

Isonimia y cromosoma Y: ¿concordancia o conflicto?

Tesis Doctoral

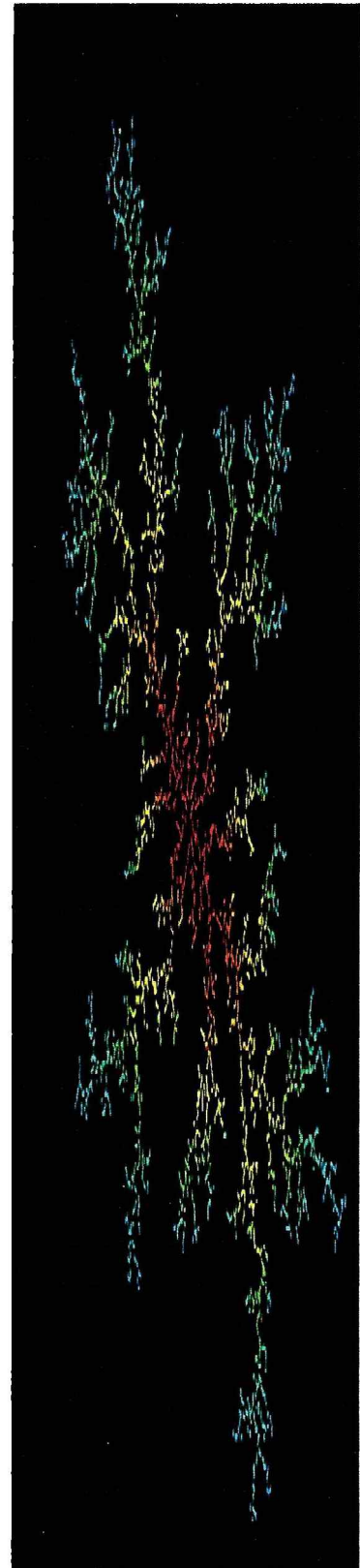
Doctorando: Marina Muzzio

Director: Graciela Bailliet

Codirector: Claudio Marcelo Bravi

Fecha: 30/09/2009

La Plata, Facultad de Ciencias Naturales y Museo,
UNLP





BIBLIOTECA

Facultad de Ciencias Naturales y Museo

A mis abuelos Johny y Quique, quienes luego de contar cómo huían de la escuela aclaraban: “pero vos estudiá, nena”

Agradecimientos

“...Oh I get by with a little help from my friends,
I get high with a little help from my friends,
Oh I'm gonna try with a little help from my friends...”

The Beatles, “With a little help from my friends”

A mis directores, Graciela Bailliet y Claudio M. Bravi por permitirme formar parte de este equipo, además de su continuo apoyo tanto profesional como personal.

A mis compañeros (y amigos) del “Equipo B-B”: Josefina, Rita, Virginia, Juan, Boris y Santiago, que me alentaron y también toleraron mis excéntricos hábitos.

A las damitas, el Comité de Damas y la Comisión de Eventos Gastronómicos, por hacer del IMBICE el lugar tan particular y agradable que es.

A Emma, por su ayuda y aliento.

A Rosana, Cecilia, Marta y Marina por su apoyo e interés.

A la Dra. Susana Salceda por contactarme con Graciela y Claudio.

A las autoridades del IMBICE por permitirme realizar este trabajo allí; a la ANPCyT y al CONICET por brindar los fondos.

A los participantes que donaron las muestras y a los equipos de bioquímicos, extraccionistas y técnicos que posibilitaron la obtención de las mismas.

A mi familia por su apoyo y la clarificación de mis reiteradas consultas sobre métodos estadísticos y modelos matemáticos.

A Gris y familia, Feli, Celeste, Ayelén y Yanel, que de distintas formas me han alentado.

A Carlos, por animarme y convivir con mi antropolocura.

ありがとうございます！

Índice

| | |
|---|-----|
| Resumen | 7 |
| Abstract | 8 |
| Introducción..... | 9 |
| 1.1 ¿Por qué el Y?..... | 13 |
| 1.1.1- Haplogrupos y haplotipos | 13 |
| 1.1.2- Haplotipos en el mundo | 17 |
| 1.2 ¿Por qué los apellidos? | 18 |
| 1.3 ¿Por qué el Noroeste Argentino?..... | 22 |
| Objetivos..... | 24 |
| Materiales y Métodos | 26 |
| 2.1 Poblaciones seleccionadas y muestra | 27 |
| 2.2 Extracción del ADN, tipificación de SNP y de STR..... | 28 |
| 2.2.1 Ciclos:..... | 30 |
| 2.3 Asociaciones entre haplogrupos y apellidos | 32 |
| 2.4 Análisis de haplotipos | 34 |
| 2.5 Distancias temporales..... | 35 |
| 2.6 Comparaciones a nivel mundial | 37 |
| 2.7 Estimaciones de la proporción de transmisión no paterna del apellido.... | 37 |
| Nuevos Desarrollos y Validaciones Metodológicas | 41 |
| 3.1 Desarrollo de la nueva propuesta para el pesado de caracteres | 42 |
| 3.1.1 Estimaciones de calidad..... | 44 |
| 3.2 Ensayos de fiabilidad de los programas de predicción de haplogrupo..... | 48 |
| 3.2.1 ¿Cómo evaluar la fiabilidad de los programas? | 49 |
| 3.2.2 ¿Qué obtuvimos? | 51 |
| 3.2.3 ¿Qué significan todos estos datos?..... | 53 |
| 3.2.4 ¿Por qué estos programas muestran niveles de precisión tan bajos?..... | 54 |
| Resultados | 56 |
| 4.1 Asociaciones entre haplogrupos y apellidos | 57 |
| 4.2 Asociaciones entre haplotipos y apellidos | 60 |
| 4.3 Redes Medianas | 64 |
| 4.3.1 Red del Haplogrupo Q1a3a..... | 64 |
| 4.3.2 Red del Haplogrupo R..... | 65 |
| 4.4 Búsqueda en YHRD..... | 70 |
| 4.5 Transmisión no paterna del apellido..... | 71 |
| 4.6 Modelos de transmisión irregular del apellido | 73 |
| Discusión | 79 |
| 5.1 Apellidos y ancestralidad | 80 |
| 5.2 El caso de los apellidos frecuentes | 85 |
| 5.3 Haplotipos modales Ibérico y Europeo..... | 85 |
| 5.4 Concordancias, discordancias..... | 86 |
| 5.5 ¿Por qué los haplotipos modales se esparcen sobre diferentes apellidos? . | 87 |
| 5.6 Transmisión no paterna del apellido..... | 89 |
| Conclusión | 91 |
| Bibliografía | 94 |
| Anexo..... | 105 |

Índice de Figuras

| | |
|---|----|
| Figura 1: Distribución contemporánea de los haplogrupos del cromosoma Y en 22 regiones..... | 14 |
| Figura 2: Esquema del modelo de a un paso..... | 16 |
| Figura 3: Poblaciones incluidas en YHRD..... | 18 |
| Figura 4: Mapa de la República Argentina..... | 27 |
| Figura 5: Extracción de ADN..... | 29 |
| Figura 6: Y DNA Thermal Cycler, PERKIN ELMER..... | 31 |
| Figura 7: Programmable Thermal Controller, MJ Research, Inc..... | 31 |
| Figura 8: Siembra de un gel de acrilamida al 6%..... | 32 |
| Figura 9: PowerPac 3000, BIO-RAD..... | 32 |
| Figura 10: Red Mediana calculada y dibujada con pesos uniformes..... | 45 |
| Figura 11: Red Red Mediana calculada y dibujada con la asignación de pesos según el método de Qamar y col. (2002)..... | 46 |
| Figura 12: Red Mediana calculada y dibujada con los pesos asignados según la formula desarrollada en la presente tesis..... | 47 |
| Figura 13: Red Mediana del haplogorupo Q1a3a..... | 65 |
| Figura 14: Red Mediana del haplogrupo R..... | 66 |
| Figura 15: Tiempo de divergencia entre muestras del mismo apellido, el caso del haplogrupo R..... | 68 |
| Figura 16: Tiempo de divergencia entre muestras del mismo apellido, el caso del haplogrupo Q..... | 68 |

Índice de Tablas

| | |
|--|----|
| Tabla 1: Tasas Mutacionales y pesos relativos por microsatélite | 35 |
| Tabla 2: S, e, LR+ y LR- totales en cada punto de corte para el Haplogroup Predictor | 51 |
| Tabla 3: Frecuencias de haplogrupo por tipificación de SNP, el H. Predictor y el H. Classifier, y los valores de s, e, LR+ y LR- | 52 |
| Tabla 4: Distribución de los individuos por provincia, haplogrupos y origen del apellido | 59 |
| Tabla 5: Apellidos y haplotipos establecidos | 60 |
| Tabla 6: Tiempos de divergencia entre muestras del mismo apellido | 69 |
| Tabla 7: Proporciones de traspaso no paterno por localidad y tiempo. | 71 |
| Tabla 8: Porcentajes por localidad | 73 |
| Tabla 9: Resultados del modelo | 78 |

Resumen

En las sociedades con transmisión paterna de los apellidos, la distribución de los mismos debería seguir el mismo recorrido que los linajes del cromosoma Y. Se evalúa esta hipótesis mediante la definición de haplotipos de microsatélites del cromosoma Y en ciento setenta y cinco individuos del Noroeste Argentino, que abarcan noventa y cinco apellidos.

Se compararon estas dos fuentes de información, buscando concordancias entre y dentro de apellidos, trazando redes medianas, estimando el tiempo de divergencia y comparando los resultados obtenidos con la base de datos internacional YHRD.

Se establecieron dos haplotipos nodales, uno nativo americano y otro de origen europeo, rodeados por el resto de los haplotipos. En un solo caso se encontró concordancia entre apellidos y los perfiles establecidos mediante microsatélites del cromosoma Y, y también se reconocieron linajes genéticos que involucraban diferentes apellidos.

A su vez, para definir si las divergencias se debían a polifiletismo o a mutaciones recientes, se estimó la distancia temporal entre las muestras del mismo apellido y haplogrupo, encontrando que en cerca de la mitad de estos casos esta profundidad temporal excedía el origen de los apellidos.

Estos datos sugieren que la transmisión irregular del apellido en el Noroeste Argentino fue más elevada que aquella encontrada en otros países. Asimismo, las similitudes entre ambos tipos de linajes no debe ser dada por sentado.

Abstract

In societies with a paternal transmission of surnames the distribution of these should follow the same trail of the Y-chromosome lineages. We evaluate this hypothesis by establishing the Y-STR haplotypes of a hundred and seventy five individuals from the Argentinean Northwest, distributed among ninety five surnames.

We compared these two sources of information, searching for matches within and through surnames, establishing median-joining networks, estimating their time of divergence and comparing our results with the international database YHRD.

We established two nodal haplotypes (one Native American and the other European), surrounded by the rest. In only one case we found a match between surnames and Y-STR, and we also recognized genetic lineages involving different surnames.

Likewise, to define if these divergencies were due to poliphyletism or recent mutations, we estimated the time distance between the samples of the same surname and haplogroup, finding that in nearly half of these cases the time depth exceeded the surname origin.

This data suggest that the irregular transmission of surnames in the Argentinean Northwest has been higher than the one found in other countries. Likewise, similarities between both kinds of lineages should not be taken for granted.

Introducción

“Tis but thy name that is my enemy;-
Thou art thyself though, not a Montague.
What’s a Montague? It is nor hand, nor foot.
Nor arm, nor face, nor any other part
Belonging to a man. O, be some other name!
What’s in a name! that which we call a rose
By any other name would smell as sweet;”

W. Shakespeare, Romeo and Juliet, Acto II, Escena 1

Una tesis sobre apellidos y ADN, combinación poco ortodoxa a simple vista, que levanta esa clásica inquietud surgida cada vez que se estudian fenómenos de diferentes caminos. ¿Apellidos y cromosoma Y? Sí, linajes sociales y linajes genéticos masculinos, los dos tienen reglas específicas, con un punto clave en común: la transmisión de ambos sigue la línea paterna.

Aquello que la gente dice y sus diferencias con lo que hace es uno de los principales intereses de la Antropología. Llevó al desarrollo de la observación participante, característico de la disciplina, porque fue una de las soluciones que se encontraron para poder comprender qué reglas se respetan, y cuándo se las deja de lado.

En Occidente, ya a fines del siglo XVI se ven ejemplos de desafío a las pautas establecidas, como en el caso de la escena a la que pertenece la cita del inicio de este apartado, datada en el 1595 d. C. Sobre esa época (y también antes en el tiempo) nos habla el historiador francés Jean Verdon en su libro L'amour au Moyen Age. La chair, le sexe et le sentiment (2008), describiendo desde las restricciones existentes por la Iglesia Católica, y cómo eran sorteadas en esa época, tanto desde la literatura como con ejemplos históricos.

De todas formas, si bien un estudio de este tema enfocado desde la Historia o la Sociología de la Cultura sería interesante, aquí se empleará otra estrategia para analizar en qué grado no se siguió la transmisión paterna del apellido. Se definirán linajes del cromosoma Y en poblaciones actuales y se compararán sus similitudes dentro y entre apellidos, como tomando una fotografía instantánea a partir de la cual se tratará de reconstruir el pasado de los involucrados. Esta reconstrucción requiere de inferencias y razonamientos, que son fáciles de entender cuando se comprenden una serie de conceptos que se detallan a continuación.

La variabilidad genética humana es tanto el resultado de la selección natural, como de los llamados “procesos neutrales”: mutación, migración y deriva génica. Los dos primeros aumentan la variabilidad, mientras que el último la disminuye (Zerjal y col. 2003).

La selección natural, se debe a diferencias entre probabilidades de supervivencia a la edad reproductiva y fertilidad heredadas. Remueve variantes deletéreas (selección purificadora), mantiene los polimorfismos (selección balanceada) y produce tendencias (selección direccional) (Jobling y col. 2003).

La deriva génica también compete al número de individuos en etapa reproductiva de una población y la variación en el número de descendientes, siendo independiente de la selección natural. Como cada generación subsiguiente representaría una muestra de la parental, ocurren variaciones estocásticas entre las frecuencias alélicas de las generaciones, que causan la fijación y eliminación de alelos a lo largo del tiempo (Cavalli-Sforza, 2007).

La migración es el tercero de estos factores, y en este caso es importante distinguir dos tipos de migración, el primero es la migración individual, y el segundo es la migración grupal que lleva a la colonización de nuevas áreas.

Con respecto a la migración individual, tradicionalmente se la propuso como limitada a distancias cortas (aunque los medios de transporte disponibles actualmente hacen necesaria una revisión de esta característica), y se debe al desplazamiento de por lo menos uno de los miembros de la pareja. Su efecto es reducir las diferencias entre poblaciones producidas por la deriva o selección natural (para el caso de ambientes distintos).

La migración grupal que conlleva la colonización de nuevas áreas, puede tener el efecto opuesto de la migración individual, ya que es propensa a generar una nueva deriva mediante el “efecto fundador”, debido a que estos grupos suelen ser pequeños, al igual que su subsiguiente generación. Además, da lugar a nuevos patrones de variación genética resultante de la selección natural en los nuevos ambientes de asentamiento (Cavalli-Sforza, 2007).

El cuarto factor, la mutación, es difícil de estimar puesto que es poco frecuente, y su medición precisa requeriría mayores números que los que comúnmente se pueden manejar, aunque a medida que aumentan los estudios genéticos y desarrollos bioinformáticos se esperan mejores estimaciones. Es la

única fuente de alelos nuevos y provee la “materia prima” (Jobling y col. 2003) sobre la cual las demás fuerzas evolutivas actúan.

Al ver a las mutaciones como eventos azarosos, se ha propuesto tomar las medidas de diversidad genética para datar la divergencia entre linajes (Zukerkandl y Pauling, 1965). Las porciones no-recombinantes del genoma (ADN mitocondrial y la región no recombinante del cromosoma Y) permiten el trazado de una única filogenia parsimoniosa, de manera tal que es posible datar los nodos dentro de esa filogenia. Existen diferentes métodos para estimar el tiempo al ancestro común más cercano (TMCRA), siempre teniendo en cuenta que puede ser más joven que la edad de la mutación que generó esa variante, debido a que se data la convergencia de aquellas ramas sobrevivientes en la actualidad, no todas las que se generaron.

Toda nuestra información biológica está almacenada dentro de las células, escrita en un código especial: el ADN. Recibimos esta información de nuestros padres, mitad de nuestra madre y mitad de nuestro padre, con pequeños re-arreglos que ocurren durante un proceso conocido como “recombinación genética”, aumentando la variabilidad genética. Esta recombinación ocurre durante la meiosis, cuando los cromosomas homólogos se alinean entre sí, y luego producen un intercambio de segmentos entre las cromátidas homólogas.

Existen dos excepciones para este modo de transmisión, una sigue la línea materna de ancestría (el ADN mitocondrial), y la otra se pasa de padre a hijo varón (el ADN de la región no recombinante del cromosoma Y). Debido a su rol en la determinación del sexo, el cromosoma Y no posee un cromosoma homólogo con quién recombinar, y su mayor parte (>90%) corresponde a la región no recombinante (NRY por sus siglas en inglés). De todas formas, durante la meiosis masculina, algunas regiones sí recombinan con el cromosoma X, son aquellas que todavía preservan identidad con este cromosoma, pueden ser heredadas de cualquiera de los padres y se las denomina regiones pseudoautosómicas.

Así, cualquier diferencia dentro de estos tipos de ADN sólo puede atribuirse a las mutaciones, permitiendo a los investigadores develar desde

estructuras y flujo génico (Lell y col. 1997, Karafet y col. 1999, Kittle y col. 1998, Demarchi y Mitchell, 2004), al origen de la humanidad (Cann y col. 1987).

1.1 ¿Por qué el Y?

En los humanos, el sexo genético está determinado por la configuración XX (femenino) o XY (masculino) del par de cromosomas sexuales. El Y es exclusivo de los varones e incluye una región que no recombina (NRY), pasando su información casi perfectamente a lo largo del linaje paterno y brindando de esta forma un registro de su historia más sencillo que los otros cromosomas.

A su vez, es más sensible a los fenómenos de deriva y flujo génico, puesto que su número efectivo es un cuarto del de los autosomas, un tercio del X y similar al del ADN mitocondrial. Esta mayor susceptibilidad acelera la diferenciación en este cromosoma entre poblaciones, que ayuda a investigar estos eventos en el pasado.

1.1.1- Haplogrupos y haplotipos

Como ya se mencionó, los únicos cambios observados en la región no recombinante del cromosoma Y son producidos por mutaciones. Dos tipos de marcadores, resultantes de diferentes tipos de mutaciones, son de interés para el presente trabajo: los SNP (por su nombre en inglés: “Single Nucleotide Polymorphism”) y los STR (“Short Tandem Repeats”).

Los SNP se deben a mutaciones de punto, es decir, el cambio de una base en la secuencia nos brinda dos alelos posibles, el ancestral y el derivado (por esto también son llamados marcadores binarios). Estas sustituciones presentan tasas mutacionales bajas, de ahí que fueron los marcadores elegidos para trazar la filogenia de los linajes del cromosoma Y (Y Chromosome Consortium, 2002; Karafet y col. 2008). Los conglomerados

definidos de esta forma son conocidos como haplogrupos y se los codifica con letras y números, desde el haplogrupo A hasta el T (Karafet y col op. cit.).

Underhill y col. (2001) asociaron estos haplogrupos con sitios geográficamente específicos, permitiendo a los investigadores encontrar el origen geográfico de un individuo a través de un análisis de laboratorio. Así, es posible estimar la composición étnica de una población dada (Bravi y col. 1997, DiPierri y col. 1998, Martínez Marignac y col. 2004) y la reconstrucción de rutas migratorias (Luis y col. 2004).

En la figura 1 puede apreciarse la distribución geográfica de los haplogrupos descrita por estos autores.

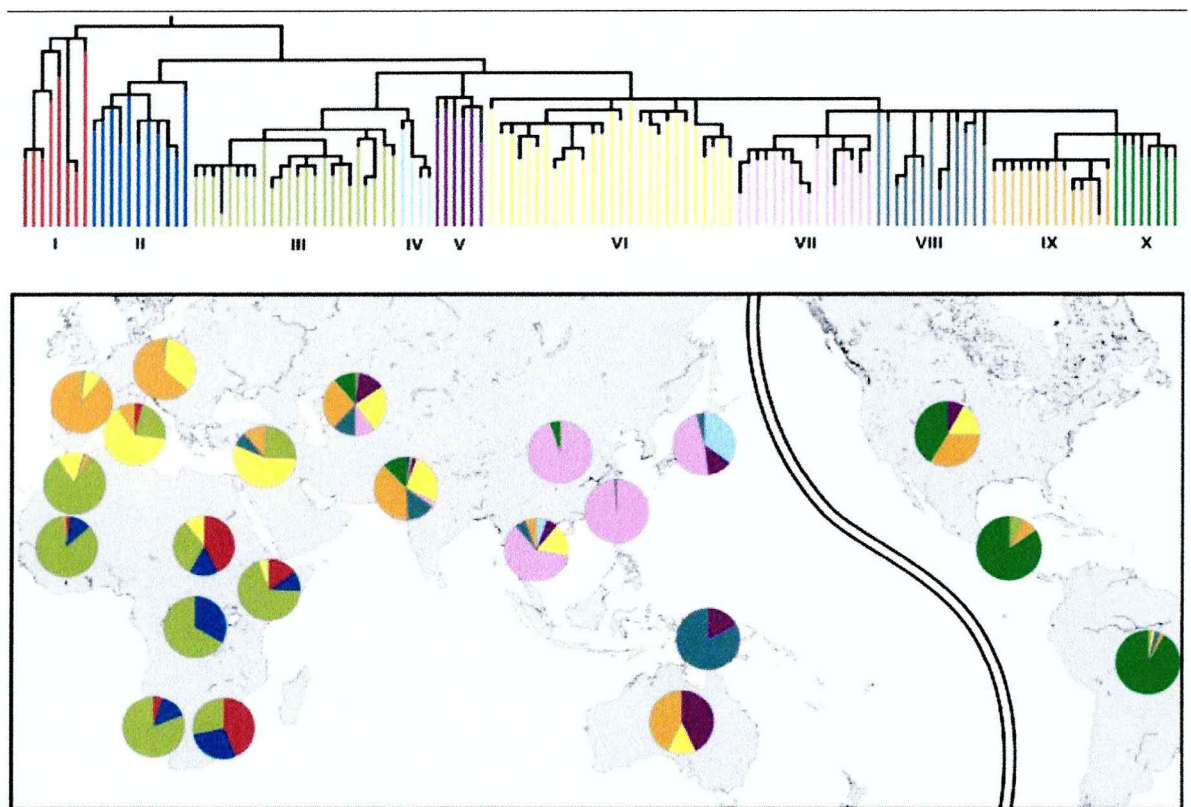


Figura 1: Distribución contemporánea de los haplogrupos del cromosoma Y en 22 regiones.

Cada haplogrupo está representado por un color diferente. Los sectores coloreados representan las frecuencias del haplogrupo. Fuente: Underhill y col. (2001)

La mutación M242 define al haplogrupo Q, cuya distribución abarca el principalmente el Norte de Eurasia y grupos siberianos (Seielstad y col. 2003, Karafet y col 2002) y es uno de los linajes mayoritarios entre los nativos americanos, con Q1a3a prácticamente restringido a América (Underhill y col. 1996, Zegura y col. 2004). Entre los nativos americanos del sur de América, sólo se encuentran los haplogrupos Q* y Q1a3a, de manera tal que cuando alguno de los individuos analizados en nuestra muestra pertenece a alguno de estos dos haplogrupos se sabrá que su ancestría patrilineal es nativa americana.

Otros haplogrupos presentes en el sur de nuestro continente se deben a migraciones peri y post conquista, originarios tanto de Europa (como el caso de los R, aunque también F, P) como de Medio Oriente (K y F) y África (DE*). Estos haplogrupos son de particular importancia, dado que actualmente son los más frecuentes en nuestro país. Principalmente, el haplogrupo R, que es el más representado tanto en el territorio argentino como en la Península Ibérica, una de las principales fuentes de migrantes hacia esta zona desde la conquista (Dipierri y col. 1998).

Los STR, también conocidos como microsatélites son marcadores constituidos por repeticiones en tandem de fragmentos cortos (de 1 a 6 pares de bases) cuyos alelos se denominan por la cantidad de repetidos. Estas mutaciones son producidas por un deslizamiento de la polimerasa durante la replicación, y su tasa mutacional, aunque varía según qué microsatélite se analice, siempre es mucho más elevada que la de los SNPs.

Según el tamaño de la unidad de repetición, los STRs tienen diferentes propiedades, distribución y uso. Aquellos con un único par de bases como repetido no suelen ser usados ya que, al ser las diferencias en el tamaño de los alelos tan pequeñas, producen problemas a la hora de definir el alelo presente debido al “tartamudeo” de la PCR (PCR es la reacción en cadena de la polimerasa, el mecanismo a través del cual se obtienen copias *in vitro* del segmento a analizar): además de obtener una banda del tamaño esperado, surgen otras bandas adicionales, típicamente un repetido más extensas y un

repetido más cortas, resultantes de errores de deslizamiento de la polimerasa durante la síntesis.

Los STRs dinucleotídicos (su repetido con una extensión de dos pares de bases) fueron usados en los primeros estudios ya que su descubrimiento es sencillo, pero actualmente no se usan tanto por el problema del “tartamudeo”. Los tri y tetranucleotídicos son los más usados, puesto que el “tartamudeo” es reducido y los alelos se discriminan fácilmente, en el caso de los penta y hexanucleotídicos, su uso es raro en vista de su relativa escasez.

Se ha propuesto un conjunto básico de 7 STRs para estandarizar las tipificaciones: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393. (Pascali y col. 1999, Kayser y col. 1997), esto permite comparar poblaciones a lo largo del mundo, entre investigadores. El detalle de estos marcadores se presenta en el Anexo.

Para los STRs, se suele aceptar al modelo de mutación de a un paso (stepwise mutation model), originalmente planteado por Ohta y Kimura (1973), donde el alelo modal posee un tamaño intermedio y los demás alelos se distribuyen en forma simétrica a ambos lados del mismo (Fig. 2). Cada vez que ocurre una mutación se aumentaría o disminuiría el tamaño en un repetido, aunque se podría ocasionalmente dar el caso de dos pasos en un único evento mutacional. Además, es importante tener en cuenta que existiría una correlación entre la longitud de las alelos y la mutabilidad, con preferencia por las deleciones en los alelos más grandes y expansiones en los más cortos.

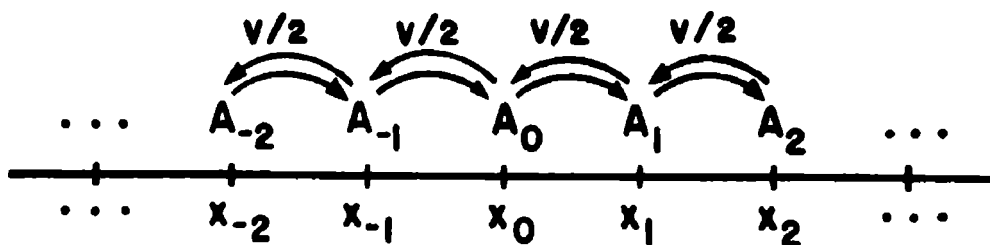


Figura 2: Esquema del Modelo de a un paso.

Fuente: Kimura y Ohta, 1978.

Cada combinación de alelos de STR heredados en bloque constituye un haplotipo. Esta es una categoría de análisis con una resolución más fina que la

del haplogrupo, puesto que cada haplogrupo abarca numerosos haplotipos posibles. Dada la alta variabilidad de STR, y el elevado número de haplotipos posibles, su análisis en el cromosoma Y es un excelente camino para establecer linajes masculinos, permitiendo la identificación de varones emparentados con precisión, tal que son rutinariamente empleados en el ámbito forense (Willuweit y Roewer 2007).

1.1.2- Haplotipos en el mundo

La colaboración tanto de investigadores como de forenses, junto con la creciente complejidad de los desarrollos informáticos, permitió la creación de bases de datos de haplotipos de todo el mundo, siendo la más importante YHRD (Willuweit y Roewer 2007). Esta base abarca 72082 haplotipos y es de acceso público, disponible en internet (<http://www.yhrd.org/>), permitiendo tanto buscar las frecuencias de haplotipos específicos como información por microsatélite (frecuencia por población, tasa mutacional, etc).

Los haplotipos almacenados en YHRD están clasificados por metapoblación de origen:

- Euroasiática: Comprende las poblaciones europea (occidental, oriental, sudeste europeo), altaica, caucásica, urálica-yukaghir, indo-iraní, india.
- Asia del Este: Abarca las poblaciones coreana, japonesa, sino-tibetana, austroasiática, tailandesa, austronesia, indo-pacífica y dravidiana.
- Aborigen Australiano
- Africana: Subdividida en sub-sahariana, afro-caribeña y afro-americana.
- Amerindia
- Eskimo Aleut
- Afroeuroasiática: Incluye las poblaciones semítica, bereber y cushitica.
- Mezclada

En la Fig. 3 se observan las poblaciones presentes en esta base de datos.

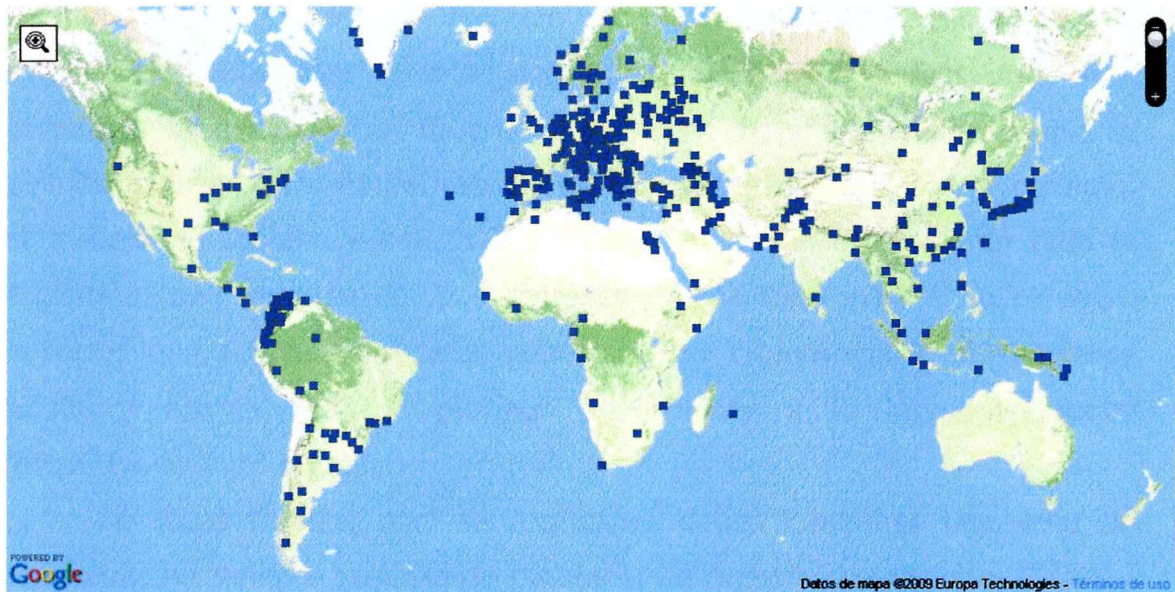


Figura 3: Poblaciones incluidas en YHRD

Si bien esta base no cuenta con la información del haplogrupo al cual corresponden los haplotipos presentes, puede dar una aproximación a en cuales regiones geográficas es más frecuente, y si se trata de un haplotipo muy común o raro.

1.2 ¿Por qué los apellidos?

Los apellidos y las genealogías establecidas a partir de ellos han tenido una gran importancia en múltiples culturas a lo largo de su historia. Generalmente se transmiten por una única línea, por ejemplo la paterna, permitiendo asumir que los individuos con un mismo apellido deben presentar relaciones de parentesco genético, es decir, que provienen de un mismo linaje con un antecedente común. Sobre esta asunción se basa la isonimia, que avala construir estas genealogías y estimar el grado de contacto entre poblaciones mediante el escudriñamiento de archivos de distinto origen, con el objetivo de dilucidar las líneas de parentesco que permitan establecer las genealogías, como expresan Castilla y Adams (1990, 1996).

Su aplicación en la bioantropología se basa en que es una fuente de información de fácil acceso y económica (se puede emplear tanto la guía

telefónica como los padrones electorales), además de que posibilitan no sólo estudios de poblaciones actuales sino que también de históricas. Asimismo, permiten obtener muestras de gran tamaño, mucho mayores que las que se podrían analizar a través del estudio de marcadores genéticos. Por ejemplo, el trabajo de Dipierri y col. 2005a analiza el padrón electoral de toda la Argentina (una muestra de 22.6 millones de electores para todo el país), mientras que la base de datos genéticos del cromosoma Y más extensa existente abarca 72082 haplotipos, distribuidos a lo largo del globo (<http://www.yhrd.org/>).

