión 6.0.6). La asignación de haplogrupos para cada muestra se realizó de manera manual.

d) Caracterización a través de un Chip comercial

Se utilizó como modelo de genotipificación un microarray de alta densidad: Axiom Latino. Este Chip cuenta con un total de 817,810 marcadores, de los cuales 123 SNPs corresponden al ADN mitocondrial y 234 SNPs al Cromosoma Y. Se determinó el linaje materno y paterno a través del set de SNPs del ADN mitocondrial y del Cromosoma Y presentes en el chip a 55 individuos (para el Cromosoma Y se analizaron 28 individuos masculinos). La estrategia utilizada para este proceso se llevó a cabo con el lenguaje estadístico R (R Core Team 2013) mediante la propuesta de dos algoritmos desarrollados en este estudio, los cuales toman como datos de entrada dos matrices: una matriz con los haplogrupos descritos por el Chip, y otra matriz con la información del Chip para los SNPs del ADN mitocondrial y Cromosoma Y para cada individuo analizado en este estudio (Figura 3). La construcción de la matriz de haplogrupos en base al Chip será explicada más adelante. En la Figura 4 se puede apreciar el diagrama de flujo para los dos algoritmos junto con el detalle de cada operación. Para el primer algoritmo se puede ver que se realizó una comparación pareada entre estas dos matrices, donde se entregó como output sólo aquellos elementos de la matriz de haplogrupos del Chip que fueron idénticos a los elementos de la matriz de individuos. El segundo algoritmo realizó también esta comparación pareada desde la matriz de haplogrupos del Chip hacia la matriz de individuos, entregando esta vez para cada comparación dos indicadores: el primero corresponde básicamente a un valor que

muestra los SNPs en común que tiene cada elemento (haplogrupo) de la matriz del Chip con cada individuo de la otra matriz, mientras que el segundo indicador corresponde a la división entre el número de SNPs que posee cada individuo con cada haplogrupos de la matriz del Chip. Se entregó para cada individuo aquellos haplogrupos de la matriz del Chip con el mayor valor para el primer y segundo indicador.

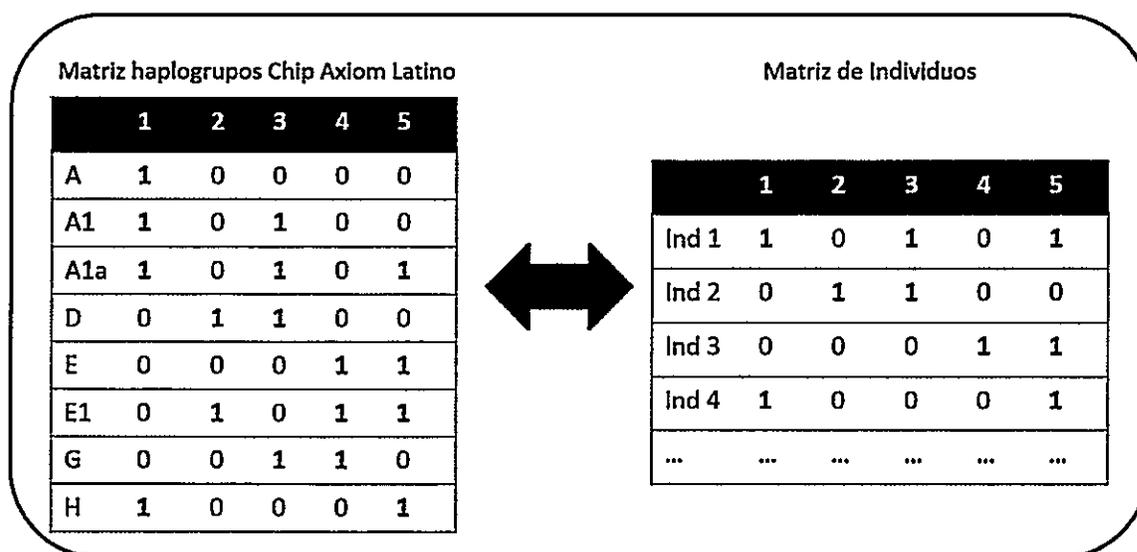


Figura 3. Esquema de comparación entre la matriz del Chip Axiom Latino con la matriz de los individuos. Estas dos matrices corresponden a los datos de entrada para el análisis de los algoritmos.

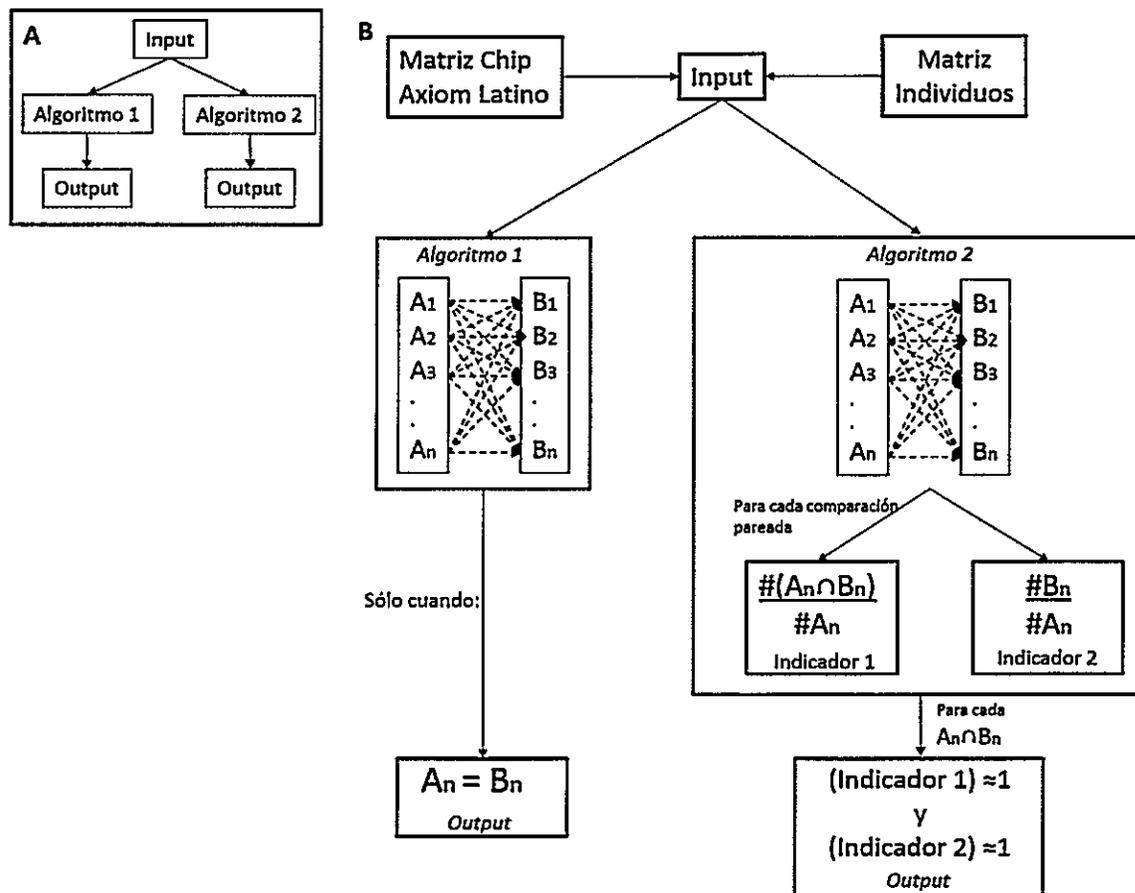


Figura 4. Figura con el diagrama de flujo para cada algoritmo. (A) diagrama básico para los dos algoritmos con los datos de entrada (input) y datos de salida (output). (B) diagrama en detalle para cada caja de flujo. Como datos de entrada (input) se tomaron la matriz del Chip con la matriz de individuos. Para el algoritmo 1 se realizó una comparación entre cada haplogrupo de la matriz del Chip ($A_1, A_2, A_3, \dots, A_n$) con cada individuo de la otra matriz ($B_1, B_2, B_3, \dots, B_n$), donde se entregó como output aquellos elementos (haplogrupos) de la matriz del Chip que son idénticos en todos sus SNPs a elementos de la matriz de individuos. El segundo algoritmo realizó la misma comparación que el algoritmo anterior, con la diferencia que entrega para cada comparación dos indicadores. El primer indicador calcula la intersección (número de SNPs en común) entre el haplogrupos de la matriz del chip con el individuo, luego cada valor se dividió por la longitud (número de SNPs totales de un linaje) del haplogrupo al que se comparó el individuo. Este indicador genera valores del cero al uno, siendo cero cuando no existe ningún SNP en común entre el individuo con un haplogrupo específico, y uno cuando todos los SNPs de un haplogrupo

del chip están presentes en la información para un individuo. El segundo indicador corresponde a la división entre la longitud del individuo (número de SNPs totales que presenta) con la longitud del haplogrupo al que se está comparando. Este indicador se utilizó para corregir los resultados del primer indicador en caso que existan haplogrupos incorrectamente asignados. El análisis de los indicadores se realizó de manera conjunta y dependiente, entregando como datos de salida aquellos haplogrupos que tuvieran el valor más alto para los dos indicadores para cada muestra.

d.1) Construcción de la matriz a través del Chip Axiom Latino

Para la construcción de la matriz del chip se realizó un análisis de los SNPs presentes en la filogenia del ADN mitocondrial y Cromosoma Y que consta de tres etapas (Figura 5). La primera etapa describe la cantidad de SNPs del chip comercial que se encuentran asociados a uno o más haplogrupos dentro de la filogenia del ADN mitocondrial y del Cromosoma Y comparándolas con dos bases de datos: <http://phylotree.org> y <http://isogg.org>.

En una segunda etapa se realizó una transposición de estos datos, y se confeccionó una matriz de haplogrupos con todos los linajes mitocondriales y del cromosoma Y. Estas matrices fueron convertidas a datos binarios, donde cero indica la ausencia del SNP sobre un haplogrupo determinado, mientras que el valor de uno señala presencia. La finalidad de esta conversión es para los análisis matemáticos posteriores. Finalmente, en la tercera etapa se procedió a analizar estos haplogrupos dentro de la matriz, con la intención de evaluar si hay linajes que estén representados por los mismos SNPs. Esta última etapa se analizó con el lenguaje estadístico de R (R Core Team 2013), a través de un algoritmo que permite analizar diferentes elementos de una matriz, en este caso, permite realizar una comparación múltiple de todas las filas (haplogrupos) entre sí, arrojando como resultado de salida diferentes combinaciones pareadas con el valor lógico de "TRUE" de todos aquellos haplogrupos que estén definidos por la misma combinación de SNPs.

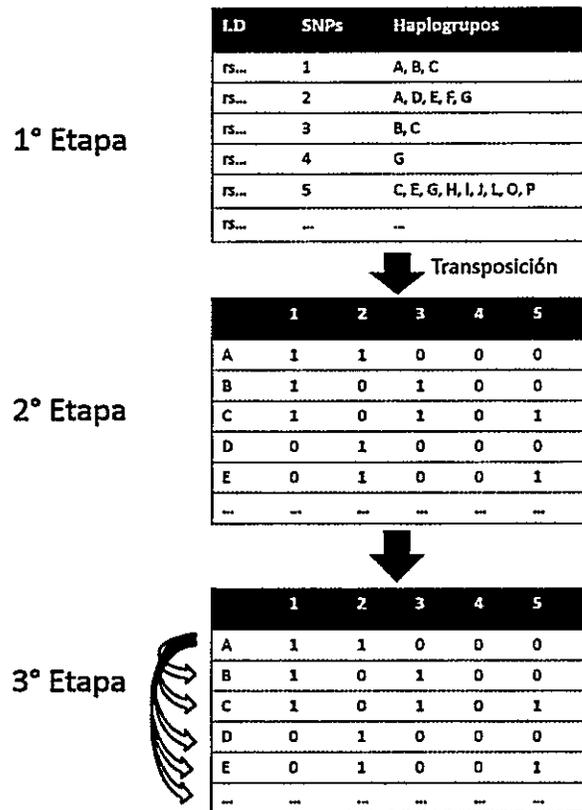


Figura 5. Esquema general con el procedimiento realizado para cada etapa durante el análisis del Chip.

e) Asignación de haplogrupos mediante el set mínimo de SNPs

Finalmente, con el sub-set de SNPs que se desarrolló a partir del set optimizado para poblaciones mestizas latinoamericanas se caracterizaron aquellos individuos de los cuales se poseía información genómica para ADN mitocondrial y Cromosoma Y. Los algoritmos utilizados fueron los mismos que en la caracterización a través del Chip, y se tomaron como datos de entrada dos matrices: la matriz con el sub-set de SNPs mínimos (optimizada), y otra matriz con la información de cada SNPs optimizado para cada individuo.

RESULTADOS

I.- Confección Tabla de Haplogrupos

1.- ADN mitocondrial

La búsqueda en la base de datos (<http://phylotree.org>) permitió seleccionar los linajes mitocondriales más frecuentes en cada continente. Los resultados muestran que en total se registraron 351 haplogrupos distribuidos en: América (38), África (38), Europa (72), Asia (132), y Oceanía (33). Cabe mencionar que hay 78 macro-haplogrupos en esta tabla que debido a la amplia distribución geográfica que poseen se encontraron en más un continente. En la figura 6 podemos ver la proporción de haplogrupos seleccionados para cada continente para el ADN mitocondrial. En el material anexo adjunto podemos ver en detalle la totalidad de los haplogrupos seleccionados.

