



FACULTADE DE MEDICINA
E ODONTOLOXÍA

DEPARTAMENTO DE ANATOMÍA
PATOLÓXICA E CIENCIAS FORENSES

INSTITUTO DE CIENCIAS FORENSES
LUIS CONCHEIRO

Análisis de linajes uniparentales en población Boliviana

Memoria que presenta para optar el grado de doctor:
Jorge Mario Cárdenas Paredes
Santiago de Compostela, Septiembre 2014



Portada inspirada en el portal web del *Y Chromosome Human Reference Database* y en la bandera insignia de los grupos étnicos de la cordillera de los Andes en América del Sur, *Wiphala*.

El Dr. *D.* Ángel Carracedo Álvarez, Catedrático de Medicina Legal de la Facultad de Medicina de la Universidade de Santiago de Compostela, el Dr. *D.* Antonio Salas Ellacuriaga Profesor Titular de Medicina Legal de la Facultad de Medicina de la Universidade de Santiago de Compostela y la Dra. *Dña.* Vanesa Álvarez Iglesias, investigadora en la Universidade de Santiago de Compostela

CERTIFICAN:

Que la presente memoria que lleva por título “**Análisis de linajes uniparentales en población Boliviana**” realizada por el licenciado en Biología *D.* Jorge Mario Cárdenas Paredes, ha sido realizada bajo nuestra dirección, en el Instituto de Ciencias Forenses – Luis Concheiro – considerándola en condiciones para optar el Grado de Doctor y autorizándola para su presentación ante el Tribunal correspondiente.

Y para que así conste, firmamos la presente en Santiago de Compostela a 25 de Septiembre de 2014.

Fdo.: Prof. Dr. Ángel Carracedo Álvarez

Fdo.: Prof. Dr. Antonio Salas Ellacuriaga

Fdo.: Dra. Vanesa Álvarez Iglesias





*"La vida no es lo que uno vivió, sino lo que recuerda,
y cómo la recuerda para contarla".*

Gabriel García Márquez.

(Vivir para contarla, 2002)



Agradecimientos

Después de cuatro años de vivir experiencias con muchas personas en un país extranjero y “solo”, evidentemente escribir este apartado no es ninguna tarea fácil, a este momento la secuenciación del genoma mitocondrial parece ser más fácil...sin embargo en estas *terras galegas* he aprendido que *es de bien nacidos ser agradecidos*, así que primeramente gracias a Jehová Dios por permitirme tener la vida y haber contado con su espíritu durante todo este tiempo.

Muchas gracias Ángel, por tu sencillez, modestia y generosidad que te caracterizan que no tienen igual y que indudablemente han sido las que han permitido que haga parte de tu equipo de investigación.

Maviki, muchas gracias siempre por tu apoyo incondicional e interés personal que te caracterizan porque si no fuera por tu ayuda este trabajo nunca terminaría y es gracias a ti que puedo terminar todo satisfactoriamente.

Toño, para ti no tengo las palabras de agradecimiento que pudieran expresar toda mi gratitud hacia ti, porque más que un director haz sido un amigo y colega que siempre ha estado dispuesto en todo tiempo a ayudar y que has estado hasta en lo más mínimo de este trabajo sin esperar nada a cambio, realmente muchas gracias Toño.

Vane, a ti también te debo muchas gracias porque desde el primer momento que empecé a trabajar contigo, sentí la guía y la dirección que siempre desee de un director y que a pesar que el “mundo del ADNmt” es una locura, siempre me orientaste con mucha paciencia y has estado siempre en el momento indicado.

A Cris, muchas gracias porque me has hecho parte de tus proyectos y mucho de los trabajos de esta tesis se deben a ti.

También tengo mucho que agradecer a todos mis compañer@s del Lab, que durante todos estos años han estado brindándome su compañía y amistad y aún más... me han soportado.

Quiero que siempre lean este apartado, recuerden que han hecho parte de mi vida emocional y profesional, y que no será fácil olvidarlos; Gracias Ana Freire, por tu calidad humana y por tu sencillez, que te hacen grande, sin duda tendrás una vida exitosa; Gracias Ana Mosquera por tu infinita paciencia e instrucción y también por el tema de las alícuotas!; Gracias Fonde por enseñarme una manera distinta de vivir la ciencia. A mis colegas del *cuartucho*, A Jens Heinrich Karl muchas gracias por tu sincera amistad por tu disponibilidad y siempre por tu amabilidad has sido un gran compañero y apoyo; A Miguel y Danel gracias chicos por vuestra ayuda y entrenamiento en el mundo del cromosoma Y, sin vosotros este trabajo no hubiese sido posible; A Jacobo, gracias porque siempre has tenido tiempo para resolver mis dudas y que mucho del trabajo gráfico de esta tesis tiene tu sello. A Carlinha, gracias porque siempre tienes muy buenas ideas porque siempre estás disponible y porque has sido una gran compañera y colega. A María de la Puente, gracias mujer por sopórtame eres una chica estupenda y espero algún día escucharte decir que cuando yo estaba en el Lab *molaba*. A Tanja, gracias porque siempre has compartido tus conocimientos conmigo y has sido una

gran compañera de batalla en todos los trabajos. A Niaz, gracias porque eres un excelente amigo y confidente te deseo siempre muchos éxitos en tu vida. También gracias a los demás chicos que han estado en el Lab durante mi formación, a Luis, a Olalla, a Paula, a Yarimar, como también a Raqui, y a Meli que en su debido momento siempre se mostraron atentas para resolver mis problemas técnicos.

También muchas gracias a los chicos de fuera que he conocido y han sido grandes compañeros, a Ilker, Özlem y Müga (Turquía), a Fabián (Chile), A Fabio, Stephanie y Balsam (UK), a Tolga (Chipre), Adriana (Colombia), Gabrielle y Anna Bárbaro (Italia) y Jeremy (Australia). Así mismo muchas gracias a Patricia (Bolivia) porque gracias a tu detallado muestreo, ha sido posible esta tesis. Igualmente gracias a Simona (Italia) porque has sido una excelente compañera y colega en mis últimos días en el Lab y resultado una muy buena consejera de diseño. Por último y no menos importante muchas gracias a Laura M. (Italia) por tu compañía en el día a día, por hacerme reír, por distraerme, por tus ideas de diseño, por las largas conversaciones y porque me has enseñado que todo en esta vida puede ser *Micidiale!*

Igualmente muchas gracias a todas aquellas personas que me han ayudado fuera del Lab y han hecho mi vida más sencilla, a Rosa y Javier por ser excelentes personas y por quienes tuve un espectacular lugar donde vivir y compartir mi vida con los demás. A Lore y Luisa porque a pesar de que nuestros futuros tomaron caminos distintos, fueron un apoyo al comenzar esta aventura juntos. También a José, Ana y Tamara porque han sido mis amigos y familia desde que llegue a Santiago y porque su hospitalidad no tiene fin. Así mismo a Santi y Flora que han sido como unos padres para mí, nunca tendré como pagarles. A Javier, Pili, Joel y Yony que son mis amigos de siempre y que son mi familia en segundo grado; gracias a Pablo, Fina y José Ramón por su amistad incondicional y porque más que amigos han sido mis hermanos. También gracias a Manuel Paredes, porque gracias a él conocí este emocionante mundo de la genética forense y ha sido un verdadero profesor desde que lo conocí. Gracias a mis amigos de Brasil, a Fernando por tu sincera amistad y porque siempre aprenderé cosas nuevas cada vez que nos veamos, y a Cristiani porque más que una compañera has sido una gran amiga te deseo lo mejor para tu vida y recuerda que la motivación todo lo puede.

A mi Papá y mi Mamá muchas gracias porque es por ellos que no solo he tenido una educación académica si no una educación con valores que es la que me ha permitido afrontar la el día a día de esta vida. También a mis hermanos Lino y Luis porque a la distancia me han apoyado y aconsejado en todo tiempo y porque sin ellos mi vida será una complicación.

Por último, nada de esto hubiese sido posible sin ti Vidita, porque desde que he estado lejos de ti, siempre me has apoyado a la distancia, me has escuchado y aconsejado y aunque no ha sido fácil para ti estar sola, siempre has sido más fuerte y es esa fuerza es la que me ha mantenido durante todo este tiempo; Naty espero compensarte todo este tiempo ausente y que algún día te sientas orgullosa de mi.

Y a los que olvide, perdonar, muchas Gracias a todos!



A mi Esposa, A Naty

A Mis Padres



Índice

Abreviaturas.....	xiii
I. Introducción.....	1
1. El Genoma Humano, la genética de poblaciones y forense	3
2. El cromosoma Y	6
2.1. Herencia	8
2.2 Polimorfismos del cromosoma Y	10
2.2.1 Polimorfismos binarios Y-SNPs.....	10
2.2.1.1 Nomenclatura.....	11
2.2.1.2 Diversidad de los haplogrupos de cromosoma Y.....	14
2.2.2 STRs del cromosoma Y.....	17
2.2.2.1 Rapid Mutating Y-STRs	21
2.3 Aplicaciones generales del cromosoma Y en genética forense	23
2.3.1 Estudios en Y-SNPs & Y-STRs en América del Sur	26
2.4 Poblamiento de América y su importancia en la población boliviana	29
2.4.1 Poblaciones fundadoras de América	29
2.4.2 Bolivia, una población mayormente nativo americana	32
II. Justificación y Objetivos	35
III. Materiales y Métodos	39
3.1 Muestras.....	41
3.2 Extracción de ADN	41
3.3 Cuantificación espectrofotométrica de ADN total.....	42
3.4 Análisis de microsatélites de cromosoma Y	42
3.4.1 Amplificación Y-STRs - Yfiler®	43
3.4.2 Amplificación Y-STRs – PowerPlex® Y23	44
3.4.3 Amplificación Y-STRs – Rapidly Mutating Y-STRs	44
3.4.4 Secuenciación	46
3.5 Polimorfismos Binarios de cromosoma Y (Y-SNPs).....	46
3.5.1 Reacción de PCR Multiplex – Amplificación de los SNPs.....	47
3.5.2 Purificación post-PCR, Reacción de mini-secuenciación y Purificación post- extensión	50
3.5.3 Electroforesis Capilar.....	51
3.6 Análisis Estadísticos.....	52
IV. Resultados y Discusión	55
4.1 Variabilidad genética del Yfiler® en población boliviana	57
4.1.1 Diversidad molecular de los Y-STRs	57
4.1.2 Diversidad de los haplogrupos de Y-SNPs	58
4.1.3 Análisis de varianza molecular (AMOVA) de los Y-STRs	60
4.1.4 Patrones espaciales de Y-STRs y de la variación en Y-SNPs	62

4.1.5 Análisis de componentes de ancestralidad	63
4.1.6 Análisis del sesgo por género entre los polimorfismos de cromosoma Y, ADN mitocondrial y AIMS	64
4.1.7 Importancia del estudio de la variabilidad del cromosoma Y en Bolivia	66
4.2. Valoración de la variabilidad de los Rapidly Mutating Y-STRs en Bolivia en el contexto de las poblaciones mundiales	68
4.2.1 Diversidad Global de los haplotipos de RM-YSTRs en diferenciación de los linajes masculinos	68
4.2.2 Análisis global de la sub-estructura geográfica con los Rapidly Mutating Y-STRs	68
4.2.3 Mejoramiento en la diferenciación masculina con RM-YSTRs	70
4.2.4 Habilidad de los RM-YSTRs para detectar relaciones paternas no visibles.	74
4.2.5 Bolivia y su aporte a la variabilidad de los RM-YSTRs en el contexto regional	74
4.2.5.1 Bolivia, el grupo regional nativo americano y la Sub-estructura geográfica con los RM-YSTRs	74
4.2.5.2 Bolivia en el grupo regional nativo americano en el mejoramiento en la diferenciación masculina con RM-YSTRs	75
4.2.6 Capacidad de diferenciación entre individuos masculinos emparentados.	77
4.2.7 Importancia del estudio de la variabilidad con RM-YSTRs en Bolivia y el grupo nativo americano	78
4.3. Valoración de la variabilidad del PowerPlex® Y23 (PPY23) en Bolivia en el contexto de las poblaciones mundiales	79
4.3.1 Generalidades del análisis comparativo de parámetros forenses de discriminación en PPY23 con Yfiler	79
4.3.2 Análisis comparativo de la estructura poblacional con los paneles Yfiler y PPY23	81
4.3.3 Las poblaciones Latino-Americanas y Nativo americanas y su efecto en el análisis haplotípico con Yfiler y PPY23.	83
4.3.4 Estructura poblacional de las poblaciones Latino-Americanas	84
V. Consideraciones finales.....	87
VI. Conclusiones.....	93
VII. Bibliografía	99
VIII. Índice de Figuras	117
IX. Índice de Tablas	121



Abreviaturas



ADN: Ácido desoxirribonucleico

ADNmt: ADN mitocondrial

AIM: *Ancestry Informative Marker*

Indels: *Insertion/deletion polymorphisms*

CrY: Cromosoma Y

KYA: *Thousand Years Ago*

Mb: Mega Base

MHT: *Minimum Haplotype*

MSY: *Male Specific Region of the Y chromosome*

NRV: *Non recombining region*

PAR: *Pseudoautosomic region*

Pb: Pares de Bases

PCR: *Polymerase Chain Reaction*

RFLP: *Restriction Fragment Length Polymorphism*

RM Y-STRs *Rapidly mutating Y-STRs*

SNP: *Single Nucleotide Polymorphism*

STR: *Short Tandem Repeats*

UEPs: *Unique Events Polymorphisms*

VNTR: *Variable Number of Tandem Repeat*

Yp: Brazo corto del cromosoma Y

Yq: Brazo largo del Cromosoma Y





I. Introducción



1. El Genoma Humano, la genética de poblaciones y forense

Desde la publicación del primer borrador y la posterior secuencia completa del genoma humano constituido por aproximadamente 3.000 millones de pares de bases de ADN y con alrededor de 20.000-25.000 genes que codifican para proteínas, la genética moderna ha experimentado un cambio de rumbo sin precedentes, y esto sin lugar a dudas ha tenido un profundo efecto en diversas disciplinas, tales como la investigación biomédica, la práctica clínica, la genética forense, la genética evolutiva, etc.

Desde la luz de la genética de poblaciones y la genética forense, la publicación del genoma humano ha supuesto una ventana grande hacia la información, a través de la cual se puede observar con mayor detalle la historia demográfica de un individuo en particular y de las poblaciones humanas en general. La principal motivación de estas disciplinas de la genética es encontrar elementos, en forma de polimorfismos genéticos, que permitan inferir diferencias entre los individuos y/o las poblaciones que puedan dar respuestas a preguntas muy diversas así como plantear nuevos retos. Así por ejemplo, la genética clínica podría estar interesada en conocer los polimorfismos genéticos o mutaciones patogénicas que pudieran estar relacionadas con una determinada patología, la genética de poblaciones podría plantearse preguntas tales como el origen evolutivo de un grupo poblacional o sus procesos demográficos de mestizaje históricos, mientras que la genética forense podría estar interesada en estudiar aquellos marcadores más apropiados para llevar a cabo una determinada pericia genético-forense. Si bien las preguntas planteadas parecen corresponder a campos de investigación muy diversos, se da la circunstancia de que todos ellos comparten un enorme fondo de conocimiento común, que en pocas palabras se podría definir como la necesidad de entender la variabilidad genética humana.

El estudio de los marcadores uniparentales (linajes paternos del cromosoma Y, maternos del ADN mitocondrial o ADNmt) ha tenido un enorme protagonismo en muchas de estas disciplinas de la genética. En particular, han jugado un papel esencial en el marco de la genética de poblaciones humanas y la genética forense desde hace aproximadamente dos décadas.

Los marcadores del ADNmt y del cromosoma Y representantes respectivamente de las líneas maternas y paternas han sido los preferidos para este proyecto de tesis. Debido a sus características de ausencia de recombinación y herencia haploide, hacen posible analizar y representar de una forma sencilla las relaciones evolutivas existentes entre los individuos y

las poblaciones en una filogenia. Estos marcadores se comportan como bloques de herencia sencillos y únicos que se transmiten como tales de generación en generación a través del padre (cromosoma Y) o de la madre (ADNmt); a diferencia del ADN nuclear donde la recombinación distorsiona los patrones de evolución lineal que caracteriza a los marcadores uniparentales. Por otro lado, estos polimorfismos uniparentales muestran una distribución geográfica más específica que los marcadores autosómicos. Esto es debido a que el tamaño efectivo poblacional es menor que los marcadores autosómicos lo que favorece los procesos de deriva génica. La mayor diferenciación genética existente en los marcadores uniparentales permite inferir diferencias entre las poblaciones con mayor resolución que los marcadores autosómicos (cuando estos son considerados de manera individual). Asimismo, el análisis de esta variación genética en el ADNmt y el cromosoma Y en las poblaciones humanas da vida a la filogeografía, es decir, la ciencia que estudia la distribución de los linajes o haplogrupos (conjunto de secuencias o haplotipos relacionados evolutivamente en un tiempo reciente; clado) en el espacio geográfico. Esta información es vital para entender la historia de las poblaciones humanas y ha tenido un fuerte impacto también en la genética forense (i.e. conocimiento de la estructura poblacional y sus consecuencias en la valoración de la prueba; monitorización de errores de genotipado y secuenciación, etc.).