De todas formas, para ser utilizados deben cumplirse una serie de supuestos, no siempre posibles de satisfacer en nuestras poblaciones (Dipierri y col. 1991). Por un lado, es necesario que los apellidos sean monofiléticos (es decir, que provengan de un único ancestro común). Este requisito es uno de los más difíciles de cumplir puesto que el polifiletismo en muchos apellidos, como aquellos que derivan de apodos o profesiones, es altamente probable.

Otro supuesto es que su transmisión sea vertical, es decir, que el apellido se transmita en sentido progenitor-descendiente, lo que no contemplaría las adopciones e ilegitimidad.

Asimismo también es necesaria la neutralidad de selección de los apellidos, que se distribuyan homogéneamente a lo largo de la descendencia y su constancia a lo largo del tiempo. Todos estos requisitos son difíciles de cumplir cuando se tienen en cuenta los fenómenos sociales que pueden afectar tanto la ortografía de los apellidos como la neutralidad de los mismos, dado que se asocian a linajes sociales y es posible que en algunos casos existan ventajas para los portadores de apellidos específicos.

Trabajos relevantes de isonimia han sido realizados tanto fuera como dentro de nuestro país, en vista a lo expresado por Colantonio y col. (2003) en su revisión, y diferentes modelos matemáticos para explicar la relación entre la diversidad genética presente en una población a partir de los apellidos, son analizados en el trabajo de Manrubia y Zanette (2007).

En la escala local, ya desde Eric Boman, quien dentro de sus investigaciones del NOA incluyó un relevamiento de nombres y apellidos de esa región (Boman, 1908), Dipierri y col. (1991) estimaron el coeficiente de

endogamia en la población histórica de la Quebrada de Humahuaca a partir de este método, Dipierri y Alfaro (1996) evaluaron la dinámica poblacional de comunidades de altura en la provincia de Jujuy a partir de estimaciones de aislamiento y sedentarización obtenidas de la frecuencia y distribución de los apellidos, Albeck y col. (2005) registraron la perduración y dispersión de nombres o apellidos autóctonos de Casabindo a través de archivos históricos que cubren períodos temporales desde 1557 al año 2000, donde 21 de estos nombres se mantuvieron desde mediados del siglo XVII al presente, Dipierri y col. (2005a) evaluaron la estructura genética, la migración reciente, el aislamiento y las relaciones entre las poblaciones que integran el NOA a través de los apellidos.

Los estudios moleculares han permitido trazar desde perfiles moleculares, estructura y flujo génico de poblaciones nativas (Catanesi y col. 2001, Demarchi y Mitchell, 2004), identificar ancestrías, (Fejerman y col. 2005), hasta árboles familiares desde otra perspectiva, dando lugar a nuevos estudios que abarquen la información del archivo celular. Un ejemplo es la identificación de los restos del último Zar de Rusia y su familia, realizado a partir de su ADN mitocondrial y comparándolo con linajes maternos actuales de su genealogía (Gill y col. 1994). También cabe mencionar el estudio de los Lemba, un grupo de lengua bantú que se reconoce como descendiente de judíos de Sena (que habría estado ubicada en Yemen, Judea, Egipto o Etiopía), donde se contrastó su tradición oral con estudios con marcadores de cromosoma Y, encontrando linajes hebreos dentro del grupo (Thomas y col. 2000). En una amplia región de Asia Central se identificó un grupo de linajes que representaron el 8% de los cromosomas (0.5% mundial total), cuyo ancestro común más reciente se remonta alrededor de 1.000 años atrás, y que putativamente derivan del prolífico Genghis Khan (Zerjal y col. 2003). De esta forma es que se cuenta con dos estrategias diferentes, la isonimia y los estudios genéticos, para analizar esta problemática.

Cuando los apellidos se heredan por línea paterna, los mismos se distribuyen junto con el cromosoma Y, de ahí que algunos investigadores (Manni y col. 2005, Mikerezi y col. 2003, Yuan y col. 2000) hayan tomado a ambas estrategias, la molecular y la histórica como intercambiables. Sin

embargo, surgen dudas respecto a si es cierto que los apellidos idénticos representan una ancestralidad común (Jobling, 2001), en respuesta a esto Sykes e Irven (2000) correlacionaron al apellido Sykes con los haplotipos del cromosoma de sus portadores, encontrando más de un linaje paterno para este apellido.

Otros trabajos también testearon esta hipótesis (Immel y col. 2006; King y col. 2006), resultando en un cierto grado de concordancia entre linajes genéticos y sociales.

Soodyall y col. (2003) compararon los linajes fundadores de la isla de Tristan da Cunha basándose en las genealogías tradicionales con linajes establecidos a partir de polimorfismos del cromosoma Y, encontrando siete que se remontaban a aquellos ancestros y dos nuevos, uno originado por un único evento mutacional en un microsatélite y otro foráneo, no reflejado en los registros históricos. Trumme y col. (2004) buscaron reconstruir la relación entre tres líneas familiares diferentes con el apellido Zierdt, incluyendo candidatos de varias ramas familiares identificadas por registros escritos. Al analizar los haplotipos en dos líneas familiares se encontró que pertenecían al mismo linaje paterno, mientras que la tercer línea reveló un haplotipo desviado cuya relación podría excluirse, además de hallar cuatro casos de paternidades mal atribuidas que fueron puestas en contexto histórico.

De esta forma, al cruzar ambas fuentes de información se establece un lazo entre las genealogías tradicionales y el análisis de ADN, develando la historia silenciosa de lo que la gente hizo, en vez de sólo sus registros orales e históricos, con una profundidad cronológica que excede la memoria humana. Este enfoque dio lugar a una nueva disciplina llamada “Heráldica Molecular”, lanzada por el Laboratorio de Genética Molecular Poblacional del IMBICE, con el estudio de muestras de individuos pertenecientes a una genealogía de la localidad de Aicuña en La Rioja. Se encontró concordancia entre la genealogía histórica convencional (Castilla y Adams, 1990, 1996) y la reconstrucción de los linajes ancestrales para la genealogía, en un 85% de los casos (Bailliet y col. 2001). Trabajos más recientes en esta área fueron presentados en la Reunión Anual de la Sociedad Argentina de Genética 2006 (Muzzio y col.

2006), 2007 (Muzzio y col. 2007a), en las Octavas Jornadas Nacionales de Antropología Biológica de la Asociación de Antropología Biológica Argentina a realizadas en octubre del año 2007 (Alfaro y col. 2007, Bailliet y col. 2007, Muzzio y col. 2007b,) y en el X Congreso de la Asociación Latinoamericana de Antropología Biológica (Muzzio y col. 2008a, 2008b, 2008c).

Además de los casos comentados, debe tenerse en cuenta que la evangelización y la ocupación extranjera, de la misma forma que los nacimientos extramaritales, pueden desvirtuar el supuesto subyacente al método isonímico (es decir, que los apellidos idénticos representan ancestralidad común), como fue expresado en el trabajo de Dipierri y col. (1998), donde hallaron cromosomas Y de origen español en poblaciones amerindias, además de cromosomas Y amerindios en personas con apellidos hispánicos.

Otros trabajos interesantes que nos señalan las posibles discordancias entre la trasmisión de apellidos y los linajes del cromosoma Y son el de Gil Montero (2007) y el de Robinchaux (2007), que exponen la complejidad de las familias americanas, su alta variabilidad, el importante número de hijos extramaritales entre otros factores.

De esta forma habría una disparidad entre los métodos isonímico y molecular, que genera incertidumbre sobre la exactitud de las genealogías construidas con el primero de los mismos. En respuesta a esta problemática es que se buscará cuantificar con exactitud las coincidencias y disparidades entre estos métodos.

1.3 ¿Por qué el Noroeste Argentino?

El Noroeste Argentino (NOA) es un área que comprende cinco provincias: Catamarca, Jujuy, Salta, Tucumán y Santiago del Estero, abarcando una gran diversidad de ambientes. Estuvo sujeta a complejos procesos de disrupción étnica, desde el poblamiento americano, el desarrollo de diferentes sociedades prehispánicas, a la llegada y asentamiento de los españoles. Estos procesos incluyeron extinciones, traslocaciones, migraciones y conversiones religiosas, en muchos casos forzadas, que modificaron el

paisaje demográfico. Por estos motivos, el NOA tiene características particulares y compartidas que la vuelven un objeto interesante para estudios antropológicos.

Fue una de las áreas más densamente pobladas en épocas precolombinas (Martínez-Sarazola, 1999), la historia del contacto con los europeos se remonta a la fundación de la ciudad de Santiago del Estero en 1533 (Binayán Carmona, 1999), además de que existen numerosos registros históricos que permiten interpretar los linajes paternos y trabajos sobre isonimia, (ver apartado “¿Por qué los apellidos?”), de ahí que se cuenta con información inherente a esta disciplina, tales como la frecuencia de los apellidos en las distintas regiones y cuáles fueron utilizados como apellidos tipo durante la evangelización, la distinción entre apellidos nativos según el deme de origen, y consideraciones sobre la compleja matriz de poblaciones nativas que cohabitaron el NOA.

Objetivos

“Y que yo me la llevé al río
creyendo que era mozuela,
pero tenía marido...”

F. García Lorca Romancero Gitano, “La casada infiel”

El objetivo general del presente trabajo es contribuir a la comprensión de los fenómenos de configuración del acervo genético de poblaciones aborígenes y urbanas argentinas, conformado por los componentes étnicos aborígenes, africanos y europeos, así como también complementar la información sobre la compleja red de parentescos denotadas por estudios previos de las disciplinas de isonimia y genealogía.

Teniendo en cuenta que los apellidos y los sistemas específicos del cromosoma Y se transmiten por línea paterna, el objetivo específico de este trabajo es evaluar la similitud de los linajes genéticos tanto de individuos con un mismo apellido como de aquellos que no lo comparten, en el NOA. De ahí que se construirán haplotipos del cromosoma Y, luego de identificar su ancestría a través de la determinación de haplogrupos.

También se examinará si existe asociación entre apellidos y ancestrías, es decir, si los individuos con un apellido nativo también pertenecen a un linaje nativo americano.

Además, se analizará si, en el caso de encontrar disparidades, las mismas se corresponden a mutaciones recientes o si se trata de casos de polifiletismo y/o transmisión irregular de los apellidos, posible a través de la estimación del tiempo al ancestro común más reciente: si es mayor que el origen de los apellidos se trata de polifiletismo, si es menor probablemente se debe a mutaciones recientes.

Materiales y Métodos

“... Sólo los Constructores, los Formadores, los Dominadores, los Poderosos del Cielo, los Procreadores, los Engendrados, estaban sobre el agua, luz esparcida. (Sus símbolos) estaban envueltos en las plumas, las verdes; sus nombres (gráficos) eran, pues, Serpientes Emplumadas. Son grandes Sabios. Así es el cielo, [así] son también los Espíritus del Cielo; tales son, cuéntase, los nombres de los dioses...”

Popol-Vuh, Anónimo. Traducción de M. A. Asturias y J. M. González de Mendoza.

2.1 Poblaciones seleccionadas y muestra

Se recolectaron muestras de 175 varones procedentes de las capitales de las provincias de Tucumán (n=9), Salta (n=47), Catamarca (n=76) y Jujuy (n=43)(Figs 4). Corresponden a muestras de sangre, donadas voluntariamente y codificadas, de manera tal que están irreversiblemente disociadas de una persona identificable, respetando la privacidad de los donantes.

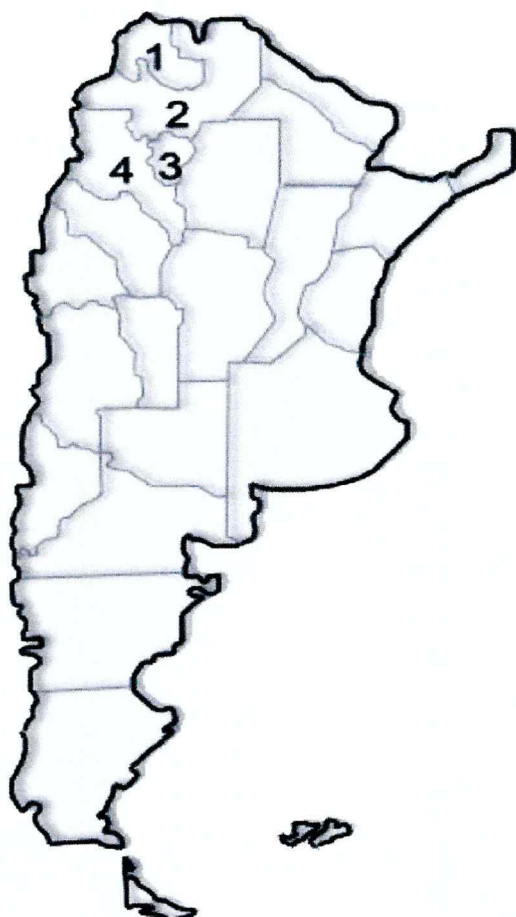


Figura 4: Mapa de la República Argentina donde se señalan las provincias muestreadas.

Referencias:

1. Jujuy
2. Salta
3. Tucumán
4. Catamarca

El proyecto de esta tesis contó con la aprobación del Comité de Ética en Investigaciones Biomédicas del IMBICE y del Comité Provincial de Bioética de la provincia de Jujuy. Se adjuntan en el Anexo la Cartilla de Explicación para los donantes, así como el modelo del Consentimiento Informado escrito.

Al tratarse de una selección de varones al azar, el criterio de inclusión de apellidos que constituyeron la muestra se limita a ser aquellos apellidos que los donantes poseían. De ahí que se incluyeron 94 apellidos, 8 nativos y 72 de origen español (de los cuales 4 eran vascos), 3 de otros orígenes mientras que 12 no pudieron ser clasificados. Los apellidos de origen nativo y español son los más comunes en el NOA (Dipierrri y col. 2005a), y permiten evaluar el impacto de los bautismos puesto que fueron empleados por los misioneros en el proceso de evangelización.

Cincuenta y dos apellidos estaban representados por un solo individuo (apellidos únicos), pero también se trabajó con un apellido que abarcaba 9 individuos en la muestra (Cruz). El promedio de individuos por apellido es de 2, con un desvío estándar de 1.

Se clasificó a los apellidos en dos grupos:

- Comunes: aquellos postulados dentro de los cien más frecuentes en Argentina (Dipierrri y col. 2005a). Dentro de esta categoría incluimos un total de 35 apellidos.
- Raros: aquellos no incluidos dentro de los cien más comunes. Este grupo abarcó 60 apellidos.

En el Anexo se presenta la Tabla 1 (numeración del Anexo) que indica la cantidad de individuos portadores de los apellidos estudiados, en Salta, Jujuy, Catamarca y Tucumán, según en padrón electoral del año 2001, proporcionado por la Cámara Nacional Electoral.

2.2 Extracción del ADN, tipificación de SNP y de STR

Se extrajo ADN a partir de las muestras de sangre periférica, mediante el empleo del kit de extracción “QIAamp DNA Blood Kit, QIAGEN” (Fig. 5). Una fracción de cada muestra se diluyó en agua millique para conformar alícuotas de trabajo (1:10), con la intención de preservar las muestras de los daños por cambios de temperatura, además de evitar el derroche de material.



Figura 5: Extracción de ADN

Los haplogrupos fueron determinados previamente, a través de análisis de polimorfismos de un sólo nucleótido (SNP), en el trabajo de Ramallo y col. (2005), de manera tal que se pudo separar los cromosomas Y de origen nativo de aquellos extranjeros.

Para el establecimiento de los haplotipos se eligieron los microsatélites (STR) del conjunto básico propuestos por Kayser y col. (1997) y Pascalli y col. (1999), amplificados a través de los cebadores y ciclos detallados, realizándose la reacción en múltiplex en aquellos casos en que las temperaturas de ciclado coincidieran:

- DYS19: Es un tetranucleótido, con 10 alelos conocidos. Cebadores: 5' GAC TAC TGA GTT TCT GCA ATA GTG 3'. 5' AGC ACT GCA CCT GGA AAT AG 3'.

- DYS389 I: Es un tetranucleótido, con 9 alelos conocidos. Cebadores: 5' CCA ACT CTC ATC TGT ATT ATC TAT G 3'. 5' TCT TAT CTC CAC CCA CCA GA 3'.
- DYS389 II : Es un tetranucleótido, con 9 alelos conocidos. Cebadores: Utiliza los mismos que DYS 389 I.
- DYS390: Es un tetranucleótido, con 12 alelos conocidos. Cebadores: 5' TAT ATT TTA CAC ATT TTT GGG CC 3'. 5' TGA CAG TAA AAT GAA CAC ATT GC 3'.
- DYS 391: Es un tetranucleótido, con 9 alelos conocidos. Cebadores: 5' CTA TTC ATT CAA TCA TAC ACC CA 3'. 5' GAT TCT TTG TGG TGG GTC TG 3'.
- DYS 392: Es un trinucleótido, con 11 alelos conocidos. Cebadores: 5' TCA TTA ATC TAG CTT TTA AAA ACA A 3'. 5' AGA CCC AGT TGA TGC AAT GT 3'.
- DYS393: Es un tetranucleótido, con 9 alelos conocidos. Cebadores: 5' GTG GTC TTC TAC TTG TGT CAA TAC 3'. 5' AAC TCA AGT CCA AAA AAT GAG G 3'.

2.2.1 Ciclos:

- DYS 19: 94°C por 3'. Después 35 ciclos de: 94°C por 1', 57°C por 1', 72°C por 1'. Elongación final 72°C 5'
- DYS 389, DYS 390, DYS 391: 94°C por 2'. Después 40 ciclos de: 94°C por 45'', 54°C por 1', 72°C por 90''. Elongación final 72°C 5'
- DYS 392, DYS 393: 94°C por 2'. Después 40 ciclos de: 94°C por 45'', 51°C por 1', 72°C por 90''. Elongación final 72°C 5'.

Los equipos empleados como cicladores fueron PTC Programmable Thermal Controller (MJ Research, Inc.) e Y DNA Thermal Cycler (PERKIN ELMER). (Fig 6, 7).



Figura 6: Y DNA Thermal Cycler, PERKIN ELMER



Figura 7: Programmable Thermal Controller, MJ Research, Inc.

Las reacciones se verificaron a través de electroforesis en geles de poliacrilamida al 6%, a un voltaje de 230 voltios durante una hora y media (para DYS19, 392 y 393) o dos horas y media (para DYS 389, 390 y 391) (Fig 8). Luego se tipificaron los alelos mediante electroforesis en geles de poliacrilamida desnaturizante a un voltaje de 40 vatios por cinco (DYS 393, 392) seis (DYS 389) y siete horas y media (DYS 19, 390, 391). Las fuentes de poder empleadas fueron PowerPac 3000 (BIO-RAD) y Model 250 BRL (Life

Technologies, Inc., Research Products Division). (Fig 9). En todos los casos, la visualización fue posible a través del proceso de tinción en plata.

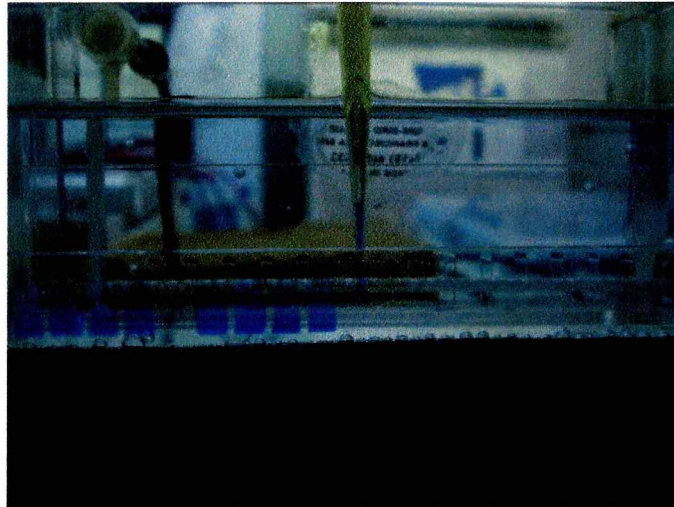


Figura 8: Siembra de un gel de acrilamida al 6%



Figura 9: PowerPac 3000, BIO-RAD

2.3 Asociaciones entre haplogrupos y apellidos

En una primera instancia se quiso ver si había haplogrupos compartidos dentro de cada apellido, aún a la baja resolución de nuestro método, como una forma de estimar a gran escala el impacto de los bautismos realizados por los misioneros españoles. La proporción de correspondencia de haplogrupo dentro de cada apellido fue comparada con aquella encontrada en Inglaterra

por otros investigadores (King y col., 2005), usando un ensayo estadístico de diferencias de proporciones (Walpole y Myers, 1992):

$$Z = \frac{p_1 - p_2}{\sqrt{pq [(1/n_1) + (1/n_2)]}}$$

donde

$$p_1 = \frac{x_1}{n_1}$$

$$p_2 = \frac{x_2}{n_2}$$

$$p = \frac{x_1 + x_2}{n_1 + n_2}$$

x_1 = cantidad de correspondencias de haplogrupo dentro de los apellidos en la muestra.

n_1 = cantidad de casos en la muestra.

x_2 = cantidad de correspondencias de haplogrupo dentro de los apellidos en la bibliografía.

n_2 = cantidad de casos en la bibliografía.

Este mismo ensayo fue usado para ver si la proporción de individuos con un apellido nativo era equivalente a la de individuos con un haplogrupo nativo (Q1a3a o Q242).

Se estimó el Riesgo Relativo para analizar si existía una asociación entre poseer un apellido nativo y un haplogrupo americano, así como también entre tener un apellido extranjero y un haplogrupo foráneo.

El Riesgo Relativo del suceso del grupo A frente al del grupo B se calcula como:

| | Grupo A | Grupo B |
|-----------|---------|---------|
| Suceso | a | b |
| No suceso | c | d |
| Total | N A | N B |

El riesgo en el Grupo A es el cociente a/N_A mientras que en el Grupo B es el cociente b/N_B . El Riesgo Relativo es el cociente entre estos dos riesgos, indicando cuánto más probable es que ocurra el suceso en el primer grupo frente al segundo.

2.4 Análisis de haplotipos

Con los haplotipos obtenidos se construyeron redes medianas (Bandelt y col. 1999), una para cada haplogrupo. Esta metodología permite reconstruir filogenias a nivel intraespecífico, con datos poblacionales libres de recombinación, sin resolver “empates” entre árboles y además es aplicable a caracteres de estados múltiples con gran velocidad de cálculo. El análisis fue llevado a cabo mediante el programa NETWORK 4.5 (fluxus-engineering.com), que admite el pesado de caracteres. Esto último es necesario, ya que permite tomar en cuenta la importancia diferencial de cada marcador a la hora de establecer la red mediana. De esta forma se señalan las conexiones más plausibles, siendo más sencillo interpretar la red resultante.

Así, a los eventos poco probables (en este caso los microsatélites con tasas mutacionales bajas) se les asigna un peso más elevado, en vista de que cuando suceden son significativos, mientras que a los eventos más probables (los microsatélites con tasas mutacionales altas) se les da un peso menor.

Para este trabajo, se atribuyeron pesos diferenciales a los STR, siguiendo las tasas mutacionales descritas en la base de datos internacional YHRD (Willuweit y Roewer, 2007) y resumidas en la Tabla 1. La estimación del peso de los caracteres se realizó mediante la fórmula descrita en el Capítulo “Nuevos Desarrollos y Validaciones Metodológicas”, apartado 3.1.

Tabla 1: Tasas Mutacionales y pesos relativos por microsatélite

| STR | N. de meiosis | N. de mutaciones | Tasa mutacional (x10 ⁻³) | Años por mutación | Peso relativo estimado |
|------------|---------------|------------------|--------------------------------------|-------------------|------------------------|
| DYS 19 | 9390* | 23* | 2.45 (1.55-3.67 95% CI)* | 408 | 4 |
| DYS 389 I | 7594* | 18* | 2.37 (1.41-3.74 95% CI)* | 422 | 4 |
| DYS 389 II | 7581* | 26* | 3.43 (2.24-5.02 95% CI)* | 292 | 3 |
| DYS 390 | 8872* | 21* | 2.37 (1.47-3.62 95% CI)* | 422 | 4 |
| DYS 391 | 8821* | 25* | 2.83 (1.83-4.18 95% CI)* | 353 | 3 |
| DYS 392 | 8785* | 4* | 0.45 (0.12-1.17 95% CI)* | 2222 | 10 |
| DYS 393 | 7574* | 6* | 0.79 (0.29-1.72 95% CI)* | 1266 | 7 |

*Fuente: YHRD database

La búsqueda de haplotipos comunes fue realizada por conteo directo, tanto dentro como a través de los apellidos, y posteriormente corroborada mediante el programa informático Arlequin (Schneider y Excoffier, 1999).

2.5 Distancias temporales

Las estimaciones de distancia temporal al ancestro común más cercano se realizaron con el programa NETWORK 4.5, asumiendo las tasas mutacionales descritas en la Tabla 1, y 25 años por generación. De esta manera, el parámetro de años por mutación es redefinido según qué microsatélite denota la rama (véase la Tabla 1 para estos valores convertidos), y en el caso de haber diferentes marcadores se empleó un promedio entre ellos.

Este software emplea el estadístico ρ (“rho”), que mide la edad de un nodo ancestral en unidades mutacionales que luego son convertidas en años mediante la multiplicación con la tasa mutacional.

$$\rho = \mu t$$

Donde:

μ = tasa mutacional

t = tiempo

La edad de un nodo es a su vez una estimación temporal mínima para la mutación precedente. El desvío estándar sigma también es provisto por el programa, basado en un árbol estimado dentro de la red.

Una ventaja del estadístico ρ es ser inherentemente no sesgado, porque es libre de modelo. Si bien las demografías del pasado influenciaron la forma del árbol evolutivo, esto sólo influencia en el error de la estimación temporal en lugar de sistemáticamente aumentarla o disminuirla.

Con la intención de poder establecer si las diferencias entre los haplotipos de un mismo apellido se debían a mutaciones posteriores al origen de los mismos o a casos de polifiletismo, se evaluó la divergencia temporal entre individuos del mismo apellido y haplogrupo, tomándolos de a pares y siguiendo las mismas condiciones anteriormente mencionadas. Para el haplogrupo R la fecha considerada como origen de los apellidos fue el siglo IX, mientras que para el Q1a3a se consideró el siglo XVII, que se supone como el momento en el cual los misioneros expandieron sus bautismos en esta región (Dipierri y col. 1991, 1996, 2005a y 2005b).

Se eligió el criterio de 25 años por generación ya que es la edad más joven propuesta por la bibliografía. Si bien la mayoría de los autores concuerdan con una edad de 35 años para el caso del cromosoma Y (Jobling y col. 2003), elegimos la menor edad de manera tal que nos estime el menor lapso posible. De esta forma, cuando la distancia temporal es mayor al origen de los apellidos, se descartaría la posibilidad de que el tiempo generacional

hubiera sido considerablemente menor al criterio elegido, reasegurando el polifiletismo.

2.6 Comparaciones a nivel mundial

Además, se buscaron los haplotipos en YHRD (Willuweit y Roewer, 2007), que abarca 72082 haplotipos y es de acceso público, disponible en internet. De esta forma, se compararon las frecuencias encontradas con aquellas presentes en la base de datos, permitiendo ver si los haplotipos representados por más de un individuo en la muestra se corresponden a haplotipos modales a gran escala.

2.7 Estimaciones de la proporción de transmisión no paterna del apellido

Se analizaron encuestas genealógicas relevadas en el marco del Proyecto “Pasado Indígena, presente criollo. Diversidad filética y estructura geográfica de los linajes maternos aborígenes presentes en poblaciones criollas argentinas”, cuyo investigador responsable es el Dr. Claudio Bravi, PICT 2005 N°32450, FONCyT en las poblaciones de:

- Aguaray, Macueta (Salta): 44 encuestas, abarcando 110 traspasos.
- Belén y Londres (Catamarca): 158 encuestas con 434 traspasos.
- Calingasta (San Juan): 70 encuestas que incluyen 187 traspasos.
- Chepes (La Rioja): 75 encuestas, incluyendo 210 traspasos.
- Curuzú-Cuatiá y Perugorría (Corrientes): 196 encuestas, abarcando un total de 547 traspasos.
- Galeguaychú (Entre Ríos): 216 encuestas, 603 traspasos.
- La Paz (Entre Ríos): 72 encuestas, 190 traspasos.
- La Quiaca (Jujuy): 224 encuestas con 597 traspasos.
- La Rioja Capital: 239 encuestas que abarcan 667 traspasos.
- Lavalle (Mendoza): 115 encuestas con un total de 318 traspasos.
- Maimará (Jujuy): 200 encuestas con 534 traspasos.

- Malargüe (Mendoza): 151 encuestas con 432 traspasos.
- Mendoza Capital: 145 encuestas, con 388 traspasos.
- Reconquista (Santa Fe): 135 encuestas, con un total de 364 traspasos.
- San José (Catamarca): 22 encuestas, con 59 traspasos.
- San Juan Capital: 112 encuestas, abarcando 320 traspasos.
- Santa María (Catamarca): 151 encuestas, 411 traspasos.
- Tartagal (Salta): 225 encuestas, incluyendo 583 traspasos.

Cada encuesta cuenta con información del ego (G1), sus padres (G2) y los abuelos y abuelas del ego (G3). Una copia de la encuesta genealógica se adjunta en el Anexo.

Estas encuestas fueron copiadas en planillas de Microsoft Excel, y se contabilizaron los traspasos de apellido de la G2 a la G1 por un lado (transmisión ego) y de la G3 a la G2 (transmisión parental), clasificados en “paterno” cuando provenía del padre, “materno”, cuando era el de la madre, “otro” adopciones y casos especiales, y “sin datos”. Se establecieron las proporciones de cada una de estas categorías para cada población y transmisión, y se comparó la proporción de traspasos no paternos en la transmisión ego con la de la transmisión parental, en cada población, con el objetivo de revisar si había diferencias en esta proporción a lo largo del tiempo. La prueba empleada en este caso fue la de diferencia entre dos proporciones (fórmula op. cit.) de Walpole y Myers (1992).

Además, se comparó la proporción de traspaso no paterno entre todas las poblaciones involucradas, mediante la prueba para varias proporciones (Walpole y Myers, 1992), con el estadístico chi cuadrada:

$$\chi^2 = \sum_i \frac{(o_i - e_i)^2}{e_i} \quad \text{Donde:} \quad \begin{array}{l} o: \text{proporciones observadas} \\ e: \text{proporciones esperadas} \end{array}$$

con,

v = grados de libertad

k = cantidad de poblaciones

$$v = (2 - 1)(k - 1) = k - 1$$

También se estimó el Riesgo Relativo para estudiar la posible asociación entre un traspaso materno en la transmisión ego y la parental (fórmula op. cit.).

Se planteó una serie de modelos para analizar el efecto de las transmisiones irregulares del apellido, estimando en cuántas generaciones el porcentaje de linajes que mantuvieron la transmisión paterna se reduce a la mitad.

El más sencillo de estos modelos asume una cantidad infinita de apellidos, y no considera el crecimiento poblacional. Proviene de la regla multiplicativa de probabilidad (Walpole y Myers, 1992), que permite calcular la probabilidad de que dos (o más) eventos sucedan. Según la misma, la probabilidad de que estos eventos sucedan equivale al producto de las probabilidades de los mismos.

En la generación 1 contamos con un único evento:

$$P_1 = (p)$$

Donde:

$$P = \frac{n}{N}$$

n = cantidad de los resultados que corresponden al evento.

N = cantidad total de resultados.

En la generación 2 se requiere el evento en la generación 1 y en esta generación, por eso:

$$P_2 = (p) \times (p) = p^2$$

En la generación 3 es necesario que se mantuviera el evento en las dos generaciones anteriores, además de que ocurra en ella:

$$P_3 = (p) \times (p) \times (p) = p^3$$

De esta forma, en la generación “n” sería:

$$P_n = (p)^n$$

En el caso a modelar:

p = la probabilidad de un individuo de recibir el apellido de su padre (obtenida a partir de las encuestas genealógicas).

q = la probabilidad de un individuo de recibir un apellido distinto del de su padre (se suman los traspasos “maternos” y los “otros”).

A_1 = Probabilidad de que un apellido de la generación 1 mantenga la transmisión paterna.

A_n = Probabilidad de que un apellido de la generación n haya mantenido la transmisión paterna desde la generación 1.