2.- Cromosoma Y

Para el linaje paterno pudimos identificar un total de 97 haplogrupos principales distribuidos alrededor del mundo. Esta información proviene de la base de datos de la "International Society of Genetic Genealogy" (ISOGG, <http://isogg.org>), y comprende a América (11), Europa (38), Asia (19), África (17) y Oceanía (10). Al igual que con la tabla de haplogrupos mitocondriales, se encontraron 23 haplogrupos que se encuentran distribuidos sobre extensas áreas geográficas, específicamente macro-haplogrupos, por lo que aparecen en más de un continente. En la figura 6 podemos ver la proporción de haplogrupos seleccionados para cada continente para el Cromosoma Y. En el material anexo adjunto podemos ver en detalle la totalidad de los haplogrupos seleccionados.

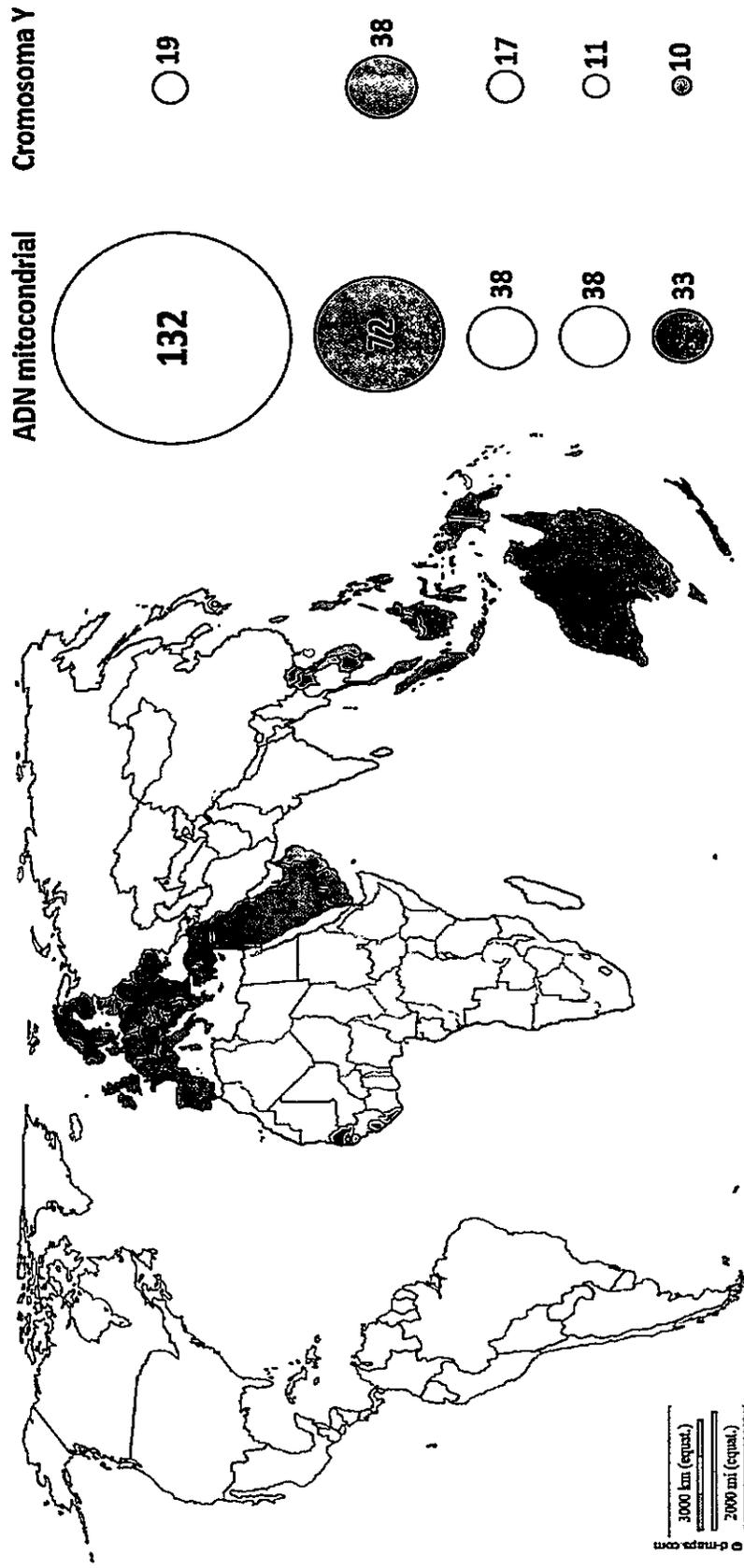


Figura 6. Distribución y frecuencia de los haplogrupos seleccionados alrededor del mundo para el ADN mitocondrial y Cromosoma Y. Se agruparon por cinco áreas: América, África, Europa-Medio Oriente, Asia y Oceanía.

II.- Confección de un set mínimo de SNPs para el ADN mitocondrial y Cromosoma Y.

1.- Optimización de un set mínimo de SNPs

a) ADN mitocondrial

La matriz construida a partir de la elaboración de la tabla de haplogrupos tiene 269 filas que corresponden a los haplogrupos descritos y 724 columnas que corresponden a los SNPs totales que los definen. El número de haplogrupos disminuyó debido a que se corrigieron aquellos haplogrupos que se encuentran en más de un continente. En la Figura 7 podemos ver la cantidad mínima de SNPs posibles que se eligieron para cada una de las permutaciones en el método de exclusión independiente. El rango varía de 183 a 207 SNPs, siendo 196 SNPs el set que apareció en más ocasiones (11.6%). En toda la muestra de haplogrupos se detectó un total de 143 SNPs de carácter homoplásico, los cuales están representados parcialmente en la Figura 8 donde vemos que las frecuencias absolutas de algunos SNPs poseen proporciones desiguales dentro de linajes (H8, L2b, L2c), lo que corresponden a SNPs de carácter homoplásicos. Para el resto de los haplogrupos se puede apreciar que las frecuencias absolutas de los SNPs son similares.

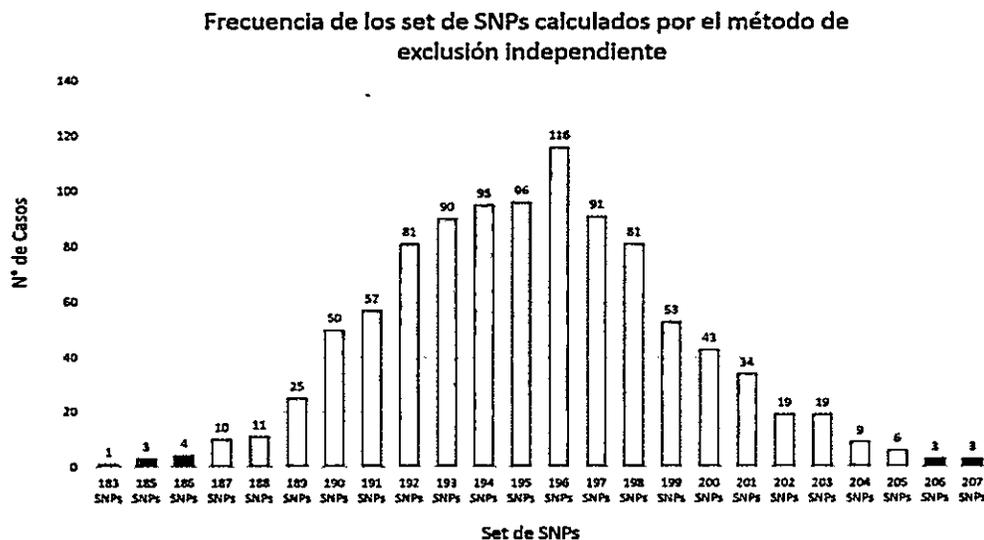


Figura 7. Frecuencia de los distintos set de SNPs calculado por el método de exclusión independiente en cada análisis de permutaciones.

b) Cromosoma Y

En lo que respecta al Cromosoma Y, se construyó una matriz con 71 haplogrupos (corregidos) caracterizados por un total de 301 SNPs. El número de SNPs optimizados para estos haplogrupos de acuerdo a las permutaciones para el método de exclusión independiente arrojó en un 8% (80 veces) a 68 SNPs, 39.5% (395 veces) a 69 SNPs y en un 52.5% (525 veces) a 70 SNPs. Solamente se detectaron 3 SNPs originados independientemente en clados monofiléticos distintos. La Figura 9 muestra la frecuencia de los SNPs dentro de cada haplogrupo (no todos están representados en la figura). Para la mayoría de los casos vemos que las frecuencias absolutas de los SNP dentro de un haplogrupo son similares, salvo aquellos haplogrupos que poseen SNPs compartidos producto de homoplasias como E1b1a1a1e/K y E/R1a1a.

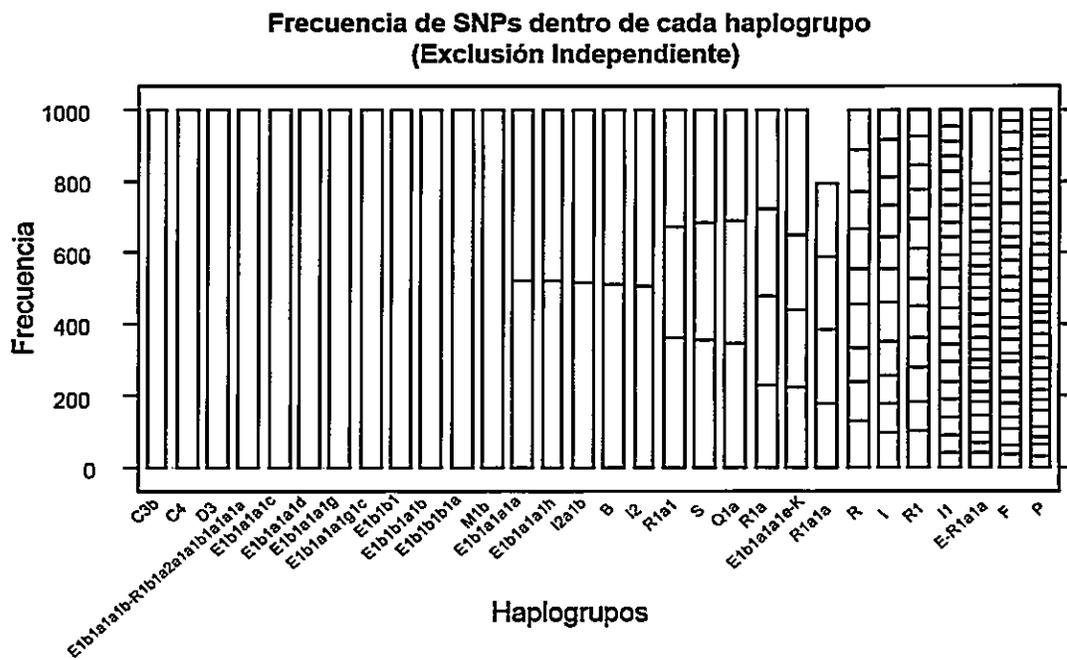


Figura 9. Gráfico con la frecuencia de cada SNP dentro de cada haplogrupo del cromosoma Y. La cantidad de colores dentro de cada haplogrupo representa el número de SNPs que lo definen, el tamaño de cada color corresponde a la frecuencia absoluta del SNP, esto quiere decir el número de veces que fue elegido ese SNPs por el algoritmo en las 1000 permutaciones.

Para el árbol de decisiones se seleccionaron en un 3% a 68 SNPs y en un 97% a 69 SNPs. En la Figura 10 vemos la frecuencia de los SNPs dentro de algunos haplogrupos. Como podemos ver existen SNPs que se encuentran sobre-representados sobrepasando las 1000 permutaciones.

Frecuencia de SNPs dentro de cada haplogrupo (Árbol de Decisiones)

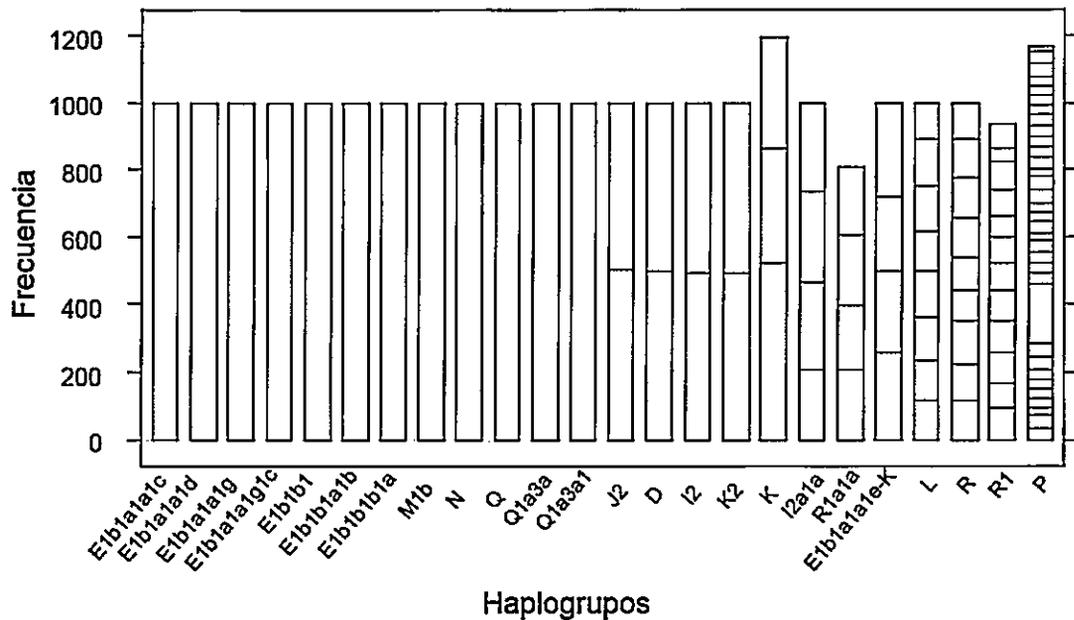


Figura 10. Gráfico con la frecuencia de cada SNP dentro de cada haplogrupo del cromosoma Y. La cantidad de colores dentro de cada haplogrupo representa el número de SNPs que lo definen, el tamaño de cada color corresponde a la frecuencia absoluta del SNP, esto quiere decir el número de veces que fue elegido ese SNPs por el algoritmo en las 1000 permutaciones.