Lo anterior ha permitido, en el área de genética forense la creación de bases de datos a nivel regional y global que permiten determinar la composición genética de las poblaciones y de los individuos. Estas bases de datos pueden ser usadas para identificar los orígenes y las relaciones de un individuo o de una familia con una evidencia encontrada por ejemplo en una escena del crimen o como prueba para identificar unos restos óseos en la casuística forense. Igualmente la identificación de los grupos de marcadores de ADN, o haplotipos, que son únicos a una población, se puede usar para determinar los orígenes geográficos específicos posibles y los vínculos de los individuos con un linaje.

Por lo tanto podemos decir que la ciencia forense es uno de los campos de la investigación científica que ha adoptado a la genética como disciplina desde hace tan solo dos décadas, y que en muy poco tiempo se ha erigido como un área de investigación fundamental para la resolución de un gran número de casos judiciales. La genética forense tiene entre sus muchos objetivos, el de identificar individuos en base al análisis de los polimorfismos del ADN, o por ejemplo hacer inferencias que tienen repercusiones importantes en la resolución de otros muchos casos judiciales (análisis biológico de la paternidad) o que pueden representar una

ayuda importante para la investigación policial (inferencia del origen geográfico de un perfil genético o de las características físicas del donante de una muestra de ADN).

Remontándonos a la historia de los polimorfismos, el primer locus de ADN polimórfico fue descrito por Wyman and White (1980) cuando, usando una sonda de ADN arbitraria observó en geles de agarosa fragmentos de más de 15 longitudes diferentes en una pequeña muestra de individuos. Unos años más tarde se reportaron otros *loci* hipervariables en la secuencia del gen de la insulina humana, específicamente en el oncogén “ras”, en el pseudogen de la zeta-globina y en el gen de la mioglobina por (Durrin *et al.* 1984). Estos *loci* hipervariables constaban de repeticiones en tándem de una secuencia de oligonucleótidos (11 a 60 pb), de manera que las diferentes longitudes de los fragmentos originados dependían del número de dichas repeticiones y se les denominó VNTR (*Variable Number of Tandem Repeat*). Tras la descripción de los primeros VNTRs estos se aplicaron a la medicina forense y sustituir a los marcadores clásicos.

Para aquel tiempo la genética poseía las herramientas necesarias para el estudio del ADN, sin embargo la aplicación en la resolución de casos judiciales no se produjo sino hasta 1985, cuando el Ministerio del Interior Británico solicitó la ayuda de Alec J. Jeffreys, profesor de genética de la Universidad de Leicester, para identificar al agresor sexual y asesino de dos chicas en Narborough (Leicestershire) usando la técnica “DNA fingerprint” o “huella genética multilocus”. Jeffreys encontró secuencias en ciertas regiones del ADN, que consistían en un determinado número de repeticiones en tándem de una secuencia central, conocidas como VNTRs; también observó que el número de secciones repetidas presentes en una muestra podía diferir de un individuo a otro. La técnica usada para examinar los VNTRs se le denominó RFLP (*Restriction Fragment Length Polymorphism*) porque implica el uso de enzimas de restricción que cortan en determinadas regiones que flanquean a los VNTRs. Estas serían las bases del sistema de identificación por “DNA fingerprint” desarrollado por Jeffreys.

Posteriormente, el siguiente avance tecnológico de gran relevancia para la genética forense así como para otros muchos campos fue el descubrimiento de la reacción en cadena de la polimerasa (PCR) por *Kary Mullis* en 1987, técnica que permite la amplificación de secuencias específicas del ADN de manera cíclica.

Con la implementación de la PCR y la posterior descripción de los STRs (*Short Tandem Repeats* - polimorfismos caracterizados por altos niveles de diversidad y simplicidad de tipado), actualmente se pueden realizar los análisis de identificación con gran rapidez amplificando simultáneamente varios STRs en la misma reacción de PCR (PCR Multiplex).

Así mismo, más recientemente el campo forense ha incluido el análisis de polimorfismos simples denominados SNP (*Single Nucleotide Polymorphism*) que a pesar de tener un poder de discriminación inferior al de los STRs, permiten complementar en algunos casos la información proporcionada por los STRs.

Finalmente, la genética de poblaciones ha recurrido al uso de estos polimorfismos para poder observar la diversidad genética en las poblaciones y analizar cómo éstas cambian a través del tiempo. Esta tesis se centrará fundamentalmente, pero no exclusivamente, en las aplicaciones de los STRs y SNPs del cromosoma Y sobre todo en estudios genético poblacionales y forenses.

2. El cromosoma Y

El cromosoma Y es una molécula acrócentrica de ADN de aproximadamente 60 Mb de longitud y que representa alrededor del 2% del genoma humano. Desde la citogenética se puede dividir a este cromosoma en dos grandes regiones, la heterocromática su proporción de mayor tamaño está localizada en el brazo largo distal (Yp) con una longitud de ~30 Mb y una pequeña región de 1Mb localizada en centrómero. La segunda región está representada por la región eucromática que tiene un tamaño aproximado de 23 Mb, dividida en un fragmento de 8 Mb el brazo largo Yp y en otro de ~14.5 Mb en el brazo corto Yq (**Figura 1b**). Esta última región es de principal interés ya que está compuesta por secuencias altamente repetidas, como son DYZ1 y DYZ2, que no son codificantes.

Respecto a la región eucromática, es importante anotar que esta región contiene secuencias homólogas al cromosoma X, además de secuencias altamente polimórficas, como también de genes codificantes y funcionales, para este cromosoma (Skaletsky *et al.* 2003). El cromosoma Y presenta una ausencia de recombinación en prácticamente su totalidad, excepto en las regiones PAR (Regiones Pseudoautosómicas) situadas en ambos extremos del cromosoma; la primera es PAR1 situada en el brazo corto (Yp) con ~2.6 Mb de longitud y la segunda es PAR2 en el brazo largo (Yq) con una longitud de ~0.32 Mb. (**Figura 1a**). La porción localizada entre estas dos regiones, fue denominada en un principio NRY (*non recombining region*) por lo que esta región cromosómica se hereda en bloque, constituyendo un grupo de ligamiento.

Sin embargo, algunos investigadores han observado que en esta región se producen numerosos eventos de recombinación intracromosómica (Skaletsky *et al.* 2003), y de esta manera a esta región comprendida entre las dos regiones pseudoautosómicas, ha pasado a denominarse más acertadamente, *Male Specific Region of the Y chromosome* (MSY).

Con relación a la región MSY, esta se encuentra compuesta de regiones de heterocromatina y eucromatina (Foote *et al.* 1992, Skaletsky *et al.* 2003). La región de heterocromatina contiene largas secuencias repetitivas y está presente en 3 regiones cromosómicas: (I) en la región centromérica, (II) en la parte distal del brazo largo con un tamaño variable y (III) en la parte proximal del brazo largo (**Figura 1b**).

Con respecto a la región eucromática, ésta se ha categorizado en tres clases de secuencias distintas, conocidas como: los X-transpuestos, los X-degenerados y los segmentos amplicónicos (**Figura 2**).

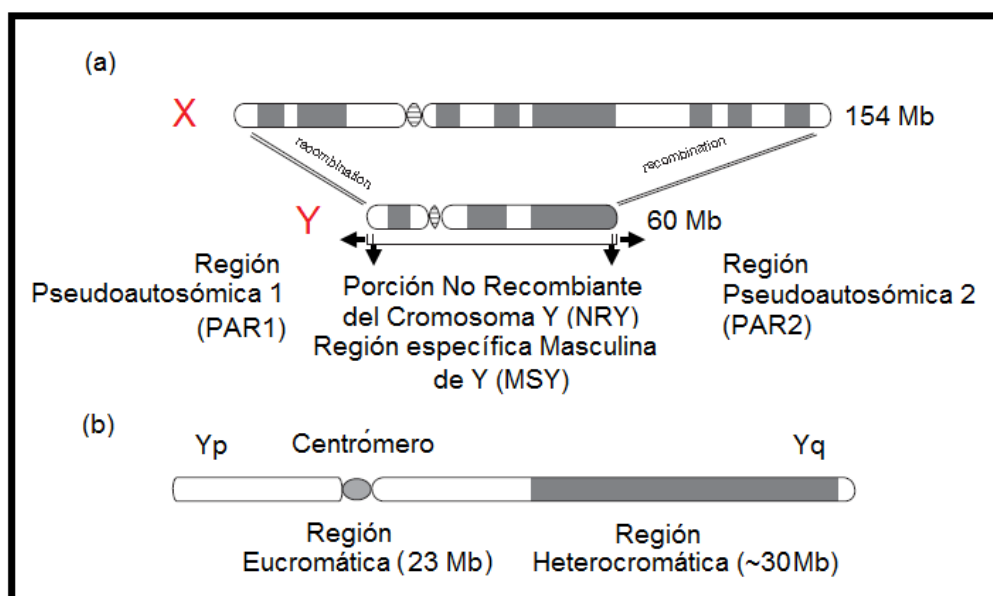


Figura 1. (a) Representación esquemática de los Cromosomas sexuales X e Y; en los extremos del cromosoma Y, se encuentran las regiones pseudoautosómicas, PAR 1 y PAR 2, y el resto del cromosoma lo ocupa la MSY o NRY. (b) El Cromosoma Y está compuesto de dos grandes regiones, una eucromática y otra heterocromática (Tomado y modificado de (Butler 2005)).

Las secuencias X-transpuestas, tienen un 99% de homología con el cromosoma X, y se cree que son el resultado de la trasposición entre los cromosomas X e Y. Estas secuencias tienen una longitud de alrededor 3.4 Mb y de las clases de secuencias eucromáticas descritas

anteriormente, esta clase es la de menor densidad génica, conteniendo solo 2 genes (Skaletsky *et al.* 2003).

Las secuencias X-degeneradas presentan una homología de menor grado respecto al cromosoma X que las secuencias X-transpuestas, ya que se ha sugerido que estas regiones representan remanentes de los autosomas a partir de los que X e Y co-evolucionaron.

Con respecto a los segmentos amplicónicos, se ha descrito que estos segmentos tienen más de un 99.9% de similitud con otras secuencias de la región MSY. Estas secuencias están dispersas a lo largo del brazo largo y la parte proximal del brazo corto del cromosoma Y con una longitud aproximada de 10.2 Mb. La importancia de esta región, radica en la presencia de ocho palíndromos, de los cuales seis se han reconocido como genes que codifican a proteínas que se expresan específicamente en los testículos.

Finalmente, en la eucromatina se expresan al menos 27 proteínas distintas y familias de genes, entre los cuales 12 son expresados de manera ubicua y 22 son expresados exclusivamente o mayormente en los testículos.

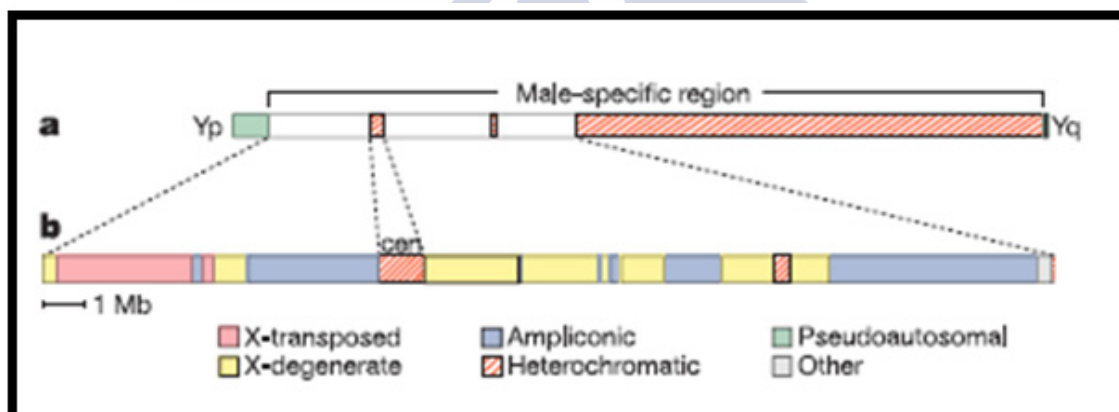


Figura 2. (a) Diagrama Completo del Cromosoma Y (b) Detalle de la eucromatina de la región MSY y las clases de repeticiones. La barra 1 Mb indica la escala del diagrama (Tomado y modificado de (Skaletsky *et al.* 2003))

2.1. Herencia

A excepción de las regiones PAR 1 y PAR 2 (regiones pseudo-autosómicas) el cromosoma Y no recombina, por lo tanto se comporta como una entidad haploide que permite determinar un linaje paterno que se hereda fielmente de padre a hijo y de hijo a todos sus descendientes varones y así sucesivamente, exclusivamente por línea paterna; esto permite que se guarde su historia genética intacta, en forma de haplotipo. Esta herencia se mantiene intacta a no ser que ocurra un evento mutacional (únicamente intra-alélico), que de por sí es la única fuente de variación en el cromosoma. (Jobling *et al.* 1997)

El tamaño de población efectiva del cromosoma Y es un cuarto comparado con el de los cromosomas autosómicos y un tercio del cromosoma X. Por esta razón, este cromosoma es más susceptible al efecto de la deriva genética (Jobling and Tyler-Smith 2003); este hecho, es el que permite acelerar la manera en la que los cromosomas se diferencian unos de otros y como resultado generar diversidad entre las poblaciones. Como consecuencia, el análisis del cromosoma Y resulta muy útil para estudiar eventos demográficos que hayan ocurrido durante la historia las poblaciones humanas (Cinnioglu *et al.* 2004). Adicionalmente, las poblaciones humanas frecuentemente acumulan una mayor diferenciación genética a nivel del cromosoma Y, permitiendo así de esta manera el mejor entendimiento de los orígenes de los humanos modernos (Jobling and Tyler-Smith 1995).

Paralelamente a los patrones descritos de herencia del cromosoma Y, la variación genética de las poblaciones también puede acumular diferencias cuando se tiene en cuenta el componente migratorio. El fenómeno de patrilocalidad, es la explicación que diversos autores han propuesto para explicar la mayor diferenciación genética que existe en las poblaciones para el cromosoma Y con respecto al ADN mitocondrial (ADNmt) (Oota *et al.* 2001). Refiriéndonos a este fenómeno de migración, podemos mencionar que la patrilocalidad, se refiere al fenómeno que ocurre cuando un núcleo familiar se mantiene geográficamente cerca del mismo linaje masculino a través del tiempo, o también explicado cuando la mujer se traslada a la residencia de su pareja; y matrilocalidad, se refiere a la situación inversa, donde un núcleo familiar se mantiene cerca de la misma área del linaje femenino o el hombre se traslada a la residencia de su pareja.

Se ha estimado que aproximadamente en el 70% de la población mundial se practica la patrilocalidad (Burton *et al.* 1996). Esto implica que el ADN mitocondrial, se tiende a desplazarse principalmente entre las poblaciones con el paso de las generaciones, mientras la variabilidad del cromosoma Y tiende a permanecer más estable dentro de cada población. Teniendo en cuenta que los cambios migratorios, resultan en la reducción de la diferenciación poblacional, esto parece explicar el por qué el grado de diferenciación genética entre las poblaciones en el cromosoma Y es mayor, en comparación con las de ADN mitocondrial; esto según algunos autores (Seielstad *et al.* 1998, Kayser *et al.* 2001), se debe a la alta tasa de migración femenina en el pasado. Estos trabajos concluyen que, las pequeñas migraciones tienen mayor peso, desde el enfoque de los patrones geográficos de la diversidad genética, que por procesos migratorios asociados al recorrido de grandes distancias. Teniendo presente lo anterior, tanto las migraciones como el bajo tamaño de población efectiva de cromosoma

Y, nos permite entender la baja variabilidad que existe dentro de una población para este cromosoma (Oota *et al.* 2001).

En último lugar, el hecho de que este cromosoma tenga especificidad masculina y su herencia se presente sin recombinación, en la mayoría de los casos, permite que este cromosoma sea de alto interés para la genética forense. Casos donde la paternidad no es clara por la ausencia del presunto padre y se trate de hijos varones, el estudio de la línea paterna es muy útil para poder resolver este estilo de casos. Igualmente esta característica es de utilidad en el área de criminalística donde el análisis de material genético masculino en casos de agresión sexual, permiten discriminar el ADN masculino de mezclas hombre-mujer y poder lograr identificar el linaje perteneciente al individuo sospechoso.

2.2 Polimorfismos del cromosoma Y

Se han descrito distintos tipos de polimorfismos en el cromosoma Y entre ellos, duplicaciones, inversiones, inserciones/deleciones (Indels), Single Nucleotide Polymorphisms (SNPs) (polimorfismos de un solo nucleótido), Microsatélites o STRs y recientemente RM-YSTRs o Y-STRs de alta tasa de mutación (*Rapidly Mutating Y-STRs*).