De manera tal que

$$A_n = (p)^n$$

Si la probabilidad de transmisión paterna varía entre las generaciones, y se cuenta con esas probabilidades, se aplica la misma regla multiplicativa, sólo que en lugar de elevar una probabilidad a un número equivalente de generaciones, se multiplican cada una de las probabilidades.

$$A_n = p_1 \times p_2 \times \dots \times p_n$$

Esta adaptación del modelo sólo será empleada en aquellos casos que exista una diferencia significativa entre la proporción de traspasos no paternos de las generaciones ego y parental.

Nuevos Desarrollos y Validaciones Metodológicas

“(...)

Adieu l'Antoine je vais mourir
C'est dur de mourir au printemps tu sais
Mais je pars aux fleurs la paix dans l'âme
Car vu que tu étais son amant
Je sais que tu prendras soin de ma femme

(...)

Adieu ma femme je vais mourir
C'est dur de mourir au printemps tu sais
Mais je pars aux fleurs les yeux fermés ma femme
Car vu que je les ai fermé souvent
Je sais que tu prendras soin de mon âme.

(...)”

Le Moribond, J. Brel

3.1 Desarrollo de la nueva propuesta para el pesado de caracteres

La importancia de definir un peso para cada carácter en el trazado de las Redes Medianas ya fue detallado en el apartado Materiales y Métodos. De todas formas, Bandelt y col. (1999) no proponen de qué forma es más adecuado establecer estos pesos, y los sistemas de asignación seguidos en la bibliografía no resultan satisfactorios.

El método de asignación de pesos más citado es el de Qamar y col. (2002), empleado por Bolnick y col (2006), King y col. (2007), Malhi y col. (2008), entre otros. Este método consiste en un esquema fijo basado en la varianza de cada microsatélite. Estos autores proponen asignar pesos de 1 a 5, de la siguiente manera:

- Varianza entre 0 y 0.09: peso = 5.
- Varianza entre 0.1 y 0.19: peso = 4.
- Varianza entre 0.2 y 0.49: peso = 3.
- Varianza entre 0.5 y 0.99: peso = 2.
- Varianza ≥ 1 : peso = 1.

Los mismos autores de esta clasificación comentan en su trabajo que, a pesar de emplearla, sus Redes Medianas resultantes contenían muchos cubos de altas dimensiones (cabe aclarar que los mismos ocurren cuando la red no logra ser trazada de una forma clara y dificultan la comprensión), motivo por el cual necesitaron, además, utilizar el algoritmo de Media Reducida antes del de Red Mediana. Más allá de esto, el método de Qamar y col. (2002) presenta otras complicaciones no discutidas ni planteadas por sus autores:

- Basar el peso en la variación de microsatélites requiere haberla estimado previamente. En el caso de estos autores, la calculan con los mismos datos con los cuales luego trazan la red.
- Esta variación puede verse reducida por procesos de deriva, de manera tal que, en una población dada, podríamos encontrar una varianza baja

para un microsatélite altamente polimórfico en otras, como resultado de un cuello de botella reciente.

- Aún si las varianzas de los microsatélites fueran estimadas a partir de extensas bases de datos (como YHRD), cuando se tiene en cuenta el modelo mutacional de a un paso, específicamente su característica de presentar tamaños mínimos y máximos posibles, se evidencia como en casos de mutaciones altamente recurrentes la varianza resulta subestimada.
- Se trata de un rango de pesos muy acotado, poco útil para casos en los cuales se empleen muchos microsatélites distintos.

Por estas razones fue necesario elaborar para la presente tesis una fórmula nueva para el pesado de caracteres en la realización de las Redes Medianas:

$$P(x) = M (\text{tasa mutacional}_y / \text{tasa mutacional}_x)^a$$

Donde:

“P(x)” es el peso de un microsatélite dado.

“M” es el valor de peso máximo que el investigador decide emplear.

“tasa mutacional_y” es la tasa mutacional más baja en el set de microsatélites analizados.

“tasa mutacional_x” es la tasa mutacional del microsatélite para el cual el investigador desea estimar el peso.

“a” es el valor que le permite al investigador adaptar la escala resultante al rango deseado.

Para estimarlo:

$$\text{Menor valor de peso} = M (\text{tasa mutacional}_y / \text{tasa mutacional}_x)^a$$

Que es convenientemente pasado a una forma logarítmica, lineal en “a”, que es la incógnita, y simplifica su cálculo:

$$\text{Log(menor valor de peso)} = \text{Log}(M) + a\{\text{Log}(\text{tasa mutacional}_y) - \text{Log}(\text{tasa mutacional}_x)\}$$

De manera que:

$$A = \frac{\{\text{Log}(\text{menor valor de peso}) - \text{Log}(M)\}}{\{\text{Log}(\text{tasa mutacional}_y) - \text{Log}(\text{tasa mutacional}_x)\}}$$

Se buscó un “a” que permitiera que el valor mínimo asignado fuera 3 y el valor máximo 10. Entonces, tomando las tasas mutacionales de YHRD:

$$3 = 10 (0.45 \cdot 10^{-3} / 3.43 \cdot 10^{-3})^a$$

$$a = 0.593$$

Los pesos así calculados fueron redondeados a números enteros en vista de que el programa no admite valores decimales.

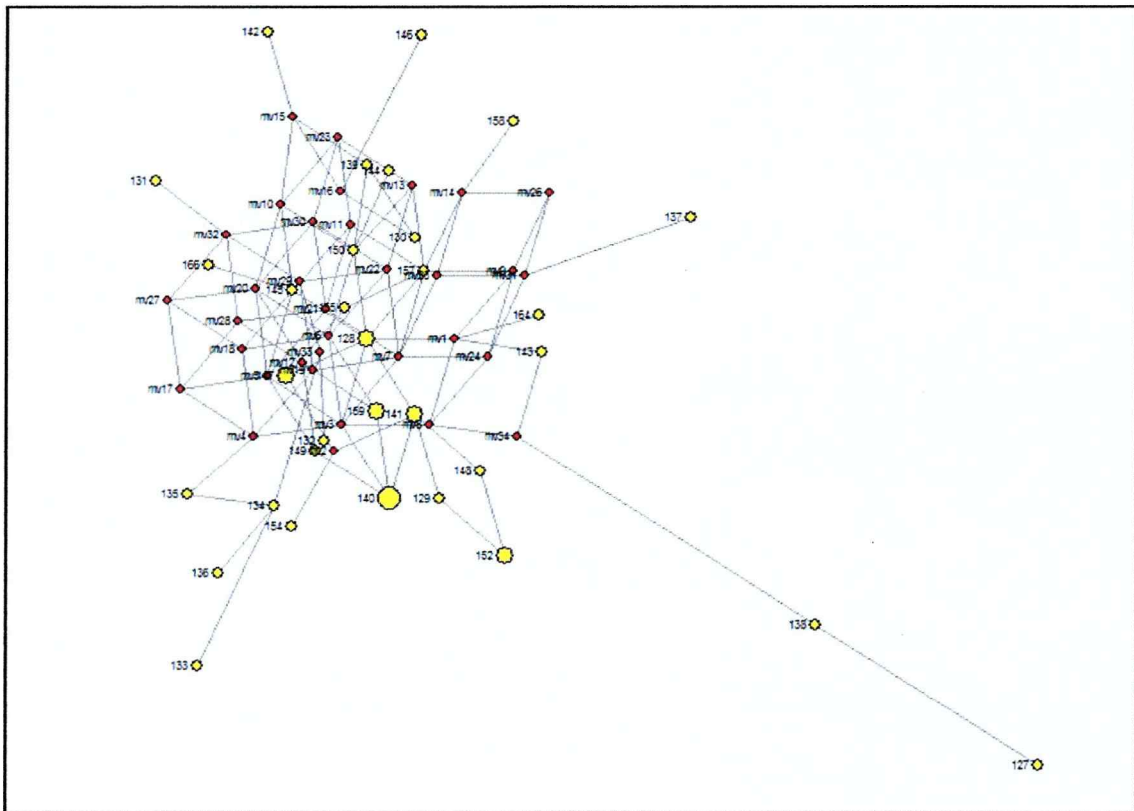
3.1.1 Estimaciones de calidad

Para estimar la calidad de nuestra propuesta comparada con la de Qamar y col. (2002) o una asignación uniforme de pesos, se empleó la información publicada de muestras provenientes del haplogrupo E de una población libanesa (Zalloua y col. 2008). Se eligieron los mismos microsatélites que se analizaron en la presente tesis (DYS 19, DYS 389 I y II, DYS 390, DYS391, DYS 392, DYS 393), y con estos datos se calcularon y dibujaron tres redes medianas con el programa Network 4.5, una para cada sistema de asignación de pesos.

En cada red, se contabilizó la cantidad de vectores medianos y la distancia máxima al nodo, y la presencia o ausencia de cubos de altas dimensiones. En los tres casos el haplotipo nodal es el mismo, aunque sus topologías difieren.

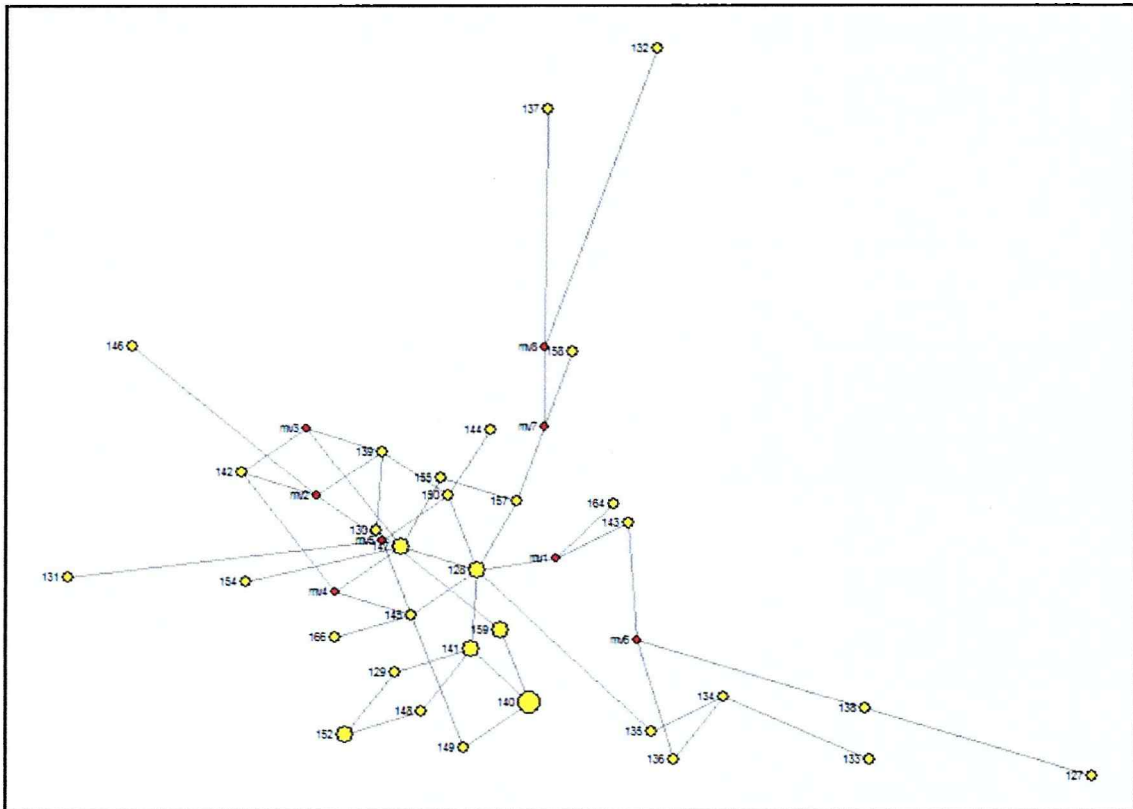
La figura 10 representa la red calculada con un sistema uniforme, de peso 10 para cada microsatélite (el “default” del programa). Como puede verse, una alta cantidad de vectores medianos está presente (en total 34), al igual que cubos de altas dimensiones y una distancia máxima de 10 pasos mutacionales.

Figura 10: Red Mediana calculada y dibujada con pesos uniformes.



La figura 11 muestra la red estimada siguiendo el método de Qamar y col. (2002). En ella hay 8 vectores medianos, formas de altas dimensiones y una distancia al nodo máxima de 10 pasos mutacionales.

Figura 11: Red Mediana calculada y dibujada con la asignación de pesos según el método de Qamar y col. (2002).



La Figura 12 presenta la red computada y trazada con la presente propuesta. En ella hay un total de 6 vectores medianos, una distancia máxima de 8 pasos mutacionales, y ausencia de cubos de altas dimensiones.

Figura 12: Red Mediana calculada y dibujada con los pesos asignados según la formula desarrollada en la presente tesis.



La propuesta de esta tesis permite a los investigadores obtener redes más parsimoniosas que aquellas computadas con otros esquemas de pesos, sin emplear pre o post procesamientos, dado que la red resultante contuvo menos vectores medianos, una menor distancia máxima al nodal, y no presentó figuras de altas dimensiones. Notese que todas las figuras (Fig. 10, 11 y 12) fueron copiadas en forma exacta de aquellas que produjo el programa, sin modificar la topología en ninguna forma que pudiera simplificar o alterar los gráficos.

La fórmula desarrollada para esta tesis toma de base las tasas mutacionales de los microsatélites, en lugar de su varianza. Así, aquellos marcadores con tasas elevadas reciben pesos bajos, mientras que a los que presentan tasas bajas se les asigna un peso alto, ya que son eventos menos probables que los primeros.

De todas formas, cuando no se cuenta con tasas mutacionales confiables, pero la variación de los microsatélites si lo es (como podría ser el

caso de nuevos microsátelites), esta fórmula puede modificarse tomando en cuenta la misma en lugar de las tasas mutacionales:

$$P(x) = M (\text{variación}_y / \text{variación}_x)^a$$

Otra ventaja es que permite al investigador establecer el rango de valores que le interesa abarcar, a través del cálculo de “a”. Esto facilita el análisis de muchos microsátelites en simultáneo, puesto que si la cantidad de marcadores es elevada, se puede ampliar. En esta tesis se eligió el peso mínimo de 3 y el máximo de 10, pero el programa Network posibilita la selección de un mínimo de 0 y un máximo de hasta 100, de manera que el rango posible es amplio.

Cabe destacar que la fórmula no permite valores de 0, pero eso no es un problema ya que pesos iguales a 0 son inútiles por definición.

Además, la flexibilidad de esta fórmula permite su aplicación para otros marcadores y características cuyo análisis requiera una asignación de peso. Sencillamente, el investigador substituiría en la fórmula las tasas mutacionales por la probabilidad de un cambio de estado para los caracteres de interés. Así, una escala fácil de calcular y matemáticamente precisa se encuentra disponible para diferentes campos de la Antropología, más allá de la Antropología Molecular.

3.2 Ensayos de fiabilidad de los programas de predicción de haplogrupo

Como ya se mencionó en la Introducción, los linajes del cromosoma Y se establecen a través de polimorfismos de un único nucleótido (SNP) y repeticiones cortas en tandem (STR), que proveen los correspondientes haplogrupo y haplotipo. En el estudio de la genética poblacional humana, la determinación del haplogrupo es de particular interés puesto que revela las relaciones filogenéticas por descendencia.

Considerando los hallazgos de Bosch y col. (1999) y de Behar y col. (2004), donde la variabilidad de los STR es particionada por los haplogrupos de

una forma más abarcativa que por las poblaciones, existe un interés en aumento por unificar estas dos fuentes de información, y de encontrar formas de predecir el haplogrupo de un haplotipo dado cuando los datos de SNP no están disponibles.

Uno de estos programas es el *Haplogroup Predictor* de Whit Athey (Athey 2005, Athey 2006) (<http://home.comcast.net/~hapest5/index.html>), empleado en publicaciones de Salas y col. (2008), Mertens (2007) y Goff y Athey (2006) para estimar la composición étnica de diferentes poblaciones y diagnosticar valores de STR para un haplogrupo dado. El otro es el *Haplogroup Classifier* (Schlecht y col. 2008) (<http://bcf.arl.arizona.edu/haplo>), que consiste en algoritmos de aprendizaje de la computadora que requieren de modelos previos y haplotipos con un haplogrupo conocido para entrenar al programa.

Como se acaba de mencionar, si bien algunos autores ya estaban empleando estos programas para predecir los haplogrupos de los perfiles de STRs que habían establecido (Salas y col. 2008, Mertens, 2007, Goff y Athey 2006), en el presente trabajo se optó por no confiar ciegamente en ellos, sino en evaluar su fiabilidad a partir de nuestros perfiles completos. De esta forma, si de estos ensayos se obtenía que los programas presentaban buenos poderes de predicción, los mismos serían empleados para estimar el haplogrupo en aquellos casos en los cuales no se contaba con los datos de SNP.

3.2.1 ¿Cómo evaluar la fiabilidad de los programas?

Se seleccionaron 119 de las muestras analizadas, aquellas que tuvieran tanto el haplogrupo como haplotipo, considerando una única muestra por haplotipo, y que se contara con por lo menos 10 muestras con haplotipo completo para ese haplogrupo.

Estos haplotipos fueron introducidos en el *Haplogroup Predictor*, con priors iguales, obteniendo probabilidades para los haplogrupos inferidos. En el caso del *Haplogroup Classifier*, seguimos los modelos, archivos de árbol y

datos públicos provistos por los autores del programa. Los haplogrupos determinados por SNP fueron comparados con los del software.

Se calcularon la sensibilidad (s), especificidad (e), likelihood ratio positivo $[(LR+=s/(1-e)]$ y likelihood ratio negativo $[(LR-=(1-s)/e]$ (McAlister y col., 1999) de ambos programas por haplogrupo y total. “S” representa la probabilidad de corroboración de un haplogrupo predicho cuando el haplotipo fue determinado como de ese haplogrupo a través de análisis de SNP, mientras que “e” es la probabilidad de confirmación de otro haplogrupo cuando el haplotipo correspondía, según nuestra tipificación, a otro haplogrupo. Se ha definido que un ensayo es adecuado cuando su especificidad y sensibilidad son altas (Greenhalgh, 1997), y un LR+ de por lo menos 10, de ahí que tomamos este criterio.

En el caso del *Predictor*, se consideraron diferentes categorías de precisión de asignación: por ser el primer haplogrupo en el ranking (rk), y también puntos de corte de 50 a 95% con intervalos de a 5, mientras que para el *Classifier* sólo se contaron los casos donde presentó un haplogrupo acordado, aunque de esta forma se redujo el número de casos considerados en los cálculos ($N=62$; el 52,1% de la muestra original).

Para obtener estimaciones cuantitativas de las predicciones realizadas por los programas, se calculó el coeficiente de incertidumbre de y , $U(y|x)$, con la subrutina *cntab2* de Press y col. (1994) que también proveen una explicación detallada del significado y cómo computar este coeficiente. Tomando los resultados de los programas como la variable x y aquellos de la tipificación por SNP como la variable y , $U(y|x)$ da la fracción de información de tipificación de SNP que se pierde cuando el resultado del software ya es conocido. Los casos más extremos son: 1) $U(y|x) = 1,00$ (o 100%), que implicaría que el programa da respuestas perfectas (toda la información que proveería la tipificación por SNP ya fue provista por el software); 2) $U(y|x) = 0,00$ (o 0%), que implicaría que el software no provee información (toda la información se obtendría por tipificación de SNP).

3.2.2 ¿Qué se obtuvo?

Los s , e , $LR+$ y $LR-$ por punto de corte del *Haplogroup Predictor* se resumen en la Tabla 2. El s del *Haplogroup Classifier* equivale a 0,45, el e es 0,92, el $LR+$ es 5,99 y el $LR-$ es 0,59. Es importante resaltar que cerca de la mitad de los haplotipos (57 de un total de 119) no pudieron ser asignados por el *Haplogroup Classifier* a un haplogrupo acordado (la categoría “no clasificado” de la Tabla 3).

Tabla 2: S , e , $LR+$ y $LR-$ totales en cada punto de corte para el *Haplogroup Predictor*

| Punto de corte | s1 | e2 | LR+3 | LR-4 |
|----------------|-----|------|------|------|
| RK | 0,5 | 0,88 | 4,21 | 0,6 |
| 50 | 0,5 | 0,89 | 4,52 | 0,5 |
| 55 | 0,5 | 0,89 | 4,54 | 0,6 |
| 60 | 0,5 | 0,89 | 4,46 | 0,6 |
| 65 | 0,5 | 0,9 | 4,65 | 0,6 |
| 70 | 0,5 | 0,91 | 5,12 | 0,6 |
| 75 | 0,4 | 0,91 | 4,59 | 0,7 |
| 80 | 0,4 | 0,92 | 5,22 | 0,7 |
| 85 | 0,4 | 0,93 | 5,81 | 0,7 |
| 90 | 0,3 | 0,94 | 5,93 | 0,7 |
| 95 | 0,3 | 0,96 | 8,00 | 0,7 |

Referencias: 1: sensibilidad; 2: especificidad; 3: Likelihood ratio positivo; 4: Likelihood ratio negativo.

Cuando se considera el $LR+$ por haplogrupo, en el caso del Q tanto como para el *Haplogroup Predictor* como *Classifier*, y también para el DE* en el *Classifier*, es más alto que 10.

El haplogrupo con mayor proporción de falsos positivos fue R, significativamente más alto que las proporciones de falsos positivos de los demás haplogrupos.

Tabla 3: Frecuencias de haplogrupo por tipificación de SNP, el *H. Predictor* y el *H. Classifier*, y los valores de s, e, LR+ y LR-.

| | | Por tipificación de SNP | | | | | Total Programa | Falsos Positivos | Falsos Positivos % | |
|-----------------------|----------------|-------------------------|------|------|------|------|----------------|------------------|--------------------|-------|
| | | Haplogrupo | | | | | | | | |
| | | F | K | Q | R | DE* | | | | |
| Haplogroup Predictor | Haplogrupo | F | 9 | 0 | 1 | 4 | 5 | 19 | 10 | 52,6 |
| | | K | 0 | 0 | 2 | 0 | 0 | 2 | 2 | 100,0 |
| | | Q | 0 | 3 | 17 | 0 | 0 | 20 | 3 | 15,0 |
| | | R | 6 | 12 | 16 | 30 | 1 | 65 | 35 | 53,8 |
| | | DE* | 2 | 1 | 2 | 3 | 5 | 13 | 8 | 61,5 |
| | Total por SNP | | 17 | 16 | 38 | 37 | 11 | 119 | 58 | 48,7 |
| | Sensibilidad | | 52,9 | 0,0 | 44,7 | 81,1 | 45,5 | 51,3 | | |
| | Especificidad | | 90,2 | 98,1 | 96,3 | 57,3 | 92,6 | | | |
| | LR+ | | 5,4 | 0,0 | 12,1 | 1,9 | 6,1 | | | |
| | LR- | | 0,5 | 1,0 | 0,6 | 0,3 | 0,6 | | | |
| Haplogroup Classifier | Haplogrupo | F | 4 | 0 | 1 | 4 | 3 | 12 | 8 | 66,7 |
| | | K | 0 | 1 | 3 | 1 | 0 | 5 | 4 | 80,0 |
| | | Q | 0 | 1 | 2 | 0 | 0 | 3 | 1 | 33,3 |
| | | R | 2 | 8 | 9 | 19 | 0 | 38 | 19 | 50,0 |
| | | DE* | 0 | 0 | 1 | 1 | 2 | 4 | 2 | 50,0 |
| | No clasificado | | 11 | 6 | 22 | 12 | 6 | 57 | | |
| | Total por SNP | | 17 | 16 | 38 | 37 | 11 | 119 | 34 | 54,8 |
| | Sensibilidad | | 66,7 | 10,0 | 12,5 | 76,0 | 40,0 | 23,5 | | |
| | Especificidad | | 92,9 | 96,3 | 99,0 | 79,8 | 98,2 | | | |
| | LR+ | | 9,4 | 2,7 | 12,9 | 3,8 | 22,8 | | | |
| LR- | | 0,4 | 0,9 | 0,9 | 0,3 | 0,6 | | | | |

Los resultados para $U(y|x)$ son 0,244 (o 24,4%) para el *Haplogroup Predictor* y 0,207 (o 20,7%) para el *Haplogroup Classifier*. En otras palabras, el porcentaje de información por tipificación de SNP que es perdida si los resultados de estos programas son previamente conocidos es sólo entre un 20% y 25%.

3.2.3 ¿Qué significan todos estos datos?

Estos resultados representan una alta probabilidad de error en las predicciones, y un sesgo hacia el haplogrupo R, de manera tal que es probable que resultados basados en el uso de estos programas sean débiles, de ahí que se optó por no emplearlos para los casos en los cuales no se contaba con el haplogrupo por SNP.

En aquellos trabajos donde se estimaron diferencias de los aportes por sexo en poblaciones multiétnicas a través de estos programas (Salas y col. 2008; Meterns, 2007), una sobrestimación del componente Europeo es esperada. La determinación de haplogrupo por análisis de SNP sigue siendo el mejor método, considerando la baja precisión de predicción de los programas disponibles. Para el caso de Salas y col. (2008), su porcentaje de linajes Q (~2%) es tres veces menor que las estimaciones más bajas (~6%). Como pudo verse en la Tabla 3, el *Haplogroup Predictor* reconoció 16 de 38 muestras del haplogrupo Q como R, mientras que sólo 17 fueron asignadas correctamente, de manera que es factible que en la muestra de estos investigadores más linajes Q que los detectados por el programa estuvieran presentes.

Los LR+ adecuados para los haplogrupos Q y DE* se pueden explicar por una diversidad más baja dentro de cada grupo. Especialmente en el caso del Q, que es un haplogrupo relativamente reciente, la homogeneidad es un resultado de su edad evolutiva joven, dado que el lapso de tiempo en el cual los haplotipos se alejan del fundador del haplogrupo es corto (Bianchi y col. 1998).

Considerando que las muestras a partir de las cuales se estimaron las frecuencias de calibración pertenecen mayoritariamente a poblaciones europeas (o de ancestralidad europea), la alta proporción de falsos positivos en el haplogrupo R puede ser un reflejo de ese error de muestreo, dado que es el haplogrupo más común en esas poblaciones.

Para tener una mejor idea de lo que significan los resultados de $U(y|x)$, comparemos la tipificación de SNP a la tirada de un dado, es decir, los resultados Q, R, etc. como equivalentes a obtener un 1, un 2, etc., y veamos los resultados de los programas como aquellos de un proceso que nos

permitiría saber si la tirada daría un resultado par o impar. En este caso, se obtiene $U(y|x) = 0,387$ (38,7%), de manera tal que es posible concluir que la información provista por los programas, en relación a la provista por tipificado de SNP, es menor que la información que el resultado de tirar un dado sea par o impar, relativo a aquella provista por tirar el dado.

3.2.4 ¿Por qué estos programas muestran niveles de precisión tan bajos?

Se proponen dos explicaciones: podría decirse que no hay suficientes perfiles de STRs con haplogrupos asociados para calibrar adecuadamente a estos programas, pero, además, dadas las tasas mutacionales de los microsatélites y la profundidad temporal de las ramificaciones de los haplogrupos (Karafet y col. 2008), es posible encontrar casos de convergencia, es decir, el mismo haplotipo en muestras de diferente haplogrupo (un ejemplo de esto fue hallado en la presente tesis). Por ejemplo, el haplogrupo más reciente, el R1, tiene un tiempo al ancestro común más cercano de 18.500 (12.500-25.700) años, y el clado más reciente, el CT, fue estimado en 70.000 años (Karafet y col. op. cit.), mientras que las tasas mutacionales de los STRs varían entre 6,35 (4,19-9,22 95% CI) $\times 10^{-3}$ para DYS 439 a 0,45 (0,12-1,17 95% CI) $\times 10^{-3}$ en el caso de DYS 392 (www.yhrd.org), de manera tal que la repetición de las combinaciones de alelos entre clados evolutivamente divergentes no es rara, dados los eventos de meiosis que se acumularon a lo largo de una profundidad temporal tan importante.

Una forma rápida de ver esto es buscar las frecuencias para cada locus de STR entre las metapoblaciones de la base YHRD, que muestra como los mismo alelos están presentes en todas. Aunque sus frecuencias pueden diferir (y considerando que esta base no provee información de haplogrupos), dada la asociación geográfica de los haplogrupos evidenciada por otros autores (Underhill y col. 2001), si esa asociación entre alelos de STR y haplogrupos fuera lo suficientemente fuerte como para permitir la predicción de los últimos a partir de los primeros, estos mostrarían una distribución geográfica similar a aquella encontrada en los haplogrupos.

Así, se observó que los valores predictivos de estos programas informáticos no eran adecuados, y por ello no fueron empleados para proponer un haplogrupo en aquellos casos donde sólo fue posible definir el haplotipo, ya que por el momento la definición mediante SNP es la única forma adecuada de establecer a qué haplogrupo pertenece el cromosoma Y de un individuo dado.

Resultados

“I have no name
I am but two days old. -
What shall I call thee?
I happy am
Joy is my name, -
Sweet joy befall thee!”

Infant Joy, Songs of Innocence, Songs of Innocence and of Experience. W.
Blake

4.1 Asociaciones entre haplogrupos y apellidos

El análisis de haplogrupos permitió clasificar las muestras en 44 nativo americanas (43 Q1a3a y 1 Q*), 111 foráneas (14 DE*, 19 F, 61 R, 10 K y 7 P) mientras que 20 no pudieron ser clasificadas. En la Tabla 4 se presenta la distribución de los mismos por provincias, así como origen del apellido. Allí puede verse claramente cómo predominan los apellidos de origen español, seguidos en importancia por los apellidos nativos.

La proporción de individuos con un haplogrupo nativo americano (0.25) fue significativamente más alta que aquella de individuos con un apellido nativo (0.08) ($P < 0.01$).

De los 41 apellidos con más de un individuo, 24 (58.54%) presentaron diferentes haplogrupos entre los individuos del mismo apellido, y sus frecuencias absolutas en el NOA presentan un mínimo de 79 y un máximo de 12296 portadores (Tabla 10, Anexo). De estos apellidos, 15 incluyeron haplogrupos foráneos mientras que cada uno del resto incluyó un haplogrupo nativo americano y uno foráneo. Asimismo, 18 de los apellidos polifiléticos forman parte del grupo de los apellidos más comunes en la región y 6 son más raros.

Hubo coincidencia de haplogrupo y apellido en 10 casos (24.39%):

- 1 del haplogrupo K
- 1 del haplogrupo Q1a3a
- 1 del haplogrupo DE*
- 1 del haplogrupo F
- 6 del haplogrupo R

Las frecuencias absolutas de estos 10 apellidos en el NOA oscilan entre un mínimo de 14 y un máximo de 4773 (Tabla 10, Anexo). En 3 de estos casos, el apellido pertenece al grupo de los 100 más comunes de la Argentina, mientras que el resto posee frecuencias menores.

Además es importante resaltar que hubo 7 casos (17.07%) en los cuales no se pudo excluir que abarquen más de un haplogrupo, 5 pertenecientes al grupo de los apellidos más comunes y 2 al de los menos comunes.

Estos hallazgos referentes a la relación entre haplogrupos y apellidos señalan que esta concordancia es menor que aquella encontrada para Inglaterra (43%) (King y col. 2006) con una diferencia altamente significativa ($p < 0.01$) aún considerando la menor resolución de nuestro método para determinar haplogrupos.

El Riesgo Relativo de un individuo con un apellido nativo americano para presentar un haplogrupo nativo americano fue igual a 4.21 (1.34 - 13.22) ($p = 0.013$), mientras que aquel de que un individuo con apellido extranjero presente un haplogrupo foráneo fue de 1.2 (1.2 - 1.14) ($p = 0.013$).

Tabla 4: Distribución de los individuos por provincia, haplogrupos y origen del apellido

| Provincia | Número de individuos | Haplogrupo | | | | | | | | | | Origen del apellido | | |
|-----------|----------------------|-------------------|-------|-------|---------|-------|-------|-------|----------------|-------------|-----------|---------------------|---------------------|--|
| | | No confirmado (%) | F (%) | K (%) | DE* (%) | P (%) | Q (%) | R (%) | N de apellidos | Español (%) | Vasco (%) | Autóctono (%) | Otro extranjero (%) | |
| Catamarca | 76 | 15,78 | 14,47 | 1,32 | 10,53 | 3,95 | 9,21 | 44,74 | 66 | 90,9 | 1,52 | 1,52 | 6,06 | |
| Jujuy | 43 | 6,98 | 4,65 | 18,6 | 2,33 | 0 | 39,53 | 27,91 | 31 | 70,96 | 9,68 | 16,13 | 3,23 | |
| Salta | 47 | 10,64 | 6,38 | 2,13 | 8,51 | 8,51 | 40,43 | 23,4 | 47 | 79,17 | 8,33 | 12,5 | 0 | |
| Tucumán | 9 | 0 | 22,22 | 0 | 22,22 | 0 | 11,11 | 44,45 | 9 | 88,9 | 0 | 0 | 11,1 | |

4.2 Asociaciones entre haplotipos y apellidos

Se definieron 162 haplotipos, de los cuales 153 resultaron únicos (es decir, representados por un solo individuo) y 9 compartidos por más de un individuo (Tabla 5). De estos últimos se destacan por presentar asociación entre haplotipo y apellido los siguientes:

- H94: del haplogrupo F, con dos individuos del mismo apellido (*Prellis*), el único caso encontrado de concordancia perfecta entre apellido y haplotipo.
- H1: incluyó 3 individuos del mismo apellido (*Cruz*) y 1 de otro apellido (*Gregorio*), todos pertenecientes al haplogrupo Q1a3a y a la provincia de Jujuy.