Sobre estos dos algoritmos, realizamos la selección de los SNPs optimizados para el set de haplogrupos descritos a nivel mundial. Se tomó en cuenta que no fueran de carácter homoplásico, salvo un SNP (T21765281G) el cual aparece en los haplogrupos E1b1a1a1b y R1b1a2a1a1b1a1a1a y es único para cada linaje, por lo que se tuvo que incluir. El resultado para esta elección consta de un total de 70 SNPs optimizados para la caracterización de 71 haplogrupos del Cromosoma Y.

2.- Ajuste del set de SNPs optimizado para poblaciones mestizas de Latinoamérica

Para el ADN mitocondrial se seleccionó una sub-muestra de 79 haplogrupos definidos a través de 73 SNPs. La distribución de estos haplogrupos está concentrada en América (37), Europa (29) y África (13). Para el Cromosoma Y se seleccionó un total de 65 haplogrupos distribuidos en América (11), Europa (37) y África (17), mediante la caracterización de 64 SNPs.

IV.- Desarrollo Experimental

1.- ADN mitocondrial

a) Determinación de la composición Genética de la muestra mediante PCR-RFLP

Se caracterizó mediante PCR-RFLP a los 55 individuos incluidos en este estudio, las técnicas de amplificación y el uso de las enzimas de restricción se realizó para los cuatro macrohaplogrupos fundadores de América: A, B, C y D. Estos resultados arrojan un 5.45% (n=3) para el haplogrupo A, un 34.54% (n=19) para el haplogrupo B, un 20% (n=11) para el haplogrupo C y un 30.90% (n=17) para el haplogrupo D. Además, se encontraron cinco individuos no amerindios (9.09%). En la tabla V se encuentran las comparaciones entre cuatro técnicas distintas de caracterización del linaje mitocondrial para estos 18 individuos.

b) Caracterización de la región control

Se pudo secuenciar la región control para estos 47 individuos: dentro del haplogrupo A, se detectaron tres A2. Para el haplogrupo B se encontraron 11 B4b y seis B2i2. Dentro del haplogrupo C, se encontraron dos C1, un C1b, un C1d y cuatro C1b13. Finalmente dentro del haplogrupo D, pudimos encontrar seis D1, 10 D1g y un D4h3a5 (Tabla V).

c) Caracterización de haplogrupos mediante secuenciación completa

Sólo 18 individuos presentan información para sus secuencias genómicas. El resultado de la caracterización de haplogrupos a través de secuencias genómicas para ADN mitocondrial, a través del árbol filogenético *mtDNA Build Tree 15 (30 Sep 2012)* (van Over y Kaiser, 2009) nos revela que estos 18 individuos caen dentro de cinco clados filogenéticos principales, estos corresponden a A2, B2, C1, D1 y D4h. Dos individuos se encuentran dentro del clado A2, uno de ellos pertenece al haplogrupo A2af1a. Siete individuos pertenecen al haplogrupo B2, de los

cuales únicamente dos al linaje B2i2. Dentro de los cuatro individuos pertenecientes al clado C1, se encontró un individuo correspondiente al haplogrupos C1d1 y uno al C1b. De los cuatro individuos pertenecientes al haplogrupo D1, encontramos un D1g3 y un D1g2. Por último, sólo un individuo se encuentra dentro del clado D4h, específicamente del haplogrupo D4h3a5 (Tabla V).

d) Microarreglo de alta densidad: Chip Axiom Latino

d.1) Descripción del linaje materno dentro del Chip

De los 123 SNPs del chip, sólo 115 pudieron ser caracterizados en todos los individuos. De la totalidad de los SNPs mitocondriales analizados, sólo 5 (4,35%) no se encuentran descritos para algún linaje mitocondrial. Los 110 SNPs restantes (95,65%) aparecen en la filogenia del ADN mitocondrial ("mtDNA tree Build 15 (12 Sep 2012)". Dentro de estos SNPs sólo 23 (20%) se encuentran asociados a un solo haplogrupo, dejando un total de 87 (75,65%) SNPs que se van reiterando en más de un linaje mitocondrial (Tabla III).

Tabla III. Representación parcial de la asignación de los distintos haplogrupos en donde se encuentra asociado cada SNPs.

Tipos de Asociaciones entre los SNPs del chip comercial y los haplogrupos mitocondriales*			
ID	SNP	Haplogrupos	Casos Totales
rs3937039	A663G	A	23 (20%)
rs28721398	6481	-	5 (4,35%)
rs28411821	T9824C	M7, D4b2	
b36_MT_1050	C1048T	L0, M29, D5b	87 (75,65%)
rs35452858	T16325C	L3e1b, M18b, C1, G1a1, M17a, M36b, M76a, M80, D1, ...	
rs...	

(*). En esta tabla sólo se muestra la información de algunos SNPs de los 115 SNPs en total. En el primer caso vemos que el SNP 663 se encuentra presente únicamente en el haplogrupo A, se encontró un total de 23 casos con esta característica. La posición 6481 no se encontró en la filogenia del ADN mitocondrial junto con otros 4 SNPs. Los SNPs T9824C, C1048T y T16325C

aparecen asociados a más de un haplogrupo junto con otros 84 SNPs. La tabla completa tiene una dimensión de 117 filas y 4 columnas y la podemos encontrar en el material anexo.

La transposición de la matriz en la segunda etapa permitió ordenar la información en relación a cuantos haplotipos se encuentran representados por el chip (Tabla IV). El resultado de esta base de datos nos arroja un total de 423 haplotipos descritos bajo los 110 SNPs del chip. Cabe mencionar que dentro de estos 423 linajes mitocondriales se encuentran tanto haplogrupos como linajes terminales (haplotipos) (Figura 11).

Tabla IV. Matriz parcial de haplogrupos construido a partir de la figura anterior. Sólo se muestra el clado A (*).

	663	1598	1736	2706	3308	3927	8027	11365	11914	12007	12720	13708	15629	16325
A	A663G		A1736G											
A4b	A663G		A1736G								A12720G			
A4f	A663G		A1736G	G2706A										
A2	A663G		A1736G				G8027A			G12007A				
A2b	A663G		A1736G				G8027A	T11365C		G12007A				
A2e	A663G		A1736G				G8027A			G12007A		G13708A		
A2f3	A663G		A1736G				G8027A		G11914A	G12007A				
A2h1	A663G	G1598A	A1736G				G8027A			G12007A				
A2l	A663G		A1736G		T3308C		G8027A			G12007A				T16325C
A2l	A663G		A1736G				G8027A			G12007A			T15629C	
A2o	A663G		A1736G			A3972G	G8027A			G12007A				

(*). En esta tabla sólo se muestra la información de algunos haplogrupos. El resultado de la transposición completa se encuentra en material anexo. La tabla contiene una dimensión de 424 filas (423 haplotipos) y 116 columnas (115 SNPs).

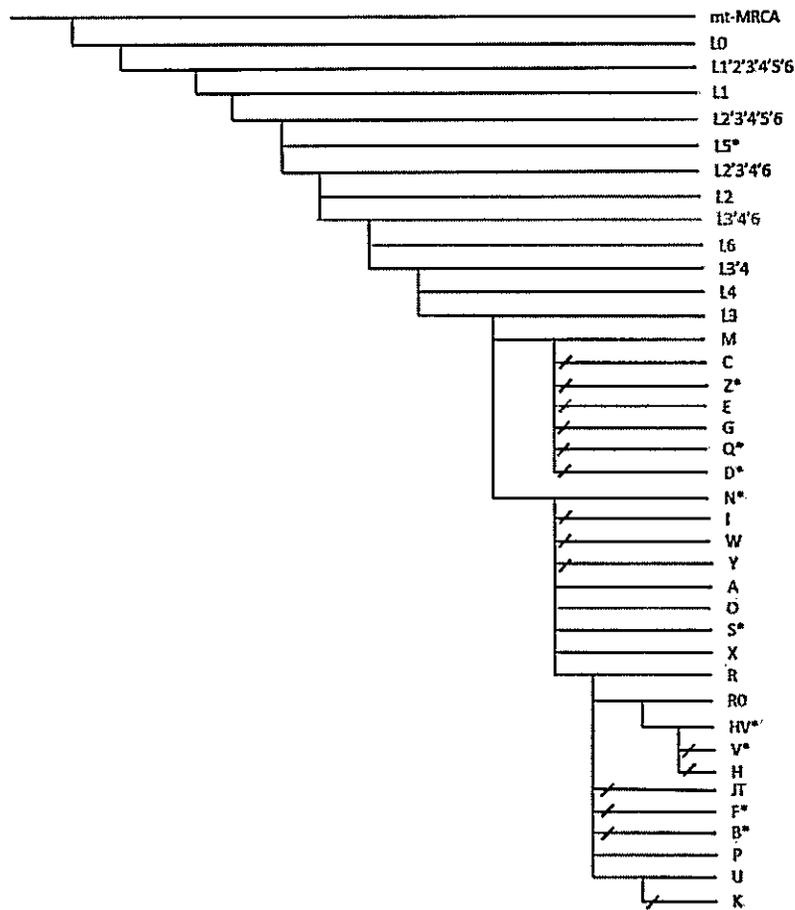


Figura 11. Caracterización de los macro-haplogrupos mitocondriales representados por el Chip Axiom Latino. En negro están los linajes representados por el Chip. Aquellos macro-haplogrupos que aparecen con un asterisco (*) no se encuentran representados por el chip, pero sí algunos de sus sub-haplogrupos. Se colapsaron aquellos macro-haplogrupos que no venían directamente del macro-haplogrupo que los precede en la figura con una barra en diagonal (/). La filogenia completa se encuentra en la página web <http://phylotree.org>

Los resultados para el tercer paso arrojaron la existencia de un total de 94 linajes en donde el set de SNPs que define a cada uno de estos, se encuentra a lo menos presente en otro linaje. En la mayoría de los casos existe una combinación de SNPs igual para dos haplogrupos, sin embargo también podemos encontrar combinaciones idénticas para tres, cuatro, e incluso para 5 haplogrupos distintos.

d.2) Caracterización de haplogrupos asociados al Chip

Para esta etapa, se analizaron a 45 individuos. Los dos algoritmos utilizados para la caracterización de haplogrupos dieron los mismos resultados sobre la muestra. Lo primero que podemos observar es que aparecen 15 individuos que fueron asignados a más de un haplogrupo, 14 de estos cayeron en dos resultados posibles: D1 y D4h2. La Figura 12 ilustra esta situación en donde vemos a ambos linajes caracterizados bajo los mismos SNPs de acuerdo al Chip. El otro individuo correspondiente a D4h3a5 fue caracterizado dentro de cuatro haplogrupos: M34, D4, D4c1b1, D1g2a. El resto de los individuos fueron caracterizados dentro de un linaje mitocondrial: tres individuos dentro del haplogrupo A2, 16 dentro del haplogrupo B2, siete dentro del haplogrupo C1, y finalmente uno dentro de D1g3. Por último, los dos individuos correspondientes a los haplogrupos L1b1 y L3e1a1 fueron caracterizados dentro de los haplogrupos L1b y L3e respectivamente. Dos individuos no se caracterizaron, dado que no se pudieron genotipificar en una primera instancia con el chip (Tabla V).

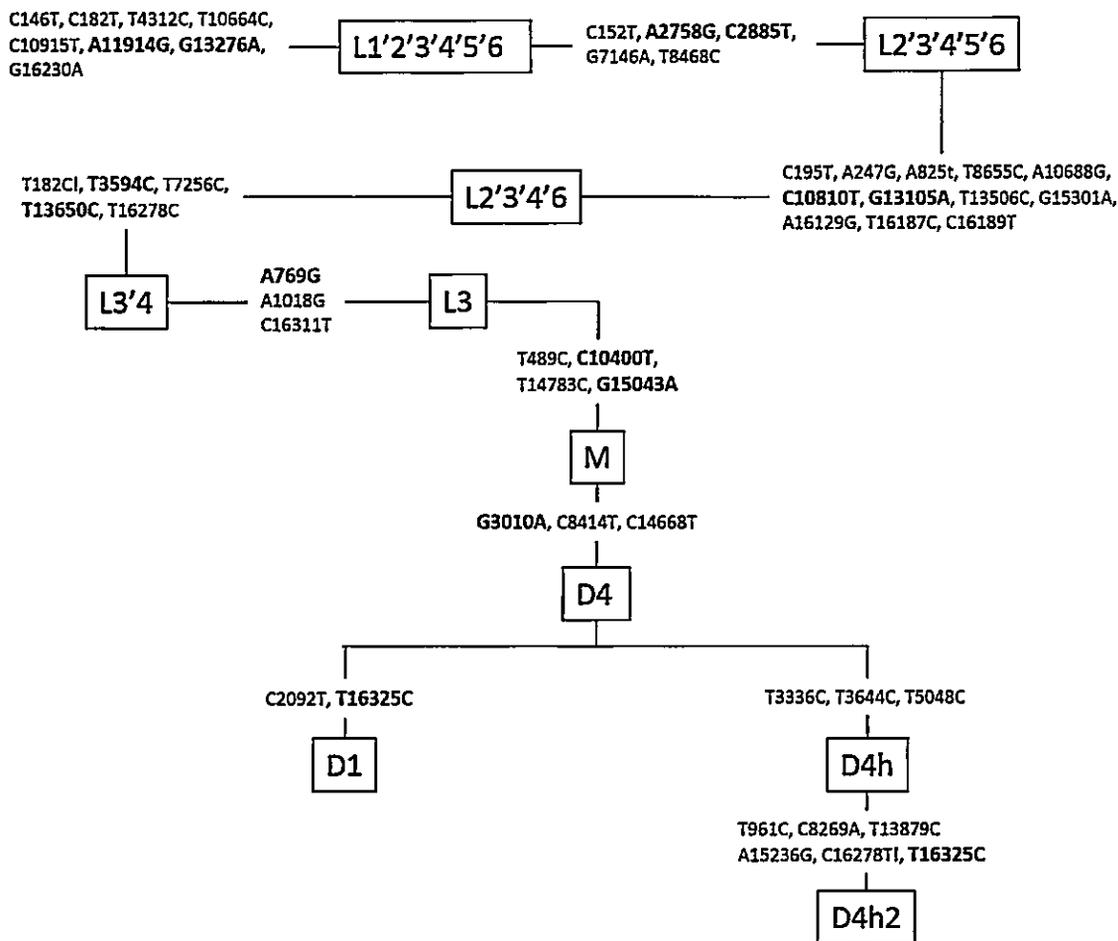


Figura 12. Esquema que muestra a dos haplogrupos en el chip caracterizados bajo el mismo set de SNPs. En los recuadros aparecen los haplogrupos, en rojo aquellos SNPs que aparecen en el chip Axiom Latino.

e) Asignación de haplogrupos por el set mínimo de SNPs

La caracterización de los 18 individuos con el set de SNPs optimizados que se construyó en este estudio nos muestra que cada sujeto fue asignado a su haplogrupos de manera correcta de acuerdo a lo visto en los diferentes métodos en la Tabla V: dos individuos fueron asignados al haplogrupo A2, cuatro a B2, dos a B2i2, tres individuos caracterizados a los haplogrupos C1, C1b y C1d respectivamente, un sujeto a D1, tres a D1g, y finalmente uno a D4h3a.