Éstos polimorfismo pueden ser categorizados dependiendo de su tasa mutacional, en (I) marcadores lentos o binarios que incluyen a los Indels y a los SNPs y (II) en los de mutación rápida o multialélicos, que incluyen los minisatélites, los STRs y los RM Y-STRs (Jobling and Tyler-Smith 1995, Jobling and Gill 2004, Ballantyne *et al.* 2010, Ballantyne *et al.* 2012).

En los diferentes estudios presentados en esta tesis se han efectuado análisis del cromosoma Y, mediante el tipado de marcadores binarios tipo SNPs y marcadores tipo microsatélites Y-STRs y RM Y-STRs. A continuación se describirá de manera general los marcadores empleados en el presente trabajo.

2.2.1 Polimorfismos binarios Y-SNPs

Los polimorfismos tipo SNPs son sitios en el ADN que cambian en una sola base y se consideran polimórficos si el alelo de menor frecuencia se encuentra entre el 1 y 5 % de la población, por lo que no es correcto considerarlos como mutaciones raras. La tasa mutacional de estos polimorfismos se ha estimado en 10^{-8} eventos mutacionales por generación en comparación con tasas de 10^{-3} - 10^{-5} en regiones microsatélites (Thorisson 2003), lo que los hace más estables en términos de herencia; por esta razón se han considerado como eventos

que surgen una única vez en la evolución, acuñando el término en inglés, “*unique events polymorphisms*” (UEPs).

Debida a esta baja tasa de mutación y ausencia de recombinación en el cromosoma, los SNPs de cromosoma Y (Y-SNPs) son útiles en la identificación de linajes masculinos y en la reconstrucción de *pools* genéticos que permiten profundizar en la evolución e historia de las poblaciones humanas. Los haplogrupos o grupos de linajes relacionados de Y-SNPs, son generalmente más frecuentes en los lugares geográficos de origen, por lo que son de particular interés en estudios poblacionales, para los análisis de flujo genético, mezcla poblacional y eventos de subestructura poblacional (Onofri *et al.* 2006).

Hasta la fecha, un gran número de polimorfismos binarios se han descrito para cromosoma Y, es por esto que desde 2002 se han hecho esfuerzos para estandarizar la nomenclatura de los SNPs de cromosoma Y. En concreto, la primera aproximación fue por parte del Y Chromosome Consortium (YCC) cuando publicó el árbol filogenético de máxima parsimonia de cromosoma Y que incluyó 153 haplogrupos pertenecientes a 18 clados, basados en 245 marcadores bialélicos (Y-Chromosome-Consortium 2002). Luego, de la misma manera, en 2003 Jobling y Tyler-Smith publicaron una versión actualizada del árbol filogenético e introdujeron pequeñas modificaciones (**Figura 3**).

Aun así, solo hasta el año 2008 y aún vigente, Karafet y cols., introdujeron la última actualización del árbol filogenético en una versión revisada que cuenta con 311 haplogrupos dentro de 20 clados principales, que incluye 2 nuevos clados de 2 nuevos haplogrupos, el S y T y alrededor de 600 marcadores bialélicos (**Figura 4**). Sin embargo cabe anotar que el portal YCC (<http://www.isogg.org/tree/>), ha permanecido en la actualización de la nomenclatura desde 2008, publicando en línea cada año una versión revisada de la filogenia. A pesar de esto, para los efectos de esta tesis se usará la nomenclatura sugerida por Karafet y colaboradores, ya que hasta el momento de la escritura de este trabajo es la nomenclatura oficial de la base de datos <http://www.yhrd.org>, usada en genética forense.

2.2.1.1 Nomenclatura

La nomenclatura actual para los polimorfismos tipo SNP consiste, en la estructura inicial que propuso el YCC en 2002, donde la agrupación de linajes o Haplogrupos que comparten un estado derivado de uno o varios marcadores polimórficos, se asigna a una división principal del árbol filogenético (clados). El diseño de este árbol de máxima parsimonia no fue complicada, ya que los marcadores están localizados, como se ha mencionado anteriormente,

en una región no recombinante, de tasa de mutación baja y existe poca presencia de mutaciones recurrentes; así mismo, la asignación de la raíz del árbol filogenético de cromosoma Y, se ha llevado a cabo mediante la comparación de las regiones homologas *MSY* de especies cercanas (chimpancés, gorilas y orangutanes), las cuales se han secuenciado y determinado el estado ancestral de los polimorfismos.

Respecto al nombramiento de los haplogrupos, se ha establecido la regla de usar letras mayúsculas; esta asignación comienza por la letra A, que corresponde al haplogrupo teóricamente más antiguo hasta la letra T. Aquellos linajes que no son definidos en base a un carácter derivado, pero pertenecen a un clado, pero aun así a ningún sub-clado, son haplogrupos potencialmente parafiléticos; éstos en la nomenclatura se distinguen con un Asterisco (*) y se han denominado paragrupos. Igualmente la nomenclatura actual ha mantenido la flexibilidad respecto a los 2 sistemas de nomenclatura existentes; el primer sistema consiste en la definición de sub-clados dentro de un haplogrupo por medio de un código alfanumérico (por ejemplo, E1, E1a, E1a1, etc.). El segundo, es el sistema corto o abreviado que consiste en nombrar el haplogrupo por la mutación terminal que esta define (por ejemplo, E-M81). A su vez, los paragrupos son marcados con asterisco siguiendo esta misma regla (E*, E1a*, E1a1*, etc.). Finalmente, para casos en los que no se haya tipado todos los marcadores de un clado, se recomienda indicar los linajes o haplogrupos ausentes usando una x, que significa “excluidos”; esta última regla puede ser usada indistintamente para los 2 sistemas mencionados anteriormente (por ejemplo, P*(xQ*); P*- 92R7(xQ*-M242)).

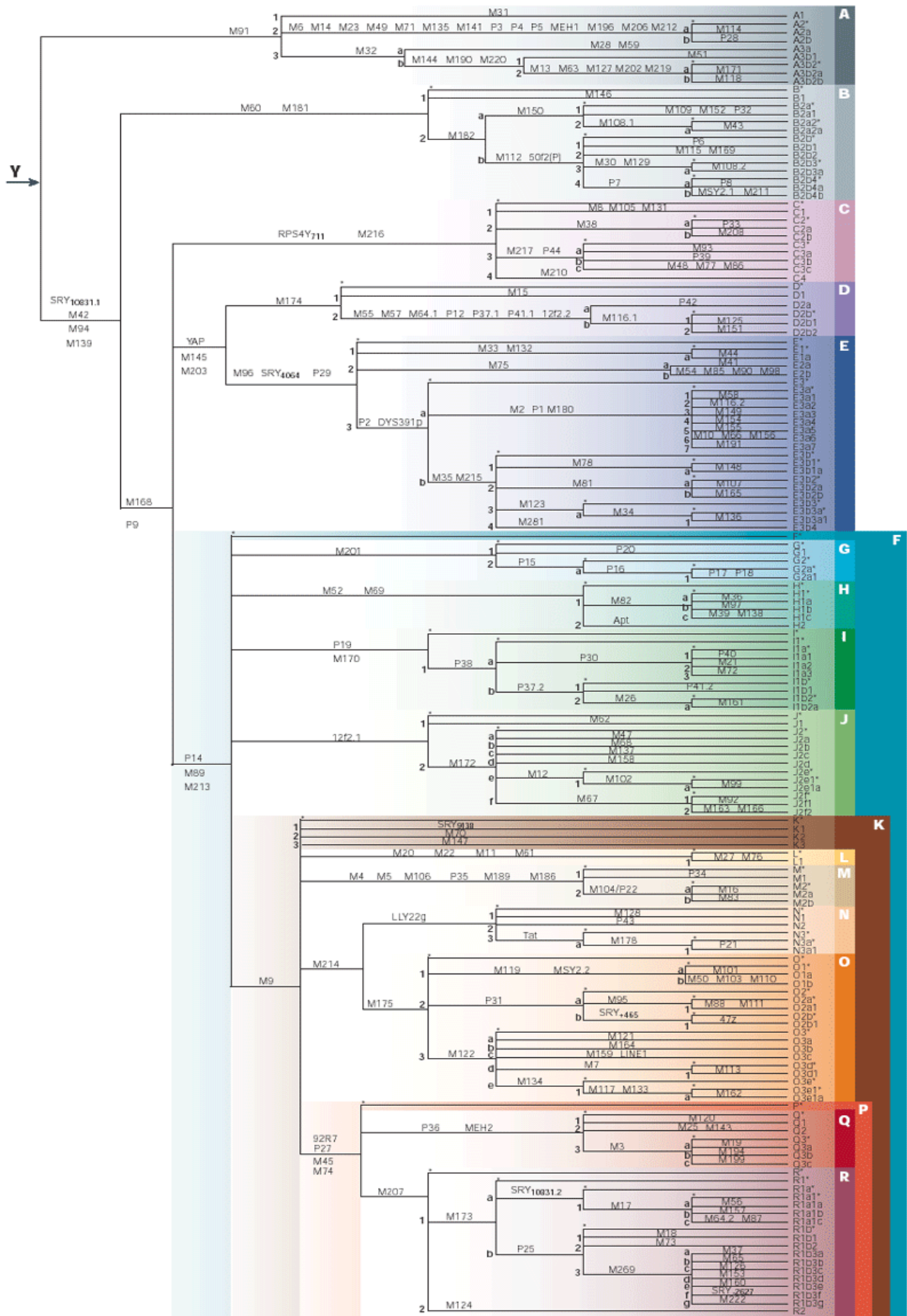


Figura 3. Árbol filogenético de los 18 haplogrupos principales de Cromosoma Y (A-R); los eventos mutacionales se presentan en las ramas del árbol (Tomado y modificado de (Jobling and Tyler-Smith 2003))

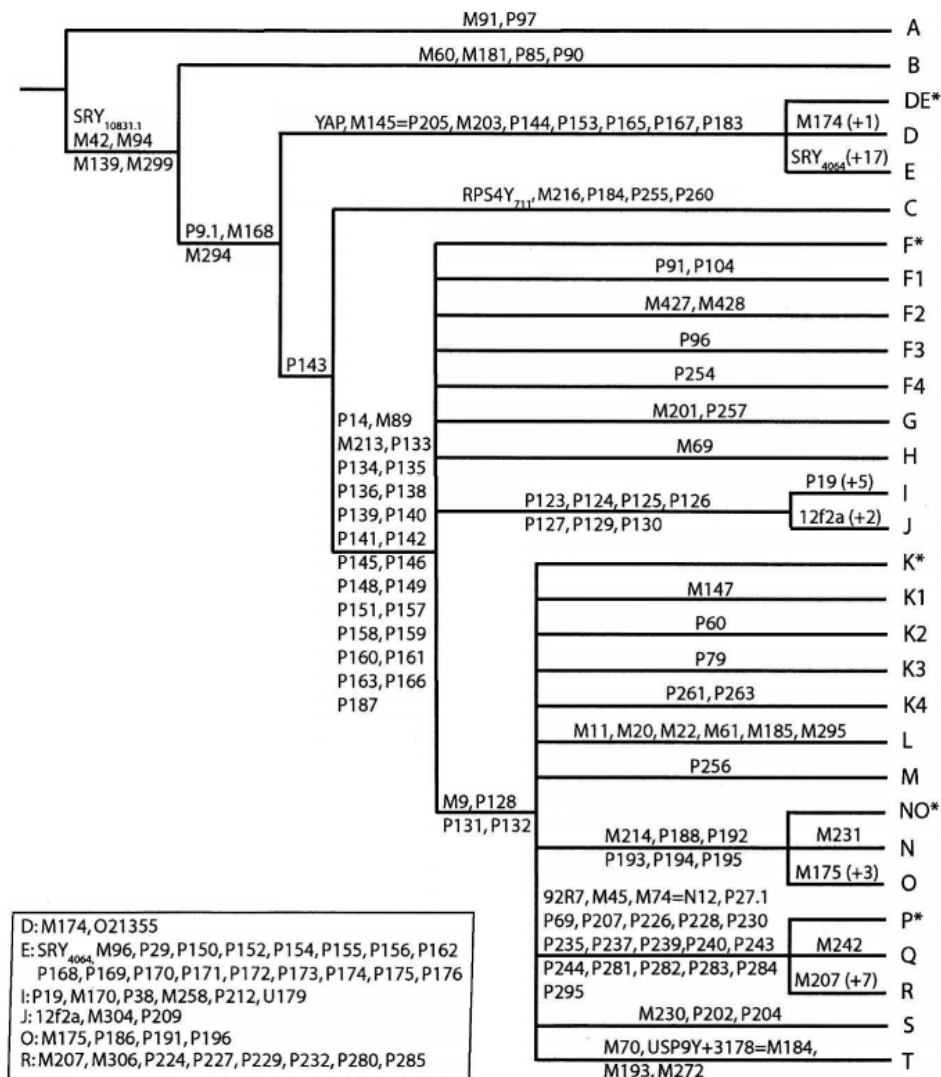


Figura 4. Mapa actual del árbol de máxima parsimonia, donde se resaltan los 20 clados o Haplogrupos mayores (A-T) (Tomado de (Karafet *et al.* 2008))

2.2.1.2 Diversidad de los haplogrupos de cromosoma Y

Seguindo las huellas de los distintos haplogrupos del árbol filogenético presentados por Jobling con Tyler-Smith y el de Karafet y colaboradores y analizando la presencia de éstos en las diferentes localidades actuales, es posible descifrar la distribución geográfica de los haplogrupos. De esta manera, a partir de los movimientos poblacionales que han ocurrido desde y en África a lo largo de la historia de la humanidad, como se ha sugerido por los autores, pueden reconocerse los siguientes haplogrupos en las siguientes localidades (**Figura 5**) (Underhill *et al.* 2001, Jobling and Tyler-Smith 2003, Karafet *et al.* 2008):

- El Clado A o haplogrupo A, es mayormente frecuente en poblaciones etíopes, sudanesas y de población khoisánida.
- El haplogrupo B es muy frecuente en los pigmeos africanos, siendo algunos sublinajes casi restringidos a ese grupo étnico.

- Se ha sugerido que el haplogrupo C ha tenido origen en Asia, y es frecuentemente encontrado en poblaciones Asiáticas, Australiana y oceánicas, y últimamente se ha reportado en África Subsahariana.
- El haplogrupo D se ha descrito para en el Sureste y parte Central de Asia y entre las poblaciones de Japón y las Islas de Andamán.
- El haplogrupo E se encuentra ampliamente disperso a través de África, Oriente Medio, Europa del Sur y Asia (Underhill *et al.* 2001, Cruciani *et al.* 2002).
- En el subcontinente hindú, en los países bajos, en Asia del este y Sri Lanka se ha descrito el haplogrupo F;
- El haplogrupo G tiene una muy baja frecuencia pero se encuentra distribuido entre el Mediterráneo, Oriente Medio y las montañas del Cáucaso;
- El haplogrupo H se encuentra restringido al subcontinente hindú;
- El haplogrupo I, es particularmente uno de los 2 mayores haplogrupos europeos, y se encuentra distribuido en toda Europa y aparentemente está ausente en otros continentes;
- El haplogrupo J se encuentra en altas frecuencias en, Oriente Medio, el norte de África, Asia central, Pakistán e India, estando J2 (definido por M172) presente en Europa, y J1 (definido por M267) presente en el norte de África y Oriente Medio;
- El haplogrupo K, se ha observado en India, Oceanía, Indonesia y Australia; Es importante aclarar que según la nueva nomenclatura, el haplogrupo K2 es ahora haplogrupo T;
- El haplogrupo L se encuentra en el subcontinente hindú en muy alta frecuencia;
- En Oceanía y en Indonesia del Este se encuentra presente de manera restringida el haplogrupo M;
- El haplogrupo N es principalmente encontrado en el noreste de Eurasia;
- El haplogrupo más frecuente en Asia del este es el haplogrupo O;
- El haplogrupo P está dividido en dos linajes principales, Q y R; el haplogrupo Q se encuentra distribuido a través del norte de Eurasia, y Q3 (definido por M3) está restringido a las Américas. El haplogrupo R incluye la mayoría de los linajes de cromosoma Y en las poblaciones europeas.
- El haplogrupo S se ha encontrado en Oceanía e Indonesia;
- El haplogrupo T ha sido observado en Oriente Medio, África y Europa en baja frecuencia (Karafet *et al.* 2008).