Además, en el haplogrupo R se observaron 2 haplotipos modales (aquellos que incluyeron la mayor cantidad de individuos):

- H41: de tres individuos con diferentes apellidos (es decir, un haplotipo trans-apellido).
- H42: abarcó tres individuos, trans-apellido.

El resto de los haplotipos compartidos incluyeron:

- 4 del haplogrupo R (H49, H54, H78 y H86)
- 1 del haplogrupo Q1a3a (H22)

Estos últimos cinco son haplotipos trans-apellidos, con dos individuos cada uno.

Tabla 5: Apellidos y haplotipos establecidos

| Apellido | H. N° | Haplogrupo | DYS19 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 |
|-----------|-------|------------|-------|---------|----------|--------|--------|--------|--------|
| CRUZ | 1 | Q1a3a | 13 | 13 | 31 | 24 | 10 | 14 | 13 |
| CRUZ | 1 | Q1a3a | 13 | 13 | 31 | 24 | 10 | 14 | 13 |
| CRUZ | 1 | Q1a3a | 13 | 13 | 31 | 24 | 10 | 14 | 13 |
| GREGORIO | 1 | Q1a3a | 13 | 13 | 31 | 24 | 10 | 14 | 13 |
| LOPEZ | 2 | Q1a3a | 11 | 13 | 30 | 21 | 10 | 13 | 13 |
| VILTE | 3 | Q1a3a | 12 | 12 | 29 | 24 | 10 | 14 | 13 |
| BERMUDEZ | 4 | Q1a3a | 12 | 13 | 31 | 22 | 11 | 16 | 14 |
| RODRIGUEZ | 5 | Q1a3a | 12 | 13 | 29 | 24 | 11 | 13 | 13 |
| BOLIVAR | 6 | Q1a3a | 13 | 12 | 29 | 21 | 10 | 12 | 13 |

| | | | | | | | | | |
|----------------|----|-------|----|----|----|----|----|----|----|
| FLORES | 7 | Q1a3a | 13 | 12 | 31 | 24 | 10 | 14 | 13 |
| PASTRANA | 8 | Q1a3a | 13 | 12 | 29 | 24 | 11 | 15 | 13 |
| CRUZ | 9 | Q1a3a | 13 | 13 | 28 | 24 | 10 | 14 | 13 |
| ROMERO | 10 | Q1a3a | 13 | 13 | 28 | 24 | 11 | 14 | 14 |
| CRUZ | 11 | Q1a3a | 13 | 13 | 29 | 24 | 10 | 14 | 13 |
| GOMEZ | 12 | Q1a3a | 13 | 13 | 29 | 24 | 10 | 15 | 13 |
| OLIVERA | 13 | Q1a3a | 13 | 13 | 31 | 24 | 12 | 14 | 13 |
| BOVARINOS | 14 | Q1a3a | 13 | 13 | 31 | 24 | 10 | 13 | 13 |
| LOPEZ | 15 | Q1a3a | 13 | 13 | 31 | 25 | 11 | 13 | 13 |
| PAZ | 16 | Q1a3a | 13 | 13 | 30 | 24 | 11 | 14 | 13 |
| PEREYRA | 17 | Q1a3a | 13 | 13 | 30 | 21 | 11 | 12 | 13 |
| QUIPILDOR | 18 | Q1a3a | 13 | 13 | 30 | 23 | 10 | 13 | 13 |
| DIAZ | 19 | Q1a3a | 13 | 13 | 30 | 24 | 10 | 15 | 13 |
| DIAZ | 20 | Q1a3a | 13 | 13 | 30 | 24 | 9 | 16 | 14 |
| MORENO | 21 | Q1a3a | 13 | 13 | 30 | 24 | 10 | 14 | 13 |
| GUTIERREZ | 22 | Q1a3a | 13 | 13 | 30 | 25 | 10 | 14 | 13 |
| CRUZ | 22 | Q1a3a | 13 | 13 | 30 | 25 | 10 | 14 | 13 |
| CALAPIÑA | 23 | Q1a3a | 13 | 14 | 29 | 23 | 10 | 14 | 13 |
| CHAYLE | 24 | Q1a3a | 13 | 14 | 32 | 23 | 10 | 15 | 14 |
| CRUZ | 25 | Q1a3a | 13 | 14 | 29 | 25 | 10 | 14 | 13 |
| DELGADO | 26 | Q1a3a | 13 | 14 | 29 | 23 | 9 | 15 | 13 |
| QUISPE | 27 | Q1a3a | 13 | 14 | 28 | 23 | 11 | 15 | 14 |
| ROMERO | 28 | Q1a3a | 13 | 14 | 29 | 24 | 11 | 14 | 14 |
| GARCÍA RAMÍREZ | 29 | Q1a3a | 13 | 14 | 30 | 24 | 11 | 14 | 13 |
| TORRES | 30 | Q1a3a | 13 | 14 | 31 | 24 | 11 | 15 | 14 |
| AVILA | 31 | Q1a3a | 14 | 13 | 30 | 25 | 9 | 13 | 13 |
| CARPIO | 32 | Q1a3a | 14 | 13 | 32 | 25 | 11 | 14 | 13 |
| GUTIERREZ | 33 | Q1a3a | 14 | 13 | 30 | 24 | 11 | 14 | 13 |
| MARTINEZ | 34 | Q1a3a | 14 | 13 | 29 | 23 | 10 | 14 | 13 |
| PAZ | 35 | Q1a3a | 14 | 13 | 30 | 23 | 10 | 14 | 13 |
| PEREZ | 36 | Q1a3a | 14 | 13 | 29 | 24 | 10 | 13 | 14 |
| FERRIL | 37 | Q1a3a | 14 | 14 | 31 | 24 | 10 | 15 | 14 |
| LACCI | 38 | Q1a3a | 15 | 14 | 31 | 23 | 10 | 16 | 14 |
| LEIVA | 39 | Q | 15 | 13 | 30 | 23 | 10 | 14 | 13 |
| GALAN | 40 | Q1a3a | 16 | 12 | 31 | 24 | 11 | 14 | 13 |
| PACHECO | 41 | R* | 14 | 13 | 30 | 24 | 11 | 13 | 13 |
| URZAGATI | 41 | R* | 14 | 13 | 30 | 24 | 11 | 13 | 13 |
| JUAREZ | 41 | R* | 14 | 13 | 30 | 24 | 11 | 13 | 13 |
| VERA | 42 | R* | 14 | 13 | 28 | 24 | 11 | 13 | 13 |
| QUIROGA | 42 | R* | 14 | 13 | 28 | 24 | 11 | 13 | 13 |
| LEIVA | 42 | R* | 14 | 13 | 28 | 24 | 11 | 13 | 13 |
| CAMPOS | 43 | R* | 13 | 13 | 29 | 23 | 11 | 11 | 13 |
| LOPEZ | 44 | R* | 13 | 13 | 30 | 24 | 10 | 11 | 13 |
| CARDOZO | 45 | R* | 13 | 14 | 30 | 24 | 12 | 13 | 13 |
| MERCADO | 46 | R* | 13 | 14 | 31 | 24 | 11 | 13 | 13 |
| CARABAJAL | 47 | R* | 14 | 12 | 30 | 22 | 10 | 10 | 14 |
| CARABAJAL | 48 | R* | 14 | 12 | 30 | 24 | 11 | 13 | 13 |
| ARTEAGA | 49 | R* | 14 | 13 | 30 | 23 | 10 | 13 | 13 |
| BARRERA | 49 | R* | 14 | 13 | 30 | 23 | 10 | 13 | 13 |
| FLORES | 50 | R* | 14 | 13 | 30 | 23 | 11 | 13 | 13 |
| CONDORI | 51 | R* | 14 | 13 | 30 | 24 | 9 | 12 | 13 |
| MERLO | 52 | R* | 14 | 13 | 30 | 24 | 9 | 13 | 13 |
| MERCADO | 53 | R* | 14 | 13 | 30 | 24 | 11 | 12 | 13 |
| FERNANDEZ | 54 | R* | 14 | 13 | 30 | 24 | 11 | 13 | 12 |

| | | | | | | | | | |
|---------------|-----|----|----|----|----|----|----|----|----|
| BARRIONUEVO | 54 | R* | 14 | 13 | 30 | 24 | 11 | 13 | 12 |
| ROBIN | 55 | R* | 14 | 13 | 30 | 25 | 11 | 13 | 13 |
| HERRERA | 56 | R* | 14 | 13 | 31 | 23 | 11 | 13 | 14 |
| AGUIRRE | 57 | R* | 14 | 13 | 31 | 24 | 11 | 13 | 13 |
| SAAVEDRA | 58 | R* | 14 | 13 | 29 | 24 | 10 | 12 | 14 |
| LEIVA | 59 | R* | 15 | 12 | 29 | 23 | 11 | 15 | 13 |
| SORIA | 60 | R* | 15 | 12 | 32 | 24 | 11 | 11 | 14 |
| LOSSI | 61 | R* | 15 | 13 | 30 | 24 | 11 | 12 | 13 |
| VALLE | 62 | R* | 16 | 14 | 30 | 25 | 11 | 12 | 13 |
| PINERO | 63 | R* | 12 | 13 | 30 | 24 | 10 | 14 | 13 |
| RODRIGUEZ | 64 | R* | 12 | 14 | 27 | 24 | 13 | 14 | 12 |
| TOLEDO | 65 | R* | 12 | 14 | 28 | 25 | 11 | 10 | 13 |
| CAMPOS | 66 | R* | 14 | 13 | 28 | 24 | 9 | 11 | 13 |
| ALARCÓN | 67 | R* | 14 | 14 | 31 | 24 | 13 | 13 | 13 |
| MARTINEZ | 68 | R* | 14 | 14 | 30 | 23 | 11 | 13 | 13 |
| CARRIZO | 69 | R* | 14 | 12 | 29 | 21 | 10 | 13 | 13 |
| ALVAREZ | 70 | R* | 14 | 13 | 29 | 22 | 11 | 13 | 13 |
| GARCIA | 71 | R* | 14 | 13 | 29 | 23 | 11 | 13 | 13 |
| FLORES | 72 | R* | 14 | 13 | 29 | 24 | 11 | 13 | 13 |
| ARTEAGA | 73 | R* | 14 | 13 | 28 | 25 | 11 | 13 | 12 |
| BAZAN | 74 | R* | 14 | 13 | 28 | 24 | 11 | 13 | 12 |
| BAZAN | 75 | R* | 14 | 13 | 28 | 24 | 10 | 13 | 12 |
| GÓMEZ | 76 | R* | 14 | 13 | 28 | 24 | 10 | 11 | 12 |
| LOPEZ | 77 | R* | 14 | 13 | 28 | 23 | 11 | 12 | 13 |
| MORALES | 78 | R* | 14 | 13 | 28 | 23 | 11 | 13 | 14 |
| CARRIZO | 78 | R* | 14 | 13 | 28 | 23 | 11 | 13 | 14 |
| NAVARRO | 79 | R* | 14 | 13 | 28 | 25 | 10 | 13 | 13 |
| QUIROGA | 80 | R* | 14 | 13 | 28 | 24 | 12 | 13 | 13 |
| RODRIGUEZ | 81 | R* | 14 | 13 | 28 | 24 | 9 | 12 | 14 |
| SOSA | 82 | R* | 14 | 13 | 28 | 24 | 11 | 11 | 16 |
| GONZALEZ | 83 | R* | 14 | 14 | 29 | 23 | 10 | 13 | 13 |
| MERCADO | 84 | R* | 14 | 14 | 29 | 24 | 11 | 12 | 13 |
| MERCADO | 85 | R* | 14 | 14 | 29 | 24 | 10 | 12 | 13 |
| VERGARA | 86 | R* | 14 | 14 | 28 | 24 | 11 | 13 | 13 |
| QUIROGA | 86 | R* | 14 | 14 | 28 | 24 | 11 | 13 | 13 |
| COBACHO | 87 | R* | 14 | 14 | 28 | 24 | 11 | 14 | 13 |
| ARTEAGA | 88 | R* | 15 | 13 | 28 | 24 | 11 | 11 | 12 |
| NAVARRO | 89 | R* | 15 | 13 | 30 | 25 | 10 | 15 | 13 |
| RIOS | 90 | R* | 15 | 14 | 28 | 25 | 12 | 13 | 14 |
| TELLO | 91 | R* | 15 | 14 | 30 | 22 | 11 | 12 | 13 |
| MORENO | 92 | R* | 16 | 13 | 29 | 24 | 11 | 13 | 13 |
| PEREZ | 93 | R* | 16 | 14 | 28 | 25 | 10 | 14 | 13 |
| PRELIS | 94 | F | 14 | 12 | 32 | 21 | 10 | 11 | 15 |
| PRELI | 94 | F | 14 | 12 | 32 | 21 | 10 | 11 | 15 |
| DELGADO | 95 | F | 12 | 12 | 29 | 23 | 9 | 9 | 14 |
| LEGUIZAMON | 96 | F | 12 | 12 | 29 | 22 | 10 | 11 | 13 |
| PEREA | 97 | F | 12 | 13 | 31 | 25 | 11 | 12 | 13 |
| CRUZ | 98 | F | 13 | 13 | 28 | 23 | 10 | 11 | 13 |
| MORENO | 99 | F | 13 | 14 | 27 | 25 | 9 | 9 | 12 |
| PACHECO ROJAS | 100 | F | 14 | 12 | 29 | 23 | 10 | 11 | 13 |
| NIEVES | 101 | F | 14 | 13 | 31 | 21 | 11 | 10 | 13 |
| RIOS | 102 | F | 14 | 13 | 33 | 23 | 12 | 10 | 12 |
| TEJERINA | 103 | F | 14 | 13 | 30 | 23 | 12 | 12 | 12 |
| DIAZ | 104 | F | 14 | 14 | 28 | 24 | 11 | 9 | 13 |

| | | | | | | | | | |
|----------------|-----|-----|----|----|----|----|----|----|----|
| CRUZ | 105 | F | 15 | 12 | 29 | 22 | 10 | 11 | 12 |
| FERNANDEZ | 106 | F | 15 | 12 | 29 | 22 | 10 | 11 | 14 |
| GONZALEZ | 107 | F | 15 | 12 | 29 | 23 | 10 | 11 | 12 |
| COBACHO | 108 | F | 16 | 13 | 31 | 23 | 10 | 12 | 13 |
| FERNANDEZ | 109 | F | 16 | 14 | 27 | 25 | 11 | 11 | 13 |
| FERNANDEZ | 110 | F | 17 | 13 | 30 | 23 | 9 | 11 | 12 |
| VILLARREAL | 111 | F | 17 | 13 | 31 | 23 | 10 | 10 | 12 |
| ARGAÑARAZ | 112 | K | 13 | 12 | 31 | 24 | 11 | 14 | 13 |
| GREGORIO | 113 | K | 13 | 13 | 31 | 24 | 11 | 14 | 13 |
| ALBERTO | 114 | K | 13 | 14 | 32 | 23 | 11 | 15 | 14 |
| APAZA | 115 | K | 13 | 14 | 30 | 23 | 11 | 16 | 13 |
| CHOQUE | 116 | K | 14 | 13 | 31 | 24 | 12 | 12 | 13 |
| CHUMACERO | 117 | K | 14 | 13 | 31 | 23 | 10 | 13 | 13 |
| GONZALEZ | 118 | K | 15 | 12 | 29 | 24 | 10 | 14 | 13 |
| CHUMACERO | 119 | K | 15 | 13 | 31 | 24 | 10 | 13 | 13 |
| POLO | 120 | K | 15 | 13 | 30 | 24 | 11 | 13 | 13 |
| SOSA | 121 | K | 15 | 13 | 28 | 24 | 10 | 11 | 18 |
| LOPEZ | 122 | P | 13 | 13 | 28 | 24 | 11 | 14 | 13 |
| GONZALEZ | 123 | P | 14 | 13 | 29 | 24 | 10 | 13 | 13 |
| FLORES | 124 | P | 15 | 12 | 29 | 22 | 11 | 11 | 14 |
| CHACUR | 125 | P | 15 | 13 | 31 | 22 | 10 | 11 | 13 |
| CONDORI | 126 | P | 15 | 13 | 28 | 24 | 10 | 13 | 13 |
| MARTINEZ | 127 | P | 16 | 13 | 28 | 24 | 10 | 11 | 13 |
| LEIVA | 128 | P | 16 | 14 | 29 | 21 | 10 | 11 | 13 |
| NAVARRO | 129 | DE* | 12 | 14 | 28 | 22 | 9 | 11 | 13 |
| VEGA | 130 | DE* | 12 | 14 | 28 | 23 | 9 | 11 | 13 |
| ABREGU | 131 | DE* | 13 | 13 | 31 | 24 | 10 | 13 | 13 |
| ALMARA | 132 | DE* | 13 | 13 | 31 | 25 | 9 | 11 | 15 |
| VERGARA | 133 | DE* | 13 | 13 | 31 | 25 | 9 | 12 | 13 |
| ROMERO | 134 | DE* | 13 | 13 | 28 | 24 | 10 | 11 | 13 |
| VARELA | 135 | DE* | 13 | 13 | 29 | 24 | 10 | 11 | 13 |
| DIAZ | 136 | DE* | 14 | 13 | 29 | 23 | 11 | 11 | 14 |
| GUTIERREZ | 137 | DE* | 14 | 13 | 29 | 22 | 10 | 12 | 13 |
| DIAZ | 138 | DE* | 15 | 13 | 28 | 24 | 9 | 12 | 16 |
| FLORES | 139 | DE* | 15 | 13 | 30 | 23 | 10 | 11 | 13 |
| MOYANO | 140 | DE* | 15 | 13 | 32 | 22 | 10 | 11 | 13 |
| GARCIA | 141 | DE* | 16 | 13 | 29 | 21 | 10 | 11 | 13 |
| VARELA | 142 | DE* | 16 | 13 | 30 | 22 | 11 | 11 | 13 |
| SORIA GALBARRA | 143 | -- | 12 | 12 | 30 | 23 | 10 | 11 | 13 |
| TOMASSI | 144 | -- | 12 | 14 | 30 | 22 | 10 | 14 | 13 |
| VILTE | 145 | -- | 13 | 13 | 31 | 23 | 11 | 14 | 13 |
| CERDA | 146 | -- | 13 | 14 | 31 | 22 | 10 | 13 | 13 |
| MOYANO | 147 | -- | 13 | 14 | 31 | 24 | 10 | 13 | 13 |
| PACHECO | 148 | -- | 13 | 14 | 31 | 23 | 10 | 14 | 13 |
| PEREZ | 149 | -- | 13 | 14 | 31 | 23 | 9 | 12 | 14 |
| AGÜERO | 150 | -- | 14 | 13 | 30 | 22 | 10 | 12 | 14 |
| BERON | 151 | -- | 14 | 13 | 31 | 23 | 10 | 11 | 12 |
| CANOVA | 152 | -- | 14 | 13 | 28 | 25 | 9 | 15 | 13 |
| TOLEDO | 153 | -- | 14 | 13 | 28 | 24 | 10 | 11 | 13 |
| QUIROGA | 154 | -- | 14 | 13 | 31 | 25 | 9 | 12 | 13 |
| MORALES | 155 | -- | 14 | 14 | 29 | 23 | 10 | 11 | 16 |
| CORTEZ | 156 | -- | 15 | 12 | 29 | 22 | 11 | 15 | 14 |
| MORALES | 157 | -- | 15 | 13 | 30 | 22 | 10 | 11 | 13 |
| JEREZ | 158 | -- | 15 | 13 | 30 | 24 | 11 | 13 | 12 |

| | | | | | | | | |
|----------|--------|----|----|----|----|----|----|----|
| FLORES | 159 -- | 15 | 14 | 29 | 22 | 11 | 12 | 13 |
| HERRERA | 160 -- | 16 | 13 | 30 | 23 | 10 | 13 | 13 |
| MARTINEZ | 161 -- | 16 | 13 | 30 | 24 | 10 | 13 | 13 |
| MARIOT | 162 -- | 16 | 14 | 32 | 24 | 11 | 14 | 13 |

Se resaltan los haplotipos compartidos

4.3 Redes Medianas

Se trazaron dos redes medianas, una para el haplogrupo Q1a3a y otra para el haplogrupo R.

4.3.1 Red del Haplogrupo Q1a3a

En la red del haplogrupo nativo americano (Fig. 13), se encontró un haplotipo nodal (H21) que concuerda con el haplotipo del cromosoma Y nativo americano ancestral previamente establecido por Bianchi y col. (1998). Este haplotipo se encuentra rodeado por 5 de nuestra muestra a una distancia de un paso mutacional, y 5 a una distancia de 2 pasos.

Dos haplotipos se encuentran representados por más de un individuo, uno por 2 y otro por 4. Este último es el ya mencionado H1, mientras que el primero es H22, un haplotipo trans-apellido que abarca individuos de diferentes provincias (Jujuy y Tucumán).

El resto de los haplotipos se distribuye en la red de forma radiada.

El tiempo de divergencia entre el nodal y los haplotipos cercanos al mismo fue estimado, obteniendo que el de H11 corresponde a 146 (± 146) años, H16 a 177 (± 177), H1 a 234 (± 234), H22 a 281 (± 281) y H19 a 1111 (± 1111) años.

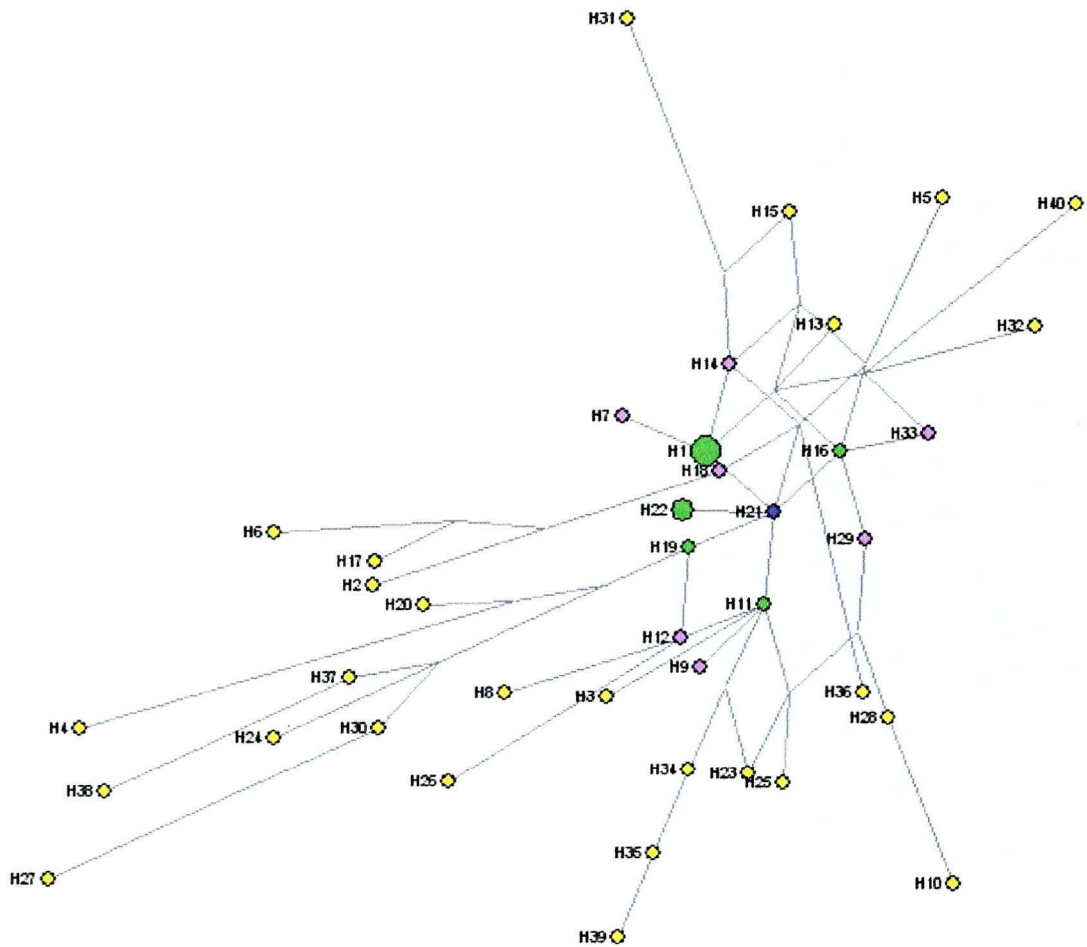


Figura 13: Red Mediana del Haplogrupo Q1a3a.

En azul se señala el nodal, en verde los haplotipos que se encuentran a un paso mutacional, en rosa aquellos a dos pasos mutacionales y en amarillo aquellos a más de dos pasos mutacionales. Las distancias gráficas son proporcionales a la cantidad de pasos mutacionales y el tamaño de los círculos lo es a la cantidad de muestras que abarca ese haplotipo.

4.3.2 Red del Haplogrupo R

En la red del haplogrupo R (Fig. 14), también se encontró un haplotipo nodal (H72). A su alrededor, la red muestra 3 haplotipos separados por un paso mutacional (H41, H42, H71), 12 por 2 (H48, H50, H53, H54, H55, H57, H70, H74, H80, H84, H86 y H92). El resto de los haplotipos se distribuye en forma radial.

Las estimaciones temporales de distancia al nodal, calculadas dentro del núcleo de esta red, revelaron que H41 y H42 difirieron de este centro hace 317 (± 317) años, H71 211 (± 211), H48 y H55 hace 357 (± 179), H80 388 (± 194), H92 408 (± 289), H50, H57 y H70 422 (± 211), H86 476 (± 238), H74 779 \pm (390), H54 1125 (± 563), y H84 y H53 hace 1322 (± 661) años.

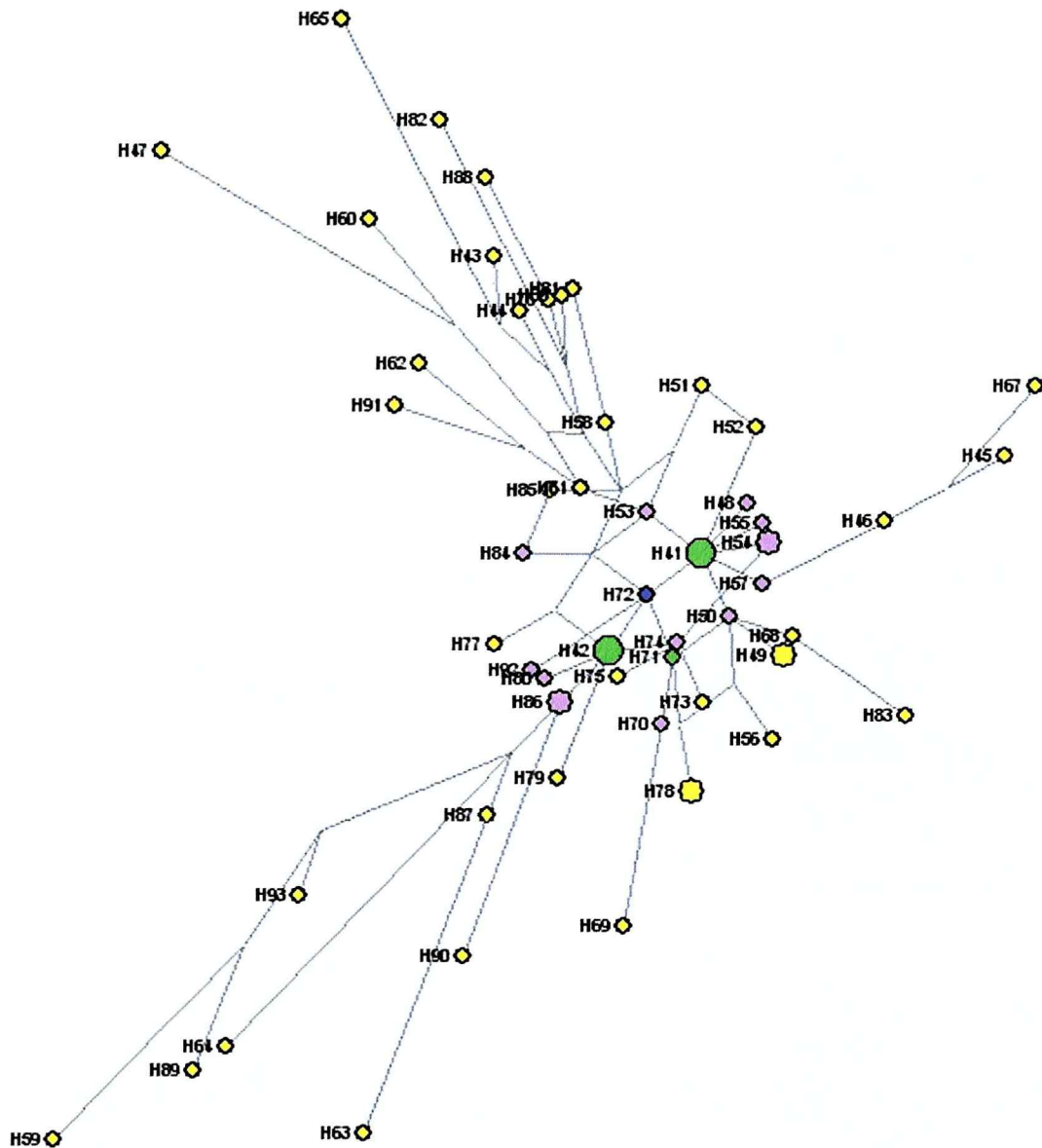


Figura 14: Red Mediana del Haplogrupo R.

En azul se señala el nodal, en verde los haplotipos que se encuentran a un paso mutacional, en rosa aquellos a dos pasos mutacionales y en amarillo aquellos a más de dos pasos mutacionales. Las distancias gráficas son

proporcionales a la cantidad de pasos mutacionales y el tamaño de los círculos lo es a la cantidad de muestras que abarca ese haplotipo.

En el caso de las estimaciones del tiempo de divergencia entre las muestras de un mismo apellido y haplogrupo, las mismas se encuentran detalladas en la Tabla 6, así como su comparación con la distancia al nodal de cada red en los casos que corresponde.

Veintitres de los 43 pares presentan una distancia temporal menor que el origen de los apellidos y, cuando esta distancia se compara con aquella al nodal de su respectiva red mediana, (la cantidad de pares se reduce a 36 puesto que este análisis se limita a los haplogrupos R* y Q1a3a), 19 de los mismos poseen una distancia al nodal mayor que aquella encontrada al haplotipo más cercano que a su vez pertenezca al mismo apellido.

Las Fig. 15 y 16 esquematizan estos tiempos con sus desvíos en relación a la implementación del uso de los apellidos para Europa y América. La línea señala el origen de los apellidos, de manera que todos aquellos casos que se encuentran dentro del área gris se corresponden a mutaciones posteriores a la implementación de los apellidos en esa zona.

Figura 15: Tiempo de divergencia entre muestras del mismo apellido, el caso del haplogrupo R.

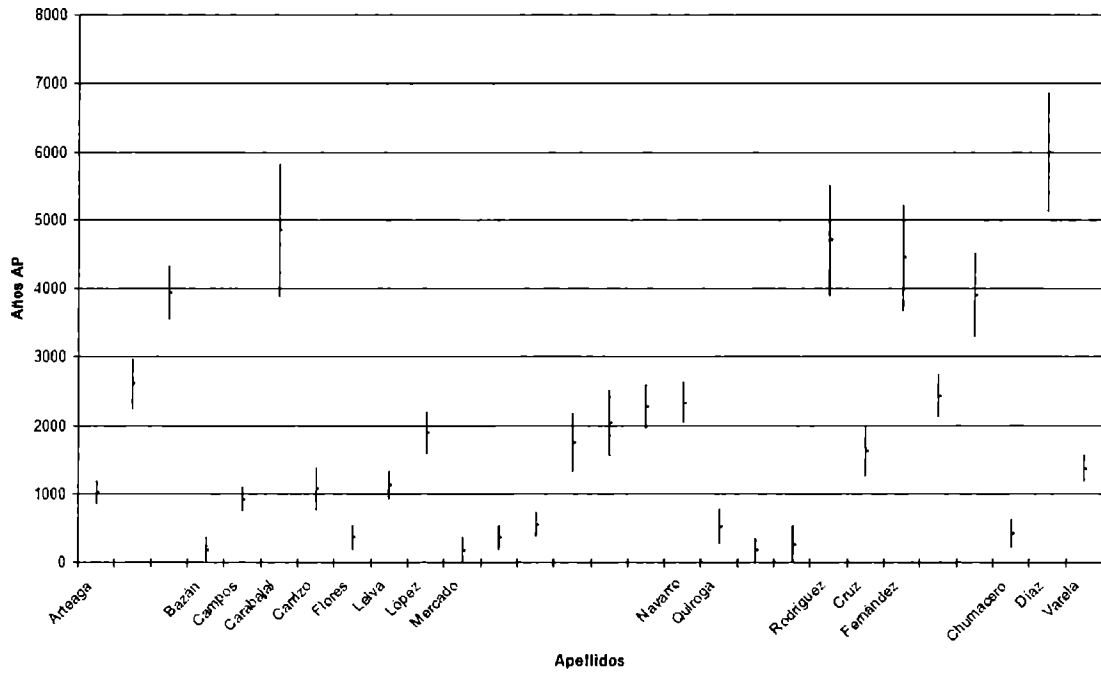


Figura 16: Tiempo de divergencia entre muestras del mismo apellido, el caso del haplogrupo Q.