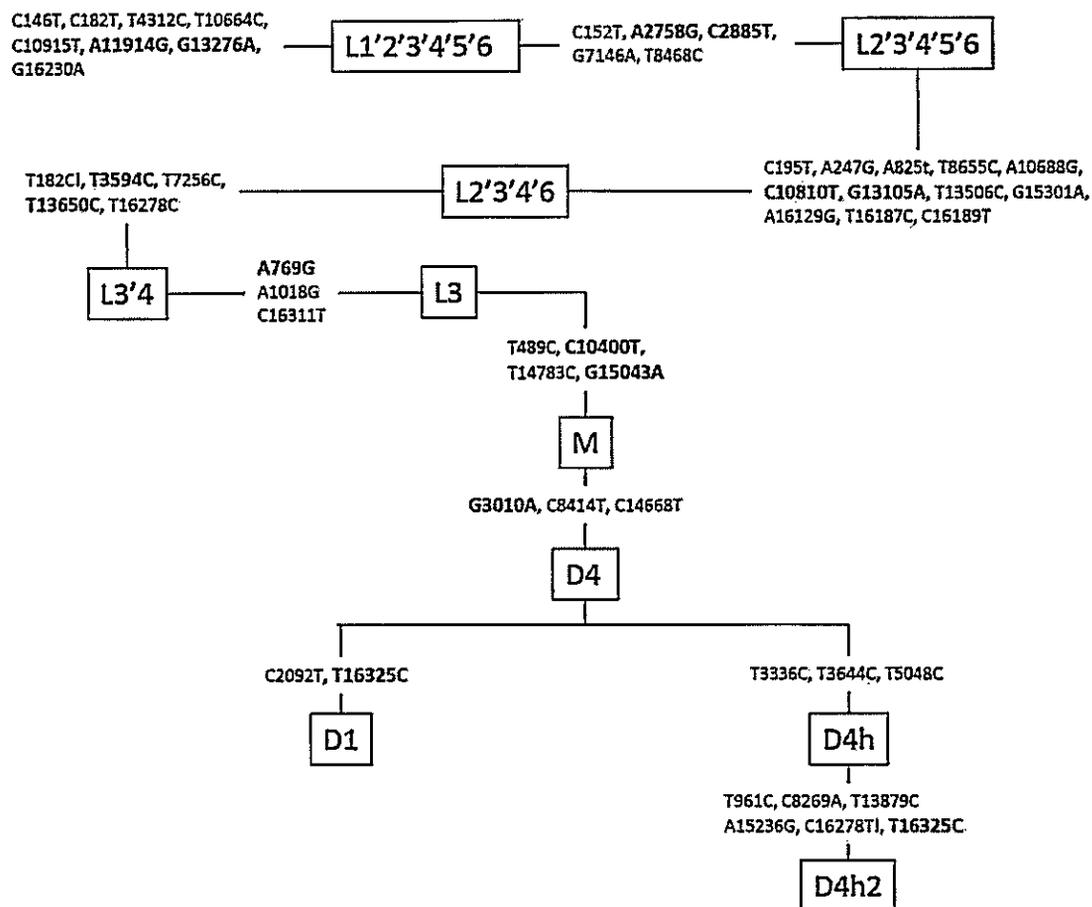


Figura 12. Esquema que muestra a dos haplogrupos en el chip caracterizados bajo el mismo set de SNPs. En los recuadros aparecen los haplogrupos, en rojo aquellos SNPs que aparecen en el chip Axiom Latino.

e) Asignación de haplogrupos por el set mínimo de SNPs

La caracterización de los 18 individuos con el set de SNPs optimizados que se construyó en este estudio nos muestra que cada sujeto fue asignado a su haplogrupos de manera correcta de acuerdo a lo visto en los diferentes métodos en la Tabla V: dos individuos fueron asignados al haplogrupo A2, cuatro a B2, dos a B2i2, tres individuos caracterizados a los haplogrupos C1, C1b y C1d respectivamente, un sujeto a D1, tres a D1g, y finalmente uno a D4h3a.

Tabla V. Comparación de cinco métodos diferentes para la caracterización de linajes mitocondriales sobre una muestra de 18 individuos.

Caracterización Linajes ADN mitocondrial				
Set de SNPs mínimo (16)	PCR-RFLP (55)	Región Control (47)	Chip Axiom Latino (45)*	Secuenciación Genómica (18)
A2 (2)	A (3)	A2 (3)	A2 (3)	A2 (1) A2af1a (1)
B2 (4)	B (19)	B4b (11)	B2 (16)	B2 (5)
B2i2 (2)		B2i2 (6)		B2i2 (2)
C1 (1)	C (11)	C1 (2)	C1 (7)	C1 (2)
C1b (1)		C1b (1)		C1b (1)
		C1b13 (4)		
C1d (1)		C1d (1)		C1d1 (1)
D1 (1)	D (17)	D1 (6)	D1,D4h2 (15)	D1 (1)
D1g (3)		D1g (10)		D1g (1)
				D1g2 (1)
D4h3a (1)		D4h3a (1)	D1g3 (1)	D1g3 (1)
	NA (5)	M34, D4, D4c1b1, D1g2a (1)		D4h3a5 (1)
		L1b1 (1)	L1b (1)	L1b1 (1)
		L3e1a1 (1)	L3e (1)	L3e1a1 (1)

* Los resultados de ambos algoritmos para la caracterización de haplogrupos fueron idénticos, por lo que se muestran en sólo una columna. Los valores entre paréntesis () indican el número de individuos dentro de ese haplogrupo o técnica de caracterización. El número de individuos entre cada método van variando de acuerdo a la disponibilidad de los resultados.

2.- Cromosoma Y

a) Composición Genética de la muestra mediante PCR-RFLP

Se caracterizaron mediante PCR-RFLP a 28 individuos. Las técnicas de amplificación y el uso de las enzimas de restricción se realizaron para los haplogrupos descritos en la tabla II. Los resultados arrojan un 46.43% (n=13) para el haplogrupo Q1a3a1-M3, un 50% (n=14) para el haplogrupo R-M207 y un 3.57% (n=1) para el haplogrupo E-M40. En la tabla VI se pueden ver los diferentes resultados de caracterización de acuerdo a las distintas técnicas empleadas en este estudio.

b) Caracterización de haplogrupos mediante Secuenciación Completa

La caracterización de haplogrupos a través de las secuencias obtenidas para estos 12 individuos concuerda con los resultados obtenidos en el paso anterior. Nueve individuos fueron caracterizados dentro del linaje Q1a3a1-M3. El haplogrupo final para los dos individuos pertenecientes al clado R-M207 corresponde al haplogrupo derivado R1b1a2a1a1-L11 (Tabla VI). El individuo perteneciente al clado E-M40 perteneció finalmente al haplogrupo E1b1a1a1

c) Microarreglo de alta densidad: Chip Axiom Latino

c.1) Descripción del linaje paterno dentro del Chip

Del total de los SNPs en este estudio, sólo 120 (51.1%) se encuentran asociados a la filogenia del Cromosoma Y. De estos, sólo uno (rs9786819) está relacionado a un marcador (P189) presente en dos haplogrupos distintos (E1b1a1, N1a), mientras que cada uno de los SNPs restantes están ligados a un solo haplogrupo respectivamente. Luego, se construyó una matriz en donde se transpusieron los datos y se ordenaron los diferentes haplogrupos descritos anteriormente. Los resultados muestran que los SNPs del Cromosoma Y presentes en el chip nos permiten caracterizar a 43 haplogrupos distintos, los cuales se distribuyen en su mayoría en el clado R (14 haplogrupos) y E (9 haplogrupos) (Figura 13). Por último, analizamos la matriz mediante un algoritmo implementado en R encontrando que todos estos se definen por set único de SNPs.



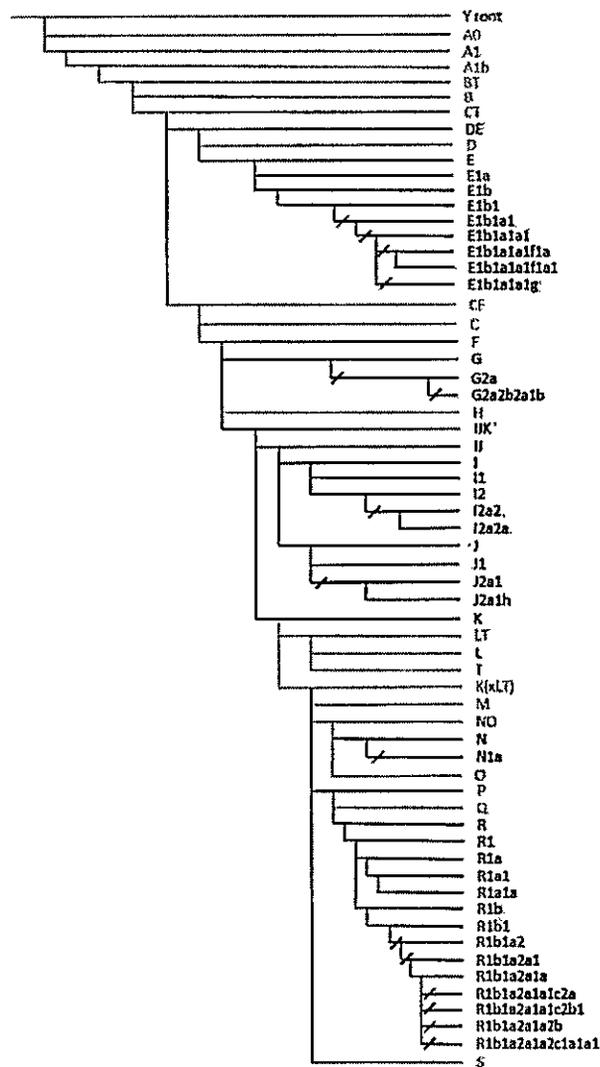


Figura 13. Filogenia del Cromosoma Y (correspondiente al Y-DNA Haplogroup Tree 2012, <http://isogg.org>), en negro los linajes representados por el chip comercial de Affymetrix. Aquellos linajes con una barra diagonal (/) señalan que no provienen directamente del haplogrupo que los precede en la figura.

c.2) Caracterización de haplogrupos asociados al Chip

Los resultados que arrojaron los dos algoritmos empleados en esta fase fueron idénticos para cada individuo. Los 13 individuos que de acuerdo a la secuenciación genómica pertenecen a Q1a3a1-M3 fueron asignados al haplogrupo P. Los 14 individuos pertenecientes al haplogrupo

R1b1a2a1a1-L11 fueron asignados a R1b1a2a1a-L51. Por último, el individuo perteneciente al haplogrupo E-M40 fue caracterizado como E1b1a1a1g-U175 (Tabla VI).

d) Asignación de Haplogrupos por el set de SNPs optimizados

La caracterización para el Cromosoma Y a través del set mínimo de SNPs nos muestra resultados muy similares a la secuenciación genómica y el Chip (Tabla VI). Nueve individuos fueron caracterizados dentro del haplogrupos Q1a3a1-M3, un sujeto a E1b1a1a1g, y dos a R1b1a2a1a.

Tabla VI. Comparación de cuatro métodos diferentes para la caracterización de linajes mitocondriales sobre una muestra de 12 individuos.

Caracterización Linajes Cromosoma Y			
Set de SNPs optimizados (12)*	PCR-RFLP (28)	Chip Axiom Latino (28)*	Secuenciación Genómica (12)
Q1a3a1-M3 (9)	Q1a3a1-M3 (13)	P-M45 (13)	Q1a3a1-M3 (9)
E1b1a1a1g-U175 (1)	E-M40 (1)	E1b1a1a1g-U175 (1)	E1b1a1a1-M180 (1)
R1b1a2a1a-M412 (2)	R-M207 (14)	R1b1a2a1a-M412 (14)	R1b1a2a1a1-L11 (2)

* Los resultados de ambos algoritmos para la caracterización de haplogrupos fueron idénticos, por lo que se muestran en sólo una columna. Los valores entre paréntesis () indican el número de individuos pertenecientes al haplogrupos caracterizado.