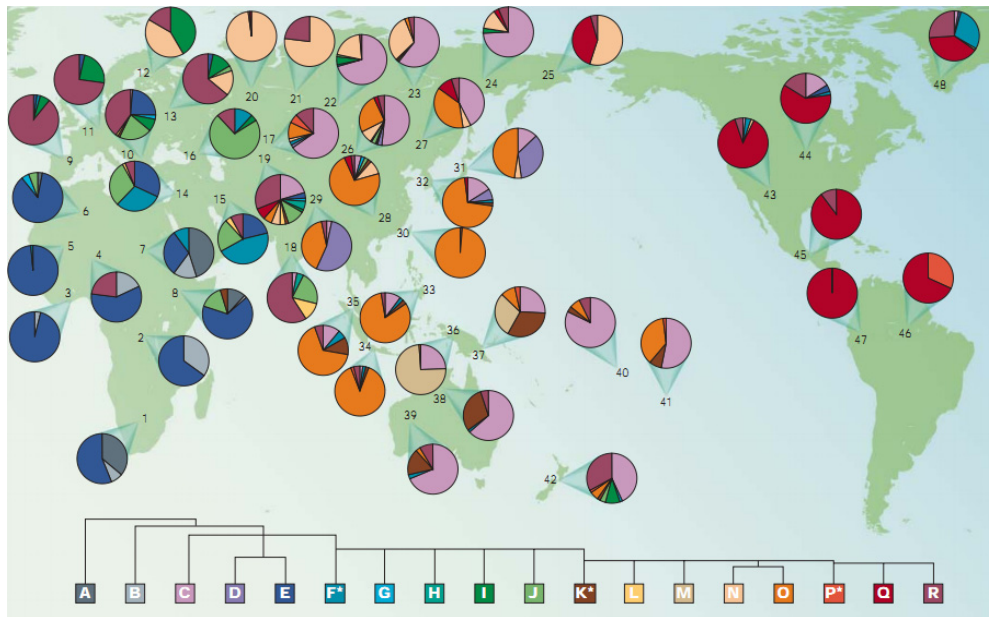


Figura 5. Distribución geográfica global de los 18 haplogrupos principales de Cromosoma Y (Jobling y Tyler-Smith, 2003). La distribución demuestra una tendencia migración de los haplogrupos desde África hacia América y Oceanía. (Poblaciones tomadas de Jobling y Tyler-Smith, 2003)

Así de esta manera puede concluirse que los haplogrupos con mayor distribución son el A, B y E; éstos dos primeros son haplogrupos exclusivos de África Subsahariana también presentes en algunas otras poblaciones pero con muy baja frecuencia (Jobling and Tyler-Smith 2003). Estos dos haplogrupos (A y B) representan los haplogrupos más antiguos del árbol (Underhill *et al.* 2001, Semino *et al.* 2002). El patrón de distribución de estos muestra la temprana diversificación de ellos desde y en África, explicado por el *out of africa*.

Respecto al haplogrupo E, la mayoría de las poblaciones africanas presentan sub-haplogrupos que pertenecen a este haplogrupo; los haplogrupos más basales dentro de este clado son el E1 y E2, definidos por las mutaciones P147 y M47 respectivamente.

En Europa, los cuatro haplogrupos más frecuentes son el E, I, J y R, representando el 80% de los haplogrupos de esta población. Las dos principales ramas del haplogrupo R (R1a y R1b1) muestran frecuencias opuestas, siendo R1b el más frecuente en Europa occidental y R1a el más frecuente en Europa oriental; se ha sugerido que este fenómeno tiene relación con las últimas glaciaciones (10.000 mya).

En el caso del continente americano, se han descrito dos linajes fundadores representados en las poblaciones nativas americanas. Los haplogrupos son definidos por el marcador M3 que representa al haplogrupo Q (también descrito como Q1a3a) y por el Y-SNP M217 que representa al haplogrupo C3* (Zegura *et al.* 2004, Geppert *et al.* 2011)

Por último, los métodos que son usados para la datación filogenética de estos haplogrupos del cromosoma Y, consideran la diversidad intrínseca de los Y-STRs presentes en cada linaje masculino en cuestión. Debido a que los SNPs son “*Unique Event Polymorphisms*”, el estudio de Y-STRs estará relacionado en términos de variabilidad genética con los haplogrupos, es decir, a medida que sea más antiguo el linaje, mayor será la variabilidad genética de los Y-STRs.

Por último, la presencia diferencial de haplogrupos específicos de cada población, tiene sin lugar a dudas una gran utilidad forense ya que en términos de identificación, esto puede ser una herramienta para determinar el origen geográfico a la que puedan pertenecer, restos óseos, partes corporales producto de un desastre masivo o evidencias donde estén involucradas personas de distintas procedencias geográficas.

2.2.2 STRs del cromosoma Y

Los STRs o microsatélites son polimorfismos, como se ha descrito anteriormente, de la categoría de tasa de mutación alta o rápida. Estos marcadores, también son clasificados entre los polimorfismos de longitud, debido a su estructura; su frecuencia es alta en el cromosoma Y (más de 220 identificados), sin embargo son muy pocos los usados en la literatura. La distribución de estos marcadores es amplia a lo largo del genoma, con promedio de un STR por cada 5000 o 10000 pares de bases, entre regiones no codificantes del ADN como en regiones génicas e intergénicas.

Respecto a su estructura molecular los STRs son *loci* cuya secuencia está compuesta por repeticiones en tándem de un mismo motivo repetitivo o secuencia núcleo; estos motivos de repetición varían dependiendo del STRs, pero generalmente son de 2 a 7 nucleótidos, y son estos motivos completos los que se repiten en tándem. La extensión total de estos polimorfismos puede variar entre 100 y 400 pb, las cuales incluyen los motivos de repetición *per se* y las regiones flanqueantes. Entre las características que los hacen apropiados para su uso están, que tienen un alto nivel de discriminación, que tienen un amplio rango alélico, que son reproducibles y que tienen baja tendencia a formar bandas tipo *stutter* o tartamudas. Estas últimas son bandas generadas por deslizamientos (*Slippage*) de la polimerasa durante la amplificación, éstas diferencian el alelo real por uno con unidad de repetición; este fenómeno es muy frecuente en STRs de motivos de repetición de 2 nucleótidos o dinucleótidos.

Así mismo, el número de repeticiones de los STRs puede variar entre los individuos, lo cual los hace apropiados como herramienta para la identificación humana. Adicionalmente estos marcadores presentan un grado de diversidad mucho mayor en comparación con los polimorfismos binarios, como se ha mencionado anteriormente, debida su alta tasa de mutación estimada, entre 10^{-3} y 10^{-4} por locus por generación. Así mismo por la naturaleza haploide del cromosoma Y, todos los alelos Y-STRs de un individuo son heredados en bloque; Es esta transmisión hereditaria, la que ha de permitir definir el término haplotipo, que se refiere a la distinta combinación y organización de diferentes *loci* Y-STRs en un individuo dentro de una población que son heredados en bloque; por el alto grado y naturaleza recurrente de mutación en los Y-STRs, los haplotipos pueden ser idénticos por descendencia o por estado, y pueden generarse más de una vez y hallarse en distintas poblaciones, a diferencia de los haplogrupos, que mayormente son típicos de una región en particular (Ver apartado anterior).

Por el gran número de Y-STRs descritos, se ha considerado necesario seleccionar un set de *loci* universal para toda la comunidad científica para identificar un individuo, el cual se ha denominado el haplotipo mínimo (*Minimal Haplotype - MHT*), frecuentemente útil para las bases de datos forenses; entre ellos se cuentan 9 sistemas STRs o microsatélites (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392 y DYS393). La ocurrencia de un haplotipo en particular puede ser interrogada, respecto a su frecuencia en una población específica, si con anterioridad se ha estudiado la población en cuestión y se ha determinado la frecuencia de los diferentes haplotipos presentes en la misma. También puede consultarse la frecuencia del haplotipo en cuestión cuando se crea una base de datos que contenga muchos haplotipos descritos para distintas poblaciones. Este es el caso de la base de datos consultada por la comunidad forense denominada la “Y-STR Haplotype Reference Database” que se encuentra disponible en línea en el sitio web <http://www.yhrd.org> la cual permite establecer la frecuencia del haplotipo en la población e incluso el posible origen geográfico del mismo.

El uso de los Y-STRs en genética forense tiene un papel muy importante, principalmente cuando se trata de casos de agresión sexual o casos de paternidad o cuando el presunto el presunto padre está ausente. (Jobling *et al.* 1997). De esta misma manera, también son útiles, en estudios genealógicos (Jobling 2001), y en estudios evolutivos de los linajes masculinos (Jobling and Tyler-Smith 2003).

Adicionalmente, los microsatélites de cromosoma Y, pueden ser usados para dos fines, 1) distinguir linajes; que estará sujeto al número de marcadores usados y la variabilidad que estos expresen en términos de discriminación y 2) para obtener información acerca de relaciones entre linajes; lo cual dependerá, igualmente, del número de marcadores y el grado de consistencia de éstos cuando se hagan las inferencias. (Kayser *et al.* 2004)

Como se ha mencionado anteriormente, las bases de datos usan un set de 9 *loci*, sin embargo para algunos estudios este número de marcadores no es suficiente; por ejemplo en poblaciones pakistaníes el 14% de la población comparte un mismo haplotipo de 16 Y-STRs (Mohyuddin *et al.* 2004), Así mismo en gran parte de Asia, aproximadamente el 4% de los individuos comparte el mismo haplotipo de 15 Y-STRs (Zerjal *et al.* 2003) y en población finlandesa, el 13% de los varones comparten el mismo haplotipo de 16 Y-STRs (Hedman *et al.* 2004); incluso en casos excepcionales, los haplotipos entre un padre y un hijo pueden ser diferentes (Kayser and Sajantila 2001). Por lo tanto para inferencias y resultados consistentes, tanto en estudios evolutivos como incluso de paternidad, un número mayor de microsatélites se aconseja usar, como también deben entenderse las tasas mutacionales para cada *locus*.

Respecto a la característica mutacional de los STRs, se han considerado diferentes procesos responsables de la variabilidad detectada en el número de unidades repetitivas de un *locus* determinado. El mecanismo principal con el que actualmente se cree que genera este tipo de variación es llamado el “*slippage*” replicativo o *Slipped Strand mispairing*. (Heyer *et al.* 1997, Brinkmann *et al.* 1998); Refiriéndonos al *slippage* replicativo, se puede decir que este es un fenómeno intracromosómico que tiene lugar en el proceso de replicación; al producirse un apareamiento incorrecto de las hebras complementarias de la doble hélice, se provoca un deslizamiento de una hebra sobre la otra. Por este deslizamiento, la ADN polimerasa incorpora las bases complementarias con un emparejamiento erróneo, formando un bucle que no tiene apareamiento. Por lo tanto, si el bucle de no apareamiento se produce en la hebra naciente que se está replicando, la secuencia repetitiva habrá ganado una unidad de repetición y se genera una inserción, por el contrario si este bucle se forma en la hebra molde, se producirá y se generará una delección (**Figura 6**) (Sia *et al.* 1997).

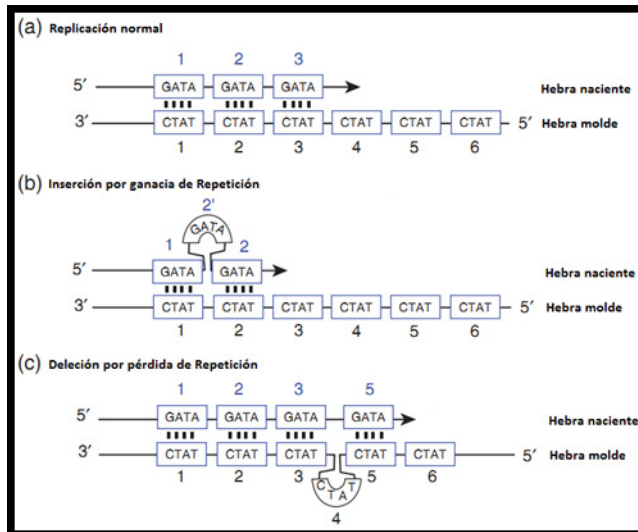


Figura 6. *Slippage* replicativo. a) Replicación normal sin pérdida ni ganancia de repetición. b) Inserción por ganancia de repetición en la hebra naciente. c) Delección por pérdida de repetición en la hebra naciente.

Respecto a lo anterior, el modelo mutacional que asume el paso mutacional de ganancia o pérdida de una unidad repetitiva en un *locus* STRs, se explica por el modelo de “*Stepwise Mutation*” (Ohta and Kimura 1973). Este modelo explica el comportamiento de la mayoría de los *loci* STRs, sin embargo otros autores también han propuesto otros modelos en los que puede haber una ganancia o pérdida de más de una unidad de repetición, aunque consideran que estas variaciones a gran escala son poco frecuentes (Di Rienzo *et al.* 1994).

Relacionando esto con en el apartado anterior, puede concluirse que es muy útil una buena estima de la tasa de mutación de los Y-STRs, porque es fundamental para dar soporte a las inferencias y datación de los linajes de cromosoma Y definidos por los Y-SNPs, como también para la interpretación de pruebas de pericia forense.

Respecto a esto último, a la fecha los estudios de tasa de mutación en Y-STRs, varían ligeramente en sus estimas; uno de los estudios más amplios (Gusmão *et al.* 2005), analizó 16 Y-STRs en 3026 casos padre/hijo y estimó la tasa de mutación en 1.998×10^{-3} un valor que es más o menos parecido al reportado en un estudio anterior (Kayser *et al.* 2000) con un valor de 2.8×10^{-3} y otro posterior a este, estimado en 2.31×10^{-3} (Dupuy *et al.* 2004). Más recientemente (Ballantyne *et al.* 2010), analizó 186 Y-STRs en cerca de 2000 casos confirmados padre/hijo, y obtuvo una estima de 3.35×10^{-3} por marcador por generación; igualmente en este estudio se encontraron una serie de marcadores que tienen una tasa de mutación por encima a 1×10^{-2} , a los cuales se denominaron RM-YSTRs (*Rapidly Mutating Y-STRs*) por la alta tasa de mutación. A pesar de los grandes esfuerzos que se han efectuado por estimar las tasas de mutación, continúa el debate sobre cuál debería ser la estima apropiada para ser usada en estudios evolutivos o genealógicos, lo cual podría suponer una

discrepancia como propone (Zhivotovsky and Underhill 2005); Aun así recientemente (Pinto *et al.* 2014), ha sugerido una serie de parámetros que deberían seguirse a la hora de estimar las tasas de mutación con fines de filiación, entre los que se incluye que en la casuística debe incorporarse una razón de verosimilitud específica para los alelos transitorios observados y que las estimas deben ir acompañadas de las frecuencias de transmisión absolutas de los alelos mutados y no mutados. De la misma manera (Wei *et al.* 2013) sugiere que para casos de estimas evolutivas, los modelos de predicción en datos re-secuenciados, demuestran que la tasa de mutación descrita por (Zhivotovsky *et al.* 2004) tienen una mayor correlación que otras estimas. De esta manera el debate continúa y parece indicar que dependiendo el análisis que se requiere hacer, evolutivo o genealógico, debe ser utilizada una u otra tasa mutacional.

2.2.2.1 Rapid Mutating Y-STRs

A pesar de lo útiles que resultan ser los Y-STRs, estos marcadores tienen algunas limitaciones en el campo de investigación forense a la hora de discriminar entre individuos; refiriéndonos a los Y-STRs tipados por el kit AmpFISTR® *Yfiler*® (Applied Biosystems) los marcadores en este multiplex en términos generales, muestran una diversidad haplotípica alta que en promedio está sobre 0.995 y en algunas poblaciones puede ser algo mayor (Mulero *et al.* 2006, Kovatsi *et al.* 2009), sin embargo la habilidad para discriminar individuos de un mismo linaje paterno o discriminar individuos en poblaciones donde haya ocurrido una rápida expansión poblacional precedida de un grupo reducido, o en poblaciones reducidas donde se practique la patrilocalidad, (factores que limitan la diversidad en el cromosoma Y, por alto índice de endogamia), la discriminación puede resultar un reto. Aun así, este panorama se ha abordado en muchos casos adicionando más *loci* microsatélite permitiendo una mayor diferenciación de linajes masculinos (Hanson and Ballantyne 2004, Hedman *et al.* 2004, Hanson and Ballantyne 2007, Rodig *et al.* 2008) y como resultado un haplotipo completo de Y-STRs (entre 25 – 50 *loci*) genera una mayor discriminación y resulta en una buena alternativa. Aun así, desde el punto de vista forense, estos marcadores continúan siendo incapaces de excluir parientes cercanos o lejanos de un mismo linaje paterno, en una mezcla de material biológico que incluya al sospecho; a esto puede sumarse que la tasa de mutación de los Y-STRs que se usan de rutina, no es suficientemente alta por *locus* por generación para generar variación entre familiares cercanos. Por lo tanto, desde el punto de vista del análisis forense surge la necesidad de poder usar las bondades del cromosoma Y, como es la de discriminar linajes masculinos, con la capacidad de poder

discriminar al mayor grado posible al punto de individualizar, con el fin de ser aplicados la casuística de agresión sexual o identificación humana.