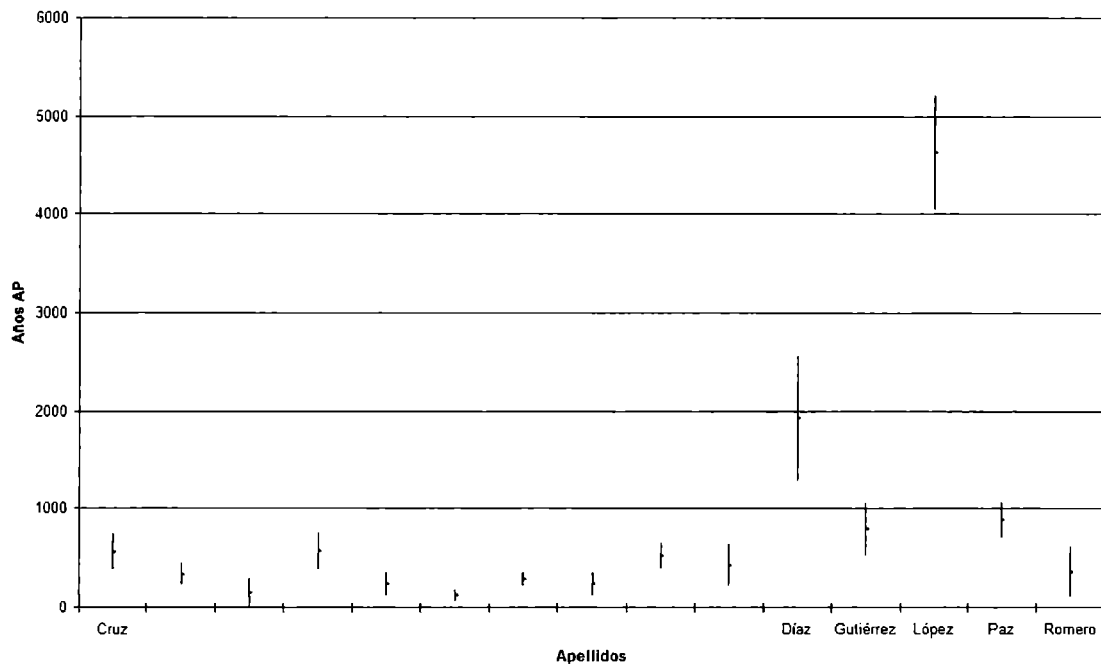


Tabla 6: Tiempos de divergencia entre muestras del mismo apellido

| Haplogrupo | Apellido | Muestras | Tiempo de divergencia (en años) | Tiempo de divergencia relativo al nodal (TDAN) | |
|---------------|-----------|-----------------|---------------------------------|--|----------------------|
| R | Arteaga | Sal141 - Sal124 | 1016 (\pm 169) | Más alto que el TDAN | |
| | | Sal141 - Sal109 | 2603 (\pm 368) | Más alto que el TDAN | |
| | | Sal109 - Sal124 | 3939 (\pm 394) | Más alto que el TDAN | |
| | Bazán | Cat10 - Cat97 | 177 (\pm 177) | Menor que el TDAN | |
| | Campos | Sal102 - Cat57 | 915 (\pm 183) | Menor que el TDAN | |
| | Carabajal | Tuc35 - Cat5 | 4840 (\pm 984) | Más alto que el TDAN | |
| | Carrizo | Cat33 - Cat62 | 1060 (\pm 306) | Menor que el TDAN | |
| | Flores | Sal113 - Juj113 | 357 (\pm 179) | El mismo | |
| | Leiva | Cat35 - Cat4 | 1122 (\pm 204) | Menor que el TDAN | |
| | López | W10 - Cat158 | 1884 (\pm 314) | Menor que el TDAN | |
| | Mercado | Cat78 - Cat140 | 177 (\pm 177) | Menor que el TDAN | |
| | | Cat140 - Cat149 | 357 (\pm 179) | Menor que el TDAN | |
| | | Cat78 - Cat149 | 543 (\pm 178) | Menor que el TDAN | |
| | | Cat131 - Cat149 | 1738 (\pm 435) | Más alto que el TDAN | |
| | | Cat131 - Cat140 | 2028 (\pm 478) | Más alto que el TDAN | |
| | | Cat78 - Cat131 | 2268 (\pm 324) | Más alto que el TDAN | |
| | Navarro | Cat136 - Cat102 | 2321 (\pm 298) | Más alto que el TDAN | |
| | Quiroga | Cat65 - Cat46 | 517 (\pm 256) | Más alto que el TDAN | |
| | | Cat65 - Cat47 | 169 (\pm 169) | Menor que el TDAN | |
| | | Cat46 - Cat47 | 265 (\pm 265) | Menor que el TDAN | |
| | Rodríguez | Cat61 - Cat23 | 4697 (\pm 822) | Más alto que el TDAN | |
| | Q1a3a | Cruz | Sq17 - Sq26 | 561 (\pm 187) | Más alto que el TDAN |
| | | | Sq17 - Sq4 | 335 (\pm 112) | Más alto que el TDAN |
| Sq17 - Sq6 | | | 146 (\pm 146) | Menor que el TDAN | |
| Sq17 - Sal154 | | | 569 (\pm 190) | Menor que el TDAN | |
| Sq26 - Sq4 | | | 238 (\pm 119) | Más alto que el TDAN | |
| Sq26 - Sq6 | | | 117 (\pm 58) | Menor que el TDAN | |
| Sq26 - Sal154 | | | 286 (\pm 71) | Menor que el TDAN | |
| Sq4 - Sq6 | | | 238 (\pm 119) | Menor que el TDAN | |
| Sq4 - Sal154 | | | 520 (\pm 130) | Menor que el TDAN | |
| Sq6 - Sal154 | | | 422 (\pm 211) | Menor que el TDAN | |
| Díaz | | W123 - Sal108 | 1920 (\pm 640) | Menor que el TDAN | |
| Gutiérrez | | Tuc26 - Cat118 | 788 (\pm 263) | Más alto que el TDAN | |
| López | | Cat145 - Juj35 | 4632 (\pm 579) | Más alto que el TDAN | |
| Paz | | Sal126 - Cat95 | 883 (\pm 177) | Más alto que el TDAN | |

| | | | | |
|-----|-----------|----------------|---------------------------|-------------------|
| | Romero | Sal100 - Sal67 | 357 (± 252) | Menor que el TDAN |
| F | Cruz | Sal143 - Sal73 | 1608 (± 379) | -- |
| | Fernández | Cat66 - Cat96 | 4440 (± 785) | -- |
| | | Cat66 - Cat2 | 2420 (± 303) | -- |
| | | Cat96 - Cat2 | 3897 (± 612) | -- |
| K | Chumacero | Juj39 - Juj34 | 415 (± 208) | -- |
| DE* | Díaz | Sal147 - Cat9 | 5994 (± 881) | -- |
| | Varela | Cat92 - Cat59 | 1362 (± 195) | -- |

Los valores que sobrepasan el tiempo estimado para el origen de los apellidos en su región están resaltados

4.4 Búsqueda en YHRD

Dentro de las concordancias entre haplotipos y apellidos, el haplotipo de Preli/s estuvo ausente en la base, mientras que el haplotipo modal del haplogrupo Q1a3a (que reúne individuos del mismo apellido y otros) tuvo una frecuencia absoluta de 43, mayoritariamente en las metapoblaciones Amerindias (18) y Mezclada (20).

El otro haplotipo compartido Q1a3a, H22, tuvo una frecuencia absoluta de 16 individuos, de los cuales 6 se encontraban en la metapoblación Amerindia y 6 en la Mezclada.

Uno de las haplotipos modales, H41, presentó una frecuencia absoluta de 380 individuos (de los cuales 296 eran de la metapoblación Europea), mientras H42 se presentó en 113 individuos de la base (93 europeos).

En el caso del resto de los haplotipos compartidos, H49 estaba representado por 60 individuos (51 pertenecientes a la metapoblación Europea), H54 por 121 individuos (117 europeos), H78 solamente 8 individuos (5 de Europa) y H86 tenía solamente 2 individuos, ambos europeos.

4.5 Transmisión no paterna del apellido

Los porcentajes de traspasos no paternos del apellido por localidad y tiempo se detallan en la Tabla 7.

No se encontraron diferencias significativas ($p>0.01$) entre las generaciones ego y parental en ninguna de las localidades analizadas, lo que nos permitió agruparlas y obtener un porcentaje total de traspaso no paterno del apellido para cada localidad.

Tabla 7: Proporciones de traspaso no paterno por localidad y tiempo.

| Localidad | Porcentaje de traspaso no paterno del apellido en la Transmisión Ego | Porcentaje de traspaso no paterno del apellido en la Transmisión Parental | Porcentaje total de traspaso no paterno del apellido |
|---|--|---|--|
| Aguaray, Macueta (Salta) | 18,8% | 10,6% | 16,6% |
| Belén y Londres (Catamarca) | 20% | 25% | 24,7% |
| Calingasta (San Juan) | 14,3% | 27,4% | 22,5% |
| Chepes (La Rioja) | 17,3% | 14,1% | 15,2% |
| Curuzú-Cuatiá y Perugorría (Corrientes) | 11,7% | 18,5% | 16,1% |
| Gualeguaychú (Entre Ríos) | 9,4% | 12,8% | 11,6% |
| La Paz (Entre Ríos) | 11,1% | 21% | 17,4% |
| La Quiaca (Jujuy) | 13,1% | 17,3% | 15,8% |
| La Rioja Capital | 12,1% | 12,9% | 12,6% |
| Lavalle | 2,6% | 4,9% | 4,1% |

| | | | |
|----------------------------|-------|-------|-------|
| (Mendoza) | | | |
| Maimará (Jujuy) | 14,5% | 20,4% | 18,2% |
| Malargüe (Mendoza) | 7,3% | 9,9% | 9% |
| Mendoza Capital | 6,9% | 8,6% | 8% |
| Reconquista (Santa Fé) | 6,7% | 12,2% | 10,2% |
| San José (Catamarca) | 13,6% | 29,7% | 23,7% |
| San Juan Capital | 9% | 10,1% | 9,7% |
| Santa María (Catamarca) | 13% | 21% | 18% |
| Tartagal (Salta) | 13,8% | 10,1% | 11,5% |

En cuanto a las comparaciones entre localidades, cuando se compararon todas entre sí mediante el ensayo para varias proporciones del estadístico chi cuadrado, la diferencia fue altamente significativa ($p < 0.01$). Como esta prueba no permite distinguir qué localidades son las que se diferencian, se optó por revisar si había diferencias dentro de grupos más pequeños, definidos geográficamente:

- NOA: Santa María, San José y Belén (provincia de Catamarca), Maimará y La Quiaca (provincia de Jujuy). Las diferencias resultaron significativas ($p < 0.01$), por eso no se pudieron agrupar. Con sucesivos ensayos pudo verse que la única localidad que difería era La Quiaca, por eso se la presenta en las tablas como una localidad aparte y el resto sí se agrupó.
- NEA: Tartagal y Aguaray, (Noreste de la Provincia de Salta), La Paz y Gualguaychú (provincia de Entre Ríos), Reconquista (provincia de Santa Fé) y Curuzú Cuatiá (provincia de Corrientes). También se pudo estimar un porcentaje para esta región ya que las diferencias entre las localidades resultaron no significativas ($p > 0.01$).

- Provincia de Mendoza: Malargüe, Lavalle y Mendoza Capital. Estas localidades se agruparon en un único porcentaje puesto que sus diferencias fueron no significativas ($p > 0.01$).
- La Rioja y San Juan: Chepes y La Rioja Capital (provincia de La Rioja), Calingasta, San Juan Capital (provincia de San Juan). La diferencia en las proporciones de transmisiones no paternas del apellido entre estas localidades resultaron significativas ($p < 0.01$). Mediante sucesivos ensayos se pudo observar que la localidad que difería del resto era Calingasta, de manera que se agruparon las localidades de provincia de La Rioja junto con San Juan Capital y Calingasta permanece separada.

En la Tabla 8 se exponen las agrupaciones resultantes.

Tabla 8: Porcentajes por localidad

| Grupo | Encuestas | Transmisiones | Porcentaje no paterno |
|---------------------|-----------|---------------|-----------------------|
| Mendoza | 411 | 1138 | 7,3% |
| NOA | 531 | 1441 | 20% |
| NEA | 888 | 1793 | 12,9% |
| La Rioja y San Juan | 426 | 2396 | 12% |
| La Quiaca | 224 | 597 | 15,8% |
| Calingasta | 70 | 187 | 23% |
| Totales | 2183 | 5921 | --- |

4.6 Modelos de transmisión irregular del apellido

Considerando que sólo fue posible agrupar por regiones geográficas ya que se observaron diferencias significativas entre las localidades, las estimaciones fueron realizadas con las proporciones de cada grupo y no con una proporción promedio total.

Para el modelo propuesto, con una cantidad infinita de apellidos, cuya fórmula era:

$$A_n = (p)^n$$

se detallan los cálculos de los modelos de transmisión no paterna del apellido.

Grupo Mendoza:

$$A_1 = (p)^1 = 0.927$$

$$A_2 = (p)^2 = 0.859$$

$$A_3 = (p)^3 = 0.797$$

$$A_4 = (p)^4 = 0.71$$

$$A_5 = (p)^5 = 0.66$$

$$A_6 = (p)^6 = 0.61$$

$$A_7 = (p)^7 = 0.57$$

$$A_8 = (p)^8 = 0.52$$

$$A_9 = (p)^9 = 0.49$$

Tomaría nueve generaciones para que sólo la mitad de los apellidos retengan el linaje paterno.

Si consideramos 25 años por generación:

$$9 \times 25 = 225 \text{ años}$$

Si consideramos 35 años por generación (según Jobling y col. 2007):

$$9 \times 35 = 315 \text{ años}$$

Así, en el caso donde se encontró la menor proporción de transmisiones no paternas del apellido, en un máximo de 385 años la mitad de los apellidos perdieron su carácter de seguir una línea paterna perfecta.

Grupo NOA:

$$A_1 = (p)^1 = 0.80$$

$$A_2 = (p)^2 = 0.64$$

$$A_3 = (p)^3 = 0.51$$

$$A_4 = (p)^4 = 0.41$$

En cuatro generaciones menos de la mitad de los linajes mantienen la transmisión paterna.

Si consideramos 25 años por generación:

$$5 \times 25 = 100 \text{ años}$$

Si consideramos 35 años por generación (según Jobling y col. 2007):

$$4 \times 35 = 140 \text{ años}$$

En esta región tomaría un máximo de 140 años para que sólo menos de la mitad de los apellidos hayan seguido la línea paterna.

Grupo NEA:

$$A_1 = (p)^1 = 0.87$$

$$A_2 = (p)^2 = 0.76$$

$$A_3 = (p)^3 = 0.66$$

$$A_4 = (p)^4 = 0.57$$

$$A_5 = (p)^5 = 0.49$$

Tomaría cinco generaciones para que sólo la mitad de los apellidos retengan el linaje paterno.

Si consideramos 25 años por generación:

$$5 \times 25 = 125 \text{ años}$$

Si consideramos 35 años por generación (según Jobling y col. 2007):

$$5 \times 35 = 175 \text{ años}$$

Así, en un máximo de 175 años la mitad de los apellidos perdieron su carácter de seguir una línea paterna perfecta.

Grupo La Rioja y San Juan:

$$A_1 = (p)^1 = 0.88$$

$$A_2 = (p)^2 = 0.77$$

$$A_3 = (p)^3 = 0.68$$

$$A_4 = (p)^4 = 0.60$$

$$A_5 = (p)^5 = 0.53$$

$$A_6 = (p)^6 = 0.46$$

En este caso se requerirían seis generaciones.

Si consideramos 25 años por generación:

$$6 \times 25 = 150 \text{ años}$$

Si consideramos 35 años por generación (según Jobling y col. 2007):

$$6 \times 35 = 210 \text{ años}$$

Tomaría un máximo de 210 años.

La Quiaca:

$$A_1 = (p)^1 = 0.84$$

$$A_2 = (p)^2 = 0.71$$

$$A_3 = (p)^3 = 0.59$$

$$A_4 = (p)^4 = 0.49$$

Al igual que en el grupo NOA, en sólo cuatro generaciones menos de la mitad de los linajes mantienen la transmisión paterna.

Si consideramos 25 años por generación:

$$5 \times 25 = 100 \text{ años}$$

Si consideramos 35 años por generación (según Jobling y col. 2007):

$$4 \times 35 = 140 \text{ años}$$

En esta localidad tomaría un máximo de 140 años para que la mitad de los apellidos no hayan seguido la línea paterna.

Calingasta:

$$A_1 = (p)^1 = 0.77$$

$$A_2 = (p)^2 = 0.59$$

$$A_3 = (p)^3 = 0.46$$

De esta forma, en donde se encontró la mayor proporción de transmisiones no paternas del apellidos, sólo se requieren 3 generaciones.

Si consideramos 25 años por generación:

$$3 \times 25 = 75 \text{ años}$$

Si consideramos 35 años por generación (según Jobling y col. 2007):

$$3 \times 35 = 105 \text{ años}$$

En poco más de cien años la mitad de los apellidos perdieron su carácter patrilineal.

La Tabla 9 resume los resultados de este modelo, con las estimaciones de profundidad temporal para cada caso.

Tabla 9: Resultados del modelo

| Grupo/Localidad | Proporción de traspasos no paternos | Generaciones estimadas | Tiempo (25 años por generación) | Tiempo (35 años por generación) |
|---------------------|-------------------------------------|------------------------|---------------------------------|---------------------------------|
| Mendoza | 0.073 | 9 | 225 | 315 |
| NOA | 0.2 | 4 | 100 | 140 |
| NEA | 0.13 | 5 | 125 | 175 |
| La Rioja y San Juan | 0.12 | 6 | 150 | 210 |
| La Quiaca | 0.16 | 4 | 100 | 140 |
| Calingasta | 0.23 | 3 | 75 | 105 |

Discusión

“(...)

Father, you left me but I never left you,

I needed you but you didn't need me,

So I got to tell you,

Goodbye, goodbye.

(...)”

Mother, John Lennon

Para visualizar esta asociación entre apellidos y haplotipos, recordemos las muñecas rusas conocidas como Matrioshkas. Conforman una serie de muñecas idénticas que representan un linaje, cada una integrada por dos partes: una base y una porción superior. Imaginemos que la base es un linaje de cromosoma Y (el color predominante podría ser el haplogrupo y los motivos más intrincados el haplotipo) y la porción superior el apellido. El fundador del linaje sería la muñeca más grande, aquella que queda por fuera cuando se ensamblan una adentro de la otra; si cada linaje del Y se transmite junto con un apellido, el resultado es la clásica línea con todas las muñecas iguales, la siguiente más pequeña que la anterior. En cambio, si hay interrupciones en la transmisión del apellido, en cada sucesión de bases idénticas (la descendencia del fundador), las porciones superiores irían cambiando, a veces manteniendo el color de la base pero no los motivos (en los casos que el apellido tiene el mismo origen que el haplogrupo) y en otros con un color completamente distinto (cuando el apellido y el haplogrupo tienen diferentes orígenes).

Siguiendo con la analogía, en un estudio con ADN contemporáneo, se analizarían las muñecas más pequeñas de la serie, y con ellas se trazaría la sucesión de bases hasta el fundador de ese linaje. Al haber discordancia entre la transmisión de apellidos y la del cromosoma Y, no sería posible recuperar las porciones superiores, puesto que se fueron alternando a lo largo de las generaciones.

5.1 Apellidos y ancestralidad

La baja frecuencia de apellidos nativos en la muestra analizada concuerda con la menor frecuencia de haplogrupos nativos en relación a los haplogrupos extranjeros. Esto se puede comprender fácilmente cuando consideramos los trabajos de Dipierri y colaboradores (1991, 1996, 2005a, 2005b) y Dipierri y Alfaro (1996), que explican que en el NOA los apellidos fueron asignados por los misioneros españoles. El hallazgo de haplogrupos extranjeros y nativos dentro del mismo apellido agrega evidencia en esta dirección, haciendo coincidir a los datos genéticos con los registros históricos.

Asimismo, las estimaciones de riesgo relativo señalaron que existe una asociación significativa entre tener un apellido nativo americano y pertenecer a un haplogrupo autóctono. Si bien existen casos de individuos con haplogrupo nativo y apellido extranjero, no es esperable encontrar individuos de apellido nativo y haplogrupo extranjero.

La asociación entre haplogrupos y apellidos es menor que aquella encontrada previamente en Inglaterra por King y colaboradores (2006), con una diferencia altamente significativa ($p < 0.01$) aún considerando la menor resolución de nuestro método para asignar haplogrupos. Esto se puede comprender cuando tomamos en cuenta su decisión de sólo analizar apellidos de origen inglés, y criterios de inclusión en la muestra más estrictos, resultantes en una menor diversidad étnica que la de la actual población inglesa. El caso del NOA es distinto, debido a la miríada de diferentes condiciones socio-históricas que resultaron en la configuración de sus poblaciones.

Por otra parte, Jin y col. (1999) analizaron 19 SNPs del cromosoma Y en 1500 varones *Han* de 28 provincias chinas, encontrando un origen polifilético para sus apellidos. En el caso de los *Han* de China, los apellidos comenzaron a usarse 3000 años antes que en Europa, dando una profundidad temporal mucho mayor que la de la presente tesis. De todas formas, los marcadores empleados por estos investigadores (SNPs) presentan tasas de mutación muy bajas (recuérdese que a través de su empleo definen los haplogrupos), de manera tal que el hallazgo de polifiletismo a este nivel en los apellidos es muy significativo. Si bien las diferencias culturales entre los *Han* y el NOA son vastas, resulta interesante que en ellos tampoco se mantuvo la asociación entre apellidos y linajes del cromosoma Y.

El grupo de Kim Y. J. (1999) realizó una serie de amplificaciones y restricciones 49a/TaqI mediante las cuales establecían haplotipos. Su interés era analizar el origen de la población coreana, y para ello tomaron 6.538 muestras de sangre, pertenecientes a 24 apellidos distintos, que analizaron a

través de 49a/TaqI. Estos investigadores no encontraron alelos específicos o característicos que sean típicos de ciertos apellidos, de esta forma tampoco se evidencia una asociación entre combinaciones genéticas y apellidos.

McEvoy y col. (2006) estudiaron la influencia vikinga en poblaciones irlandesas, para ello seleccionaron irlandeses de apellido de origen vikingo, además de algunos de origen local, y compararon sus perfiles trazados por microsatélites y SNPs. Ellos no encontraron evidencia convincente de un aporte substancial de ancestría escandinava en los linajes paternos de individuos portadores de apellidos putativamente nórdicos. En este caso tampoco habría una relación clara entre la ancestría genética y la del apellido.

En Alemania, Immel y col. (2006) tipificaron microsatélites de 419 varones de Halle, que subdividieron en tres grupos de acuerdo al origen de sus respectivos apellidos en “alemanes”, “eslavos” y “mixtos”. Compararon la distribución de estos haplotipos, hallando indistinguibles a los mixtos de los alemanes, pero el grupo eslavo era significativamente diferente de los alemanes y mixtos agrupados, por eso es que proponen que una forma sencilla de subdividir a las poblaciones Centro Europeas es mediante los apellidos. De todas formas, los mismos autores detallan la necesidad de realizar análisis de SNP, y compararlos con las agrupaciones por apellido, especialmente en los casos donde se quiere analizar movimientos poblacionales antiguos.

En el Reino Unido, Bryan Sykes y Catherine Irven llevaron a cabo uno de los primeros estudios de asociación entre apellido y cromosoma Y (Sykes e Irven, 2000). Ellos establecieron haplotipos a través de cuatro microsatélites del cromosoma Y (DYS 19, DYS 390, DYS 391 y DYS 393) en 48 muestras de individuos del apellido Sykes provenientes de tres condados (West Yorkshire, Lancashire y Cheshire), 21 individuos de los mismos condados pero con apellidos diferentes y 139 nativos ingleses del resto de ese país. Encontraron que el 43.8% de los individuos Sykes llevaban el mismo haplotipo, además de que el 14.8% del resto de ellos presentaban haplotipos a un paso mutacional

del mismo. Con sus datos estimaron una tasa de no paternidad de 1.3% por generación, asumiendo que pasaron 23 generaciones desde el primer antecesor común masculino.

De todas formas, para comprender mejor cómo es que estos investigadores obtuvieron estos resultados, es importante tener en cuenta que es posible que algunos de los individuos Sykes estuvieran emparentados entre sí, considerando que provienen de los mismos condados y que este apellido sólo cuenta con 9.885 votantes registrados para todo el Reino Unido. Asimismo, si bien en la época en la cual este trabajo fue realizado considerar cuatro microsatélites era difícil de llevar a cabo, tomar un número tan bajo de marcadores podría enmascarar la diversidad subyacente.

Otra comparación interesante es analizar el trabajo de Gómez y col. (2008). Este grupo también buscó asociaciones entre apellidos y haplotipos, pero para tres poblaciones de Colombia. En su caso, al igual que aquí, encontraron una elevada cantidad de haplotipos únicos (268 en 308 muestras) y los haplotipos compartidos que hallaron no presentaban apellidos compartidos.

Aún así, se preguntaron si existía una asociación entre individuos de apellido nativo americano y cromosoma Y de origen nativo, y proponen que habría un 40% de inconsistencia de linaje. En este caso es importante tener en cuenta una serie de errores metodológicos cometidos por estos autores:

- En ningún momento realizan análisis de haplogrupo mediante SNP, sino que asignan como nativo americanos a aquellos individuos que presentan el alelo 13 en el STR DYS19.
- Si bien este alelo es el más común en poblaciones nativas, otros alelos también están presentes en frecuencias no despreciables (Bianchi y col. 1998).
- El alelo 13 para DYS 19 es el segundo más frecuente en muestras del haplogrupo R, característico de poblaciones europeas, y también está presente en otras poblaciones.

- No realizaron un ensayo de asociación para analizar esta hipótesis, sino que sólo contrastaron la frecuencia de haplotipos denominados en forma ad-hoc como nativo americanos con la frecuencia de apellidos nativo americanos.
- Al no haber establecido los haplogrupos, es probable que algunos de los 20 haplotipos compartidos que describen sean sólo idénticos por convergencia en vez de por ascendencia (podrían tener la misma combinación de alelos de STR pero pertenecer a haplogrupos diferentes; entre los haplotipos establecidos en la presente tesis doctoral, con una muestra menor, se encontró un caso de convergencia, de allí que no es algo tan poco probable).

Si bien sus comparaciones entre origen del apellido y ancestralidad de los linajes masculinos no resultan satisfactorias, sí son válidos sus perfiles genéticos trazados y estos haplotipos reflejan una evidencia en contra de la clásica propuesta de “un haplotipo, un apellido”.

La bibliografía arriba detallada nos muestra panoramas muy diversos según qué región del mundo se esté estudiando. No obstante, aún en aquellos casos donde sí se ven asociaciones claras entre apellidos y linajes del cromosoma Y, las mismas suelen restringirse a agrupaciones por origen geográfico del apellido y linaje, más que a haplotipos y apellidos específicos, aún considerando las mutaciones que podrían producirse en un linaje a lo largo del tiempo transcurrido desde el origen de los apellidos hasta el día de hoy.

Como ya se mencionó anteriormente, en el NOA se encontró una asociación entre el origen del apellido y el haplogrupo (origen del linaje del cromosoma Y). De esta manera, es más probable que un individuo con un apellido nativo pertenezca un haplogrupo Q que a un haplogrupo foráneo.

5.2 El caso de los apellidos frecuentes

Los apellidos frecuentes son más propensos a tener múltiples fundadores, consecuentemente presentando diferentes haplotipos y haplogrupos (Jobling y col. 2001). Los resultados obtenidos apoyan esta hipótesis, dada la baja asociación entre haplogrupo y apellido en nuestra muestra, además de que es más común en los apellidos menos frecuentes.

Es interesante resaltar, de todas formas, que dentro del apellido con mayor frecuencia en la muestra (Cruz, uno de los apellidos más comunes en Argentina), se encontró una concordancia entre apellido y haplotipo (h1), aún cuando también incluyó a un individuo portador de otro apellido. Se debe remarcar la importancia de que estos individuos presentan un haplogrupo nativo americano, dado que los nativos de nuestra zona de interés recién comenzaron a tener apellido bajo la influencia de los misioneros españoles, quienes bautizaron a esta gente en modo arbitrario.

Existen diferentes explicaciones posibles para este hallazgo. Por un lado, no se puede descartar relaciones de parentesco entre los individuos involucrados puesto que provienen de la misma provincia. Por otra parte, este haplotipo tiene una alta frecuencia en las poblaciones Nativas Americanas (en YHRD incluyó 43 individuos, la mayoría provenientes de poblaciones definidas como Amerindias o Mezcladas), al igual que este apellido en individuos de esta ascendencia. De esta manera, la probabilidad de encontrar individuos con este apellido y haplotipo no sería tan baja.

5.3 Haplotipos Modales Ibérico y Europeo

Dado el patrón de inmigración del NOA, se esperaba encontrar los Haplotipos Modales Ibéricos e Europeos (Gusmao y col. 2003). El Haplotipo Modal Europeo fue hallado en uno de los individuos analizados (H72), y se corresponde con el haplotipo nodal en la red mediana. Como era de esperar, este haplotipo mostró las frecuencias más altas en la comparación con la base

de datos YHRD, aún cuando su frecuencia en la muestra analizada en la presente tesis es de sólo un individuo.

En relación al Haplotipo Modal Ibérico, el mismo estuvo ausente en la muestra estudiada. Si bien esto podría parcialmente explicarse por el tamaño muestral de la presente tesis, es más factible que sea el resultado del efecto fundador producido mediante las migraciones hispánicas a este continente. Las frecuencias de los linajes masculinos que arribaron al NOA serían diferentes a aquellas presentes en el lugar de origen, en vista de que los colonos pertenecían a segmentos sociales y localidades específicos, y sólo una fracción pequeña de los varones españoles se dirigió a esta región de todas las que colonizaron en América.

Estas diferencias también se pueden comprender al considerar las estimaciones de los tiempos de divergencia, ya que en la mayoría de los casos ocurre posteriormente a la llegada de los españoles, especialmente en el caso de quienes rodean al nodo por un paso mutacional. Con estos resultados es altamente probable que estas separaciones se hayan dado una vez instalados en el NOA, y esto también explicaría por qué son disímiles a las frecuencias encontradas en la base de datos internacional YHRD.

Una explicación plausible es que se esté observando una fotografía instantánea del resultado de estos fenómenos de efecto fundador y deriva, resultantes en una distribución de haplotipos distintiva para cada uno de estos continentes.

5.4 Concordancias, discordancias

Resulta interesante la baja frecuencia del haplotipo exclusivo de un apellido (H94) en YHRD, enfatizando la concordancia entre ambos sistemas para este caso en particular, mientras que los haplotipos trans-apellidos (H22, H41, H42, H49, H54, H78 y H86) son vistos no sólo como una discordancia sino también como un ejemplo de los complejos procesos que afectan los linajes sociales. Además, el apellido en el cual se da la asociación más clara entre ambos sistemas, *Preli/s*, es uno de los apellidos con frecuencias más bajas en

el NOA (14 en total, sumadas ambas variantes del mismo), y sólo presente en la provincia de Catamarca. Es probable que se trate de individuos con un parentesco cercano, aunque el mismo no pudo ser detectado cuando se los seleccionó al comienzo de esta investigación.

En el nivel de coincidencia dentro de un apellido y haplogrupo, en veintitrés de los cuarenta y tres pares, las estimaciones de tiempo de divergencia son posteriores al comienzo del uso de apellidos en Europa (para los haplogrupos R, F, K y DE*) y América (en el Q1a3a).