3.- Comparación entre la Tabla de Haplogrupos y el Microarreglo de alta Densidad

Se compararon los datos obtenidos de la revisión bibliográfica para la distribución de los linajes mitocondriales en América, África, Europa, Asia y Oceanía con los resultados de la caracterización de haplogrupos del Chip Axiom Latino, obteniendo los siguientes resultados: De un total de 38 haplogrupos encontrados en el continente americano, sólo 12 (32%) se encuentran representados en el chip. De los 38 haplogrupos del continente africano, sólo 18 aparecen en el chip (47.4 %). Para Europa existen 24 haplogrupos (34.7%) que están en el chip de un total de 72 que se describieron para este continente. De los 132 haplogrupos descritos para Asia, sólo 28

(21.2%) se encuentran en el Chip. Finalmente, de los 33 haplogrupos descritos para Oceanía, 10 (30.3%) están en el Chip.

Para el cromosoma Y se realizó esta misma comparación, pudiendo encontrar que para el continente africano sólo hay 3 (17.6%) haplogrupos de los 17 que se describieron. Para Europa se encontraron 38 haplogrupos principales, de los cuales 15 (39.4%) están en el Chip. En Asia se pudieron encontrar sólo 6 (31.6%) haplogrupos de los 18 que se describieron para este continente. Podemos notar que tanto para el continente americano como para Oceanía no se encontró ningún haplogrupo en el Chip (Figura 14).

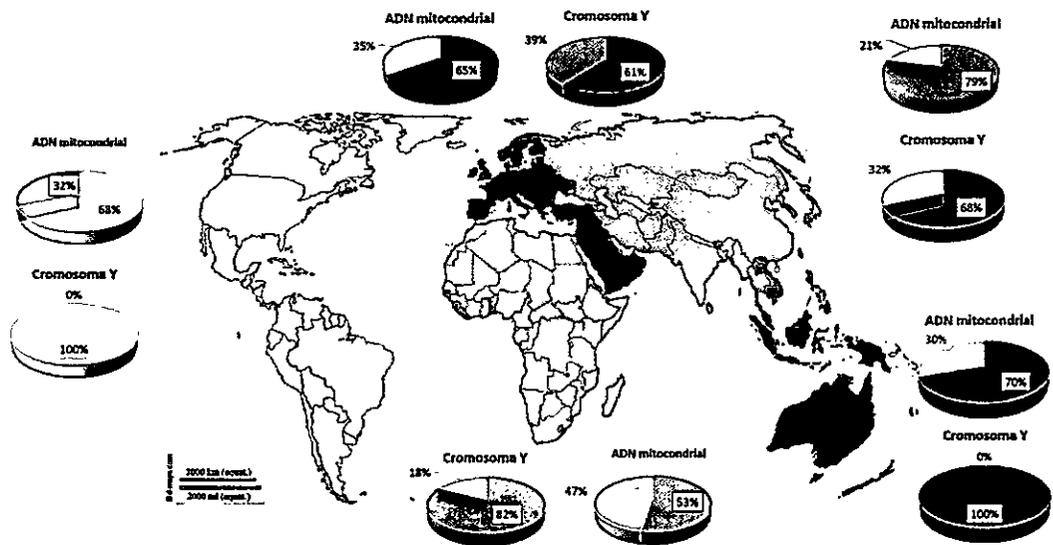


Figura 14. Comparación entre la tabla de haplogrupos con el chip Axiom Latino para ADN mitocondrial y Cromosoma Y. Para cada continente hay dos gráficos de torta: para ADN mitocondrial y Cromosoma Y. La porción en gris en cada gráfico representa el porcentaje de linajes que aparece en el Chip, la otra porción que varía de color de acuerdo al continente representa el porcentaje de linajes que sólo se encuentra en la tabla de haplogrupos.

DISCUSIÓN

I.- Elaboración Tabla de Haplogrupos

La confección de una base de datos a nivel mundial que contenga la distribución principal de haplogrupos tanto para el ADN mitocondrial como el Cromosoma Y resulta un proceso indispensable para nuestra investigación, puesto que es vital poder seleccionar la cantidad de haplogrupos necesarias para poder caracterizar de manera óptima el linaje materno y paterno a cualquier individuo de cualquier población en cualquier parte del mundo, en especial a las poblaciones mestizas en Latinoamérica, ya que son producto de la mezcla inicial de tres poblaciones principales (nativos americanos, europeos y africanos). El número total de linajes mitocondriales que se han descrito hasta Enero de este año son 4810 haplotipos (<http://phylotree.org>). Si bien es cierto que el número de linajes mitocondriales que se seleccionaron en esta etapa (351) está por debajo del 10% del total de haplotipos descritos, los linajes seleccionados están dentro de los niveles de haplogrupos y sub-haplogrupos, los cuales son capaces de representar casi la totalidad de la diversidad genética mundial.

El resultado de estas bases de datos nos muestra que para el ADN mitocondrial (n=351) se encontró un número mayor de haplogrupos que para el Cromosoma Y (n=97), lo cual puede deberse a dos motivos: 1) existe mayor información en genética de poblaciones humana para el ADN mitocondrial que para el Cromosoma Y como lo muestra la Figura 15 obtenido a partir de la base de datos en PUBMED (<http://www.ncbi.nlm.nih.gov/pubmed/>), y 2) la tasa de mutación para el ADN mitocondrial es más alta que para el Cromosoma Y, lo que se refleja en que haya una mayor diversidad genética para el linaje materno (Underhill & Kivisild 2007).

Número de Trabajos en Genética de Poblaciones Humana para ADN mitocondrial y Cromosoma Y

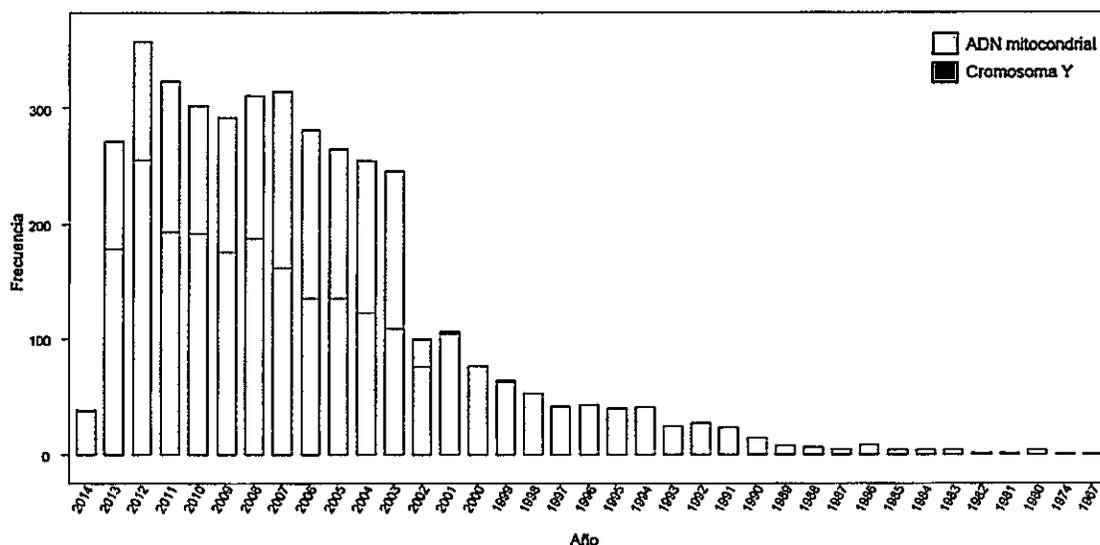


Figura 15. Número de citas para el ADN mitocondrial y Cromosoma Y en los últimos 40 años en Genética de Poblaciones Humana.

Los distintos criterios de inclusión de estos haplogrupos hacen de esta tabla una buena herramienta de caracterización para las poblaciones mestizas en Chile y el resto de Sudamérica, especialmente porque se incluyen linajes específicos y fuertemente asociados a estas poblaciones, producto de la mezcla inicial entre grupos nativos americanos y europeos, y posteriormente de grupos africanos. Para el linaje materno tenemos una completa caracterización hacia los clados nativos americanos A2, B2, C1, D1, D4h3a y X con sus respectivos sub-haplogrupos (Schurr et al., 1990; Torroni et al., 1992; Bailliet et al., 1994; Merriwether and Ferrell, 1996; Bandelt et al., 2003), adicionando especialmente una mayor profundidad de resolución para el cono sur de Sudamérica (Chile y Argentina) con los haplogrupos B2i2, C1b13 y D1g distribuidos en poblaciones del centro-sur de estos países (Bodner et al., 2012; de Saint Pierre et al., 2012) pudiendo revelar patrones de micro-diferenciación local en poblaciones mestizas actuales (Pezo et al., en revisión). Por otro lado se incorporaron aquellos haplogrupos americanos de origen africano debido al tráfico de esclavos en América (13 linajes principales) (Coelho et al., 2009; Coudray et al., 2009; Gonder et al., 2007; Non et al., 2011; Salas et al., 2004, 2005). Para la diversidad genética del linaje materno en Europa, se incorporaron los 72

haplogrupos más frecuentes, enfocados principalmente a aquellos distribuidos en la península ibérica, los cuales podrían estar en menor grado presentes en poblaciones mestizas en Latinoamérica producto de migraciones posteriores a la conquista (Maca-Meyer et al., 2003; Achilli et al., 2004; Dubut et al., 2004; Loogväli et al., 2004; Malyarchuk et al., 2006; Ingman y Gyllensten 2007; Brandstätter et al., 2008; Malyarchuk et al., 2008; Rose et al., 2008; Álvarez-Iglesias et al., 2009; Ennafaa et al., 2009; Kujanová et al., 2009; van Oven y Kayser 2009; Achilli et al., 2011).

Para el Cromosoma Y tenemos una completa representación de los principales macrohaplogrupos del viejo continente: E-M40, F-M45, G-M201, I-M258, J-M304 y R-M207 (Semino et al., 2000; Flores et al., 2004; Pericic et al., 2005; Adams et al., 2008; Karafet et al., 2008; Francalacci et al., 2010), donde además dentro de este último clado, se encuentra el haplogrupo R1b-M343 y R1b1a2a1a-M412, los linajes más frecuentes en el oeste de Europa (Francalacci y Sanna 2008; Francalacci et al., 2010; Myres et al., 2011; Niederstätter et al., 2012). Además, pusimos especial énfasis en detectar aquellos haplogrupos asociados a la península ibérica dado que a priori se espera que estén presentes en elevadas frecuencias en poblaciones mestizas de Latinoamérica dada nuestra condición inicial de mezcla asimétrica, por esta razón se incluyó por ejemplo el haplogrupo E1b1b1b1a-M81 asociado fuertemente con la influencia musulmana en España (Semino et al., 2004; Adams et al., 2008; Francalacci et al., 2010), así como también se encuentran los haplogrupos I1-M253 e I2a1a-M26 originados en España (Francalacci y Sanna 2008; López-Parra et al., 2009; Martínez-Cruz et al., 2012). Para el continente americano se incluyó la totalidad de los haplogrupos descubiertos (11) dada su baja variabilidad en comparación con el resto de los continentes, principalmente el europeo, puesto que solamente posee al haplogrupo Q1a3a1-M3 como el linaje nativo americano más frecuente en toda América (Underhill et al., 1996; Karafet et al., 2008; Geppert et al., 2011). El clado E corresponde al linaje más frecuente y diverso en África (Karafet et al., 2008), del cual se incluyeron aquellos subhaplogrupos más frecuentes en ese continente.

II.- Confección de un set mínimo de SNPs para el ADN mitocondrial y Cromosoma

Y.

En este trabajo se propone una tabla de haplogrupos con la que somos capaces de representar los mayores haplogrupos y sub-haplogrupos mundiales. La información contenida en estas tablas están asociadas a 724 SNPs para el ADN mitocondrial y 301 SNPs para el Cromosoma Y. Sin embargo, con la confección del set mínimo de SNPs somos capaces de reducir esta información en un 28% (de 724 a 196 SNPs) para el ADN mitocondrial y en un 23% (de 301 a 70 SNPs) para el Cromosoma Y manteniendo el mismo nivel de resolución. La variación en el número de SNPs que son seleccionados en cada permutación para el ADN mitocondrial nos da cuenta de la elevada variabilidad de este genoma a diferencia del Cromosoma Y, en donde el orden de los SNPs va a incidir directamente sobre el número de SNPs para cada análisis. Debido a la elevada tasa de mutación del ADN mitocondrial en comparación con el Cromosoma Y, se observó una mayor cantidad de SNPs homoplásicos. Entre los dos algoritmos que se aplicaron para este estudio, sólo el método de exclusión independiente pudo responder plenamente las exigencias y condiciones planteadas en este estudio, a pesar de que ambos métodos están igualmente enfocados hacia el objetivo principal de optimización. Esto se debe básicamente a la sensibilidad del árbol de decisiones para el tamaño de la matriz, pudiendo funcionar muy bien frente al Cromosoma Y, pero cae cuando se trata de matrices grandes (ADN mitocondrial), donde el método desarrollado en este estudio pudo llevar a cabo la tarea. Además, el método de árbol de decisiones siempre va a dejar un haplogrupo con la ausencia de todo el set de SNPs que se hayan utilizados para ese análisis. El método de exclusión independiente corresponde a un algoritmo nuevo y que fue diseñado específicamente para este trabajo, por lo cual cumple con todos los requerimientos para este estudio, analizando por separado cada columna de la matriz tomando decisiones respecto a la evaluación de las dos condiciones planteadas.