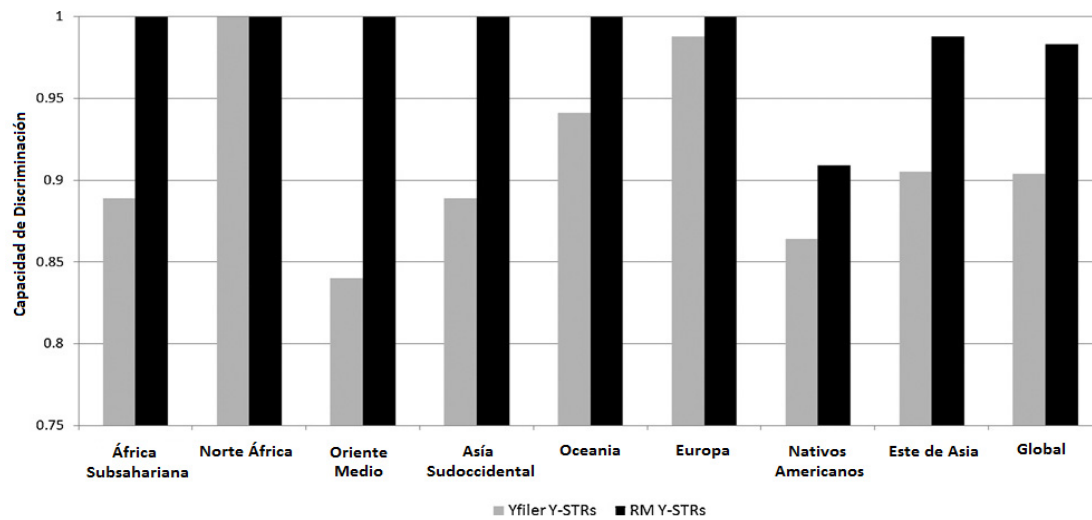


Figura 7. Capacidad de discriminación adquirida por los marcadores RM Y-STRs y *Yfiler*® en un grupo de 8 regiones geográficas y globales del panel HGDP-CEPH; el grupo de marcadores de alta tasa mutacional muestra un mayor poder de discriminación en 7 de las 8 regiones. Tomado de (Ballantyne *et al.* 2012)

En el estudio mencionado anteriormente de tasas mutacionales (Ballantyne *et al.* 2012), se identificaron un grupo de marcadores que mostraba una alta tasa mutacional mayor respecto a los Y-STRs del kit *Yfiler*® (AB). Este grupo está compuesto por 13 marcadores, denominados RM-YSTRs, tienen tasas de mutación por encima de 1×10^{-2} y en varias poblaciones del panel HGDP-CEPH han demostrado tener una diversidad haplotípica y una capacidad discriminatoria sustancialmente mayor (**Figura 7**), respecto a el kit comercial *Yfiler*. Adicionalmente estos marcadores, permiten obtener una mayor resolución, para la diferenciación de linajes paternos.

Este mismo panel, muestra menos subestructura poblacional dentro y entre las distintas regiones geográficas respecto al *Yfiler*, lo que se traduciría en mayor discriminación. Paralelamente, en un grupo de más de 300 individuos masculinos relacionados, se ha demostrado que estos marcadores, permiten separar y diferenciar, en un 66% de los casos, los individuos emparentados, por medio de eventos mutacionales, un valor que es más de cuatro veces mayor que el alcanzado por el kit mencionado anteriormente. Así mismo, este panel de marcadores tiene el poder suficiente para separar individuos estrechamente relacionados, como entre padres e hijos y entre hermanos en un casi 50% y 60% de los casos, respectivamente en comparación con los resultados obtenidos en *Yfiler*.

Por lo tanto, es visible que en un futuro cercano el análisis de RM-YSTRs en la rutina forense, pueda cambiar el panorama que ofrece el análisis del cromosoma Y actualmente, es decir podría incluirse la discriminación casi a nivel individual de parientes y/o sospechosos, gracias al alto nivel de mutación de sus marcadores.

En uno de los trabajos presentados en esta tesis se analizan estos novedosos marcadores de cromosoma Y, en una población de Bolivia y su uso forense.

2.3 Aplicaciones generales del cromosoma Y en genética forense

Es bien sabido que desde el punto de vista forense, los marcadores autosómicos, son los marcadores de primera elección cuando de resolver un caso judicial se trata, sin embargo con el paso de los años se han podido comprobar las bondades y ventajas que tiene el análisis de los polimorfismos del cromosoma Y; como se ha mencionado con anterioridad, este cromosoma es exclusivamente masculino y permite identificar linajes específicos de este origen, así mismo al tener un tamaño de población efectivo menor respecto a los marcadores autosómicos ($1/4$), lo hace más susceptible a el cambio aleatorio en las frecuencias haplotípicas por deriva genética y como resultado se produce una reducción de variabilidad de generación en generación en una población específica, lo que al mismo tiempo permite aumentar la diferenciación entre poblaciones. Es este aspecto anterior el que permite agrupar grupos de haplotipos en poblaciones definidas, es decir la frecuencia de un haplotipo puede estar ligada a una región geográfica y si existe un haplotipo masculino específico de una población esto podría permitir reducir el “abanico” de posibilidades en una investigación judicial. Por esta razón el portal <http://www.yhrd.org> se ha encargado de almacenar en una base de datos, bajo estándares estrictos de control de calidad, los haplotipos que se han descrito en la literatura, según su procedencia geográfica, organizando la información y como resultado permitiendo, que en términos judiciales ser una herramienta útil que puede ayudar a determinar el posible grupo étnico de una evidencia o de un sospechoso.

Adicionalmente, el uso de microsatélites de cromosoma Y es especialmente útil para la determinación de la paternidad, con individuos pertenecientes a la misma línea paterna y en casos donde el padre se halle ausente y se trate de hijos varones; así mismo en este tipo de situaciones, estos marcadores pueden ser analizados en todos los individuos masculinos de la misma línea paterna; cuando se obtiene una coincidencia entre perfiles, para casos como el descrito anteriormente, esta coincidencia dependerá directamente de la frecuencia de este haplotipo en particular en la población de referencia. Es de resaltar que este resultado no

excluye a ningún varón de esta línea paterna analizada y por lo tanto cualquier individuo de este linaje no se puede excluir de ser el padre biológico de la paternidad, es por esto que la individualización en muchas veces es un factor determinante para esclarecer con veracidad la paternidad.

Igualmente otra aplicación muy importante para estos marcadores es en la casuística criminal, en la que es muy común encontrar muestras en las que existe una mezcla de componentes masculino y femenino, típicas de agresiones sexuales. Para estos casos la identificación del linaje o el perfil masculino, es posible por la amplificación específica de marcadores de cromosoma Y en las evidencias, que en la mayoría de los casos se encuentra en una proporción muy baja. También son útiles cuando se tiene casos de violaciones por un individuo azoospermico, ya que el componente masculino es satisfactoriamente amplificado a partir de las células epiteliales del individuo aun estando en menor proporción que las de la víctima.

De la misma manera, en casos de violación múltiple, el análisis de los marcadores de cromosoma Y permite identificar un mínimo de agresores, e incluso es útil en la identificación de perfiles recurrentes en casos de violaciones en serie, tanto para relacionar como para excluir sospechosos. Actualmente la casuística en la mayoría de los laboratorios de genética forense que genotipan polimorfismos de cromosoma Y, está regulada por la oferta comercial por parte de dos grandes casas comerciales, que permiten la amplificación de un set de marcadores Y-STRs en kits; uno de ellos amplifica 17 sistemas en el caso del kit de Applied Biosystems AmpFLSTR® *Yfiler*®, y otro 12 sistemas microsatélite en el kit Powerplex® Y de Promega (**Figura 8**). Estos dos kits comparten entre sus marcadores, el “core” de los 9 loci que hacen parte del MHT, entre los cuales se encuentran los marcadores DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392 y DYS393.

Marcadores STR de Cromosoma Y comunmente usados para análisis forense

Haplotipo Mínimo (MHT)	Haplotipo Extendido	PowerPlex® Y	AmpFLSTR® Yfiler®
DYS19	DYS19	DYS19	DYS19
DYS385 a/b	DYS385 a/b	DYS385 a/b	DYS385
DYS389 I	DYS389 I	DYS389 I	DYS389 I
DYS389 II	DYS389 II	DYS389 II	DYS389 II
DYS390	DYS390	DYS390	DYS390
DYS391	DYS391	DYS391	DYS391
DYS392	DYS392	DYS392	DYS392
DYS393	DYS393	DYS393	DYS393
	DYS438	DYS437	DYS437
	DYS439	DYS438	DYS438
		DYS439	DYS439
			DYS448
			DYS456
			DYS458
			DYS635
			GATA H4

Figura 8. Y-STRs usados en la rutina forense. Con el fin de asegurar la compatibilidad entre los distintos grupos de marcadores, se han establecido, el haplotipo mínimo (MHT) y el extendido que incorpora los sistemas DYS438 y el DYS439, que es el recomendado por el SWGDAM (Scientific Working Group on DNA Analysis Methods; www.SWGDAM.org).

Debido a las necesidades forenses comentadas anteriormente, de individualizar y discriminar personas que pertenezcan a un mismo linaje masculino, la casa comercial Promega recientemente ha lanzado al mercado un nuevo kit que amplifica los 17 sistemas microsatélite presentes en el kit *Yfiler*® más 6 nuevos sistemas Y-STRs: DYS481, DYS533, DYS549, DYS643, DYS570 y DYS576, llamado PowerPlex Y23® (PPY23) (Promega Corporation Technical Manual PowerPlex® Y23 System Instrutions for use of products DC2305 AND DC2320 , Purps *et al.* 2014) (**Figura 9**). Estos dos últimos marcadores, DYS570 y DYS576 son sistemas RM-YSTRs, que como fueron mencionados en el apartado interior, son marcadores de alta tasa de mutación que pueden generar nuevos alelos de una generación a otra, lo cual permite discriminar y diferenciar individuos del mismo linaje en algunos casos. Varios ensayos con este nuevo kit han podido comprobar que en diferentes poblaciones el PPY23 tiene niveles de diversidad haplotípica y poder de discriminación mayor en comparación con los set de marcadores del MHT, del SWGDAM, del PowerPlex® Y 12 y del *Yfiler*® (Calderon *et al.* 2013, Coble *et al.* 2013, Davis *et al.* 2013). Entre los trabajos presentados en la presente tesis se encuentra el análisis del PPY23 para población Boliviana, el cual permite una mayor comprensión de los parámetros forenses para esta población y regiones cercanas.

Finalmente es importante resaltar, que para que todas estas estimas de inclusión (principalmente) y exclusión tengan validez es vital, una actualización constante y frecuente de la base de datos forense, (*i.e.* <http://www.yhrd.org>), en la que se incluyen los perfiles haplotípicos, bajo estrictos controles de calidad y organizados por regiones específicas. Estos controles son importantes debido a que la estima de las frecuencias haplotípicas y el posterior cálculo de probabilidad para los reportes forenses de los perfiles, dependen de estos factores.

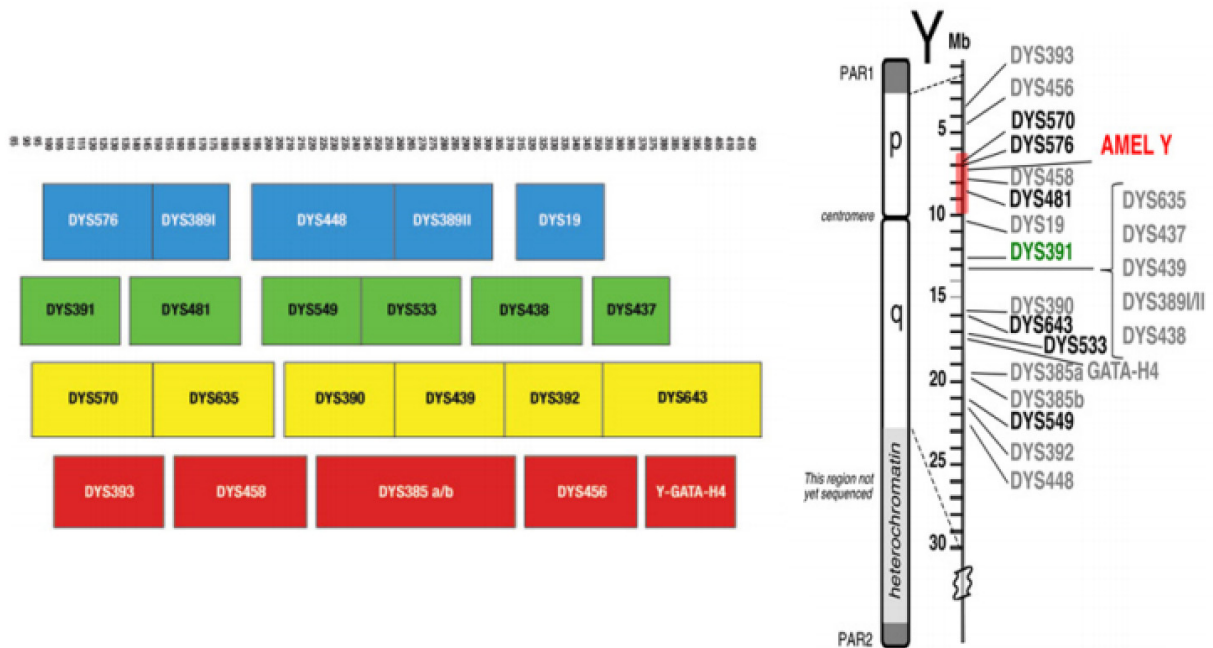


Figura 9. Marcadores incorporados en el kit PowerPlex® Y23 ® y su localización en el cromosoma Y.

Un ejemplo de identificación de perfiles específicos de una población y la actualización de la base de datos con fines poblacionales y forenses es presentado en uno de los trabajos de esta tesis, para la población Boliviana.

2.3.1 Estudios en Y-SNPs & Y-STRs en América del Sur

Desde el comienzo de la publicación masiva de las secuencias del genoma humano y de los polimorfismos tipo SNPs por los proyectos genoma humano (Lander *et al.* 2001, Venter *et al.* 2001) y HapMap (The HapMap Consortium 2003, The HapMap Consortium 2005), diferentes grupos han usado esta información para el diseño de múltiplexes con aplicaciones clínicas (Alvarez-Iglesias *et al.* 2008), poblacionales (Salas *et al.* 2005) y forenses (Jobling 2001, Jobling and King 2004, Mosquera-Miguel *et al.* 2009) debido a la gran información que pueden aportar en estas áreas, como también por su comodidad a la hora de ser implementados con la tecnología actual.

En el caso de la genética de poblaciones, como de la genética forense el uso de estos múltiplexes ha permitido el estudio extensivo de las poblaciones en términos de ancestralidad y origen geográfico, como también de las características físicas en individuos (Phillips *et al.* 2012, Maroñas *et al.* 2014). Centrándonos en los estudios de SNPs en cromosoma Y en poblaciones humanas el trabajo de Brión y colaboradores (2006) fue uno de los pioneros; este consistió en la selección y diseño de un múltiplex que incluye 29 Y-SNPs, amplificados en fragmentos de 18 – 95 pares de bases, en una sola reacción de PCR multiplex y tipados por medio del método de mini-secuenciación SNaPshot[®] (Applied Biosystems); este multiplex permite caracterizar a nivel global los principales haplogrupos masculinos de poblaciones europeas, africanas y nativo americanas, lo que permite discriminar individuos de diferentes orígenes geográficos, teniendo presente que esta predicción está limitada a la procedencia del linaje paterno que puede ser distinto al fenotipo del individuo examinado.

Más adelante, Acosta y colaboradores (2009) usando el multiplex descrito anteriormente y añadiendo el marcador M3 en una reacción singleplex, publicaron la frecuencia de los principales haplogrupos presentes en una región suroccidental de Colombia para dos distintas poblaciones colombianas, una afro-descendiente y otra caucásica-mestiza. Los resultados mostraron una marcada diferencia entre las poblaciones, donde la caucásica-mestiza muestra un importante componente Europeo (R1b1) y baja composición nativo americana (menos del 10% - Hg Q3* (Q1a3a* de acuerdo con YCC2008) y la afro-descendiente un gran aporte africano representado, como era esperando, por el conjunto de SNPs tipados para el haplogrupo E3a (E1b1a de acuerdo con YCC2008).

Más recientemente, Geppert y colaboradores (2011), estudiando poblaciones nativo americanas del Ecuador encontraron que la mayoría de los individuos analizados, pertenecían al haplogrupo C3* (C-M217), frecuente en poblaciones de Mongolia y distribuido ampliamente en lejano oriente, Rusia y en Asia oriental. Este haplogrupo, según los autores, no es resultado de una mezcla reciente, y es la primera vez que se reporta en América del sur; de la misma manera los autores proponen una serie de 3 múltiplexes donde establecen con más precisión los sublinajes dentro del haplogrupo C, específicamente del clado C3, (mediante los marcadores M217, M407, M48, P53.1, M217, P62, RPS4Y711, M93, M86, and P39); los resultados obtenidos por los autores permiten con buena resolución establecer la diversidad para este haplogrupo en cuestión en las poblaciones nativo americanas de Ecuador.

Paralelamente, Gayà-Vidal y colaboradores (2011) analizando una muestra de individuos nativo-americanos de Bolivia, encontró que en el 97% de los cromosomas Y del grupo nativo Aymara demuestra haplogrupos típicamente nativo-americanos (86% hg Q1a3a y 10% hg Q1a3*) y que en el grupo nativo Quechua esta proporción es de un 78% (Todos pertenecientes al hg Q1a3a). Así mismo reportan para los 12 marcadores Y-STRs del kit PowerPlex® Y, diversidades haplotípicas de 0.96 y 0.971 para los grupos Quechua y Aymara, respectivamente. Por esta razón los autores concluyen que existe un nivel alto de diversidad en estas dos poblaciones y que a partir de los datos publicados por otros estudios, el área central andina es relativamente homogénea.