De todas formas, cuando se estimó la distancia al nodal de los haplotipos de las redes de los haplogrupos Q1a3a y R (N=36), en 17 casos fue menor o igual que aquella a los otros haplotipos del mismo apellido. En 19 de los 36 casos la distancia al haplotipo nodal es mayor que aquella al haplotipo del mismo apellido más cercano (Tabla 8), siendo de esta forma muy probable que, en estos últimos, las diferencias a nivel del cromosoma Y se deban a mutaciones posteriores al origen de los apellidos en lugar de polifiletismo.

Así, algunos haplotipos continúan compartiendo un ancestro común hace aproximadamente entre 6 y 10 generaciones, aunque sus similitudes dentro de sus correspondientes apellidos podría explicarse por descender de ese haplotipo nodal, que a su vez es el Haplotipo Modal Europeo, esparcido a través de diferentes apellidos posteriormente a la implementación de los apellidos en Europa. Considerando la profundidad temporal estimada para esta relación, es muy factible que ese ancestro sea uno de los primeros inmigrantes europeos, probablemente español.

En los diez casos restantes de haplogrupo y apellido compartidos, el tiempo de divergencia excede ambos tipos de profundidad temporal: la distancia al haplotipo nodal y la del origen de los apellidos. Esto podría deberse a la transmisión irregular de apellidos en el caso de los apellidos menos comunes, y polifiletismo en los más comunes de Argentina.

5.5 ¿Por qué los haplotipos modales se esparcen sobre diferentes apellidos?

Ya mencionamos los mecanismos de deriva y efecto fundador como explicaciones para los datos genéticos encontrados, pero la variabilidad de

linajes sociales dentro de los haplotipos modales requiere de un análisis más detallado.

Zerjal y col. (2003) han reportado una nueva forma de selección en poblaciones humanas basada en el prestigio social. Estos autores identificaron un linaje de cromosoma Y con una serie de características inusuales: estaba presente en 16 poblaciones en una amplia región de Asia (del océano Pacífico al mar Caspio), con una frecuencia muy elevada ~8% de todos los hombres en esta región lo llevan (~0,5% del total mundial), con una variación dentro del linaje que sugería como origen a Mongolia, cerca de 1000 años antes del presente. Como estas características no pueden haberse dado por azar (especialmente una dispersión de esta magnitud realizada a tanta velocidad), deben ser el resultado de selección.

Estos investigadores ven a la selección biológica como poco probable ya que hay muy pocos genes en el cromosoma Y, además de que, al ser sus funciones altamente especializadas, provee pocas oportunidades de selección (Jobling y Tyler-Smith, 2000). En su lugar, proponen un aumento de la eficacia reproductiva transmitida socialmente, de hombres que portan el mismo cromosoma Y, que llevaría a un incremento en la frecuencia de su linaje, y que este efecto se vería a su vez aumentado mediante la eliminación de los varones no emparentados a este linaje.

El grupo de Zerjal propone que este linaje tan particular correspondería a Genghis Khan y sus descendientes, en vista de que su origen temporal y geográfico coincide con el de este personaje histórico, además de que existen registros sobre su amplia descendencia, así como de la práctica eliminar a todos los varones y someter a las mujeres de los pueblos que dominaban. Asimismo, al fundar una dinastía perdurable, este grupo de varones del mismo linaje continuó teniendo una mayor eficacia reproductiva, no basada en características biológicas sino en su estatus social.

Esta forma de selección se corresponde con el efecto fundador encontrado en el NOA y nos brinda una explicación factible a los resultados obtenidos. Recordemos que en esta región sucedieron fenómenos de caudillismo, además del efecto de los ingenios azucareros, donde existieron

casos de hijos ilegítimos entre mujeres de “la peonada” y varones de la burguesía azucarera (Dipierri, 2004).

Así, los haplotipos modales del NOA pertenecerían a la descendencia de unos pocos colonos emparentados. Teniendo en cuenta la amplia cantidad de descendientes, es muy probable que esos hijos provengan de diferentes mujeres, concebidos fuera del matrimonio o en situaciones como las de los ingenios. Dado que estas uniones no siempre fueron reconocidas, la adjudicación del apellido materno es una explicación altamente plausible para la diversidad observada.

5.6 Transmisión no paterna del apellido

Del análisis de las encuestas genealógicas relevadas en el marco del Proyecto “Pasado Indígena, presente criollo. Diversidad filética y estructura geográfica de los linajes maternos aborígenes presentes en poblaciones criollas argentinas”, se obtuvieron las proporciones de las transmisiones no paternas del apellido para diferentes localidades a lo largo del país.

Es interesante que no se observaron diferencias significativas en estas proporciones entre la generación del ego y la de sus padres, permitiendo estimar una proporción total para cada localidad. Asimismo, un problema siempre presente en este tipo de análisis es que se parte de una asunción uniformista, pero al poder tomar dos generaciones y no encontrar diferencias entre ellas se ve reducido el error que este presupuesto conlleva.

Aquí es importante tener en cuenta el trabajo sobre filiación de J. E. Dipierri (2004), que comenta como en el Primer Censo de la República Argentina (1869) el porcentaje de nacidos vivos ilegítimos era de un 18,7% para todo el país y 22.5% en la Provincia de Jujuy, porcentajes que caen dentro del rango reportado en la presente tesis.

También cabe destacar la obtención de proporciones regionales, y cómo éstas son distintas entre sí. La más baja corresponde a la provincia de Mendoza, mientras que la más elevada fue encontrada en el NOA, la región de interés de la presente tesis doctoral.

Algunos autores (King y Jobling, 2009) proponen arbitrariamente un porcentaje del 2% para las transmisiones no paternas del apellido. Esta proporción quizás sea adecuada para otras regiones del planeta, pero en ninguna de las áreas consideradas por este estudio se encontró una proporción tan baja. Aún en el área de Mendoza, nuestro porcentaje obtenido a través de las transmisiones no paternas reconocidas por los encuestados es tres veces mayor (es importante recordar que las no-paternidades biológicas ignoradas no pueden ser detectadas por el método de encuesta genealógica). El NOA fue el área donde se encontró el mayor porcentaje de transmisiones no paternas (20%), siendo el mismo diez veces mayor que el propuesto por esos autores para su estudio.

Los modelos analizados reflejaron una muy rápida pérdida de asociación entre linajes masculinos y apellidos para la mitad de la población. En el caso del NOA, en tan sólo tres o cuatro generaciones la mitad de los apellidos perdieron esa correlación, con una profundidad temporal estimada en un máximo de 140 años. Aquí también resultan de interés las estimaciones de la localidad de La Quiaca, si bien su porcentaje de transmisión no paterna del apellido era diferente a la del resto del NOA, cuando se calculó en cuántas generaciones sólo la mitad de los linajes mantenían el carácter patrilineal, también se obtuvo como resultado 4 generaciones, con un máximo de 140 años.

Este resultado es concordante con las observaciones a nivel molecular, puesto que las mismas evidencian una baja asociación entre apellidos y haplotipos. Resulta claro entonces cómo individuos del mismo apellido poseen haplotipos distintos, mientras que individuos de distintos apellidos sí comparten el haplotipo, cuando se toma en cuenta lo rápido que se pierde la equivalencia entre linaje masculino y apellido.

Conclusión

“I take thee at thy word:
Call me but love, and I'll be new baptized;”

W. Shakespeare, Romeo and Juliet, Acto II, Escena 1

En el presente trabajo se han descripto los componentes extranjeros y nativos en los linajes masculinos presentes en poblaciones del NOA, con una mayor proporción de haplogrupos de origen europeo, probablemente español, aunque igualmente persisten linajes nativos, evidenciados por la presencia de Q1a3a.

Al evaluar la similitud entre linajes genéticos de individuos con el mismo apellido y de aquellos que no lo comparten, se encontró un caso de concordancia perfecta entre apellido y haplotipo, el caso de los dos individuos correspondientes al apellido Preli/s, y un caso donde tres individuos del mismo apellido (Cruz) junto con uno de otro (Gregorio) compartían el mismo haplotipo. También se hallaron dos haplotipos modales, ambos del haplogrupo R, cada uno representado por tres individuos; así como cinco haplotipos (uno del haplogrupo Q1a3a y cuatro del R) con dos individuos cada uno, que serían resultado de fenómenos de transmisión irregular del apellido, mientras que el resto resultaron únicos. El 24,39% de los apellidos representados por al menos dos muestras evidenció una misma ancestría ya que los linajes incluidos en cada uno de ellos pertenecían al mismo haplogrupo.

Además, se encontró una asociación significativa entre ancestrías y origen del apellido, siendo más probable que individuos portadores de un apellido nativo pertenezcan a un haplogrupo americano.

Asimismo, de las estimaciones de divergencia temporal entre muestras del mismo apellido se obtuvo que, en el 53,5% de las duplas no excede a la implementación de los apellidos, no pudiéndose así descartar la posibilidad de mutaciones recientes.

Se obtuvieron tasas de transmisión no paterna del apellido para diferentes regiones del país, que se emplearon para el planteo de modelos que permitieron evaluar la velocidad con que se pierde la asociación entre apellidos y cromosoma Y para cada área. La región en la cual se observó la pérdida más rápida fue el NOA, requiriendo 4 generaciones, mientras que la menor pérdida fue estimada para la provincia de Mendoza, con 9 generaciones.

Por lo tanto, desde distintas visiones (análisis genéticos y modelizaciones matemáticas a partir de encuestas genealógicas) se encontró una baja asociación entre los linajes sociales y los biológicos, enfatizando la complejidad de estos fenómenos. De esta manera, cada caso debe ser analizado con cuidado, ya que no siempre cada apellido se corresponde con un linaje del cromosoma Y, ya sea por mutaciones recientes que enmascaran sus similitudes, como por fenómenos de polifiletismo y transmisión irregular.

Bibliografía

J: [In a pathetic voice.] Miss Prism, more is restored to you than this hand-bag. I was the baby you placed in it.

MP: [Amazed.] You?

J: [Embracing her.] Yes... mother!

MP:[Recoiling in indignant astonishment.] Mr. Worthing! I am unmarried.

J: Unmarried! I do not deny that is a serious blow. But after all, who has the right to cast a stone against one who has suffered? Cannot repentance wipe out an act of folly? Why should there be one law for men, and another for women? Mother, I forgive you. [Tries to embrace her again.]

MP: [Still more indignant.] Mr. Worthing, there is some error. [Pointing to Lady Bracknell.] There is the lady who can tell you who you really are.

O. Wilde. The Importance of being Ernest, Act III.

- Albeck, M. E., Alfaro E. L. Y Dipierri J. E. 2005. "Antroponimia e identidad en Casabindo". Revista Argentina de Antropología Biológica. Volumen 7, Número 1. Septiembre 2005.
- Alfaro, E. L., Muzzio M., Ramallo V., Dipierri J. E. y Bailliet G. 2007. "Genealogía molecular de los descendientes del Marqués de Yavi". Revista Argentina de Antropología Biológica. ISSN: 1514-7991. Volumen 9, Número 1.
- Athey, W. T. 2005. "Haplogroup prediction from Y-STR values Using an Allele- Frequency Approach". Journal of Genetic Genealogy 1:1-7.
- Athey, W.T. 2006. "Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach". Journal of Genetic Genealogy 2: 34-39.
- Baillet, G., Castilla E. E., Adams J. P., Orioli I. M., Martínez-Marignac V. L., Richard S. M., Bianchi N. O. 2001. "Correlation between molecular and conventional genealogies in Aicuña: A rural population from Northwestern Argentina". Human Heredity 51: 150-159.
- Bailliet, G., Muzzio M., Ramallo V., Alfaro E. L., Dipierri J. E., Bianchi N. O. 2007. "Linajes paternos y apellidos en genealogías y poblaciones humanas". Revista Argentina de Antropología Biológica. ISSN: 1514-7991 Volumen 9, Número 1. pg. 30.
- Bandelt, H. J., Forster P., Röhl. A. 1999. "Median-Joining Networks for Inferring Intraspecific Phylogenies" Mol. Biol. Evol. Vol 16. Issue 1. 37-48.
- Behar, D. M., Garrigan D., Kaplan M. E., Mobasher Z., et. al. 2004. "Contrasting patterns of the Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations". Hum. Genet. 114: 354-365.
- Bianchi N. O., Catanesi C. I., Baillet G., Martinez-Marignac V. L., Bravi C. M., Vidal-Rioja L. B., Herrera R. J., Lopez Camelo J. S. 1998. "Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am. J. Hum. Genet. 63: 1862-1871.

- Binayán Carmona N. 1999. en Emece (ed.) Historia Genealógica Argentina.
- Bolnick, D. A., Bolnick D. I., Smith D. G. 2006. "Asymmetric Male and Female Genetic Histories among Native Americans from Eastern North America". *Molecular Biology and Evolution* 23 (11): 2161-2174.
- Boman, Eric. 1908. Antiquités de la Région Andine de la République Argentine et du Désert d'Atacama. Vol 2. Paris : Matson.
- Bosch, E., Calafell F., Santos F., Perez-Lezaun A., et. al. 1999. "Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome". *Am. J. Hum. Genet.* 65: 1623-1638.
- Bravi, C. M., Sans M., Baillet G., Martinez-Marignac V. L., Bianchi N.O. 1997. "Characterization of mitochondrial and Y-chromosome haplotypes in an Uruguayan population of African ancestry." *Hum. Biol.* 69 (5): 641-652.
- Cann, R. L., Stoneking M, Wilson A. C. 1987. "Mitochondrial DNA and human evolution". *Nature* 1-7;325(6099):31-6.
- Castilla, E. E. y Adams J. 1990. "Migration and genetic structure in an isolated population in Argentina: Aicuña"; en Adams J. P. (ed.): Proceedings of Convergent Questions in Genetics and Demography. Oxford. Oxford University Press pp 45-62.
- Castilla, E. E. y Adams J. 1996. "Genealogical information and the use of rural Latin American populations: Reality and fantasy." *Hum. Herd.* 46:241-255.
- Catanesi, C. I., Méndez M. G., Salceda S. A. 2001. *Milenio*. pp.127-131.
- Cavalli-Sforza L. L. 2007. "Human Evolution and Its Relevance for Genetic Epidemiology". *Annu. Rev. Genomics Hum. Genet.* 8: 1-15.
- Colantonio, S. E., Lasker, G. W., Kaplan B. A., Fuster, V., 2003. "Use of Surname Models in Human Population Biology: A Review of Recent Developments." *Hum Biol.* 2003 Dec.; 75(6):785-807.
- Demarchi, D. A. y Mitchell R. J. 2004. "Genetic structure and gene flow in Gran Chaco populations of Argentina: evidence from Y-chromosome markers." *Hum Biol.* 2004 Jun; 76(3):413-29.

- Dipierri, J. E. 2004. "Filiación e historia cultural: confluencias y divergencias temáticas". Cuadernos de la Facultad de Humanidades y Ciencias Sociales, Julio, número 024. Universidad de Jujuy. San Salvador de Jujuy, Argentina. Pp 63-104.
- Dipierri, J. E., Ocampo y Russo. 1991. "An estimation of inbreeding from isonymy in the historical (1734-1810) population of the Quebrada de Humahuaca (Jujuy, Argentina)." J Biosoc Sci. 1991 Jan; 23(1):23-31.
- Dipierri, J. E. y Alfaro E. 1996. "Isolation and sedentary life indicators in populations of highland communities in the Jujuy Province (República Argentina)". Acta Cient. Venez. 1996; 47(3):178-84.
- Dipierri, J. E., Alfaro E., Martinez-Marignac V. L., Baillet G., Bravi C. M., Cejas S., Bianchi, N. O. 1998. "Paternal directional mating in two amerindian subpopulations from the northwest of Argentina". Hum. Biol. 70 (6): 1001-1010.
- Dipierri, J. E., E. L. Alfaro, Scapoli C., Mamolini E., Rodríguez-Larralde A., Barraí I. 2005a. "Surnames in Argentina: A Population Study Through Isonimy". Am. J. Phys. Anthropol. Sep; 128(1): 199-209.
- Dipierri, J. E., E. L. Alfaro, Rodríguez Larralde A., 2005b. "Estructura isonímica del Noroeste Argentino". Revista de la Sociedad Argentina de Genética, Journal of Basic & Applied Genetics. Actas XXXIV Congreso Argentino de Genética. Vol. XVII (Suplemento). Septiembre.
- Fejerman, L., F. R. Carnese, Goicochea, A. S., Avena S. A., Dejean C. B. y Ward R. H. 2005. "African ancestry of the population of Buenos Aires." Am. J. Phys. Anthropol. 2005 Sep; 128(1):164-70.
- Gil Montero, R. 2007. "¿Métodos, modelos y sistemas familiares o historias de familia?". En: Familia y Diversidad en América Latina. Estudios de casos. Robinchaux D. CLACSO, Consejo Latinoamericano de Ciencias Sociales, Buenos Aires, ISBN 978-987-1183-74-6.
- Gill, P., P. L. Ivanov, Kimpton C., Piercy R., Benson N., Tully G., Evett I., Hagelberg E., Sullivan K. 1994. "Identification of the remains of the Romanov family by DNA analysis". Nature Genetics vol. 6 Feb. 130-135.
- Gómez A., Ávila S. J., Briceño I. 2008 "De genotipos e isonimias: análisis de correlación entre el apellido y el patrimonio genético

- heredado en el cromosoma Y en la población de tres departamentos del suroccidente colombiano”. *Biomédica* 28: 357-70.
- Goff, P. G., Athey T. W. 2006. “Diagnostic STR values for haplogroup G”. *Journal of Genetic Genealogy*, 2(1):12-17.
 - Greenhalgh, T. 1997. “How to read a paper: Papers that report diagnostic or screening tests”. *BMJ*; 315:540-543.
 - Gusmao, L., Sánchez-Diz P., Alves C., Beleza S., Lopes A., Cariacedo A., Amorim A. 2003. “Grouping of Y-STR haplotypes discloses European geographic clines”. *Forensic Science International* 134: 172-179.
 - Immel, U. D., Krawczak M., Üdolph J., Richter A., Rodig H., Kleiber M., Klintschar M. 2006. “Y-chromosomal STR haplotype analysis reveals surname-associated strata in the East-German population”. *Europ. J. Hum. Genet.* 1-6.
 - Jin, B., Xiao J., Akey J., Lu D., Shen D. Chakraborty R., Tan J. 1999. “Chinese surnames are polyphyletic in origin: Evidences based on 19 Y.SNPs. *Am. J. Hum. Gen.* 65: 1136, Supp.
 - Jobling, M. A. y C. Tyler-Smith. 2000. “New uses for new haplotypes: the human Y chromosome, disease and selection. *Trends Genet* 16: 356-362.
 - Jobling, M. A. 2001. “In the name of the father: surnames and genetics”. *Trends Genet.* 17: 353-7.
 - Jobling, M. A., M. Hurles, C. Tyler-Smith. 2003. Human Evolutionary Genetics: Origins, Peoples and Disease. Garland Science.
 - Karafet, T. M., Zegura S. L., Posukh O., Osipova L., Bergen A., Long J., Goldman, D., Klitz W., Harihara S., de Kniff P., Wiebe V., Griffiths R. C., Templeton A. R., Hammer M. F. 1999. “Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes”. *Am. J. Hum. Genet.* 64: 817-831.
 - Karafet, T. M. Osipova L. P., Gubina M. A., Posukh O.L., Zegura S. L., Hammer, M. F. 2002. “High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of boreal hunter-gatherer way of life.” *Human Biology* 74: 761-789.

- 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 the resolution of the Human Y chromosomal haplogroup tree". *Genome Research* 18: 830-838.
- Kayser, M., A. Caglià, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccinini, M. Prinz, C. Schmitt, P. M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, P. de Knijff, L. Roewer. 1997. "Evaluation of Y-chromosomal STRs: a multicenter study". *Int. J Legal Med.* 110 : 125-133.
- King, T.E., Ballereau S.J., Schurer K.E., Jobling M.A. 2006. "Genetic signatures of coancestry within surnames". *Current Biology*; 16 (4): 384-8.
- King, T. E., Parkin E. J., Swinfield G., Cruciani F., Scozzari R., Rosa A., Lim S. K., Xue Y., Tyler-Smith C., Jobling M. A. 2007. "Africans in Yorkshire? The deepest-rooting clade of the Y phylogeny within an english genealogy". *European Journal of Human Genetics* 15: 288-293.
- King, T. E., Jobling M. A. 2009. "Founders, Drift and Infidelity: The Relationship between Y Chromosome Diversity and Patrilineal Surnames". *Mol. Biol. Evol.* 26, 1093-1102.
- Kim, Y. J., Paik S. G., Ahn G. S., Kim W. 1999. "49a/TaqI Haplotypes According to Surname Groups in Korean Population". *Korean J. Genet.* 21, 181-192.
- Kimura, M., Ohta T. 1978. "Stepwise mutation model and distribution of allele frequencies in a finite population". *Proc. Natl. Acad. Sci. USA*, 75 (6): 2868-2872.
- Kittles, A. R., Markus P., Peltonen L., Bergen A. W., Aragon R. A., Virkkunen M., Linnoila M., Goldman D., Long J. C. 1998. "Dual origins of Finns Revealed by Y Chromosome Haplotype Variation". *Am. J. Hum. Genet.* 62: 1171-1179.
- Lell, J. T., Brown M. D., Schurr T. G., Sukernik, R. I., Starikovskaya Y. B., Torroni A., Moore L. G., Troup, G. M. Wallace D. C. 1997. "Y

- chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes.” *Hum Genet.* 100 (5-6): 536-43.
- Luis, J. R., Rowold D. J., Regueiro M., Caeiro B., Cinnioglu C., Roseman C., Underhill P. A., Cavalli Sforza L. L., Herrera J. R. 2004. “The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations”. *Am J Hum Genet.* ;74(4):788.
 - Malhi R. S., González-Oliver A., Schroeder K. B. y col. 2008. “Distribution of Y Chromosome Among Native North Americans: a study of Athapaskan Population History”. *Am. J. of Phys. Anthropol.* 137 84): 421- 424.
 - Martínez-Sarasola C. 1999. Nuestros paisanos los indios. Ed. Emecé. Buenos Aires, Argentina.
 - McAlister, F. A., Straus S. E., Sackett D.L. 1999. Why we need large, simple studies of the clinical examination: the problem and a proposed solution. *Lancet* 354: 1721-24.
 - McEvoy, B., Brady C., Moore L. T., Bradley D. G. 2006. “The scale and nature of Viking settlement in Ireland from Y-chromosome admixture analysis.” *European Journal of Human Genetics* 14, 1288-1294.
 - Malhi, R. S., González-Oliver A., Schroeder K. B., Kemp B. M., Greenberg J. A., Dobrowski S. Z., Smith D. G., Resendez B. M., Karafet T., Hammer M., Zegura S., Brovko T. 2008. “Distribution of Y Chromosome Among Native North Americans: a study of Athapaskan Population History”. *American Journal of Physical Anthropology* 000: 000-000.
 - Manni, F., Toupance B., Sabbagh A., Heyer E. 2005. “New method for surname studies of ancient patrilineal population structures, and possible application to improvement of Y-chromosome sampling.” *Am. J. Phys. Anthropol.* 2005 Feb; 126(2):214-28.
 - Manrubia, S. C., Zanette D. H. 2007. “ At the boundary between biological and cultural evolution: The origin of surname distributions.” [arXiv:cond-mat/0201559v1 \[cond-mat.stat-mech\]](https://arxiv.org/abs/cond-mat/0201559v1).

- Martínez Marignac, V., Bertoni B., Parra E. J., Bianchi N. O. 2004 “Characterization of Admixture in an Urban Sample from Buenos Aires, Argentina, Using Uniparentally and Biparentally Inherited Genetic Markers”. *Human Biology*; 4: 543-557.
- Mikerezi, I., Pizzetti P., Lucchetti E., Ekonomi M. 2003. “Isonymy and the genetic structure of Albanian populations.” *Coll. Antropol. Dec*; 27(2):507-14.
- Mertens, G. 2007. “Y-Haplogroup frequencies in the Flemish Population”. *Journal of Genetic Genealogy*, 3(2):19-25.
- Muzzio, M., Ramallo V. Alfaro, E. L., Dipierri, J. E., Bianchi N. O., Bailliet G. 2006. “Correlación entre apellidos y linajes del cromosoma Y”. *Journal of Basic and Applied Genetics (Formerly Mendeliana)* ISSN: 1666-0390. 2007.
- Muzzio, M., Alfaro E. L., Dipierri J. E., Bianchi N. O, Bailliet G. 2007a. “Origen de los linajes paternos del NOA”. *Journal of Basic and Applied Genetics (Formerly Mendeliana)* ISSN: 1666-0390. 2007.
- Muzzio, M., Alfaro E. L., Dipierri J. E., Motti J. M. B., Bianchi N. O, Bailliet G. 2007b. “Linajes Antroponímicos y Genéticos en el NOA: Concordancias y Discrepancias”. *Revista Argentina de Antropología Biológica*. ISSN: 1514-7991. Volumen 9, Número 1. pg. 137.
- Muzzio, M., Ramallo V., Alfaro E. L., Dipierri J. E., Bailliet G. 2008a. “Herencia patrilineal en el NOA: los apellidos y el cromosoma Y.” X Congreso de la Asociación Latinoamericana de Antropología Biológica. La Plata, Argentina.
- Muzzio, M., Ramallo V., Dipierri J. E., Alfaro E. L., Motti J. M. B., Bianchi N. O., Bailliet G. 2008b. “Apellidos y linajes del cromosoma Y, dos huellas borrosas”. X Congreso de la Asociación Latinoamericana de Antropología Biológica. La Plata, Argentina.
- Muzzio, M., Ramallo V., Dipierri J. E., Alfaro E. L., Motti J. M. B., Salceda S., Bianchi N. O., Bailliet G. 2008c. “Patrones de similitud y diferenciación de linajes masculinos nativo-americanos en el Norte Argentino.” X Congreso de la Asociación Latinoamericana de Antropología Biológica. La Plata, Argentina.

- Ohta, T., Kimura M. 1973 “A model of mutation appropriate to estimate the number of electrophoretically detectable molecules in a finite population”. *Genet. Res.* 22, 201-204.
- Pascali, V. L., Dobosz M., Brinkmann B. 1999. “Coordinating Y-chromosomal STR research for the Courts”. *Int J Legal Med* 112(1), 1.
- Press, W. H., Teukolsky S. A., Vetterling W.T., Flannery B. P. 1994. Numerical Recipes in FORTRAN. The art of Scientific Computing. 2nd ed. Cambridge University Press.
- Qamar, R., Ayub Q., Mahyuddin A., Helgason A., Mazhar K, Mansoor A., Zerjal T., Tyler-Smith C., Mehdi S. Q. 2002. “Y-Chromosomal DNA Variation in Pakistan”. *Am. J. Hum. Genet.* 70: 1107-1124.
- Ramallo, V., Alfaro E. L., Dipierri J. E., Bianchi N. O., Bailliet G. 2005. “Caracterización de linajes paternos en muestras provenientes de tres provincias del NOA”. *Revista de la Sociedad Argentina de Genética, Journal or Basic & Applied Genetics. Actas XXXIV Congreso Argentino de Genética. Volumen XVII (Suplemento). Septiembre.*
- Robinchaux, D. 2007. “Sistemas familiares en culturas subalternas de América Latina: una propuesta conceptual y un bosquejo preliminar.” En: Familia y Diversidad en América Latina. Estudios de casos. Robinchaux D. CLACSO, Consejo Latinoamericano de Ciencias Sociales, Buenos Aires, ISBN 978-987-1183-74-6.
- Salas, A., Jaime J. C., Álvarez-Iglesias V., Carracedo Á. 2008. “Gender bias in the multiethnic genetic composition of central Argentina”. *Journal of Human Genetics.* 53: 662-674.
- Schlecht, J., Kaplan M. E., Barnard K., Karafet T., Hammer M. F., Merchant N. C. 2008. “Machine-Learning approaches for classifying Haplogroup from Y Chromosome STR data”. *PloS Comput Biol* 4(6): e1000093. doi:10.1371/journal.pcbi1000093. Disponible en: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2396484/>
- Schneider, S. y Excoffier L. 1999. “Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA”. *Genetics* 152: 1079-1089.

- Soodyall, H., Nebel A., Morar B., Jenkins T. 2003. "Genealogy and genes: tracing the founding fathers of Tristan da Cunha." *Eur. J. Hum. Genet.* Sep; 11(9): 705-9.
- Sykes, B. e Irven C. 2000. "Surnames and the Y Chromosome". *Am. J. Hu. Genet.* 66: 1417-1419.
- Thomas, M. Parfitt T., Weiss D., Skorecki K., Wilson M., le Roux M., Bradman N., Goldstein D. 2000. "Y Chromosomes Traveling South: The Cohen Modeal Haplotype and the Origins of the Lemba - the "Black Jews of Southern Africa". *Am. J. Hum. Gen.* 66: 674-686.
- Trumme, T., Herrmann B., Hummel S. 2004. "Genetics in genealogical research- reconstruction of a family tree by means of Y-haplotyping." *Anthropol Anz.*; 62(4):379-86.
- Underhill, P., Jin L, Zemans R., Oefner P. J., Cavalli-Sforza L. L. 1996. "A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history". *Proc. Natl. Acad. Sci. USA* 93: 196-200.
- Underhill, P., Passarino G., Lin A. A., Shen P., Mirazón Lahr M., Foley R. A. , Oefner P. J., Cavalli-Sforza L. 2001. "The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations". *Annals of Human Genetics* 65: 43-62.
- Verdon J. 2008. L'amour au Moyen Age. La chair, le sexe et le sentiment. Pour L'Histoire. Perrin.
- Walpole, R. E. y Myers R. H. 1992. Probabilidad y Estadística. Cuarta edición. McGraw-Hill.
- Willuweit, S., Roewer L., on behalf of the International Forensic Y Chromosome User Group. 2007. "Y chromosome haplotype reference database (YHRD): Update". *Forensic Science International: Genetics* 1(2) 83-87. Disponible en: <http://www.yhrd.org>
- Y Chromosome Consortium. 2002. "A nomenclature system for the tree of human Y-chromosome binary haplogroups" *Genome Research.* 12, 339-348.

- Yuan, Y. D., Zhang C., Ma Q. Y., Yang H. M. 2000. "Population genetics of Chinese surnames. I. Surname frequency distribution and genetic diversity in Chinese." *Yi Chuan Xue Bao.* 2000; 27(6):471-6.
- Zalloua P A, Xue Y, Khalife J 2008. "Y-chromosomal Diversity in Lebanon is structured by recent historical events". *Am. J. Hum. Genet.*, 82 (4): 873-882.
- 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 Evol.* 21: 164-175.
- Zerjal T. et al. 2003. "The genetic legacy of the Mongols." *Am. J. Hum. Genet.* 72, 717-721.
- Zukerkandl, E. y Pauling L. 1965. "Evolutionary divergence and convergence in proteins". En: *Evolving Genes and Proteins* (eds V Bryson, HJ Vogel). Academic Press, New York, pp. 97-166.

Anexo

Índice del Anexo

| | |
|--|-----|
| Cartilla de explicación sobre la toma de muestra, investigación y almacenamiento del material genético..... | 107 |
| Formulario de consentimiento informado sobre la toma de muestra, investigación y almacenamiento del material genético..... | 108 |
| Tabla 1: Cantidad de votantes portadores de los apellidos en las provincias de Jujuy, Salta, Tucumán y Catamarca..... | 111 |
| Microsatélites utilizados..... | 113 |
| Tabla 2: Tamaño de los productos de PCR de los alelos observados para DYS19 | 112 |
| Tabla 3: Tamaño de los productos de PCR de los alelos observados para DYS389I..... | 113 |
| Tabla 4: Tamaño de los productos de PCR de los alelos observados para DYS389II..... | 118 |
| Tabla 5: Tamaño de los productos de PCR de los alelos observados para DYS390 | 116 |
| Tabla 6: Tamaño de los productos de PCR de los alelos observados para DYS391 | 117 |
| Tabla 7: Tamaño de los productos de PCR de los alelos observados para DYS392 | 126 |
| Tabla 8: Tamaño de los productos de PCR de los alelos observados para DYS 393..... | 119 |
| Encuesta genealógica para donantes voluntarios..... | 121 |
| Comparación de las frecuencias relativas obtenidas para cada marcador con aquellas de Argentina y el mundo | 122 |
| Figura 1: Frecuencias relativas de los alelos de DYS 19..... | 122 |
| Figura 2: Frecuencias relativas de los alelos de DYS 389I..... | 122 |
| Figura 3: Frecuencias relativas de los alelos de DYS 389II | 123 |
| Figura 4: Frecuencias relativas de los alelos de DYS 390 | 123 |
| Figura 5: Frecuencias relativas de los alelos de DYS 391 | 123 |
| Figura 6: Frecuencias relativas de los alelos de DYS 392 | 124 |
| Figura 7: Frecuencias relativas de los alelos de DYS 393 | 124 |
| Software for Y-haplogroup predictions: a word of caution | 125 |
| A method for assignment of the weight of characters | 131 |

CARTILLA DE EXPLICACIÓN SOBRE LA TOMA DE MUESTRA, INVESTIGACIÓN Y ALMACENAMIENTO DEL MATERIAL GENÉTICO.