Dada la elevada cantidad de SNPs homoplásicos que se encontró para el ADN mitocondrial, pudimos optimizar el set de marcadores a un número que está muy por debajo del total de los haplogrupos descritos para este linaje materno, a diferencia de lo que ocurre con el Cromosoma Y donde el número del set de SNPs es muy similar al número de haplogrupos descritos, llegando a la asociación de un haplogrupos por cada SNP. Respecto a la selección del set de SNPs ideal para el ADN mitocondrial, se seleccionó el que más veces se repitió durante el

análisis de las permutaciones ya que representa un equilibrio entre los SNPs homoplásicos y que el set sea óptimo. Si se hubiese elegido el menor set de SNP (183) se hubiesen estado escogiendo muchos SNPs homoplásicos, lo que genera menor flexibilidad y mayores errores de caracterización al set de SNPs ideal en caso de remover algún haplogrupo.

III. Desarrollo Experimental

1.-ADN mitocondrial

Las técnicas de PCR-RFLP, secuenciación de la región control y secuenciación genómica son bastante utilizadas en la actualidad por ser bastante robustas e inequívocos para la caracterización del linaje materno. En nuestro estudio, estos tres métodos mostraron los mismos resultados, pero con niveles de resolución distintos como era de esperarse dada la naturaleza de cada técnica, siendo los métodos de secuenciación directa (especialmente genómica) los que entregan mayor detalle para la caracterización de un individuo. En relación a la muestra analizada podemos decir que tenemos una buena imagen de los principales linajes nativo americanos presentes actualmente en poblaciones urbanas de América: A2, B2, C1, D1 y D4h3, también aquellos linajes que se encuentran descritos para el cono sur de Sudamérica; B2i2 y D1g (Bodner et al., 2012; de Saint Pierre et al., 2012). El Chip posee una baja capacidad de resolución mostrando resultados similares con PCR-RFLP, además es incapaz de resolver algunos haplogrupos. Esto es muy diferente con el set mínimo de SNPs, que posee una alta resolución para la caracterización de haplogrupos, muy similar a los resultados de la secuenciación genómica.

2.-Cromosoma Y

A pesar de que la caracterización mediante la secuenciación genómica del Cromosoma Y resultó ser mucho más específica que PCR-RFLP, es este último el método más ampliamente utilizado para este propósito dado sus facilidades en la práctica y en lo económico. En nuestro estudio, ambos métodos arrojaron los mismos resultados en diferente resolución. A diferencia del ADN mitocondrial, la caracterización mediante el chip Axiom Latino mostró casi los mismo resultados que la secuenciación genómica, debido principalmente a la sobre-representación de

los clados E y R en comparación del resto de los linajes descritos por el chip, sin embargo se aprecia la carencia de representación del linaje nativo americano Q1a3a1-M3 el cual estuvo obviamente representado en el set de SNPs optimizado. Contrario a lo que se esperaba de una población mestiza, sólo se pudieron caracterizar dos individuos dentro del clado europeo R, lo cual no nos dio chance para poder evaluar otros linajes frecuentes de Europa, por ejemplo los haplogrupos I, J, F, K (Semino 2000; Flores et al., 2004; Pericic et al., 2005; Karafet et al., 2008). La razón de esta situación se debe a que las muestras analizadas se escogieron intencionalmente con una elevada mezcla indígena.

IV. Comparación del set mínimo de SNPs con el microarreglo de alta densidad y la secuenciación genómica.

Contrario a lo que se puede pensar, el elevado número de haplotipos que se pueden describir con el chip comercial para el ADN mitocondrial (423) no significa que haya una mejor resolución en la caracterización de los linajes, y se debe básicamente a que al parecer no están seleccionados bajo los criterios de representación que se usaron para la elaboración de la tabla de haplogrupos. Esto queda de manifiesto en la parte experimental donde los SNPs del chip sólo pueden identificar los grandes macro-haplogrupos A2, B2, C1 y D1, resultado similar a los niveles de resolución encontrados para el análisis de PCR-RFLP. Además, existe otro problema fundamental que radica en la incapacidad del chip en algunos casos para diferenciar entre un haplogrupo de otro debido a la elevada presencia de SNPs homoplásicos, como es el caso del haplogrupo D1 el cual no se puede distinguir del D4h2 y del haplogrupo D4h3a5 el cual es caracterizado dentro de los linajes M34, D4, D4c1b1 y D1g2a. De hecho, alrededor del 79.1% de los SNPs dentro del chip son homoplásicos y están propensos a mutar más que el resto de las posiciones. Estos *hot spots* con frecuencia causan homoplasias que complican la construcción de la filogenia (Ballard et al., 2014; Rubinoff, 2006; Underhill & Kivisild, 2007; van Oven y Kayser 2008; Soares et al., 2009). Para el Cromosoma Y esta situación cambia drásticamente, ya que la tasa de sustitución es mucho menos que el ADN mitocondrial, provocando que los árboles inferidos del Cromosoma Y sean más robustos (Underhill et al., 2000; Underhill & Kivisild, 2007) y con menores niveles de homoplasia (Jobling et al., 1996). Lo anterior –junto con la sobre-

representación del clado R- hace que la caracterización del linaje paterno con el microarreglo y con el set de SNPs optimizado se mantengan bastante similares, con excepción del haplogrupo nativo americano Q1a3a1-M3.

Para poder comprobar la baja capacidad de caracterización del chip frente a la tabla de haplogrupos se realizó un cruce entre aquellos linajes del ADN mitocondrial y Cromosoma Y del chip Axiom latino presentes en la tabla de haplogrupos (Figura 14), donde vemos que no sobrepasa del 35% en la mayoría de los casos. Por lo tanto, frente a los SNPs presentes en un microarreglo de alta densidad, el set de SNPs optimizados provenientes de la tabla de haplogrupos diseñada en este estudio resulta ser mucho más eficaz para la caracterización de los linajes uniparentales en poblaciones mestizas de Latinoamérica.

Finalmente, la baja cantidad de SNPs para caracterizar haplogrupos en el ADN mitocondrial y Cromosoma Y le confieren diferentes ventajas para su uso en aplicaciones prácticas: 1) dado el bajo número de SNPs para el ADN mitocondrial y el Cromosoma Y, se pueden añadir en plataformas actuales o futuras de genotipificación, y 2) la capacidad de flexibilidad de este set mínimo nos permite seleccionar sub-set de SNPs enfocado a las propiedades de la población en estudio, en nuestro caso las poblaciones mestizas de Latinoamérica, mediante técnicas más baratas y flexibles como la genotipificación de SNPs de Fluidigm que permite la caracterización de hasta 96 SNPs en hasta 96 muestras en forma simultánea. (<http://fluidigm.com/snp-genotyping.html>).

CONCLUSIONES

- La confección de una tabla de haplogrupos para el ADN mitocondrial y Cromosoma Y es un proceso vital e indispensable para la caracterización de linajes uniparentales mediante la optimización de un set de SNPs. Especialmente sobre poblaciones mestizas en Latinoamérica dado que contribuyeron a su formación distintas poblaciones provenientes principalmente de tres continentes (América, Europa y África).
- Los haplogrupos seleccionados para el ADN mitocondrial y Cromosoma Y cubren la mayor parte de los haplotipos descritos hasta el momento para estos linajes.
- Se logró reducir el número de SNPs para la caracterización del ADN mitocondrial y Cromosoma Y bajo el método de exclusión independiente. El método de árbol de decisiones no permite trabajar con matrices grandes.
- Para la caracterización de los individuos, el set de SNPs optimizados tanto para el ADN mitocondrial como el Cromosoma Y fueron capaces de asignar correctamente el haplogrupo a un individuo y con una resolución similar a los resultados obtenidos con la información de las secuencias genómicas.
- El microarreglo de alta densidad analizado en este estudio no es apto para la caracterización de linajes uniparentales dada su baja resolución (similar a PCR-RFLP) e incapacidad en algunos casos de diferenciar un haplogrupo del otro –con la excepción del Cromosoma Y. además, no cubre más del 35% de los linajes para ADN mitocondrial y Cromosoma Y presentes en la tabla de haplogrupos.
- Los resultados de nuestros trabajos nos permitió generar un set de SNPs optimizados con el fin de caracterizar el componente genético de una población a través del ADN mitocondrial y Cromosoma Y, mediante: 1) la elaboración de una tabla de haplogrupos, 2) el diseño e implementación de un algoritmo capaz de reducir y optimizar el número de SNPs dentro de un haplogrupo para poder caracterizar un individuo con la mínima información posible, 3) el diseño e implementación de un algoritmo de caracterización de linajes uniparentales.

BIBLIOGRAFÍA

- Abe-Sandes., K, Silva, W., Zago, M. (2004). Heterogeneity of the Y Chromosome in Afro-Brazilian Populations. *Human Biology*, 76, 77-87
- Achilli, A., Olivieri, A., Pala, M., Hooshiar Kashani, B., Carossa, V., Perego, U. a, ... Torrioni, A. (2011). Mitochondrial DNA backgrounds might modulate diabetes complications rather than T2DM as a whole. *PloS One*, 6(6), e21029. doi:10.1371/journal.pone.0021029
- Achilli, A., Perego, U. a, Bravi, C. M., Coble, M. D., Kong, Q.-P., Woodward, S. R., ... Bandelt, H.-J. (2008). The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PloS One*, 3(3), e1764. doi:10.1371/journal.pone.0001764
- Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., ... Torrioni, A. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *American Journal of Human Genetics*, 75(5), 910–8. doi:10.1086/425590
- Acuña M, Llop E, Rotthammer F. 2000. Composición genética de la población chilena: las comunidades rurales de los valles del Elqui, Limarí y Choapa. *Revista Médica de Chile* [Internet] 128(6):1-13. Available from: http://www.scielo.cl/scielo.php?pid=S0034-98872000000600004&script=sci_arttext
- Adams, S. M., Bosch, E., Balaesque, P. L., Ballereau, S. J., Lee, A. C., Arroyo, E., ... Jobling, M. a. (2008). The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *American Journal of Human Genetics*, 83(6), 725–36. doi:10.1016/j.ajhg.2008.11.007
- Avena S, Via M, Ziv E, Pe´rez-Stable EJ, Gignoux CR, et al. (2012) Heterogeneity in Genetic Admixture across Different Regions of Argentina. *PLoS ONE* 7(4): e34695. doi:10.1371/journal.pone.0034695
- Baeta, M., Núñez, C., Sosa, C., Bolea, M., Casalod, Y., González-Andrade, F., ... Martínez-Jarreta, B. (2012). Mitochondrial diversity in Amerindian Kichwa and Mestizo populations from Ecuador. *International Journal of Legal Medicine*, 126(2), 299–302. doi:10.1007/s00414-011-0656-4
- Bailliet, G., Ramallo, V., Muzzio, M., García, A., Santos, M. R., Alfaro, E. L., ... Demarchi, D. a. (2009). Brief communication: Restricted geographic distribution for Y-Q* paragroup in South America. *American Journal of Physical Anthropology*, 140(3), 578–82. doi:10.1002/ajpa.21133

- Bailliet, G., Rothhammer, F., Carnese, F. R., Bravi, C. M., & Bianchi, N. O. (1994). Founder mitochondrial haplotypes in Amerindian populations. *American Journal of Human Genetics*, *55*(1), 27–33.
- Ballard, J. W. O., Rand, D. M., & Ballard, O. (2014). THE POPULATION BIOLOGY OF MITOCHONDRIAL DNA AND ITS PHYLOGENETIC IMPLICATIONS, *36*(2005), 621–642. doi:10.1146/annurev.ecolsys.36.091704.175513
- Bandelt, H.-J., & Parson, W. (2008). Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *International Journal of Legal Medicine*, *122*(1), 11–21. doi:10.1007/s00414-006-0151-5
- Bandelt H, Herrnstadt C, Yao Y, Kong Q, Kivisild T, Rengo C, and Scozzari R. 2003. Identification of Native American Founder mtDNAs Through the Analysis of Complete mtDNA Sequences : 512–524
- Barral, I., Rodríguez-Laralde, A., Dipierri, J., Alfaro, E., Acevedo, N., Mamolini, E., ... Scapoli, C. (2012). Surnames in Chile: a study of the population of Chile through isonymy. *American Journal of Physical Anthropology*, *147*(3), 380–8. doi:10.1002/ajpa.22000
- Bedoya, G., Montoya, P., García, J., Soto, I., Bourgeois, S., Carvajal, L., ... Ruiz-Linares, A. (2006). Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(19), 7234–9. doi:10.1073/pnas.0508716103
- Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N. M., ... Villems, R. (2012). A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *American Journal of Human Genetics*, *90*(4), 675–84. doi:10.1016/j.ajhg.2012.03.002
- Blanco-Verea, A., Jaime, J. C., Brión, M., & Carracedo, A. (2010). Y-chromosome lineages in native South American population. *Forensic Science International. Genetics*, *4*(3), 187–93. doi:10.1016/j.fsigen.2009.08.008
- Bobillo, M. C., Zimmermann, B., Sala, A., Huber, G., Röck, A., Bandelt, H.-J., ... Parson, W. (2010). Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. *International Journal of Legal Medicine*, *124*(4), 263–8. doi:10.1007/s00414-009-0366-3
- Boca, S. M., & Rosenberg, N. A. (2011). Mathematical properties of F_{st} between admixed populations and their parental source populations. *Theoretical Population Biology*, *80*(3), 208–16. doi:10.1016/j.tpb.2011.05.003
- Bodner, M., Perego, U. A., Huber, G., Fendt, L., Röck, A. W., Zimmermann, B., ... Parson, W. (2012). Rapid coastal spread of First Americans: novel insights from South America's

Southern Cone mitochondrial genomes. *Genome Research*, 22(5), 811–20.
doi:10.1101/gr.131722.111