En esta misma región andina, Blanco-Verea y colaboradores (2010), estudiando población amerindia de Argentina perteneciente a 3 distintos grupos étnicos: Mapuche, Diaguita y Kolla, reportaron una frecuencia muy alta de los haplogrupos R1b1 y Q1a3a*, característicos de población europea y amerindia, respectivamente. Los autores al realizar comparaciones entre nativos argentinos y otras poblaciones de sur América pudieron corroborar que los haplogrupos amerindios son significativamente menos frecuentes que en las otras poblaciones comparadas. Concluyen que los resultados de este estudio muestran un claro componente amerindio, pero también un claro componente europeo, sugiriendo que los linajes típicos europeos, han surgido y han ido desplazando los haplogrupos amerindios genuinos de la población sur americana, debido a las recientes migraciones europeas a la región.

Respecto a estudios en población general de Bolivia, al momento de escribir este documento, solo se han descrito 3 estudios para esta población a partir de STRs de cromosoma Y.

El primero de ellos fue descrito por Lee y colaboradores (2007), quienes realizaron un análisis en población Boliviana de comunidades no emparentadas residentes en alta y baja altitud (La Paz 4000 metros a nivel del mar y El Beni 300 metros a nivel del mar). Los autores proponen un set de 11 Y-STRs (7 Y-STRs del MHT (excluyendo DYS385a/b), 2 reportados en un estudio anterior DXYS156Y, DYS413 y 2 Nuevos, DXYS241 y DXYS266), el análisis de este set reporta, que las poblaciones de alta montaña (alta altitud) son menos diversas que aquellas poblaciones del sub-trópico amazónico (baja altitud) con unas diversidades haplotípicas de 0.927 ± 0.029 y 0.996 ± 0.002 para ambas poblaciones respectivamente. Se describen 153 Haplotipos y solo 3 Haplotipos (113 baja altitud y 40 alta altitud) fueron compartidos por ambas poblaciones, lo que muestra el alto nivel de

discriminación que ofrecen esta propuesta de 11 Y-STRs analizados en el estudio que para uso forense, el cual según los autores tiene un gran impacto para esta población en particular.

En el segundo estudio Tirado y colaboradores (2009) genotipó los 17 Y-STRs del kit *Yfiler* (AB); la población de estudio incluye individuos de origen multiétnico del departamento de El Beni (Norte de Bolivia). Los autores reportan en este estudio una diversidad haplotípica para esta población de 0.9808 ± 0.0067 ; igualmente el estudio incluye comparaciones de este departamento con algunas otras muestras de la misma región mediante un análisis de MDS. Las comparaciones indican que la población Boliviana estudiada muestra una diferencia significativa con respecto a las poblaciones del alrededor (entre ellas, Brasil, Argentina, Perú, y otras regiones de Bolivia). Sin embargo, los resultados sugieren, con fuerte evidencia estadística, que existe una diferencia genética marcada entre las distintas poblaciones Bolivianas estudiadas por (Lee *et al.* 2007).

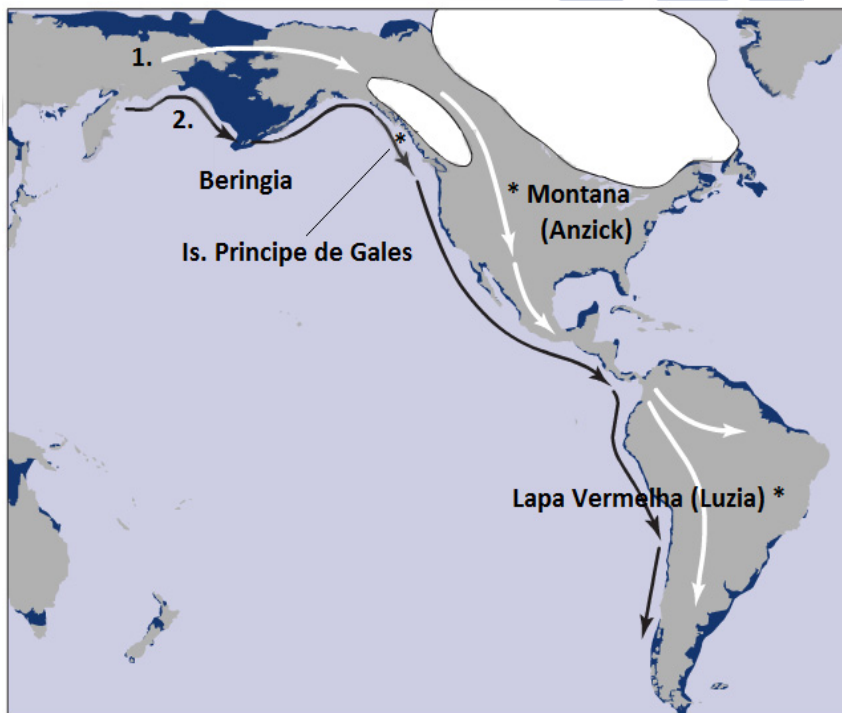
Y por último recientemente Vullo y colaboradores (2014), analizó una muestra de los departamentos de Chuquisaca y La Paz ($n = 38$ y $n = 57$, respectivamente) tipando los marcadores del *Yfiler* y 42 Y-SNPs para determinar los haplogrupos más comunes de estas dos zonas de Bolivia; los resultados indican que hay diferencias significativas entre las dos muestras, donde la frecuencia de linajes europeos es mayor en la muestra de Chuquisaca (50%) que en la de La Paz (~30%), y correspondientemente la proporción de linajes nativos americanos es mayor en la muestra de La Paz (70%) y menor en la de Chuquisaca (50%). El aporte africano detectado en ambas poblaciones fue tan solo del 1%. Con lo anterior los autores demuestran que existe una clara subestructura en estas dos poblaciones bolivianas y que estos hallazgos deben ser tomados en cuenta cuando se analice evidencia de tipo forense de este tipo.

2.4 Poblamiento de América y su importancia en la población boliviana

2.4.1 Poblaciones fundadoras de América

El estudio del poblamiento de América ha sido un tema que durante mucho tiempo ha sido objeto de debate, controversia y continuo análisis, aun así, existen una serie de evidencias en las que la comunidad científica tiene común acuerdo; entre ellas se puede decir que tanto la evidencia genética como arqueológica indica que el continente americano fue el último continente en ser habitado por los seres humanos. Se estima que este ocurrió con poblaciones de un origen siberiano que se dispersó hacia el estrecho de Bering, o Beringia que era la zona

que unía a los continentes euroasiáticos con norte américa, y que existió mientras el nivel oceánico de aquella zona fue bajo y permitió ser un corredor de movimiento migratorio, no antes de 32.000 años o 32 KYA (KYA – *Thousand Years Ago*) y luego estos grupos poblacionales migraron hacia américa hace aproximadamente unos 18 KYA (**Figura 10**). Respecto a esta última migración se han propuesto dos posibles rutas: una consiste en un puente terrestre a través de Beringia entre Siberia y Norteamérica y luego a través de un corredor libre de capas de hielo a lo largo de Norteamérica, y la segunda a lo largo de un corredor costero libre de hielo a través de Beringia de Siberia hacia Norteamérica. La importancia de cuál fue la ruta migratoria, reside en que en algunos periodos de tiempo específicos, ciertas regiones del corredor terrestre solo eran transitables y esto puede afectar las estimas en las que ocurrieron las migraciones. Sin embargo el registro fósil de restos humanos presente hasta momento, confirma que el periodo de presencia de las poblaciones humanas en el continente americano oscila entre 16 – 14 KYA, esta teoría está soportada por el fósil más antiguo encontrado hasta el momento (Cráneo – *Luzia*) en *Lapa Vermelha*,



(Minas Gerais) Brasil, que data aproximadamente en este intervalo de periodo mencionado. Por lo tanto las posibles olas de migración tuvieron que presentarse en este periodo que es justo después del último periodo glacial, o antes de este periodo entre 36- 24 KYA.

Figura 10. Posibles rutas de migración desde Siberia (Continente Euroasiático) hacia Norteamérica (Continente Americano); la ruta 1. Establece la migración desde Siberia hacia Alaska por el puente terrestre de Beringia y luego a través del corredor libre de capas de hielo en Norteamérica, y la ruta 2. Es la ruta costera que establece la migración por un corredor costero a través de Beringia; ambas rutas no se encontraban disponibles al mismo tiempo, por lo que pueden ser excluyentes una de la otra. Con asterisco (*) son indicados, los hallazgos arqueológicos con evidencia genética masculina disponible que dan soporte a la presencia humana en los periodos posteriores a 16 KYA. (Tomado y modificado de (Jobling *et al.* 2004))

La evidencia genética masculina que se tiene hasta el momento y que soporta la teoría de las migraciones humanas desde Beringia hacia Norteamérica, está basada entre otras en el hallazgo de *On Your Knees Cave*, una cueva cerca de la costa de la Isla del Príncipe de Gales (Canadá); allí se encontró una mandíbula, unas vértebras y una pelvis pertenecientes a un varón, cuya datación se estimó en ~11 KYA; análisis moleculares del cromosoma Y (Kemp *et al.* 2007), mostraron que el haplogrupo al que pertenecía el individuo es el determinado por el SNP M3-Q1a3a*, haplogrupo típico de las poblaciones nativo americanas. Así mismo recientemente (Rasmussen *et al.* 2014), durante la secuenciación del genoma completo del fósil paleo-indígena *Anzick-1* (encontrado en Montana, USA), se estimó su datación en un intervalo entre ~12.7-12.5 KYA; también los análisis de cromosoma Y indicaron que este fósil pertenecía al haplogrupo Q-L54 (xM3). Por lo tanto, estos hallazgos confirman y dan soporte, primeramente a la presencia de poblaciones en el continente americano en el periodo comprendido entre 16 KYA – 13 KYA y en segunda medida a la presencia del haplogrupo Q en estos grupos ancestrales explicando el porqué de la existencia de este haplogrupo, en las poblaciones nativo-americanas actuales.

Actualmente en el continente americano, el linaje Q1a3a-M3 es el principal haplogrupo descrito para la mayoría de los individuos nativos del continente y es muy rara la presencia de este haplogrupo en otras poblaciones de otros continentes, salvo algunas poblaciones de Siberia que se encuentran cerca al estrecho de Bering. Este hecho parece entrar en conflicto con la hipótesis de varias migraciones hacia el continente, lo que permite sugerir una sola migración desde Bering con un pool genético que ahora es característico de todos los nativos americanos. Distintos autores han estimado que este linaje está aproximadamente presente en el 58% de las poblaciones de Norte y Sur América (Karafet *et al.* 1999, Lell *et al.* 2002) y que la presencia de otros haplogrupos diferentes al Q1a3a-M3, están explicados por las recientes migraciones en el tiempo por parte de europeos (colonizador) y africanos (esclavo); por este motivo es frecuente observar que estas poblaciones nativo americanas en muchos casos presentan ancestralidades mezcladas. Adicionalmente existe la presencia de otros linajes nativos en el continente, entre ellos el C3-M217 (mencionado en el apartado anterior), que parece haber migrado un tiempo o mucho tiempo después que los linajes Q y que tiene una presencia actual en el este de Asia y con más exactitud de poblaciones provenientes del lago Baikal (Karafet *et al.* 1999, Lell *et al.* 2002); De la misma manera parece ser que las poblaciones que dieron origen de los linajes Q, en la actualidad se encuentran en poblaciones siberianas (*Altaians, Selkups y Kets*), aun así no se puede determinar con precisión si en

realidad son estas poblaciones las originarias de estos linajes, ya que las estimas se basan en la presencia de estos haplogrupos en la actualidad, o si el origen pudo haber sido por algunas otras poblaciones extintas de la region siberiana aún desconocida.

2.4.2 Bolivia, una población mayormente nativo americana

Todo lo anteriormente mencionado, tiene sin lugar a dudas un efecto directo en las poblaciones nativo americanas que encontramos en el presente; se ha estimado que la población amerindia (nativo americana) en centro y sur América, consta de alrededor de 51 millones de personas (Lizcano 2005) , y si limitamos este cálculo a américa del sur podemos decir que el 91% de esta población amerindia esta concentra en Ecuador, Perú y Bolivia.

Respecto a Bolivia, la población en la que nos centraremos en esta tesis, esta población se describe y se conoce actualmente como una población pluri-étnica; esta república se encuentra en el centro de sur América. Limita con Perú al oeste con Chile al sudoeste con Paraguay y Argentina al sur y al norte y al este con Brasil; tiene una población estimada en aproximadamente 10.200.000 habitantes, donde el ~55% de esta población tiene un componente amerindio o hace parte de un grupo indígena (Instituto Nacional de Estadística of Bolivia; INE; <http://www.ine.gob.bo/>). Los dos grupos nativos mayoritarios presentes en la actualidad en el territorio Boliviano son los Quechua y Aymara, que corresponden al ~16% y ~14%, respectivamente, del total de la población. El ~22% restante de la población indígena o nativa corresponde a grupos minoritarios, como los grupos Guaraní, Chiquitano y Mojeño (Moxeño) principalmente. El ~45% restante de la población total, está compuesto por “mestizos”, que son mayormente el resultado de la mezcla de ancestralidades entre nativos y europeos.

La historia reciente de Bolivia tiene una gran importancia ya que permite entender la composición pluri-étnica actual que encontramos en el país. En tiempos pre-coloniales, Bolivia era una región donde un número de diversas culturas indígenas se asentaron, principalmente en la región del 'altiplano' (regiones andinas y sub-andinas actuales del país), que en su conjunto constituyeron el primer imperio andino con una ciudad establecida, Tiwanacu (s. VIII). Más adelante, el Imperio Inca se levantó en el siglo XIII, y estuvo compuesto por cuatro regiones principales, entre ellas las regiones Cuntinsuyu, Chinchansuyu, Collasuyu y Antisuyu, que se expandieron a través de las regiones andinas y sub-andinas de lo que ahora se conoce como Perú, Chile y Argentina; así mismo, este imperio contaba con un idioma oficial, el Quechua. Parte actual del territorio Boliviano, pertenecía a

la provincia Collasuyu, ocupada principalmente por nativos Aymaras. El imperio de los Incas concluyó con la conquista española en 1572 y esto sentó las bases al carácter pluri-étnico de los bolivianos de hoy en día.

Durante la época de la colonización, Bolivia no fue la excepción para el destino de los esclavos africanos. Durante el período del tránsito transatlántico de esclavos (entre 1501 y 1866), los imperios europeos de la época llevaron unos 10,7 millones de esclavos a América (Eltis and Richardson 2010). En Bolivia, la mayoría de los esclavos eran obligados a trabajar en la mina real que estaba en Potosí hasta el año 1773, igualmente también eran empleados para el servicio doméstico (Lipski 2006, Lipski 2006, Lipski 2007). En la actualidad, sólo hay pequeñas comunidades de 'Afro-bolivianos', que están concentrados principalmente en la región de Nor Yungas, en el departamento de la Paz. El impacto que tuvo este comercio de esclavos Africanos en el cromosoma Y de los bolivianos de la actualidad, es todavía desconocido.

En la actualidad, Bolivia está dividida políticamente en nueve departamentos y su geografía varía desde las altas montañas de los Andes (oeste) a las tierras bajas orientales (Llanos), situadas en la cuenca del Amazonas.

El país actualmente alberga una gran diversidad cultural. El número de idiomas individuales conocidos en Bolivia asciende a un total de 45, de los cuales 37 son lenguas vivas, uno es un segundo idioma sin hablantes que la tengan como lengua materna y siete tienen hablantes no conocidos (Lewis 2009, Perego *et al.* 2012) La principal lengua hablada hoy en Bolivia es el español, pero también hay otros importantes idiomas prehispánicos, como el quechua heredado por el Imperio Inca, y hablado por nativos Quechuas y Aymaras de la actualidad.

Respecto a los pueblos Quechuas-hablantes, éstos habitan principalmente en los valles (Región sub-andina) de Cochabamba y Chuquisaca y algunas regiones de la montaña en Oruro, mientras que los Aymara-hablantes residen principalmente en el altiplano (Región Andina) de los departamentos de La Paz, Oruro, y Potosí (en el área alrededor de Lago Titicaca). También existen otros grupos étnicos asentados principalmente en los Llanos (incluyendo las áreas de la Amazonía boliviana) como son los Chiquitanos (110.000 habitantes), los Guaraníes (78.300 habitantes, presentes mayormente en la frontera con Paraguay) y los Moxeños (76.000 habitantes); Estos grupos ocupan principalmente los departamentos de Santa Cruz, Beni, Pando y Tarija (Región Llanos).

De esta manera, podemos concluir que la presencia de nativos americanos en Bolivia, es una componente importante de la población actual; como se mencionó en el apartado 2.3.1

algunos estudios se han realizado sobre algunas de estas poblaciones nativas y encajan muy bien, desde el punto de vista genético, en los linajes nativos típicos que migraron a América desde el continente euroasiático (Apartado anterior). Sin embargo hasta la fecha se desconoce, la proporción genética masculina, nativa que aún permanece entre los habitantes del país como grupo, como también la influencia del comercio de esclavos africanos ha tenido en la población actual.