El Dr. José E. Dipierri y la Lic. Emma Alfaro, del Instituto de Biología de la Altura (Universidad Nacional de Jujuy), la doctora Graciela Bailliet, investigadora del Consejo Nacional de Investigación Científica y Tecnológica (CONICET) y la licenciada Marina Muzzio, del Instituto Multidisciplinario de Biología Celular (IMBICE), lo invitan a Usted a participar en el proyecto **“El cromosoma Y, aplicaciones al estudio de poblaciones humanas, genealogías e isonimia”**.

Los apellidos y las genealogías establecidas a partir de ellos han tenido una gran importancia en múltiples culturas a lo largo de su historia. Generalmente se transmiten por una única línea, por ejemplo la paterna, permitiendo asumir que los individuos con un mismo apellido deben presentar relaciones de parentesco genético, es decir, que provienen de un mismo linaje con un antecedente común.

Toda nuestra información biológica está almacenada dentro de nuestras células, escrita en un código especial: el ADN. Recibimos esta información de nuestros padres, mitad de nuestra madre y mitad de nuestro padre. Existen dos excepciones para este modo de transmisión, una sigue la línea materna de ancestría (el ADN mitocondrial) y la otra se pasa de padre a hijo varón (el ADN del cromosoma Y). Cualquier diferencia dentro de estos tipos de ADN sólo puede atribuirse a las mutaciones, permitiendo a los investigadores develar desde estructuras y flujo génico al origen de la humanidad.

¿Por qué el Y?

Al heredarse por la línea paterna, su distribución debería empalmarse a la distribución de los apellidos en aquellas sociedades donde el linaje social sigue esa línea. Cada configuración de varias características diferentes de ese cromosoma debería corresponderse y restringirse a un apellido específico. Cualquier diferencia en este patrón se atribuye a procesos de transmisión irregular del apellido (adopciones, concepciones extramaritales, adjudicación del apellido materno) o a procesos de disrupción étnica, como la llamada “Conquista” que incluyó bautismos forzados de las poblaciones nativas, momento en el cual se les adjudicó arbitrariamente un apellido.

De esta forma resulta interesante cruzar estas dos fuentes de información: la de los apellidos y la del Y, para saber la silenciosa historia de lo que la gente realmente hizo, en lugar de acotarnos a los registros oficiales, siendo el evaluar la coherencia entre ambas el objetivo de este trabajo.

El trabajo consiste en el registro de datos genealógicos y la toma de muestras de sangre para la extracción de ADN (Acido DesoxirriboNucleico), ya que en la sangre de cada persona encontramos material genético. Lo/a invitamos a participar en esta investigación como donante de sangre, sin ningún riesgo para su salud ni costo económico alguno.

Las muestras sólo serán destinadas a este estudio y almacenadas a tal fin en el banco de ADN del Instituto Multidisciplinario de Biología Celular (IMBICE). La identidad de los donantes nunca será revelada, para preservar el anonimato las muestras se identificarán con un código y sólo personas autorizadas (Drs. Bailliet y Dipierri, Lic. Alfaro y Muzzio) del equipo tendrán acceso a los resultados.

Usted puede retirarse de este estudio: 1) cuando quiera; 2) sin tener que dar explicaciones. En caso que lo solicite, podemos comunicarle los resultados sin costo alguno. Su participación es voluntaria y limitada en forma exclusiva a los estudios detallados previamente.

Si desea colaborar firme el Término de Consentimiento abajo, después de leerlo detenidamente y aclarar todas sus dudas con la persona que la/o está atendiendo.

Para obtener cualquier información, evacuar dudas o solicitar su exclusión del proyecto como donante, por favor, comuníquese con:

Doctora Graciela Bailliet

DNI 14.739.871

Laboratorio de Genética Molecular Poblacional

Instituto Multidisciplinario de Biología Celular

526 e/10 y 11-1900 La Plata

Teléfono: 0221 – 421-0112

gbailliet@imbice.org.ar

Doctor Dipierri, José Edgardo

L.E. N° 7.628.792

Instituto de Biología de la Altura

Universidad Nacional de Jujuy

Av. Bolivia1661

(4600) San Salvador de Jujuy, Argentina

Teléfono: 0388- 422-1596

dipierri@inbial.unju.edu.ar

.....
Nombre y Firma del investigador responsable DNI

.....
Nombre y Firma de testigo DNI

Lugar y fecha

Tabla 1: Cantidad de votantes portadores de los apellidos en las provincias Jujuy, Salta, Tucumán y Catamarca. Fuente: Dipierri y col. 2005a.

| Apellido | Origen | N° de individuos | N° de votantes | | | | Total |
|-----------|---------|------------------|----------------|-------|---------|-----------|-------|
| | | | Jujuy | Salta | Tucumán | Catamarca | |
| Únicos | | 52 | | | | | |
| Arteaga | vasco | 3 | 25 | 45 | 4 | 2 | 76 |
| Bazan | español | 2 | 58 | 160 | 574 | 671 | 1463 |
| Campos | español | 2 | 159 | 454 | 767 | 134 | 1514 |
| Carabajal | español | 2 | 78 | 214 | 721 | 104 | 1117 |
| Carrizo | español | 2 | 320 | 1009 | 1873 | 1571 | 4773 |
| Chumacero | español | 2 | 18 | 1 | 5 | 0 | 24 |
| Cobacho | español | 2 | 0 | 0 | 7 | 72 | 79 |
| Condori | nativo | 2 | 655 | 567 | 145 | 27 | 1394 |
| Cruz | español | 9 | 2917 | 3264 | 915 | 109 | 7205 |
| Delgado | español | 2 | 240 | 863 | 599 | 265 | 1967 |
| Díaz | español | 5 | 1129 | 3162 | 6328 | 1275 | 11894 |
| Fernández | español | 4 | 1197 | 2208 | 3346 | 378 | 7129 |
| Flores | español | 6 | 2389 | 3751 | 1275 | 252 | 7667 |
| García | español | 3 | 754 | 1836 | 2876 | 268 | 5734 |
| Gómez | español | 2 | 591 | 1923 | 4371 | 777 | 7662 |
| Gonzalez | español | 4 | 1171 | 3003 | 5635 | 803 | 10612 |
| Gregorio | español | 2 | 76 | 11 | 4 | 5 | 96 |
| Gutierrez | español | 3 | 1590 | 2730 | 1516 | 365 | 6201 |
| Herrera | español | 2 | 474 | 1402 | 2845 | 1359 | 6080 |
| Leiva | español | 4 | 31 | 122 | 583 | 691 | 1427 |
| López | español | 5 | 1580 | 6170 | 3863 | 683 | 12296 |
| Martínez | español | 4 | 1703 | 3123 | 2425 | 655 | 7906 |
| Mercado | español | 4 | 178 | 507 | 460 | 276 | 1421 |
| Morales | español | 3 | 560 | 1643 | 1127 | 298 | 3628 |
| Moreno | español | 3 | 303 | 809 | 1436 | 627 | 3175 |
| Moyano | español | 2 | 40 | 147 | 725 | 51 | 963 |
| Navarro | español | 3 | 373 | 471 | 1627 | 383 | 2854 |
| Pacheco | español | 3 | 56 | 275 | 518 | 301 | 1150 |
| Pastrana | español | 2 | 27 | 931 | 171 | 10 | 1139 |
| Paz | español | 2 | 178 | 698 | 2005 | 128 | 3009 |
| Perez | español | 3 | 775 | 1838 | 3772 | 477 | 6862 |
| Preli/s | foráneo | 2 | 0 | 0 | 0 | 14 | 14 |
| Quipildor | nativo | 1 | 264 | 705 | 102 | 21 | 1092 |

| | | | | | | | |
|-----------|---------|---|------|------|------|------|-------|
| Quiroga | español | 3 | 461 | 871 | 886 | 474 | 2692 |
| Rios | español | 2 | 666 | 1519 | 673 | 244 | 3102 |
| Rodriguez | español | 3 | 1500 | 3694 | 5024 | 1245 | 11463 |
| Romero | español | 3 | 788 | 1762 | 1581 | 1152 | 5283 |
| Soria | español | 2 | 235 | 658 | 2248 | 765 | 3906 |
| Sosa | español | 2 | 416 | 816 | 1688 | 719 | 3639 |
| Toledo | español | 2 | 97 | 356 | 969 | 276 | 1698 |
| Varela | español | 2 | 108 | 168 | 524 | 576 | 1376 |
| Vega | español | 2 | 319 | 611 | 923 | 1182 | 3035 |
| Vergara | español | 2 | 55 | 58 | 199 | 366 | 678 |
| Vilte | nativo | 2 | 830 | 1195 | 44 | 16 | 2085 |

Microsatélites analizados:

DYS19:

Repetido: [TAGA]₃tagg[TAGA]_n = "GenBank top strand"

Tabla 2: Tamaño de los productos de PCR de los alelos observados para **DYS19**

| Alelo (N° de repetidos) | Estructura de repetidos |
|-------------------------------|--|
| 10 | [TAGA] ₃ tagg[TAGA] ₇ |
| 11 | [TAGA] ₃ tagg[TAGA] ₈ |
| 12 | [TAGA] ₃ tagg[TAGA] ₉ |
| 13 | [TAGA] ₃ tagg[TAGA] ₁₀ |
| 14 | [TAGA] ₃ tagg[TAGA] ₁₁ |
| 15 | [TAGA] ₃ tagg[TAGA] ₁₂ |
| 16 | [TAGA] ₃ tagg[TAGA] ₁₃ |

| | |
|----|--|
| 17 | [TAGA] ₃ tagg[TAGA] ₁₄ |
| 18 | [TAGA] ₃ tagg[TAGA] ₁₅ |
| 19 | 269 pares de bases |

Fuente: http://www.cstl.nist.gov/biotech/strbase/str_y19.htm

DYS389I/II:

Repetido: **DYS389I:** [TCTG]_q [TCTA]_r = “GenBank top strand”

DYS389II: [TCTG]_n[TCTA]_p[TCTG]_q [TCTA]_r = “GenBank top strand”

Tabla 12: Tamaño de los productos de PCR de los alelos observados para **DYS389 I y II**

| Alelo (N° de repetidos) | Estructura de repetidos |
|----------------------------|--|
| DYS 389 I | |
| 9(6)* | [TCTG] ₃ [TCTA] ₆ |
| 10(7) | [TCTG] ₃ [TCTA] ₇ |
| 11(8) | [TCTG] ₃ [TCTA] ₈ |
| 12(9) | [TCTG] ₃ [TCTA] ₉ |
| 13(10) | [TCTG] ₃ [TCTA] ₁₀ |
| 14(11) | [TCTG] ₃ [TCTA] ₁₁ |
| 15(12) | [TCTG] ₃ [TCTA] ₁₂ |

| | |
|------------|---|
| 16(13) | [TCTG] ₃ [TCTA] ₁₃ |
| 17(14) | |
| DYS 389 II | |
| 24(21) | |
| 25 (22) | [TCTG] ₄ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₈ |
| 26(23) | [TCTG] ₄ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₈ |
| 26(23) | [TCTG] ₅ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₈ |
| 26(23) | [TCTG] ₄ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₉ |
| 27(24) | [TCTG] ₄ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₈ |
| 27(24) | [TCTG] ₅ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₈ |
| 27(24) | [TCTG] ₄ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₉ |
| 27(24) | [TCTG] ₅ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₉ |
| 27(24) | [TCTG] ₄ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₁₀ |
| 28(25) | [TCTG] ₄ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₈ |
| 28(25) | [TCTG] ₅ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₈ |
| 28(25) | [TCTG] ₄ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₉ |
| 28(25) | [TCTG] ₅ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₉ |
| 28(25) | [TCTG] ₄ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₁₀ |
| 28(25) | [TCTG] ₅ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₁₀ |
| 28(25) | [TCTG] ₄ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₁₁ |

| | |
|--------|---|
| 29(26) | [TCTG] ₄ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₈ |
| 29(26) | [TCTG] ₅ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₈ |
| 29(26) | [TCTG] ₄ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₉ |
| 29(26) | [TCTG] ₅ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₉ |
| 29(26) | [TCTG] ₄ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₁₀ |
| 29(26) | [TCTG] ₅ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₁₀ |
| 29(26) | [TCTG] ₄ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₁₁ |
| 29(26) | [TCTG] ₅ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₁₁ |
| 29(26) | [TCTG] ₄ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₁₂ |
| 30(27) | [TCTG] ₅ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₈ |
| 30(27) | [TCTG] ₄ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₉ |
| 30(27) | [TCTG] ₅ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₉ |
| 30(27) | [TCTG] ₄ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₁₀ |
| 30(27) | [TCTG] ₅ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₁₀ |
| 30(27) | [TCTG] ₄ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₁₁ |
| 30(27) | [TCTG] ₅ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₁₁ |
| 30(27) | [TCTG] ₄ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₁₂ |
| 30(27) | [TCTG] ₅ [TCTA] ₁₀ [TCTG] ₃ [TCTA] ₁₂ |
| 31(28) | [TCTG] ₅ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₉ |
| 31(28) | [TCTG] ₄ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₁₀ |

| | |
|--------|---|
| 31(28) | [TCTG] ₅ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₁₀ |
| 31(28) | [TCTG] ₄ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₁₁ |
| 31(28) | [TCTG] ₅ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₁₁ |
| 31(28) | [TCTG] ₄ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₁₂ |
| 31(28) | [TCTG] ₅ [TCTA] ₁₁ [TCTG] ₃ [TCTA] ₁₂ |
| 32(29) | [TCTG] ₅ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₁₀ |
| 32(29) | [TCTG] ₄ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₁₁ |
| 32(29) | [TCTG] ₅ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₁₁ |
| 32(29) | [TCTG] ₄ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₁₂ |
| 32(29) | [TCTG] ₅ [TCTA] ₁₂ [TCTG] ₃ [TCTA] ₁₂ |
| 33(30) | [TCTG] ₅ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₁₁ |
| 33(30) | [TCTG] ₄ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₁₂ |
| 33(30) | [TCTG] ₅ [TCTA] ₁₃ [TCTG] ₃ [TCTA] ₁₂ |
| 34(31) | [TCTG] ₅ [TCTA] ₁₄ [TCTG] ₃ [TCTA] ₁₂ |

Fuente: http://www.cstl.nist.gov/biotech/strbase/str_y389.htm

DYS390:

Repetido: [TCTG]_n [TCTA]_m[TCTG]_p[TCTA]_q = “GenBank top strand”

Tabla 13: Tamaño de los productos de PCR de los alelos observados para
DYS390

| Alelo (N° de repetidos) | Estructura de repetidos |
|-------------------------------|--|
| 17 | |
| 18 | |
| 19 | |
| 20 | |
| 21 | [TCTG] ₈ [TCTA] ₅ ACTA[TCTA] ₂ [TCTG] ₁ [TCTA] ₄ |
| 22 | [TCTG] ₈ [TCTA] ₉ [TCTG] ₁ [TCTA] ₄ |
| 23 | [TCTG] ₈ [TCTA] ₁₀ [TCTG] ₁ [TCTA] ₄ |
| 24 | [TCTG] ₈ [TCTA] ₁₁ [TCTG] ₁ [TCTA] ₄ |
| 25 | [TCTG] ₈ [TCTA] ₁₂ [TCTG] ₁ [TCTA] ₄ |
| 26 | |
| 27 | [TCTG] ₈ [TCTA] ₁₄ [TCTG] ₁ [TCTA] ₄ |
| 28 | |

Fuente: http://www.cstl.nist.gov/biotech/strbase/str_y390.htm

DYS391:

Repetido: [TCTA]_n = “GenBank top strand”

Tabla 14: Tamaño de los productos de PCR de los alelos observados para
DYS391

| Alelo (N° de repetidos) | Estructura de repetidos |
|-------------------------------|-------------------------|
| 6 | [TCTA] ₆ |
| 7 | [TCTA] ₇ |
| 8 | [TCTA] ₈ |
| 9 | [TCTA] ₉ |
| 10 | [TCTA] ₁₀ |
| 11 | [TCTA] ₁₁ |
| 12 | [TCTA] ₁₂ |
| 13 | [TCTA] ₁₃ |
| 14 | [TCTA] ₁₄ |

Fuente: http://www.cstl.nist.gov/biotech/strbase/str_y391.htm

DYS392:

Repetido: [TAT]_n = "GenBank top strand"

Tabla 15: Tamaño de los productos de PCR de los alelos observados para
DYS392

| Alelo (N° de repetidos) | Estructura de repetidos |
|-------------------------------|----------------------------|
| 6 | [TAT] ₆ |

| | |
|----|---------------------|
| 7 | [TAT] ₇ |
| 8 | [TAT] ₈ |
| 10 | [TAT] ₁₀ |
| 11 | [TAT] ₁₁ |
| 12 | [TAT] ₁₂ |
| 13 | [TAT] ₁₃ |
| 14 | [TAT] ₁₄ |
| 15 | [TAT] ₁₅ |
| 16 | [TAT] ₁₆ |
| 17 | [TAT] ₁₇ |

Fuente: http://www.cstl.nist.gov/biotech/strbase/str_y392.htm

DYS393:

Repetido: [AGAT]_n = "GenBank top strand"

Tabla 16: Tamaño de los productos de PCR de los alelos observados para DYS 393

| Alelo (N° de repetidos) | Estructura de repetidos |
|-------------------------------|-------------------------|
| 9 | [AGAT] ₉ |
| 10 | [AGAT] ₁₀ |

| | |
|----|----------------------|
| 11 | [AGAT] ₁₁ |
| 12 | [AGAT] ₁₂ |
| 13 | [AGAT] ₁₃ |
| 14 | [AGAT] ₁₄ |
| 15 | [AGAT] ₁₅ |
| 16 | [AGAT] ₁₆ |
| 17 | [AGAT] ₁₇ |

Fuente: http://www.cstl.nist.gov/biotech/strbase/str_y393.htm

ENCUESTA GENEALÓGICA PARA DONANTES VOLUNTARIOS

DONANTE:

.....
NOMBRE Y APELLIDO LUGAR DE NACIMIENTO

LÍNEA MATERNA:

.....
NOMBRE Y APELLIDO de la madre LUGAR DE NACIMIENTO

.....
NOMBRE Y APELLIDO de la abuela materna LUGAR DE NACIMIENTO

.....
NOMBRE Y APELLIDO del abuelo materno LUGAR DE NACIMIENTO

LÍNEA PATERNA:

.....
NOMBRE Y APELLIDO del padre LUGAR DE NACIMIENTO

.....
NOMBRE Y APELLIDO de la abuela paterna LUGAR DE NACIMIENTO

.....
NOMBRE Y APELLIDO del abuelo paterno LUGAR DE NACIMIENTO

OTROS DATOS:

Comparación de las frecuencias relativas obtenidas para cada marcador con aquellas de Argentina y el mundo

A continuación, se comparan, marcador por marcador, las frecuencias relativas de los diferentes alelos, entre la muestra de la presente tesis, Argentina y el mundo. La fuente de las frecuencias Argentinas fue la base de datos del Servicio de Huellas Digitales Genéticas, de la Facultad de Farmacia y Bioquímica de la U.B.A., disponible en internet en: <http://www.ffyb.uba.ar/marcadores/referencias.asp#>

Los datos de las frecuencias relativas a nivel mundial se obtuvieron de la base de datos internacional YHRD, disponible en: <http://www.yhrd.org>

Figura 1: Frecuencias relativas de los alelos de DYS 19

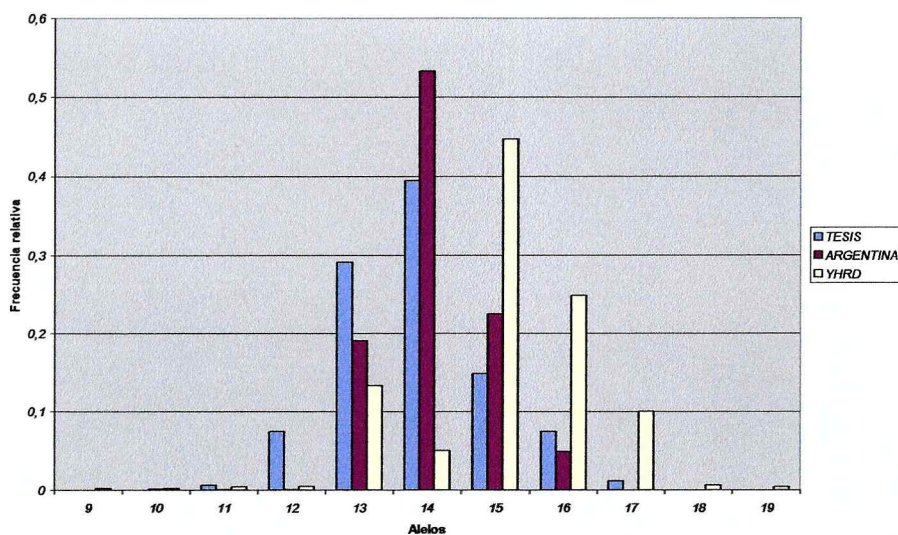


Fig. 2: Frecuencias relativas de los alelos de DYS 389 I

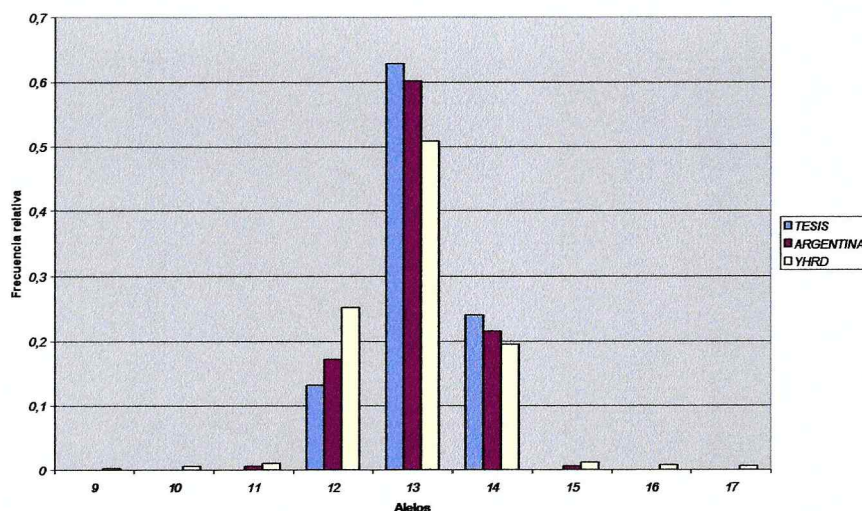


Figura 3: Frecuencias relativas de los alelos de DYS 389 II

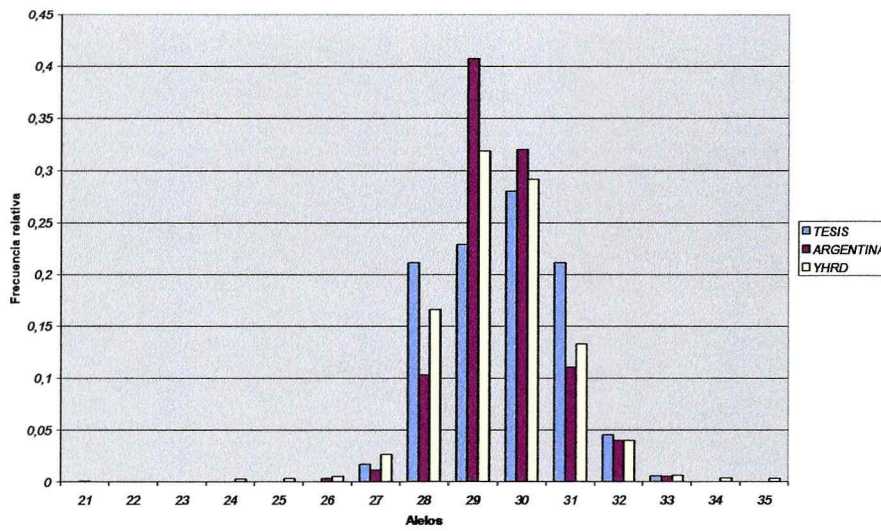


Figura 4: Frecuencias relativas de los alelos de DYS 390

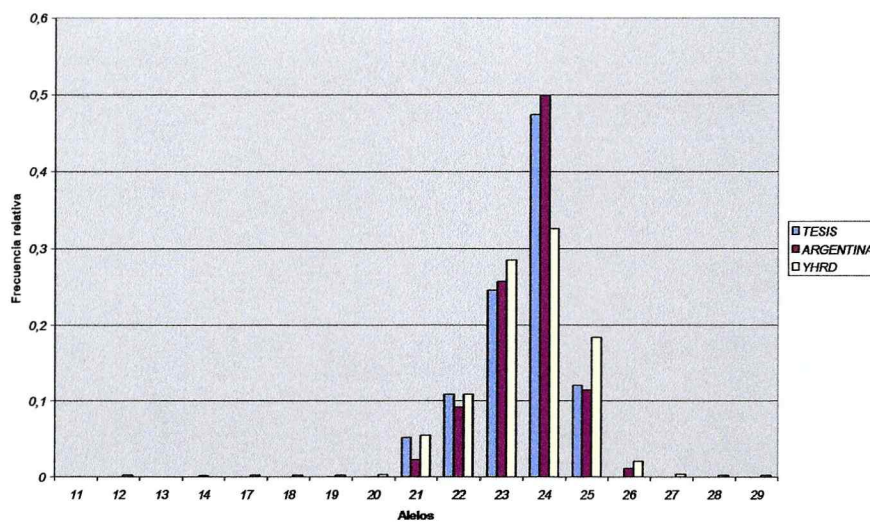


Figura 5: Frecuencias relativas de los alelos de DYS 391

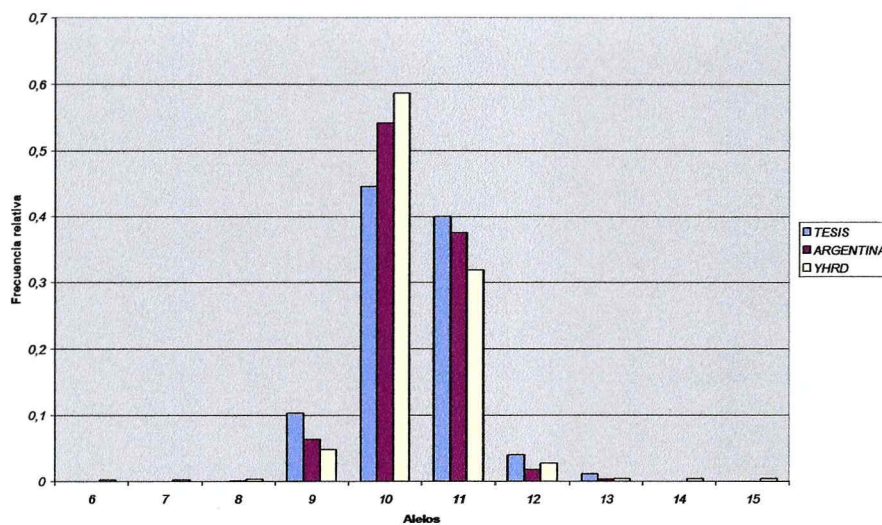


Figura 6: Frecuencias relativas de los alelos de DYS 392

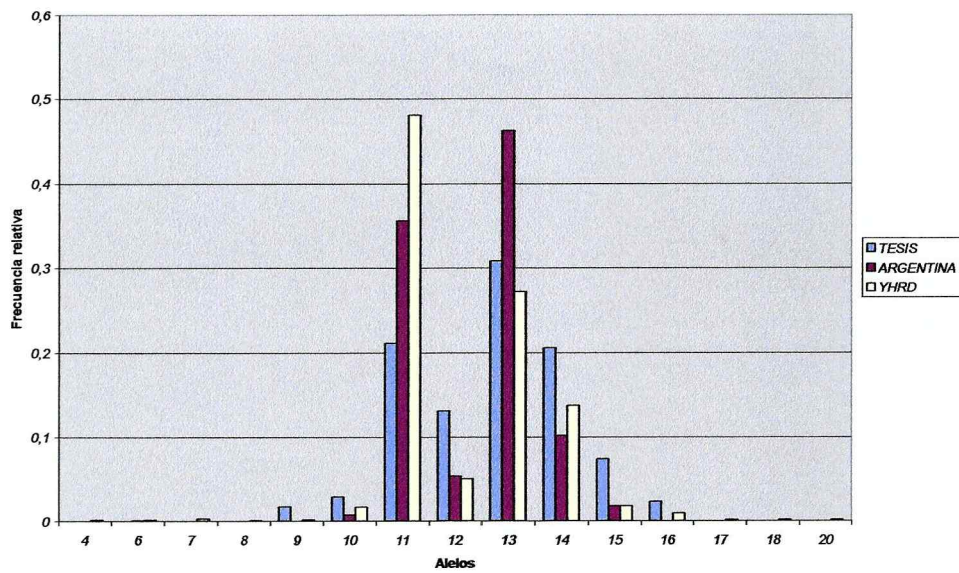
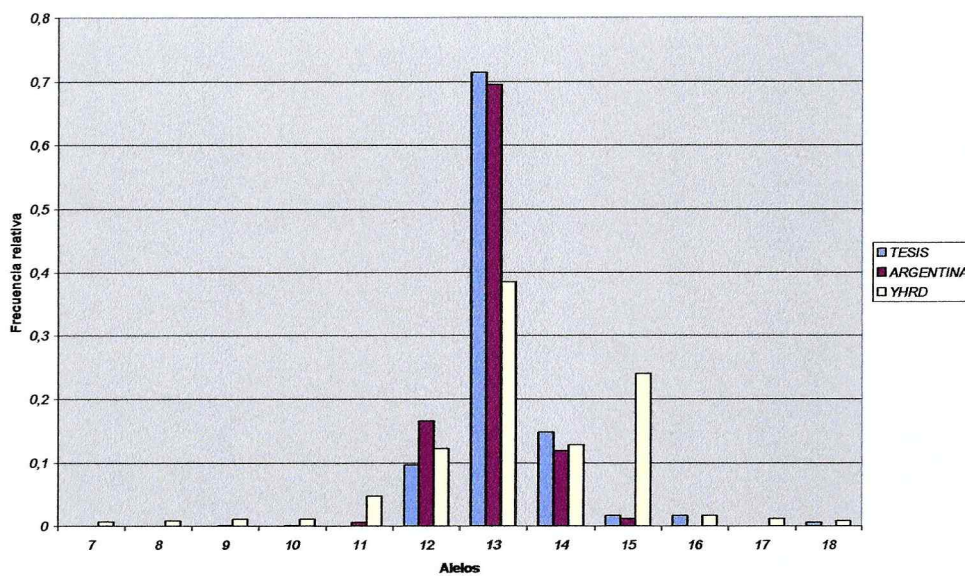


Figura 7: Frecuencias relativas de los alelos de DYS 393



Publicado durante el mes de Enero del 2010, en el International Journal of Legal Medicine (Springer)
Impact Factor: 2.574 (2008) Journal citation Reports, Thomson Reuters
Subject category: "Medicine, Legal" Rank 1 of 10
DOI: 10.1007/s00414-009-0404-1

Marina Muzzio¹, Virginia Ramallo, Josefina M. B. Motti, Maria R. Santos, Jorge S. López Camelo,
Graciela Bailliet

Software for Y-haplogroup predictions: a word of caution

Laboratorio de Genética Molecular Poblacional. Instituto Multidisciplinario de Biología Celular (IMBICE). CICPBA, CCT La Plata- CONICET. Argentina.

Corresponding author:

Marina Muzzio

IMBICE, 526 e/ 10 y 11, 1900 La Plata. Argentina

Postal Box: C.C. 403

Phone: 54-221-421 01 12 FAX: 54-221-425 33 20

e-mail address: marinamuzzio@yahoo.com.ar

Abstract The development of online software designed for genetic studies has been exponentially growing, providing numerous benefits to the scientific community. However, they should be used with care, since some require adjustments. The efficiency of two programs for haplogroup prediction was tested with 119 samples of known haplotypes and haplogroups from Argentine populations. Quantitative estimates of the predictive quality of both software systems were computed with the uncertainty coefficient; and sensitivity, specificity, positive and negative likelihood ratios were also calculated to assert the reliability of both programs, showing high probabilities of assigning an incorrect haplogroup.

*Electronic Supplementary Material * Supplementary material is available for this article at <http://dx.doi.org/10.1007/s00414-009-0404-1>

Keywords Molecular Anthropology / Population Genetics / Human Y-Chromosome / SNP/STR / Bioinformatic Approaches

Introduction

Y-chromosomal lineages are established by single nucleotide polymorphism (SNP) and short tandem repeats (STR), which provide the corresponding haplogroup and haplotype, respectively. In the study of human population genetics, haplogroup determination is of great interest, as it reveals the phylogenetic relationships by descent.