- Bortolini, M.-C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P. K., Bau, C. H. D., ... Ruiz-Linares, A. (2003). Y-chromosome evidence for differing ancient demographic histories in the Americas. *American Journal of Human Genetics*, 73(3), 524–39. doi:10.1086/377588
- Brandstätter, A., Zimmermann, B., Wagner, J., Göbel, T., Röck, A. W., Salas, A., ... Parson, W. (2008). Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East. *BMC Evolutionary Biology*, 8, 191. doi:10.1186/1471-2148-8-191
- Brisighelli, F., Capelli, C., Alvarez-Iglesias, V., Onofri, V., Paoli, G., Tofanelli, S., ... Salas, A. (2009). The Etruscan timeline: a recent Anatolian connection. *European Journal of Human Genetics : EJHG*, 17(5), 693–6. doi:10.1038/ejhg.2008.224
- Brown, P. O., & Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nature Genetics*, 21(1 Suppl), 33–7. doi:10.1038/4462
- Brown, W. M., George, M., & Wilson, a C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 76(4), 1967–71.
- Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., ... Ostrer, H. (2010). Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proceedings of the National Academy of Sciences of the United States of America*, 107 Suppl , 8954–61. doi:10.1073/pnas.0914618107
- Bumgarner, R. (2013). Overview of DNA Microarrays : Types , Applications , and Their Future, (January), 1–11. doi:10.1002/0471142727.mb2201s101
- Cardoso, S., Alfonso-Sánchez, M. a, Valverde, L., Sánchez, D., Zarrabeitia, M. T., Odriozola, a, ... de Pancorbo, M. M. (2012). Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. *Heredity*, 108(6), 609–15. doi:10.1038/hdy.2011.131
- Carvajal-Carmona, L. G., Ophoff, R., Service, S., Hartiala, J., Molina, J., Leon, P., ... Ruiz-Linares, A. (2003). Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Human Genetics*, 112(5-6), 534–41. doi:10.1007/s00439-002-0899-8
- Carvajal-Carmona, L. G., Soto, I. D., Pineda, N., Ortíz-Barrientos, D., Duque, C., Ospina-Duque, J., ... Ruiz-Linares, a. (2000). Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *American Journal of Human Genetics*, 67(5), 1287–95. doi:10.1016/S0002-9297(07)62956-5

- Carvalho, B. M., Bortolini, M. C., Santos, S. E. B. Dos, & Ribeiro-dos-Santos, Â. K. C. (2008). Mitochondrial DNA mapping of social-biological interactions in Brazilian Amazonian African-descendant populations. *Genetics and Molecular Biology*, 31(1), 12–22. doi:10.1590/S1415-47572008000100002
- Catelli, M. L., Alvarez-Iglesias, V., Gómez-Carballa, A., Mosquera-Miguel, A., Romanini, C., Borosky, A., ... Salas, A. (2011). The impact of modern migrations on present-day multi-ethnic Argentina as recorded on the mitochondrial DNA genome. *BMC Genetics*, 12(1), 77. doi:10.1186/1471-2156-12-77
- Choudhry, S., Coyle, N. E., Tang, H., Salari, K., Lind, D., Clark, S. L., ... Burchard, E. G. (2006). Population stratification confounds genetic association studies among Latinos. *Human Genetics*, 118(5), 652–64. doi:10.1007/s00439-005-0071-3
- Cifuentes, L., Morales, R., Sepúlveda, D., Jorquera, H., & Acuña, M. (2004). DYS19 and DYS199 loci in a Chilean population of mixed ancestry. *American Journal of Physical Anthropology*, 125(1), 85–9. doi:10.1002/ajpa.10380
- Coelho, M., Sequeira, F., Luiselli, D., Belez, S., & Rocha, J. (2009). On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evolutionary Biology*, 9, 80. doi:10.1186/1471-2148-9-80
- Corach, D., Lao, O., Bobillo, C., van Der Gaag, K., Zuniga, S., Vermeulen, M., ... Kayser, M. (2010). Inferring continental ancestry of argentineans from Autosomal, Y-chromosomal and mitochondrial DNA. *Annals of Human Genetics*, 74(1), 65–76. doi:10.1111/j.1469-1809.2009.00556.x
- Coudray, C., Olivieri, a, Achilli, a, Pala, M., Melhaoui, M., Cherkaoui, M., ... Dugoujon, J. M. (2009). The complex and diversified mitochondrial gene pool of Berber populations. *Annals of Human Genetics*, 73(2), 196–214. doi:10.1111/j.1469-1809.2008.00493.x
- De Saint Pierre, M., Bravi, C. M., Motti, J. M. B., Fuku, N., Tanaka, M., Llop, E., ... Moraga, M. (2012). An alternative model for the early peopling of southern South America revealed by analyses of three mitochondrial DNA haplogroups. *PloS One*, 7(9), e43486. doi:10.1371/journal.pone.0043486
- Dornelles, C. L., Bonatto, S. L., De Freitas, L. B., & Salzano, F. M. (2005). Is haplogroup X present in extant South American Indians? *American Journal of Physical Anthropology*, 127(4), 439–48. doi:10.1002/ajpa.20103
- Dubut, V., Chollet, L., Murail, P., Cartault, F., Béraud-Colomb, E., Serre, M., & Mogentale-Profizi, N. (2004). mtDNA polymorphisms in five French groups: importance of regional sampling. *European Journal of Human Genetics : EJHG*, 12(4), 293–300. doi:10.1038/sj.ejhg.5201145
- Encina, FA. (1983). Historia de Chile. Santiago, Chile: Capítulo III, Sociedad Editora Revista Ercilla

- Ely, B., Wilson, J. L., Jackson, F., & Jackson, B. a. (2006). African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. *BMC Biology*, 4, 34. doi:10.1186/1741-7007-4-34
- Fagundes, N. J. R., Kanitz, R., & Bonatto, S. L. (2008). A reevaluation of the Native American mtDNA genome diversity and its bearing on the models of early colonization of Beringia. *PLoS One*, 3(9), e3157. doi:10.1371/journal.pone.0003157
- Fagundes, N. J. R., Kanitz, R., Eckert, R., Valls, A. C. S., Bogo, M. R., Salzano, F. M., ... Bonatto, S. L. (2008). Mitochondrial Population Genomics Supports a Single Pre-Clovis Origin with a Coastal Route for the Peopling of the Americas, (March), 583–592. doi:10.1016/j.ajhg.2007.11.013.
- Flores, C., Maca-Meyer, N., González, A. M., Oefner, P. J., Shen, P., Pérez, J. a, ... Underhill, P. a. (2004). Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *European Journal of Human Genetics : EJHG*, 12(10), 855–63. doi:10.1038/sj.ejhg.5201225
- Francalacci, P., Morelli, L., Useli, A., & Sanna, D. (2010). The history and geography of the Y chromosome SNPs in Europe: an update. *Journal of Anthropological Sciences = Rivista Di Antropologia : JASS / Istituto Italiano Di Antropologia*, 88, 207–14. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20834059>
- Francalacci, P., & Sanna, D. (2008). History and geography of human Y-chromosome in Europe: a SNP perspective. *Journal of Anthropological Sciences = Rivista Di Antropologia : JASS / Istituto Italiano Di Antropologia*, 86, 59–89. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19934469>
- Galanter JM, Fernandez-Lopez JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, et al. (2012) Development of a Panel of Genome-Wide Ancestry Informative Markers to Study Admixture Throughout the Americas. *PLoS Genet* 8(3): e1002554. doi:10.1371/journal.pgen.1002554
- García, A., Pauro, M., Nores, R., Bravi, C. M., & Demarchi, D. a. (2012). Phylogeography of mitochondrial haplogroup D1: an early spread of subhaplogroup D1j from Central Argentina. *American Journal of Physical Anthropology*, 149(4), 583–90. doi:10.1002/ajpa.22174
- García, F., Moraga, M., Vera, S., Henríquez, H., Llop, E., Aspillaga, E., & Rothhammer, F. (2006). mtDNA microevolution in Southern Chile's archipelagos. *American Journal of Physical Anthropology*, 129(3), 473–81. doi:10.1002/ajpa.20297
- Geppert, M., Baeta, M., Núñez, C., Martínez-Jarreta, B., Zweynert, S., Cruz, O. W. V., ... Roewer, L. (2011). Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic

relationship of native populations in South America. *Forensic Science International Genetics*, 5(2), 100–4. doi:10.1016/j.fsigen.2010.08.016

Ginther C, Corach D, Penacino GA, Rey JA, Carnese FR, Hutz MH, Anderson A, Just J, Salzano FM, King MC. 1993. Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. *EXPERIENTIA* 67:211-219

Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 77(11), 6715–9.

Gómez-Carballa, A., Ignacio-Veiga, A., Alvarez-Iglesias, V., Pastoriza-Mourelle, A., Ruíz, Y., Pineda, L., ... Salas, A. (2012). A melting pot of multicontinental mtDNA lineages in admixed Venezuelans. *American Journal of Physical Anthropology*, 147(1), 78–87. doi:10.1002/ajpa.21629

Gonder, M. K., Mortensen, H. M., Reed, F. a, de Sousa, A., & Tishkoff, S. a. (2007). Whole-mtDNA genome sequence analysis of ancient African lineages. *Molecular Biology and Evolution*, 24(3), 757–68. doi:10.1093/molbev/msl209

González-Andrade, F., Sánchez-Q, D., & Martínez-Jarreta, B. (2003). Genetic profile of the Ecuadorian Mestizo population (Ecuador—South America) by using the Power Plex® 16 System Kit. *Forensic Science International*, 135(1), 64–66. doi:10.1016/S0379-0738(03)00165-8

Halder, I., Shriver, M., Thomas, M., Fernandez, J. R., & Frudakis, T. (2008). A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Human Mutation*, 29(5), 648–58. doi:10.1002/humu.2069

Henríquez H, Moraga Mauricio, Llop Elena, Rotthammer Francisco. 2004. Caracterización genético molecular de habitantes de caleta Paposo, último reducto Chango en Chile. *Revista Médica de Chile* [Internet] 132:663-672. Available from: http://www.scielo.cl/scielo.php?pid=S0034-98872004000600002&script=sci_arttext&tlng=en

Hoffmann, T. J., Kvale, M. N., Hesselson, S. E., Zhan, Y., Aquino, C., Cao, Y., ... Risch, N. (2011). Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics*, 98(2), 79–89. doi:10.1016/j.ygeno.2011.04.005

Hoffmann, T. J., Zhan, Y., Kvale, M. N., Hesselson, S. E., Gollub, J., Iribarren, C., ... Risch, N. (2011). Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation

- and a novel hybrid SNP selection algorithm. *Genomics*, 98(6), 422–30. doi:10.1016/j.ygeno.2011.08.007
- Hoheisel, J. D. (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nature Reviews. Genetics*, 7(3), 200–10. doi:10.1038/nrg1809
- Horai, S. (1994). [Peopling of the Americas, founded by four major lineages of mitochondrial DNA]. *Tanpakushitsu Kakusan Koso. Protein, Nucleic Acid, Enzyme*, 39(15), 2759–67.
- Ingman, M., & Gyllensten, U. (2007). A recent genetic link between Sami and the Volga-Ural region of Russia. *European Journal of Human Genetics : EJHG*, 15(1), 115–20. doi:10.1038/sj.ejhg.5201712
- Jobling, M. a, Samara, V., Pandya, a, Fretwell, N., Bernasconi, B., Mitchell, R. J., ... Tyler-Smith, C. (1996). Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. *Human Molecular Genetics*, 5(11), 1767–75. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8923005>
- Jobling, M. a, & Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews. Genetics*, 4(8), 598–612. doi:10.1038/nrg1124
- Jota, M. S., Lacerda, D. R., Sandoval, J. R., Vieira, P. P. R., Santos-Lopes, S. S., Bisso-Machado, R., ... Santos, F. R. (2011). A new subhaplogroup of native American Y-Chromosomes from the Andes. *American Journal of Physical Anthropology*, 146(4), 553–9. doi:10.1002/ajpa.21519
- Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., & Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, 830–838. doi:10.1101/gr.7172008.by
- Kemp, B. M., Malhi, R. S., Mcdonough, J., Bolnick, D. A., Eshleman, J. A., Rickards, O., ... Smith, D. G. (2007). Genetic Analysis of Early Holocene Skeletal Remains From Alaska and its Implications for the Settlement of the Americas, 621(January), 605–621. doi:10.1002/ajpa
- Leiva, Ximena. 2010. Efectos de las migraciones recientes en la composición genética de la población de Santiago de Chile. Título de Antropóloga Física. Facultad de Ciencias Sociales. Universidad de Chile. Chile. 78pp.
- Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., & Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. *Nature Genetics*, 21(1 Suppl), 20–4. doi:10.1038/4447
- Loogväli, E.-L., Roostalu, U., Malyarchuk, B. a, Derenko, M. V, Kivisild, T., Metspalu, E., ... Villems, R. (2004). Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Molecular Biology and Evolution*, 21(11), 2012–21. doi:10.1093/molbev/msh209