Es por esto que entre los objetivos enmarcados en la presente tesis se encuentra el análisis más completo que existe hasta el momento, dirigido a la caracterización del cromosoma Y de los bolivianos actuales, con el fin de describir e identificar las proporciones y los efectos genéticos que se han cuestionado anteriormente. Así mismo, entre los resultados se han comparado los haplotipos del cromosoma Y con información autosómica y de ADN mitocondrial con el fin de arrojar más luz sobre la demografía de los bolivianos según sexo y su ancestralidad total.



II. Justificación y Objetivos





El análisis de los polimorfismos del cromosoma Y, en los años recientes ha resultado ser un elemento muy importante en la resolución de casos judiciales, especialmente aquellos relacionados con agresiones sexuales o en los que se hace necesario establecer un vínculo patri-lineal entre las muestras analizadas. Aunque con limitaciones debidas al modelo de herencia, el análisis del cromosoma Y se ha utilizado también durante estos últimos años como herramienta predictiva del origen geográfico de perfiles de cromosoma Y, lo que en ocasiones podría ayudar a solucionar o guiar la investigación judicial en una dirección concreta. Para que este panorama sea real es necesario contar con una base de datos fiable y que describa de la mejor manera la población de referencia. Es por esto que generar bases de datos regionales o étnicos para un país o región siempre es el escenario ideal cuando la variabilidad genética está especialmente estratificada debido a la existencia de una importante heterogeneidad étnica, ya que las bases de datos generadas de esta manera permiten realizar estimas con mayor exactitud y precisión.

De la misma manera, el análisis de estos polimorfismos permite entender la estructura genética de una población e inferir su historia demográfica, contribuyendo al marco multidisciplinar que es necesario para entender el origen de las poblaciones humanas.

El caso de Bolivia es un caso de particularmente interesante porque es una población que desde el punto de vista de la genética poblacional y forense ha recibido poca atención, sobre todo cuando comparamos los estudios realizados en Bolivia con los realizados en otros países del continente suramericano (como Argentina, Colombia, etc). Hasta el momento, se han publicado muy pocos estudios centrados en los aspectos genéticos de la población boliviana, y casi todos ellos en los últimos cinco años (Ver por ejemplo, (Tirado *et al.* 2009, Galanter *et al.* 2012, Taboada-Echalar *et al.* 2013, Vullo *et al.* 2014). Debido a que esta población cuenta con una gran proporción de habitantes pertenecientes a grupos de origen nativo americano, es una población ideal, ya que aún, en algunos casos, permanecen linajes e individuos que todavía mantienen su carácter nativo sin interferencias de mestizaje. Es por esta razón que se analizarán los niveles de ancestralidad de linajes para Bolivia usando SNPs de Cromosoma Y, como también la estructura genética de la población por parte de 17 Y-STRs (*Yfiler*[®]) y

finalmente se analizará comparativamente esta población a partir de los marcadores de nueva generación para Cromosoma Y (PowerPlex Y23 y RM-YSTRs).

Los objetivos generales del presente trabajo de tesis son:

- Estudio de la distribución de las frecuencias de haplogrupos de cromosoma Y en Bolivia.
- Estudio de la distribución de las frecuencias de STRs de cromosoma Y (haplotipo mínimo y haplotipo *Yfiler*[®]) en Bolivia.
- Análisis de sesgo por género en población Boliviana por medio de ADN mitocondrial y cromosoma Y.
- Análisis comparativo de los linajes uniparentales con las proporciones de ancestralidad autosómicas.
- Caracterización de los microsatélites de cromosoma Y de nueva generación y reciente publicación (RM-YSTRs y PowerPlex Y23).

III. Materiales y Métodos





3.1 Muestras

Para el adelantamiento de los distintos análisis genéticos de Cromosoma Y en esta tesis, se usaron muestras tomadas de distintos orígenes geográficos de Bolivia, que a continuación se especifican.

Cromosoma Y

- Se tomaron muestras a 226 individuos varones no relacionados con hisopos bucales en 5 distintos departamentos de Bolivia entre ellos: Pando, Beni, Santa Cruz, La Paz y Cochabamba.
- Con fines de análisis estadísticos y genéticos estas muestras fueron agrupadas en tres regiones geográficas: I) *Region Andina* ($n = 109$ para Y-SNPs y $n = 107$ para Y-STRs), que incluye las muestras del departamento de La Paz, II) *Región Sub-Andina*, que incluye las muestras del departamento de Cochabamba y III) *Región Llanos* ($n = 89$ para Y-SNPs y $n = 88$ para Y-STRs), que incluye las muestras de los departamentos de Santa Cruz, Beni y Pando.

3.2 Extracción de ADN

Extracción de ADN a partir de hisopos bucales a partir de QIAamp® DNA Blood Mini Kit (QIAGEN)

Para la extracción de ADN a partir de saliva, se empleó el kit comercial *QIAamp DNA Blood Mini Kit* (QIAGEN, Alemania) un método que consiste en la extracción y purificación de ADN basado en columnas con membranas de sílica, para sangre y para saliva. Para la extracción de hisopos bucales se siguió el protocolo descrito por el fabricante bajo el apartado *Buccal Swab Spin Protocol*, y que se explica a continuación:

- Cortar hisopo sobre papel aluminio y pasar los cortes a un tubo eppendorf de 1,5 mL.
- Añadir 20 μ l de Proteínasa K.
- Añadir 400 μ l de buffer AL y mezclar con vórtex durante 15 segundos.
- Incubar a 56°C durante 10 minutos y seguidamente dar un breve *spin*, a máxima velocidad.
- Añadir 200 μ l etanol 96° y mezclar con vórtex durante 15 segundos.
- Añadir el todo el volumen anterior en una columna QIAamp y acoplarla con un tubo colector.
- Centrifugar 1 minuto a 8000 r.p.m.

- Descartar tubo colector con desecho y colocar nuevo tubo colector a la columna.
- Añadir 500 μL de WASH BUFFER AW1 a la columna y centrifugar a 8000 r.p.m. durante 1 minuto.
- Descartar el tubo con el filtrado y colocar la columna en un nuevo tubo colector.
- Añadir 500 μL de WASH BUFFER AW2 a la columna, centrifugar a 14000 r.p.m. por 3 minutos.
- Descartar el tubo colector con el filtrado y poner la columna en un tubo eppendorf nuevo.
- Añadir 200 μL de BUFFER AE y dejar incubar durante 1 minuto a temperatura ambiente.
- Centrifugar a 8000r.p.m. durante 1 minuto.
- Descartar columna y almacenar tubo eppendorf con ADN, extraído y purificado a -20°C

3.3 Cuantificación espectrofotométrica de ADN total

Algunas muestras de ADN extraídas se cuantificaron por medio de espectrofotometría usando el equipo NanoDropTM 1000 Spectrophotometer (Thermo Fischer Scientific Inc. Wilmington, DE USA.). Para la cuantificación de ADN se añadió 1 μL de muestra sin diluir en el equipo, y se tomaron mediciones por triplicado, que luego se promediaron La cuantificación consiste en la incidencia de un rayo de luz de una lámpara de xenón y un espectrofotómetro que usa una matriz CCD linear que analiza la luz después que pasa a través de la muestra. Luego el instrumento por medio de un software preinstalado en el ordenador, hace una medida aproximada de la concentración de ADN en la muestra.

3.4 Análisis de microsátélites de cromosoma Y

El estudio de los microsátélites de cromosoma Y se llevó a cabo por medio de la amplificación del material genético y posterior electroforesis capilar en secuenciador automático y análisis informático. Este protocolo se siguió en un total de 223 muestras, dadas que dos muestras del departamento de La Paz y una del departamento de Pando fallaron en el tipado.

En esta tesis se trabajó con los Y-STRs incluidos en el kit comercial AmpFISTR *YFiler*® kit de Applied Biosystems (Foster City, CA), con los Y-STRs incluidos en el Kit comercial PowerPlex® Y23 (Promega Corporation Technical Manual PowerPlex® Y23 System

Instrutions for use of products DC2305 AND DC2320) (Madison, USA) y con los 13 RM-YSTRs propuestos por (Ballantyne *et al.* 2012).

3.4.1 Amplificación Y-STRs - *Yfiler*®

Todas las muestras tomadas en el presente estudio ($n=223$) fueron amplificadas por el kit comercial AmpFISTR *YFiler*® kit de Applied Biosystems (Foster City, CA), que incluye los siguientes 17 Y-STRs: DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 y GATA H4 (Tabla 1). A continuación se explican las condiciones de amplificación usadas por muestra:

- Mezcla de Reacción: 3.0 μ L
- Set de Primers: 1.9 μ L
- AmpliTaq Gold®: 0.4 U
- H₂O: 1 μ L
- ADN: 0.7 μ L (0.5 ng/ μ L) en un volumen final de 7 μ L.

El programa de PCR utilizado en el termociclador GeneAmp® PCR System 9700 (Applied Biosystems) es el siguiente: 11 minutos a 95°C seguido de 30 ciclos a 94°C por 60 segundos, 61°C por 60 segundos y 72°C por 60 segundos, seguidos de una extensión final 80 minutos a 60°C.

Características de los Y-STR's analizados en el presente estudio				
Marcador Y-STR	Posición (Mb)	Motivo Repetición	Rango Alélico	Tasa de Mutación
DYS393	3.19	AGAT	8-17	0.08%
DYS456	4.33	AGAT	13-18	0.72%
DYS458	7.93	GAAA	14-20	1.1%
DYS19	10.13	TAGA	10-19	0.24%
DYS391	12.61	TCTA	6-14	0.29%
DYS635	12.89	TSTA	17-27	0.42%
DYS437	12.98	TCTR	13-17	0.15%
DYS439	13.03	AGAT	8-15	0.54%
DYS389 I/II	13.12	TCTR	9-17/24-34	0.19%/0.30%
DYS438	13.38	TTTTTC	6-14	0.05%
DYS390	15.78	TCTR	17-28	0.24%
GATA-H4	17.25	TAGA	8-13	0.25%
DYS385 a/b	19.26	GAAA	7-28	0.22%
DYS392	21.04	TAT	6-20	0.06%
DYS448	22.78	AGAGAT	17-24	0.18%

Tabla 1. Características de los Y-STRs analizados en el presente estudio. (Tomada y modificada de (Butler 2006))

Las muestras usadas para la amplificación del kit PowerPlex® Y23 ($n=100$) y los RM-YSTRs ($n=99$), comprende un subgrupo de muestras de la población total, que incluye muestras de los 5 departamentos (3 regiones geográficas), con ancestralidad Nativo-americana y “mestiza” confirmada por marcadores Y-SNPs tipados en el presente trabajo; el listado de muestra seleccionadas se encuentra en los apéndices anexos.

3.4.2 Amplificación Y-STRs – PowerPlex® Y23

Para las muestras amplificadas con el kit PowerPlex® Y23 (Promega) se siguió el protocolo, sugerido por la casa comercial Promega, con algunas modificaciones en los volúmenes por muestra:

- PowerPlex® Y23 5X Master Mix: 2.5 μ L
- PowerPlex® Y23 10X Primer Master Mix: 1.25 μ L
- H₂O: 7.75 μ L
- ADN: 1 μ L (0.3 ng/ μ L)

El programa de PCR ejecutado en el termociclador GeneAmp® PCR System 9700 (Applied Biosystems) tuvo los siguientes parámetros: 2 minutos a 96°C seguido de 30 ciclos a 94°C por 10 segundos, 61°C por 60 segundos y 72°C por 30 segundos, seguidos de una extensión final por 20 minutos a 60°C.

3.4.3 Amplificación Y-STRs – Rapidly Mutating Y-STRs

Las muestras tipadas con el multiplex de 13 marcadores RM-YSTRs, fueron amplificadas en una serie de tres PCR multiplex siguiendo el protocolo sugerido por (Ballantyne *et al.* 2014); cada multiplex contiene uno o más de un sistema multicopia, es decir un sistema microsátelite RM-YSTRs en el que se pueden detectar más de un alelo. El multiplex 1 contiene 2 marcadores multicopia, el DYS399S1 con posibilidades de detección de 1 a 5 alelos por sistema y el DYS387S1 donde se pueden detectar 1 o 2 alelos. El multiplex 2 cuenta con un sistema multicopia, el DYS526a/b para el cual se pueden detectar 1 alelo para a y otro alelo para b; y por último el multiplex 3 cuenta con 2 sistemas multicopia, el DYF404S1 en el que usualmente pueden detectarse 1 o dos alelos por sistema, y el DYF403S1a/b, donde en la parte a) pueden detectarse de 1 a 3 alelos y la parte b) usualmente se detecta 1 alelo.

Para cada uno de estos múltiplexes la concentración de los primers como los volúmenes de cada PCR que se muestran a continuación (Tablas 2 y 3):

Multiplex	Primers	Secuencia 5'- 3'	Concentración
RM1	DYS576 Fwd	6-FAM-TTGGGCTGAGGAGTTCAATC	0.1 μ M
	DYS576 Rev	GGCAGTCTCATTTCCTGGAG	0.1 μ M
	DYS570 Fwd	TAMRA-GAACTGTCTACAATGGCTCAG	0.3 μ M
	DYS570 Rev	TCAGCAT AGTCAAGAAACCAGACA	0.3 μ M
	DYF387S1 Fwd	JOE-GCCTGGGT GACAGAGCTAGA	0.2 μ M
	DYF387S1 Rev	GCCACAGTGTGAGAAGTGTGA	0.2 μ M
	DYF399S1 Fwd	6-FAM-GGGTTTTCCACCAGTTGCAT	0.4 μ M
DYF399S1 Rev	CCATGTTTTGGGACATTCT	0.4 μ M	
RM2	DYS626 Fwd	6-FAM-GCAAGACCCCATAGCAAAAAG	0.2 μ M
	DYS626 Rev	AAGAAGAATTTGGGACATGTTT	0.2 μ M
	DYS627 Fwd	JOE-CTAGGTGACAGCGCAGGATT	0.2 μ M
	DYS627 Rev	GGATAATGAGCAAATGGCAAG	0.2 μ M
	DYS526 Fwd	6-FAM-TCTGGTGAAGTATCCAAACC	0.2 μ M
	DYS526 Rev	GGGTACTTCGCCAGAAGGT	0.2 μ M
	DYS518 Fwd	TAMRA-GGCAACACAAGTAAAAGTGC	0.2 μ M
	DYS518 Rev	TCAGCTCTTACCATGGGTGAT	0.2 μ M
RM3	DYS612 Fwd	6-FAM-CCCCATGCCAGTAAGAATA	0.15 μ M
	DYS612 Rev	GTGAGGGAAGGCAAAAAGAAAA	0.15 μ M
	DYF404S1 Fwd	JOE-GGCTTAAGAAATTTCAACGCATA	0.15 μ M
	DYF404S1 Rev	CCATGATGGAACAATTGCAG	0.15 μ M
	DYS449 Fwd	6-FAM-TGAATATTTCCCTTAAGTGTGTGA	0.075 μ M
	DYS449 Rev	CCATTGCACTCTAGGTGGAC	0.075 μ M
	DYS547 Fwd	6-FAM-TCCATGTTACTGCAAAAATACAC	0.3 μ M
	DYS547 Rev	TGACAGAGCATAAACGTGTC	0.3 μ M
	DYF403S1 Fwd	JOE-CAAAATTCATGTGGATAATGA	0.5 μ M
	DYF403S1 Rev	ACAGAGCAGGATTCCATCTA	0.5 μ M

Tabla 2. Concentración final de primers usados para cada reacción PCR multiplex, tomado de (Ballantyne *et al.* 2014)

Multiplex	Componente Multiplex	Volumen (μ L)
RM1	QIAGEN PCR Master Mix	5
	DYS576 Primers Fwd + Rev	0.1
	DYF387S1 Primers Fwd + Rev	0.2
	DYS570 Primers Fwd + Rev	0.3
	DYF399S1 Primers Fwd + Rev	0.4
	H ₂ O	3
	DNA	1
RM2	QIAGEN PCR Master Mix	5
	DYS526 Primers Fwd + Rev	0.2
	DYS626 Primers Fwd + Rev	0.2
	DYS627 Primers Fwd + Rev	0.2
	DYS518 Primers Fwd + Rev	0.2
	H ₂ O	3.2
	DNA	1
RM3	QIAGEN PCR Master Mix	5
	DYS612 Primers Fwd + Rev	0.15
	DYF404S1 Primers Fwd + Rev	0.15
	DYS499 Primers Fwd + Rev	0.075
	DYS547 Primers Fwd + Rev	0.3
	DYF403S1 Primers Fwd + Rev	0.5
	H ₂ O	2.825
	DNA	1

Tabla 3. Volumen de componentes para las tres distintas reacciones PCR multiplex; el volumen final de reacción es de 10 μ L, tomado de (Ballantyne *et al.* 2014)

El programa de termociclaje que se siguió para las 3 distintas reacciones de PCR multiplex, es descrito a continuación: una desnaturalización inicial por 10 minutos a 94 °C seguidos de 10 ciclos de 94 °C por 30 segundos, 65 °C por 30 segundos (-1 °C/Ciclo) y 72 °C por 60 segundos, continuados por 25 ciclos a 94 °C por 30 segundos, 50 °C por 30 segundos y 72 °C por 60 segundos y finalizando con una extensión final a 60 °C por 45 minutos

3.4.4 Secuenciación

Una vez obtenido el producto amplificado para las 3 distintas amplificaciones presentadas en esta tesis (*Yfiler*®, Y23 y RM-YSTRs), se procede con la separación de los fragmentos por electroforesis capilar usando el secuenciador automático ABI PRISM 3130xl (Applied Biosystems) y para los 3 multiplex se usó polímero de corrido POP4®; para la electroforesis capilar se añade 1.5 µL de producto amplificado de muestra (*Yfiler*®, Y23 y RM-YSTRs) con 0.15 µL marcador interno de tamaño LIZ 500, para las amplificaciones con *YFiler*® como para las 3 distintas amplificaciones de RM-YSTRs, o 1.0 µL de marcador interno de tamaño CCS ILS 500 Y23, para las muestras amplificadas con Y23; en todos los casos y se mantienen las condiciones de los módulos del secuenciador por defecto. Para los 3 casos anteriores, se usaron los *ladder* alélicos provistos por las casas comerciales, como los provistos para el ejercicio colaborativo de RM-YSTRs

Los datos resultantes son posteriormente analizados con el software informático Genemapper ID 3.2.1 (Applied Biosystems, Foster City, USA), usando la matriz para Y-STRs y los Bins de tallaje específicos para cada multiplex, para asignación automática de los alelos microsátélites tanto para *Yfiler*®, Y23 y RM-YSTRs.