Considering the findings of Bosch et. al. [1] and Behar et. al. [2], where the STR variability is partitioned by haplogroups to a greater extent than by populations, there has been increasing interest in unifying these sources and finding further ways of predicting the haplogroup of a given haplotype when SNP data are unavailable. One of them is Whit Athey's Haplogroup Predictor [3, 4] (<https://home.comcast.net/~hapest5/index.html>) which has been employed in previous studies [5, 6, 7] to estimate the ethnic composition of different populations and diagnostic STR values of a given haplogroup. The other is the Haplogroup Classifier [8] (<http://bcf.arl.arizona.edu/haplo>), consisting of machine-learning algorithms that require previous models and haplotypes with a known haplogroup for training the software.

The purpose of this study was to establish the accuracy of both software systems: the Haplogroup Predictor and the Haplogroup Classifier.

Materials and methods

We analyzed a sample of 119 males from 4 provinces of the northwest of Argentina (Jujuy, Salta, Catamarca, and Tucumán), all of them with the informed consent of donors.

Haplogroups were determined in a previous report [9]; the nomenclature used followed YCC [10] recommendations and haplotypes were defined by the amplification of DYS19, DYS389 I and II, DYS390, DYS391, DYS392, and DYS393, according to methods previously published [11, 12, 13, 14, 15].

These haplotypes were submitted to the Haplogroup Predictor, with equal priors, obtaining probabilities for inferred haplogroups. In the case of the Haplogroup Classifier, we followed the models, tree files and public data provided by the authors in the downloadable version of the software; this data set consisted of 1527 Y-chromosome profiles with haplogroup and haplotype gathered from published data [16, 17]. SNP-determined haplogroups were compared with those provided by the software. It was not possible to evaluate the sample analyzed by Schlecht et al. [8] given the availability of their data.

Sensitivity (s), specificity (e), positive likelihood ratio ($LR+ = s/(1-e)$), and negative likelihood ratio ($LR- = (1-s)/e$) of the Haplogroup Predictor and the Haplogroup Classifier were calculated per haplogroup and total [18]: " s " represents the probability of corroboration for a predicted haplogroup when the haplotype was determined of that haplogroup by SNP analysis, while " e " stands for the probability of confirmation of another haplogroup when the haplotype corresponded to a different haplogroup by our typing. It has been stated that a test is adequate if it has both high sensitivity and specificity [19], and a $LR+$ value of at least 10; this is the reason why we followed these criteria.

In the case of the Predictor, we considered different precision-of-assignment categories: by being the first in the haplogroup ranking (rk), and 50-95% cut-off points by intervals of 5, whereas for the Classifier we only counted those cases showing an agreed haplogroup, thus reducing the number of cases considered in the calculations ($N = 62$; 52.1% of the original sample).

In order to get quantitative estimates of the software predictive quality, we computed the uncertainty coefficient of y , $U(y|x)$, with the subroutine `cntab2` of Press et al. [20] who also provide a detailed explanation of the meaning and way of computing that coefficient. Let us see what is the meaning of $U(y|x)$ and how it can aid us to estimate the software predictive quality. Suppose that we have a certain sample and that we want to know the result of performing the SNP typing on it, i.e., that result is all the information we want. Let us further suppose that, before performing the SNP typing, we get, for that same sample, the prediction of the software. Then, knowing beforehand the results given by the software will produce a loss of the information that we could later obtain from the SNP typing and, the better the software, the larger that loss will be. If the software was perfect and could accurately predict the result that we would later obtain from the SNP typing, then knowing the result of the former would make us lose all the information that the result of the latter would provide and, of course, it would be unnecessary to do the SNP typing. On the other hand, if the software were useless and had no predictive value, to know its result beforehand would make us lose no information at all and, accordingly, the whole information will have to come from the SNP typing. The uncertainty coefficient of y , $U(y|x)$, quantifies what we have just explained qualitatively. If we take the results from the software systems as the x variable and those from SNP typing as the y variable, then $U(y|x)$ gives the fraction of the SNP typing information that is lost if the software result is already known. As the two extreme cases described before we would have: 1) $U(y|x) = 1.00$ (or 100%), that would imply that the software gives perfect answers (i.e., all the information that would be provided by subsequent SNP

typing has already been provided by the software); 2) $U(y|x) = 0.00$ (or 0%), that would imply that the software provides no information (i.e., all the information should be obtained by subsequent SNP typing).

Results

The s , e , $LR+$, and $LR-$ values per cut-off point for the Haplogroup Predictor are summarized in Table 1. The Classifier software showed the following values: $s = 0.45$, $e = 0.92$, $LR+ = 5.99$, and $LR- = 0.59$. It is important to highlight that about half of the haplotypes (57 out of a total of 119) could not be assigned by the Classifier to an agreed haplogroup ("Unclassified" category in Table 2). The profiles of the Argentinean population sample are presented as Electronic Supplementary Material in Table 3, to allow the validation of these results.

When we consider $LR+$ per haplogroup, Q1a3a in both Classifier and Predictor, and DE* in the Classifier, values are higher than 10.

R* haplogroup showed the highest false positive proportion, significantly higher than false positive proportions of the remaining haplogroups.

Our results for $U(y|x)$ are 0.244 (or 24.4%) for the Haplogroup Predictor, and 0.207 (or 20.7%) for the Haplogroup Classifier. In other words, only about 20% and 25% of the information on the SNP typing is lost if software results are already known.

Discussion

These results represent a high probability of error, and a bias towards the R* haplogroup, so it is most likely that results based on the haplogroup predictions of these software systems are weakened. For cases in which sex bias in multiethnic populations is estimated by this method, an overestimation of the European component is expected. Haplogroup determination by SNP analysis remains the best approach, considering the low reliability of prediction of software available.

The adequate $LR+$ for the Q1a3a and DE haplogroups could be explained by a lower diversity within each group. Especially in the Q1a3a case, which is a relatively recent haplogroup, the homogeneity is the result of its young evolutionary age, given that the time lapse in which the haplotypes spread away from the haplogroup founder is rather short [14].

Considering that the samples from which the calibration frequencies are estimated belong to European (or of European descent) populations, the high false positive proportion in the R* haplogroup could be a reflection of this sampling error, as this haplogroup is the most common among those populations.

In order to get a better idea of what the $U(y|x)$ values mean, let us assume that we are interested in knowing the result of throwing a die (i.e., the information we want to know is whether we will get 1, 2, ..., or 6) and take it as the variable y . Let us further assume that we had some way to know beforehand whether the result would be an odd (1, 3, 5) or an even (2, 4, 6) number and take it as the variable x . It turns out that in this case we get $U(y|x) = 0.387$ (38.7%), that is, knowing beforehand whether the throw of the die will result in an odd or even number, produces a loss of 38.7% of the information that the actual throw will yield. If we compare this result with the values obtained above (24.4% and 20.7%) we see that, for the knowledge of the result of the SNP typing, knowing in advance the result of any of the software systems provides less information than, for the throwing of the die, knowing beforehand whether the result will be odd or even would give.

For the case of the Classifier, in which the user chooses the data to train the software, greater Y-STR profiles with associated haplogroups might improve its accuracy. Even so, this software shows haplotypes without an agreed haplogroup, given their recurrence across haplogroups, which is a better approach than suggesting a haplogroup when there is evidence that the haplotype could belong to different haplogroups. However, this software shows higher accuracy than the Predictor, given $LR+$ values.

Why do these software systems show such low accuracy levels? We propose two explanations: (1) there are not enough Y-STR profiles with associated haplogroups to calibrate the software properly; (2) given the mutation rates of the STRs (available at www.yhrd.org) and the time depth of the haplogroup ramifications [21], it is possible to find the same haplotype in samples from different haplogroups (cases of convergence). For instance, the most recent haplogroup, R1*, has a time to the most recent common ancestor (TMRCA) of 18,500 (12,500-25,700) years, and the most ancient clade (CT) is estimated at 70,000 years [20], whereas the Y-chromosome STR mutation rates vary from 6.35 (4.19 - 9.22 95% CI) $\times 10^{-3}$ for DYS 439, to 0.45 (0.12 - 1.17 95% CI) $\times 10^{-3}$ for the case of DYS 392, so that

the repetition of allele combinations among evolutionary divergent clades are not uncommon, given the meiosis events accumulated along such a vast time depth.

A simple way to confirm this is to check the allele frequencies for each STR locus among the metapopulations of the YHRD database, which show that the same alleles are present. Even though their frequencies can differ, (and taking into account that this database does not provide haplogroup information) given the geographic association of the haplogroups evidenced by other authors [22], if the association between STR alleles and haplogroups was strong enough to allow the prediction of the last based on the haplotypes, these should show similar geographic distribution to the one found on the haplogroups.

An increase in the number of STRs employed to predict the haplogroup would not enhance accuracy, considering the few reference samples available with the standard 7 STRs and associated haplogroup, while these reference samples decrease even more as the amount of STRs demanded increases. Also, there is no homogeneity in the use of more STRs, different authors chose different markers, reducing the haplotype references drastically. For example, Zalloua et al. [23] analyzed the 7 standard STRs plus DYS 388, DYS 437, DYS 438, and DYS 439, while Di Gaetano et al. [24] studied the 7 standard STRs plus DYS 385 A/B in all samples (in only a fraction of their sample other STRs were also included); these data sets are only comparable with each other on the standard STRs, not on the rest.

At present, haplogroup prediction software available do not show adequate accuracy. Thus, typing a set of SNPs which precisely define a phylogenetic branch, is the only reliable method to establish to which haplogroup a given sample belongs.

Acknowledgments

The authors thank Dr. J.E. Dipierri and E.L. Alfaro for their valuable contribution, Dr. J.C. Muzzio for his help with the uncertainty coefficient calculations, and Dr. A.N. Califano and C.M. Bravi for their suggestions.

Grant sponsorship: CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), CICIPBA (Comisión de Investigaciones Científicas de la Provincia de Buenos Aires), ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica), and Antorchas Foundation of Argentina.

G Bailliet and JS López Camelo are members of the CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Argentina.

References

- [1] Bosch E, Calafell F, Santos F et. al. (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *Am. J. Hum. Genet.* 65: 1623-1638.
- [2] Behar D M, Garrigan D, Kaplan M et. al. (2004) Contrasting patterns of the Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. *Hum. Genet.* 114: 354-365.
- [3] Athey WT (2005) Haplogroup prediction from Y-STR values using an allele- frequency approach. *J. Genet Geneal* 1:1-7.
- [4] Athey WT (2006) Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *J. Genet Geneal* 2: 34-39.
- [5] Salas A, Jaime JC, Álvarez-Iglesias V, Carracedo Á (2008) Gender bias in the multiethnic genetic composition of central Argentina. *J. Hum. Genet* 53: 662-674.
- [6] Mertens G (2007). Y-Haplogroup frequencies in the Flemish Population. *J. Genetic Genealogy*, 3(2):19-25.
- [7] Goff PG, Athey TW (2006) Diagnostic STR values for haplogroup G. *J. Genetic Genealogy*, 2(1):12-17.
- [8] Schlecht J, Kaplan ME, Barnard K., Karafet T, Hammer MF, Merchant NC (2008) Machine-Learning approaches for classifying Haplogroup from Y Chromosome STR data. *PLoS Comput Biol* 4(6): e1000093. doi:10.1371/journal.pcbi.1000093
- [9] Ramallo V, Alfaro EL, Dipierri JE, Bianchi NO, Bailliet G (2005) Caracterización de linajes paternos en muestras provenientes de tres provincias del NOA. *Revista de la Sociedad Argentina de Genética, Journal of Basic & Applied Genetics. Actas XXXIV Congreso Argentino de Genética. XVII (Suplement):* 181. Septiembre.
- [10] YCC (The Y Chromosome Consortium). 2002. A nomenclature system for the tree of Y chromosomal binary haplogroups. *Genome Res* 12:339–348.
- [11] de Knijff P, Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, et al. (1997) Chromosome Y microsatellites: Population genetic and evolutionary aspects. *Int. J. Legal Med*; 110:134-140.

- [12] Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F (1997) Evaluation of Y-chromosomal STRs: A multicenter study. *Int. J. Legal Med*; 110:125- 133.
- [13] Bravi CM, Sans M, Bailliet G, Martinez-Marignac VL, Bianchi NO (1997) "Characterization of mitochondrial and Y-chromosome haplotypes in an Uruguayan population of African ancestry." *Hum. Biol.* 69 (5): 641-652.
- [14] Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, Lopez Camelo JS (1998) "Characterization of ancestral and derived Y- chromosome haplotypes of New World native populations. *Am. J. Hum. Genet.* 63: 1862-1871.
- [15] Bailliet G, Castilla EE, Adams JP, Orioli IM., Martínez-Marignac VL, Richard SM, Bianchi NO (2001) Correlation between molecular and conventional genealogies in Aicuña: A rural population from Northwestern Argentina. *Hum Hered* 51: 150- 159.
- [16] Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, et al (2006) Polarity and temporality of high-resolution ychromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central asian pastoralists. *Am. J. Hum. Genet.* 78:202-221.
- [17] Cinnioglu C, King R, Kivisild T, Kalfoglu E, et al (2004) Excavating y-chromosome haplotype strata in Anatolia. *Hum. Genet.* 114:127-148.
- [18] McAlister FA, Straus SE, Sackett DL. (1999). Why we need large, simple studies of the clinical examination: the problem and a proposed solution. *Lancet* 354: 1721-24.
- [19] Greenhalgh T (1997) How to read a paper: Papers that report diagnostic or screening tests. *BMJ* 315:540-543.
- [20] Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1994) Numerical Recipes in FORTRAN. The art of Scientific Computing, 2nd edn. Cambridge University Press.
- [21] 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 the resolution of the Human Y chromosomal haplogroup tree. *Genome Res* 18: 830-838.
- [22] Underhill P, Passarino G, Lin A A, Shen P, Mirazón Lahr M, Foley R A , Oefner P J, Cavalli-Sforza L (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Anal Hum Genet* 65: 43-62.
- [23] Zalloua P A, Xue Y, Khalife J, Makhoul N, Debiante L, Platt D E, Royyuru A K, Herrera R J, Soria Hernanz D F, Blue-Smith J, Spencer Wells R, Comas D, Bertranpetit J, Tyler-Smith C, The Genographic Consortium. 2008. *Am. J. Hum. Genet.* April 2008. Supplementary Data.
- [24] Di Gaetano C, Cerutti N, Crobu F, Robino C, Inturri S, Gino S, Guarrera S, Underhill P A, King R J, Romano V, Cali F, Gasparini N, Matullo G, Salerno A, Torre C, Piazza A. 2008 Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. *Eur. J. Hum. Gen.* 1-9.

Table 1: Total s, e, LR+ and, LR-values at each cut-off point for the Haplogroup Predictor

| Cut-off point | s¹ | e² | LR+³ | LR-⁴ |
|----------------------|----------------------|----------------------|------------------------|------------------------|
| RK | 0.5 | 0.88 | 4.21 | 0.6 |
| 50 | 0.5 | 0.89 | 4.52 | 0.5 |
| 55 | 0.5 | 0.89 | 4.54 | 0.6 |
| 60 | 0.5 | 0.89 | 4.46 | 0.6 |
| 65 | 0.5 | 0.9 | 4.65 | 0.6 |
| 70 | 0.5 | 0.91 | 5.12 | 0.6 |
| 75 | 0.4 | 0.91 | 4.59 | 0.7 |
| 80 | 0.4 | 0.92 | 5.22 | 0.7 |
| 85 | 0.4 | 0.93 | 5.81 | 0.7 |
| 90 | 0.3 | 0.94 | 5.93 | 0.7 |
| 95 | 0.3 | 0.96 | 8.00 | 0.7 |

References: 1: sensitivity; 2: specificity; 3: Positive likelihood ratio; 4: Negative likelihood ratio

Table 2: Haplogroup frequencies by SNP typing, Haplogroup. Predictor, and Haplogroup Classifier, and s, e, LR+, and LR- values

| | | SNP typing | | | | | Total software | False positives | False positive % | |
|------------------|------------------|------------|------|-------|------|------|----------------|-----------------|------------------|-------|
| | | Haplogroup | | | | | | | | |
| | | F* | K* | Q1a3a | R* | DE* | | | | |
| Predictor | Haplogroup | F* | 9 | 0 | 1 | 4 | 5 | 19 | 10 | 52.6 |
| | | K* | 0 | 0 | 2 | 0 | 0 | 2 | 2 | 100.0 |
| | | Q1a3a | 0 | 3 | 17 | 0 | 0 | 20 | 3 | 15.0 |
| | | R* | 6 | 12 | 16 | 30 | 1 | 65 | 35 | 53.8 |
| | | DE* | 2 | 1 | 2 | 3 | 5 | 13 | 8 | 61.5 |
| | Total by SNP | 17 | 16 | 38 | 37 | 11 | 119 | 58 | 48.7 | |
| | Sensitivity | 52.9 | 0.0 | 44.7 | 81.1 | 45.5 | 51.3 | | | |
| | Specificity | 90.2 | 98.1 | 96.3 | 57.3 | 92.6 | | | | |
| | LR+ ¹ | 5.4 | 0.0 | 12.1 | 1.9 | 6.1 | | | | |
| | LR- ² | 0.5 | 1.0 | 0.6 | 0.3 | 0.6 | | | | |
| Classifier | Haplogroup | F* | 4 | 0 | 1 | 4 | 3 | 12 | 8 | 66.7 |
| | | K* | 0 | 1 | 3 | 1 | 0 | 5 | 4 | 80.0 |
| | | Q1a3a | 0 | 1 | 2 | 0 | 0 | 3 | 1 | 33.3 |
| | | R* | 2 | 8 | 9 | 19 | 0 | 38 | 19 | 50.0 |
| | | DE* | 0 | 0 | 1 | 1 | 2 | 4 | 2 | 50.0 |
| | Unclassified | 11 | 6 | 22 | 12 | 6 | 57 | | | |
| | Total by SNP | 17 | 16 | 38 | 37 | 11 | 119 | 34 | 54.8 | |
| | Sensitivity | 66.7 | 10.0 | 12.5 | 76.0 | 40.0 | 23.5 | | | |
| | Specificity | 92.9 | 96.3 | 99.0 | 79.8 | 98.2 | | | | |
| | LR+ ¹ | 9.4 | 2.7 | 12.9 | 3.8 | 22.8 | | | | |
| LR- ² | 0.4 | 0.9 | 0.9 | 0.3 | 0.6 | | | | | |

References: 1: Positive likelihood ratio; 2: Negative likelihood ratio

Aceptado con modificaciones a principios del mes de febrero en el *American Journal of Physical Anthropology*

A method for assignment of the weight of characters

Muzzio M, Bravi CM, Bailliet G

Laboratorio de Genética Molecular Poblacional. Instituto Multidisciplinario de Biología Celular (IMBICE). CICPBA, CCT La Plata- CONICET. Argentina.

Number of text pages, plus bibliography: 7

Number of figures: 3

Number of tables: 0

Abbreviated title: Assignment of character's weight

Key words: Median-Joining networks, mathematical scale, character weight, Y chromosome, microsatellites

Name, address, telephone number, fax number, and email address of the person to whom the

proof is to be sent: Marina Muzzio

IMBICE, 526 e/ 10 y 11, 1900 La Plata. Argentina

Postal Box: C.C. 403

Phone: 54-221-421 01 12 FAX: 54-221-425 33 20

e-mail address: mmuzzio@imbice.org.ar

Grant sponsorship: CONICET, CICPBA, ANPCyT, of Argentina. G. Bailliet, and CM. Bravi, are members of the Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina (CONICET).

Abstract:

The weight of characters is a crucial step in different population analyses. We propose a new formula to facilitate this while establishing a scale that follows the criteria of the probability of change in each character. This method is described for drawing of Median-Joining networks, yet it could also be used for other methods in which the weight of characters is required.

Keywords: character weight, mathematical scale, Median-Joining networks, Y chromosome, microsatellites

Median-Joining networks [1] are widely used in Molecular anthropology and population genetics since they allow estimation of phylogenies at the intra-species level, while using non recombining population data and without solving “ties” between trees. The most used software system for drawing them is NETWORK (fluxus-engineering.com) that allows the weight of characters: the less likely events should be given a higher weight, considering that when they occur it is significant, while the most probable events should have a lower weight assigned. This step is crucial because it allows giving differential importance to each mutation when building the network. This way, the more plausible connections are showed, which makes interpretation of the resulting network simpler.

In the present paper, we propose a new system for weight assignment originally designed for Y-STR data that could also be employed for other types of traits, since it is based on the probability of a change of state. In the case of Y-STR data, the most used system was proposed by Qamar et al. [2], employed by many researchers [3, 4, 5]. It consists of a fixed scheme of asymmetrical intervals, based on the variance of each STR, which gives weights from 1 to 5. These authors comment that their resulting networks contained a high amount of high-dimensional cubes (these occur when it is not possible to draw the network in a clear way and complicate its interpretation), so they also needed to use the reduced median algorithm before drawing the Median-Joining network.

Also, there are a few other complications that were not addressed by these authors: 1) Basing the weight on the STR variation requires having estimated it beforehand. 2) This variation could be reduced by drift processes, so we could find, in a given population -at

least theoretically- a low variance in a STR that is highly polymorphic in others as a result of a recent bottleneck. 3) Even if the STR variances were estimated employing independent data bases (such as YHRD, among others), when the single-step mutational model is considered, more specifically its characteristic of presenting a maximum and minimum of possible allele length, the variance of highly recurrent mutations could be underestimated. 4) It is a very restricted weight range that goes from 1 to 5, unpractical in cases where many STRs with different mutational rates are employed.

We propose the following formula for the assignment of weight per character:

$$W(x) = M (\text{mutational rate}_y / \text{mutational rate}_x)^a$$

Where:

$W(x)$ is the weight of a given STR or marker.

M is the maximum weight value that the researcher decides to employ.

Mutational rate_y is the lowest mutational rate in the analyzed STR set.

Mutational rate_x is the mutational rate of the STR for which the researcher wishes to estimate the weight.

“a” is a value that enables the researcher to adapt the resulting scale to a desired range.

To estimate it:

$$\text{Lowest weight value} = M (\text{mutational rate}_y / \text{mutational rate}_x)^a$$

Which is conveniently put in a logarithmic form, linear in “a”, which is the unknown, and simplifies its computation:

$$\text{Log}(\text{lowest weight value}) = \text{Log}(M) + a \{ \text{Log}(\text{mutational rate}_y) - \text{Log}(\text{mutational rate}_x) \}$$

So that:

$$a = \{ \text{Log}(\text{lowest weight value}) - \text{Log}(M) \} / \{ \text{Log}(\text{mutational rate}_y) - \text{Log}(\text{mutational rate}_x) \}$$

For instance, if we decide to assign a minimum value of 3 and a maximum of 10, for the minimum haplotype whose lowest mutational rate is DYS 392 (of $0.45 \cdot 10^{-3}$) and highest is DYS 389 II (of $3.43 \cdot 10^{-3}$) (www.yhrd.org):

$$3 = 10 (0.45 \cdot 10^{-3} / 3.43 \cdot 10^{-3})^a$$

a = 0.593

MATERIALS AND METHODS

To assess the quality of our scheme compared to the one of Qamar et al. [2] or a uniform weight assignment, we employed published data of samples belonging to the haplogroup E from a Lebanese population [6]. We chose the seven STRs from the minimum haplotype [7]: DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, and used these data to calculate and draw 3 median-joining networks, employing the software Network 4.5.1.0 (fluxus-engineering.com), one per each weight assignment system. In the case of the network built with our weight assignment proposal, the mutation rates were obtained from the international database YHRD (www.yhrd.org), choosing a maximum of 10 and a minimum of 3, resulting in a weight value of 3 for DYS 389 II and DYS 391; 4 for DYS 19, DYS 389 I and DYS 390; 7 for DYS 393 and 10 for DYS 392.

We estimated the amount of median vectors and maximum length to the node per network, and presence or absence of high dimension cubes.

RESULTS

The resulting networks are represented by Figures 1, 2 and 3. In Figure 1 we present the network calculated by assigning a weight of 10 for each STR, as the reader can see, a high amount of median vectors (in total 34), high dimension cubes and a maximum length of 10 mutational steps.

Figure 2 shows the network estimated with the weight assignment that follows the method of Qamar et al. [2]. There are 8 median vectors, high dimension shapes, and a maximum length to the node of 10 mutational steps.

Figure 3 presents the network computed and drawn with our proposal. There are a total of 6 median vectors, no high dimension cubes, and a maximum length to the node of 8 mutational steps.

In all three networks the ancestral haplotype remains the same, yet their topology is quite different.

DISCUSSION

Our proposal permits the researcher to obtain more parsimonious networks than those computed with other weight schemes, without employing pre or post-processing, since the resulting network contained less median vectors, a shorter maximum distance to the node and no high-dimensional cubes or figures. Note that all graphics (Fig. 1, 2, 3) were copied exactly from those produced by the software, without altering the topology in any way that could simplify or modify them.

The proposed formula considers the mutational rates of the STRs, instead of their variance. Accordingly, those markers with a high mutational rate receive low weights, while those with low rates are assigned a high weight, since they are less probable events than the former. A mutation-based weighting scheme was originally proposed by Forster et al. [8], however their suggestion only consisted in establishing two weight classes, one with high mutation rates and another with lower rates, weighted by a factor of 2:1. Our formula allows a fan of weights that gives a finer resolution according the differences in mutation rates.

Anyway, when reliable mutational rates are not available, but STR variation is (which could be the case of newly found STR markers), the formula could be modified considering the last one instead of the former:

$$W(x) = M (\text{variation}_y / \text{variation}_x)^a$$

Also, the informativeness for assignment could be employed instead of variation, and probably with better results, when there are no reliable mutational rates. This ratio is estimated by following the method proposed by Rosenberg et al. [9], with the software INFOCALC [9]. Of course, in this case the formula would be modified accordingly replacing the lowest mutational rate with the highest informative value, and the highest mutational rate with the lowest mutational value.

Another advantage is that it allows the researcher to establish the range of values desired, through the estimation of “a”. This facilitates the analysis of multiple STRs simultaneously, since it allows using a larger range when the quantity of markers is high. In this paper we chose the minimum weight of 3 and a maximum of 10 because we employed the STRs from the minimum haplotype, nonetheless, the Network 4.5.1.0

(fluxus-engineering.com) software enables the selection of a minimum weight of 0 and a maximum of 100, so the possible range is quite broad.

The formula does not allow a minimum weight of 0 value, but that is not a problem because 0 weights are useless by definition. So if one has a variable with 0 weight value, all one has to do is to exclude it from the analysis and to take the variable with the second lowest weight as that with the minimum weight.

The flexibility of this formula allows its application on other markers and features for whose analysis a weight-scheme is required. Simply, the researcher substitutes the mutational rates for a probability of a change of state for the characters of interest in the formula. Thus, an easy to calculate, yet mathematically precise, scale is available for different fields in Anthropology and Population Genetics beyond Molecular Anthropology.

ACKNOWLEDGEMENTES

The authors would like to thank Dr. Juan Carlos Muzzio for his suggestions on the formula.

LITERATURE CITED

- [1] Bandelt H J, Forster P, Röhl A 1999. "Median-Joining Networks for Inferring Intraspecific Phylogenies" *Mol. Biol. Evol.* Vol 16. Issue 1. 37-48.
- [2] Qamar R, Ayub Q, Mahyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi S Q 2002. "Y-Chromosomal DNA Variation in Pakistan". *Am. J. Hum. Genet.* 70: 1107-1124.
- [3] King T E, Parkin E J, Swinfield G, Cruciani F, Scozzari R, Rosa A, Lim S K, Xue Y, Tyler-Smith C, Jobling M A 2007. "Africans in Yorkshire? The deepest-rooting clade of the Y phylogeny within an english genealogy". *Eur. J. Hum. Genet.* 15: 288-293.
- [4] Malhi R S, González-Oliver A, Schroeder K B, et al. 2008. "Distribution of Y Chromosome Among Native North Americans: a study of Athapaskan Population History". *Am. J. of Phys. Anthropol.* 137 (4): 412-424.
- [5] Bolnick D A, Bolnick D I, Smith D G 2006. "Asymmetric Male and Female Genetic Histories among Native Americans from Eastern North America". *Mol. Biol. Evol.* 23 (11): 2161-2174.
- [6] Zalloua P A, Xue Y, Khalife J 2008. "Y-chromosomal Diversity in Lebanon is structured by recent historical events". *Am. J. Hum. Genet.*, 82 (4): 873-882.
- [7] Kayser M, Caglia A, Corach D, et al. 1997. "Evaluation of Y-chromosomal STRs: a multicenter study". *Int. J. Legal Med.* 110: 125: 133.
- [8] Forster P, Röhl A, Lünemann P, et al. 2000. "A Short Tandem Repeat-Based Phylogeny for the Human Y Chromosome". *Am. J. Hum. Genet.* 67: 182-196.
- [9] Rosenberg N A, Li L M, Ward R, Pritchard J K. 2003. "Informativeness of genetic markers for inference of ancestry". *Am. J. Hum. Genet.* 73 1402-1422.

FIGURE LEGENDS

Figure 1: Microsatellite network drawn with uniform weights. Black dots show the median vectors, gray dots the samples. Nodes are proportional to frequencies, and branches are proportional to mutational steps.

Figure 2: Microsatellite network drawn with the weights established following the method proposed by Qamar et al. (2002). Black dots show the median vectors, gray dots the samples. Nodes are proportional to frequencies, and branches are proportional to mutational steps.

Figure 3: Microsatellite network drawn with the weights established following the formula proposed in the present paper. Black dots show the median vectors, gray dots the samples. Nodes are proportional to frequencies, and branches are proportional to mutational steps.

FIGURES

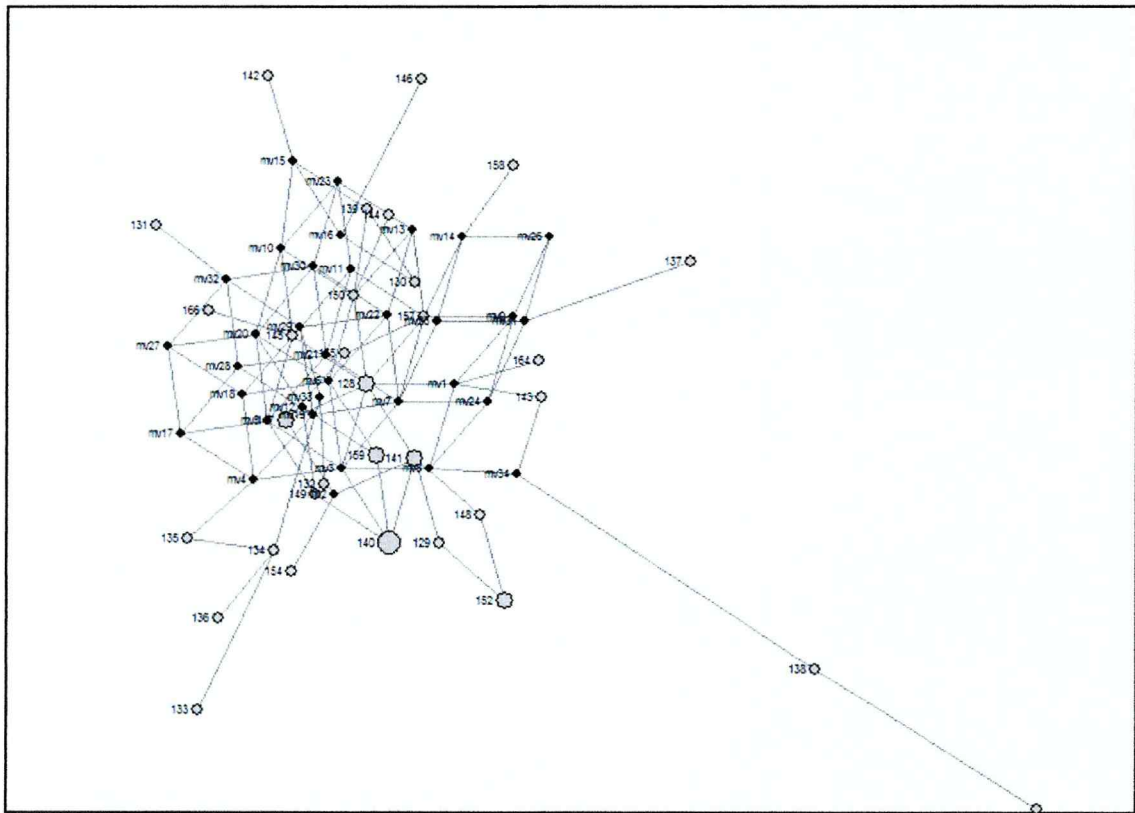


Fig. 1