- López-Parra, a M., Gusmão, L., Tavares, L., Baeza, C., Amorim, a, Mesa, M. S., ... Arroyo-Pardo, E. (2009). In search of the pre- and post-neolithic genetic substrates in Iberia: evidence from Y-chromosome in Pyrenean populations. *Annals of Human Genetics*, 73(1), 42–53. doi:10.1111/j.1469-1809.2008.00478.x
- Maca-Meyer, N., Sánchez-Velasco, P., Flores, C., Larruga, J.-M., González, a-M., Oterino, a, & Leyva-Cobián, F. (2003). Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Annals of Human Genetics*, 67(Pt 4), 329–39. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12914567>
- Malhi, R. S., Eshleman, J. a, Greenberg, J. a, Weiss, D. a, Schultz Shook, B. a, Kaestle, F. a, ... Smith, D. G. (2002). The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America. *American Journal of Human Genetics*, 70(4), 905–19.
- Malyarchuk, B. a, Perkova, M. a, Derenko, M. V, Vanecek, T., Lazur, J., & Gomolcak, P. (2008). Mitochondrial DNA variability in Slovaks, with application to the Roma origin. *Annals of Human Genetics*, 72(Pt 2), 228–40. doi:10.1111/j.1469-1809.2007.00410.x
- Marrero, A. R., Bravi, C., Stuart, S., Long, J. C., Pereira das Neves Leite, F., Kommers, T., ... Cátira Bortolini, M. (2007). Pre- and post-Columbian gene and cultural continuity: the case of the Gaucho from southern Brazil. *Human Heredity*, 64(3), 160–71. doi:10.1159/000102989
- Marrero, A. R., Das Neves Leite, F. P., De Almeida Carvalho, B., Peres, L. M., Kommers, T. C., Da Cruz, I. M., ... Bortolini, M. C. (2005). Heterogeneity of the genome ancestry of individuals classified as White in the state of Rio Grande do Sul, Brazil. *American Journal of Human Biology : The Official Journal of the Human Biology Council*, 17(4), 496–506. doi:10.1002/ajhb.20404
- Martínez-Cruz, B., Harmant, C., Platt, D. E., Haak, W., Manry, J., Ramos-Luis, E., ... Comas, D. (2012). Evidence of pre-roman tribal genetic structure in basques from uniparentally inherited markers. *Molecular Biology and Evolution*, 29(9), 2211–22. doi:10.1093/molbev/mss091
- Martínez-González, L. J., Saiz, M., Alvarez-Cubero, M. J., Gómez-Martín, A., Alvarez, J. C., Martínez-Labarga, C., & Lorente, J. a. (2012). Distribution of Y chromosomal STRs loci in Mayan and Mestizo populations from Guatemala. *Forensic Science International. Genetics*, 6(1), 136–42. doi:10.1016/j.fsigen.2011.04.003
- Mazières, S., Guitard, E., Crubézy, E., Dugoujon, J.-M., Bortolini, M. C., Bonatto, S. L., ... Salzano, F. M. (2008). Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations—implications for the region's colonization. *Annals of Human Genetics*, 72(Pt 1), 145–56. doi:10.1111/j.1469-1809.2007.00392.x

- Merriwether, D. a, & Ferrell, R. E. (1996). The four founding lineage hypothesis for the New World: a critical reevaluation. *Molecular Phylogenetics and Evolution*, 5(1), 241–6. doi:10.1006/mpev.1996.0017
- Moraga, M. L., Rocco, P., Miquel, J. F., Nervi, F., Llop, E., Chakraborty, R., ... Carvallo, P. (2000). Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. *American Journal of Physical Anthropology*, 113(1), 19–29. doi:10.1002/1096-8644(200009)113:1<19::AID-AJPA3>3.0.CO;2-X
- Moreno-Estrada, Andrés., et al. (2014). The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science*, 344, 1280. doi:10.1126/science.1251688.
- Non, A. L., Al-Meer, A., Raaum, R. L., Sanchez, L. F., & Mulligan, C. J. (2011). Mitochondrial DNA reveals distinct evolutionary histories for Jewish populations in Yemen and Ethiopia. *American Journal of Physical Anthropology*, 144(1), 1–10. doi:10.1002/ajpa.21360
- O'Rourke, D. H., & Raff, J. a. (2010). The human genetic history of the Americas: the final frontier. *Current Biology : CB*, 20(4), R202–7. doi:10.1016/j.cub.2009.11.051
- Parson, W., & Bandelt, H.-J. (2007). Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Science International. Genetics*, 1(1), 13–9. doi:10.1016/j.fsigen.2006.11.003
- Perego, U. a, Achilli, A., Angerhofer, N., Accetturo, M., Pala, M., Olivieri, A., ... Torroni, A. (2009). Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Current Biology : CB*, 19(1), 1–8. doi:10.1016/j.cub.2008.11.058
- Perić, M., Barać Lauc, L., Martinović Klarić, I., Jančićević, B., & Rudan, P. (2005). Review of Croatian genetic heritage as revealed by mitochondrial DNA and Y chromosomal lineages. *Croatian Medical Journal*, 46(4), 502–13. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16100752>
- Rangel-Villalobos, H., Muñoz-Valle, J. F., González-Martín, a, Gorostiza, a, Magaña, M. T., & Páez-Riberos, L. a. (2008). Genetic admixture, relatedness, and structure patterns among Mexican populations revealed by the Y-chromosome. *American Journal of Physical Anthropology*, 135(4), 448–61. doi:10.1002/ajpa.20765
- Rangel-Villalobos, H., Salazar-Flores, J., Dondiego, R., Anaya-Palafox, M., Nuño-Arana, I., Canseco-Ávila, L. M., ... Rubi-Castellanos, R. (2009). "South to North increasing gradient of paternal European ancestry throughout the Mexican territory: Evidence of Y-linked short tandem repeats." *Forensic Science International: Genetics Supplement Series*, 2(1), 448–450. doi:10.1016/j.fsigs.2009.08.003

- Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., ... Bandelt, H. J. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *American Journal of Human Genetics*, 67(5), 1251–76. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1288566&tool=pmcentrez&endertype=abstract>
- Rocco P, Morales C, Moraga M, Miquel J, Nervi F, Llop E, Carvallo P, Rothhammer F. 2002. Composición genética de la población chilena. Distribución de polimorfismos de DNA mitocondrial en grupos originarios y en la población mixta de Santiago. *Revista Médica de Chile* [Internet] 130(2):125-131. Available from http://www.scielo.cl/scielo.php?pid=S0034-98872002000200001&script=sci_arttext&lng=en
- Rose, G., Longo, T., Maletta, R., Passarino, G., Bruni, a C., & De Benedictis, G. (2008). No evidence of association between frontotemporal dementia and major European mtDNA haplogroups. *European Journal of Neurology : The Official Journal of the European Federation of Neurological Societies*, 15(9), 1006–8. doi:10.1111/j.1468-1331.2008.02222.x
- Rubinoff, D. (2006). Essays: Utility of Mitochondrial DNA Barcodes in Species Conservation. *Conservation Biology*, 20(4), 1026–1033. doi:10.1111/j.1523-1739.2006.00372.x
- Salas, A., Carracedo, A., Richards, M., & Macaulay, V. (2005). Charting the ancestry of African Americans. *American Journal of Human Genetics*, 77(4), 676–80. doi:10.1086/491675
- Salas, A., Richards, M., Lareu, M.-V., Scozzari, R., Coppa, A., Torroni, A., ... Carracedo, A. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. *American Journal of Human Genetics*, 74(3), 454–65. doi:10.1086/382194
- Salazar-Flores, J., Dondiego-Aldape, R., Rubi-Castellanos, R., Anaya-Palafox, M., Nuño-Arana, I., Canseco-Avila, L. M., ... Rangel-Villalobos, H. (2010). Population structure and paternal admixture landscape on present-day Mexican-Mestizos revealed by Y-STR haplotypes. *American Journal of Human Biology : The Official Journal of the Human Biology Council*, 22(3), 401–9. doi:10.1002/ajhb.21013
- Salzano, F. M., & Sans, M. (2014). Interethnic admixture and the evolution of Latin American populations. *Genetics and Molecular Biology*, 37(1 Suppl), 151–70. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3983580&tool=pmcentrez&endertype=abstract>
- Sandoval, K., Moreno-Estrada, A., Mendizabal, I., Underhill, P. a, Lopez-Valenzuela, M., Peñaloza-Espinosa, R., ... Comas, D. (2012). Y-chromosome diversity in Native Mexicans reveals continental transition of genetic structure in the Americas. *American Journal of Physical Anthropology*, 148(3), 395–405. doi:10.1002/ajpa.22062

- Sans, M. (2000). Admixture studies in Latin America: from the 20th to the 21st century. *Human Biology*, 72(1), 155–77. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10721616>
- Sans, M., Figueiro, G., & Hidalgo, P. C. (2012). A new mitochondrial C1 lineage from the prehistory of Uruguay: population genocide, ethnocide, and continuity. *Human Biology*, 84(3), 287–305. doi:10.3378/027.084.0303
- Schurr, T. G., Ballinger, S. W., Gan, Y. Y., Hodge, J. a, Merriwether, D. a, Lawrence, D. N., ... Wallace, D. C. (1990). Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *American Journal of Human Genetics*, 46(3), 613–23.
- Seldin, M. F., Tian, C., Shigeta, R., Scherbarth, H. R., Silva, G., Belmont, J. W., ... Alarconrquelme, M. E. (2007). Argentine Population Genetic Structure : Large Variance in Amerindian Contribution, 462(December 2006), 455–462. doi:10.1002/ajpa
- Semino, O. (2000). The Genetic Legacy of Paleolithic Homo sapiens sapiens in Extant Europeans: A Y Chromosome Perspective. *Science*, 290(5494), 1155–1159. doi:10.1126/science.290.5494.1155
- Semino, O., Magri, C., Benuzzi, G., Lin, A. a, Al-Zahery, N., Battaglia, V., ... Santachiara-Benerecetti, a S. (2004). Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *American Journal of Human Genetics*, 74(5), 1023–34. doi:10.1086/386295
- Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J. C., Uribe-Figueroa, L., Contreras, A., ... Jimenez-Sanchez, G. (2009). Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, 106(21), 8611–6. doi:10.1073/pnas.0903045106
- Simms, T. M., Wright, M. R., Hernandez, M., Perez, O. a, Ramirez, E. C., Martinez, E., & Herrera, R. J. (2012). Y-chromosomal diversity in Haiti and Jamaica: contrasting levels of sex-biased gene flow. *American Journal of Physical Anthropology*, 148(4), 618–31. doi:10.1002/ajpa.22090
- Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., ... Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *American Journal of Human Genetics*, 84(6), 740–59. doi:10.1016/j.ajhg.2009.05.001
- Stefflova, K., Dulik, M. C., Barnholtz-Sloan, J. S., Pai, A. a, Walker, A. H., & Rebbeck, T. R. (2011). Dissecting the within-Africa ancestry of populations of African descent in the Americas. *PloS One*, 6(1), e14495. doi:10.1371/journal.pone.0014495

- Syvänen, A.-C. (2005). Toward genome-wide SNP genotyping. *Nature Genetics*, 37 Suppl(June), S5–10. doi:10.1038/ng1558
- Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., ... Malhi, R. S. (2007). Beringian standstill and spread of Native American founders. *PLoS One*, 2(9), e829. doi:10.1371/journal.pone.0000829
- Torroni, a, Bandelt, H. J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., ... Scozzari, R. (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *American Journal of Human Genetics*, 62(5), 1137–52. doi:10.1086/301822
- Torroni, A., Schurr, G., Yang, C., Williams, R. C., Schanfield, M. S., Troup, G. A., ... Lawrence, D. N. (1992). Native American Mitochondrial DNA Analysis Indicates That the Amerind and the Nadene Populations Were Founded by.
- Toscanini, U., Gusmão, L., Berardi, G., Amorim, A., Carracedo, A., Salas, A., & Raimondi, E. (2007). Testing for genetic structure in different urban Argentinian populations. *Forensic Science International*, 165(1), 35–40. doi:10.1016/j.forsciint.2006.02.042
- Toscanini, U., Gusmão, L., Berardi, G., Gomes, V., Amorim, A., Salas, A., & Raimondi, E. (2011). Male lineages in South American native groups: evidence of M19 traveling south. *American Journal of Physical Anthropology*, 146(2), 188–96. doi:10.1002/ajpa.21562
- Underhill, P. a, & Kivisild, T. (2007). Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics*, 41, 539–64. doi:10.1146/annurev.genet.41.110306.130407
- Underhill, P. a, Shen, P., Lin, a a, Jin, L., Passarino, G., Yang, W. H., ... Oefner, P. J. (2000). Y chromosome sequence variation and the history of human populations. *Nature Genetics*, 26(3), 358–61. doi:10.1038/81685
- Van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30(2), E386–94. doi:10.1002/humu.20921
- Van Oven, M., Kayser, M. (2008). Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. *Human Mutation*, 30:E386-E394
- Verdu, P., & Rosenberg, N. a. (2011). A general mechanistic model for admixture histories of hybrid populations. *Genetics*, 189(4), 1413–26. doi:10.1534/genetics.111.132787
- Via, M., Gignoux, C. R., Roth, L. a, Fejerman, L., Galanter, J., Choudhry, S., ... Martínez-Cruzado, J. C. (2011). History shaped the geographic distribution of genomic admixture on the island of Puerto Rico. *PLoS One*, 6(1), e16513. doi:10.1371/journal.pone.0016513

- Volodko, N. V., Starikovskaya, E. B., Mazunin, I. O., Eltsov, N. P., Naidenko, P. V., Wallace, D. C., & Sukernik, R. I. (2008). Mitochondrial Genome Diversity in Arctic Siberians , with Particular Reference to the Evolutionary History of Beringia and Pleistocenic Peopling of the Americas, (May), 1084–1100. doi:10.1016/j.ajhg.2008.03.019.
- Wall, J. D., Jiang, R., Gignoux, C., Chen, G. K., Eng, C., Huntsman, S., & Marjoram, P. (2011). Genetic variation in Native Americans, inferred from Latino SNP and resequencing data. *Molecular Biology and Evolution*, 28(8), 2231–7. doi:10.1093/molbev/msr049
- Wang, S., Ray, N., Rojas, W., Parra, M. V., Bedoya, G., Gallo, C., ... Ruiz-Linares, A. (2008). Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genetics*, 4(3), e1000037. doi:10.1371/journal.pgen.1000037
- Wilkins, J. F., & Marlowe, F. W. (2006). Sex-biased migration in humans: what should we expect from genetic data? *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, 28(3), 290–300. doi:10.1002/bies.20378
- Zegura, S. L., Karafet, T. M., Zhivotovsky, L. a, & Hammer, M. F. (2004). High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Molecular Biology and Evolution*, 21(1), 164–75. doi:10.1093/molbev/msh009