3.5 Polimorfismos Binarios de cromosoma Y (Y-SNPs)

Para la detección de polimorfismos tipo SNP asociados a los diferentes haplogrupos presentes en los linajes uniparentales estudiados, se emplearon varias PCR-multiplex y luego estos polimorfismos fueron posteriormente analizados por una serie de reacciones de mini-secuenciación, por medio del método SNaPshot® (Applied Biosystems). Este protocolo se realizó en el grupo total de 226 muestras.

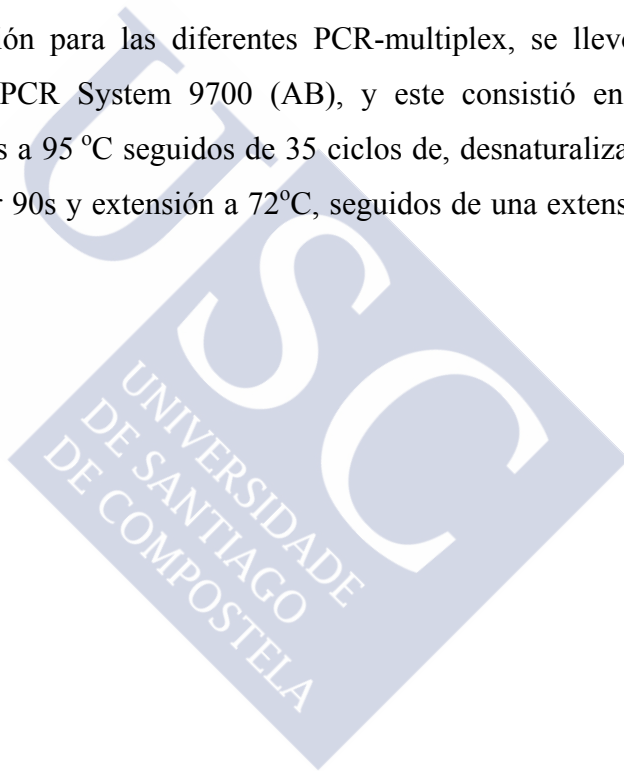
La detección de los principales linajes biogeográficos de cromosoma Y, se realizó tipando 32 Y-SNPs organizados en 5 distintos múltiplexes, descritos por Blanco-Verea y colaboradores (2006, 2010) y Ramos-Luis y colaboradores (2009). Estos 5 múltiplexes, permiten detectar 34 haplogrupos presentes en Europa, África, medio oriente, y América (**Figura 11**).

3.5.1 Reacción de PCR Multiplex – Amplificación de los SNPs

Para la amplificación de los SNPs se realizaron una serie distintas PCR-multiplex con distintas concentraciones finales de primers; se llevaron a cabo en total 5 PCR-multiplex con concentraciones de primers como se describen en la Tabla 4. Para el genotipado de Y-SNPs en la PCR multiplex se siguió el siguiente protocolo:

- PCR-Master Mix QIAGEN (QIAGEN, Hilden, Alemania): 5 μ L
- Mix de primer: 1 μ L
- ADN: 1 μ L (5-10 ng/ μ L)
- H₂O: hasta completar un volumen de 10 μ L.

El programa de amplificación para las diferentes PCR-multiplex, se llevó a cabo en un termociclador GeneAmp® PCR System 9700 (AB), y este consistió en un periodo de preincubación de 15 minutos a 95 °C seguidos de 35 ciclos de, desnaturalización a 94°C por 30s, anillamiento a 58°C por 90s y extensión a 72°C, seguidos de una extensión final a 72°C por 10 minutos.



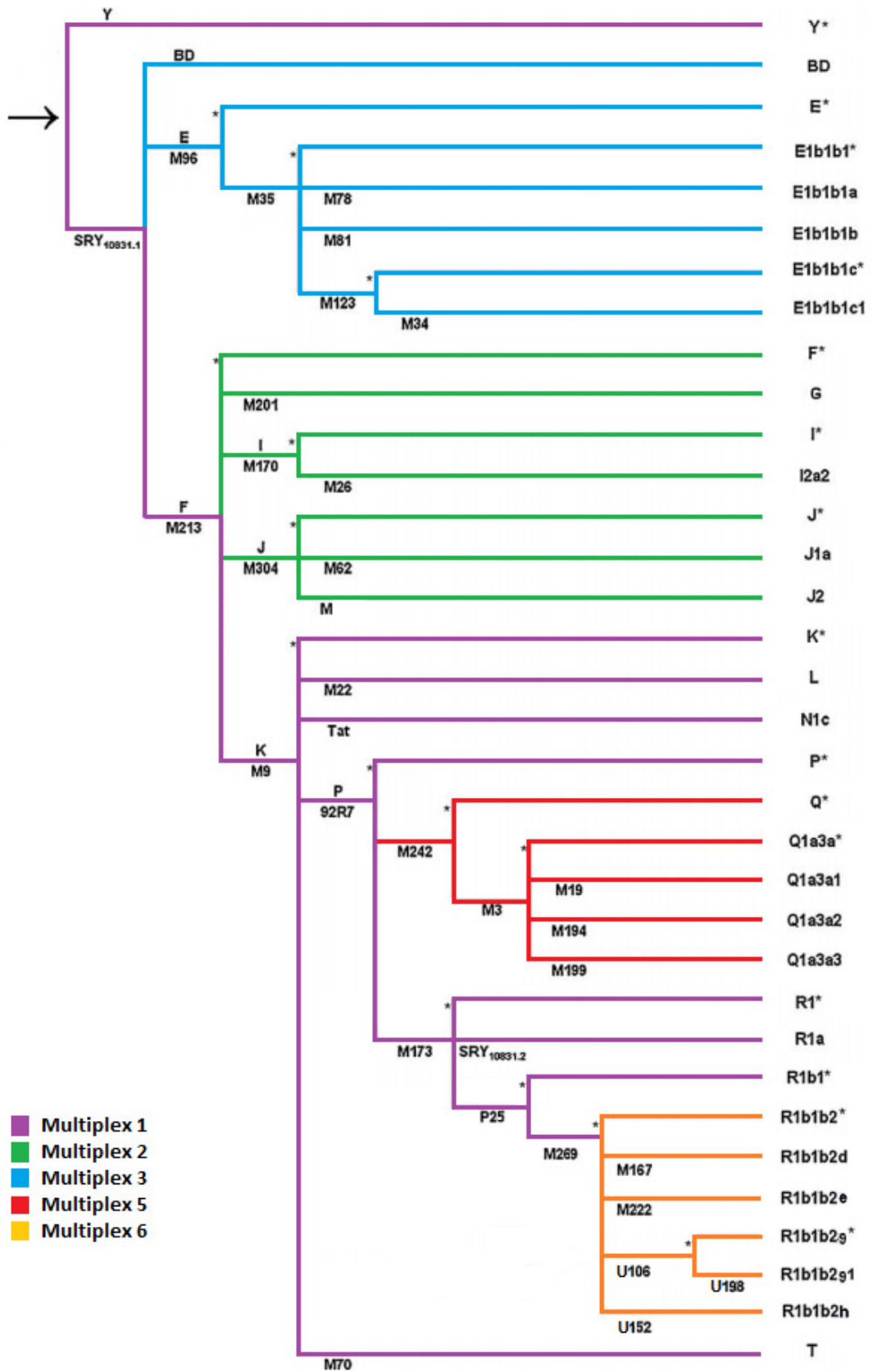


Figura 11. Árbol Filogenético de los haplogrupos de cromosoma Y detectados por los 5 distintos multiplexes usados para el tipado de Y-SNPs de acuerdo a la nomenclatura propuesta por (Karafet *et al.* 2008).

Marcador	Primer forward y reverse	Tamaño (pb)	Cambio de Base	CF (μM)	Multiplex	Diseño
SRY-1532	TCCTTAGCAACCATTAATCTGG AAATAGCAAAAAGTACACAAGGC	167	A > G	2	1	(Whitfield <i>et al.</i> 1995)
M213	GGCCATATAAAAACGCAGCA TGAATGGCAAATGATTCCA	208	T > C	2	1	Brión y cols. (2004)
M173	GCACAGTACTCACTTTAGGTTGC GCAGTTTTCCAGATCCTGA	172	A > C	2	1	Brión y cols. (2004)
M9	GCAGCATATAAACTTTCAGG AAAACCTAACTTGTCTCAAGC	340	C > G	2	1	Brión y cols. (2004)
M22	GCTGATAGTCCTGGTTCCCTA TGAGCATGCCTACAGCAGAC	106	A > G	2	1	Brión y cols. (2004)
M70	TCATAGCCCACTATACTTTGGAC CTGAGGGCTGGACTATAGGG	81	A > C	2	1	Brión y cols. (2004)
TAT	GACTCTGAGTGTAGACTTGTGA GAAGGTGCCGTAAAAGTGTGAA	112	T > C	2	1	Brión y cols. (2004)
92R7	TGCATGAACACAAAAGACGTA GCATTGTTAAATATGACCAGC	55	G > A	2	1	Brión y cols. (2004)
M173	GCACAGTACTCACTTTAGGTTGC GCAGTTTTCCAGATCCTGA	172	A > C	2	1	Brión y cols. (2004)
P25	GGACCATCACCTGGGTAAGT AGTGCTTGTCCAAGGCAGTA	121	C > A	2	1	Brión y cols. (2004)
M269	TCATGCCTAGCCTCATTCCT TCITTTGIGTGCCTTCGAGG	154	T > C	2	1	Brión y cols. (2004)
M201	TCAAATTGTGACACTGCAATAGTT CATCCAACACTAAGTACCTATTACGAA	144	G > T	2	2	Brión y cols. (2004)
M170	TGCAGCTCTTATTAAGTTATGTTTCA CCAATTACTTTCAACATTTAAGACC	158	A > C	2	2	Brión y cols. (2004)
M26	AGCAGAAGAGCAAGACAGC GACGAAATCTGCAGCAAAAA	147	G > A	2	2	Brión y cols. (2004)
M62	ACTAAAACACCATTAGAAACAAAGG CTGAGCAACATAGTGACCCC	309	T > C	2	2	Brión y cols. (2004)
M172	TCCTCATCACCTGCCTCTC TCCATGTGGTTTGGACAG	187	T > G	2	2	Brión y cols. (2004)
M96	GIGATGTGTAACCTGGAAAACAGG GGACCATATATTTGCCATAGGTT	88	G > C	2	3	Brión y cols. (2004)
M35	GCATGGTCCCTTCTATGGAT GAGAATGAATAGGCATGGGTTC	198	G > C	2	3	Brión y cols. (2004)
M78	CTTCAGGCATTATTTTTTGGT ATAGTGTTCCTCACCTTTCCTT	301	C > T	2	3	Brión y cols. (2004)
M81	TTTATAGTTTCAATCCCTCAGTAATTTT TGTTCTTCTTGGTTTGTGTGAGTA	176	C > T	2	3	Brión y cols. (2004)
M123	CACAGAGCAAGTACTCTCAAAG TCTTCCCTCAACATAGTTATCTCA	248	G > A	2	3	Brión y cols. (2004)
M34	CACAGTGTTTTCTCATGTTAATG GGGGACCCCAATAATCATAAA	92	G > T	2	3	Brión y cols. (2004)
M3	CTGCCAGGGCTTTCAAATAG AAGGGCATCTTTCATTTTAGGT	93	C > T	0.4	5	Blanco-Verea y cols. (2010)
M19	CTGGTCATAAACAAGTGGAAATC AGCTGACCACAACTGATGTAGA	171	T > A	0.4	5	Blanco-Verea y cols. (2010)
M194	GCCTGGATGAGGAAGTGTGAG ATACAGTCGTTGCCTTCTCG	127	T > C	0.4	5	Blanco-Verea y cols. (2010)
M199	CCTGGTTGGATTCTGGTCTT TGATTCAAGGATTTGTTAGTCTT	197	C > G	0.4	5	Blanco-Verea y cols. (2010)
M242	ATAGAAAAGTTTGTGCAAAAAGGTGA ACCTTACCTAGAACAACTCTGAAGC	139	C > T	0.4	5	Blanco-Verea y cols. (2010)
M167	GAGGCTGGGCCAAGTTAAGG CTTCCTCGGAACCACTACCA	130	C > T	4.37	6	Ramos-Luis y cols (2009)
M222	GCAGTGGTGTACAAAAGTGC TTTCCTGAGCAAGAAGTATGGA	91	G > A	1.87	6	Ramos-Luis y cols (2009)
U152	GAAACATCCACGCTTGAGG TGTCATGGTAGTTAATGGGAGT	107	G > A	2.5	6	Ramos-Luis y cols (2009)
U106	CTCTCATGGGCTGGACA AAATGACACCTGGAGGAAA	165	C > T	1.87	6	Ramos-Luis y cols (2009)
U198	CCCACTTGCTCACTTTGTG TGAGGCTCAGTTTCAAATGC	75	G > A	1.87	6	Ramos-Luis y cols (2009)

Tabla 4. Secuencia de primers empleados en la amplificación, diseñados por Whitfield y cols. (1995), Brión y cols. (2004), Blanco-Verea y cols. (2010) y Ramos-Luis y cols. (2009). La tercera columna indica el tamaño del amplicón; la cuarta columna indica los alelos esperados, la quinta columna indica la concentración final de la pareja de primers en el respectivo arreglo de multiplex de la sexta columna.

3.5.2 Purificación post-PCR, Reacción de mini-secuenciación y Purificación post-extensión

Para el análisis del producto amplificado, de los 5 múltiplexes de Y-SNPs, es preciso purificar el producto de PCR, con el fin de eliminar los dNTPs y primers no incorporados durante la amplificación.

Esta purificación se lleva a cabo por medio de la enzima Ilustra™ EXOSTAR™ (GE Healthcare); Esta enzima tiene una actividad exonucleasa, presente en la *Exonucleasa I*, que degrada los primers y una actividad de desfosforilación de los dNTPs no incorporados, por medio de la *fosfatasa alcalina*. La reacción enzimática se efectúa de la siguiente manera: 1 µL de producto amplificado y 0.5 µL de EXOSTAR™ a 37°C por 15 minutos (periodo de digestión enzimática) y luego 15 minutos a 85°C (periodo de desnaturalización de la enzima) en un termociclador GeneAmp® PCR System 9700 (AB). Seguida esta purificación se lleva a cabo la reacción de mini-secuenciación mediante el kit SNaPshot® (AB). En breve, esta técnica consiste en la extensión de un dideoxi-nucleótido fosfato fluoro-marcado (ddNTPs) (Tabla 5) sobre un primer no marcado; la extensión tiene lugar cuando cada primer anilla una base nitrogenada antes del polimorfismo de interés en el amplicón complementario, (previamente obtenido en la PCR), y en la presencia de los ddNTPs y la AmpliTaq® DNA FS polimerasa, incorpora un ddNTP fluoro-marcado al extremo 3', que detiene la reacción e indica el SNP de interés.

ddNTP incorporado	Fluorocromo	Señal
G	dR110	Azul
A	dR6G	Verde
C	dTAMRA™	Negro
T	dROX™	Rojo

Tabla 5. Fluorocromos asociados a los ddNTPs incorporados durante la reacción de SNaPshot™

La reacción se llevó a cabo mezclando 3 µL de *Ready Reaction Mix* (SNaPshot™), 1.5 µL mix de sondas para SNaPshot a concentraciones finales específicas para cada ensayo (Ver Tabla 6), 1.5 µL de producto purificado de PCR y 4.5 µL de H₂O, en un volumen final de 10 µL, bajo el siguiente programa de termociclaje: 25 ciclos de 96°C por 10s, 50°C por 5s y 60°C por 30s en un termociclador GeneAmp® PCR System 9700 (AB).

Posterior a la reacción de mini-secuenciación es preciso realizar una segunda purificación, en la que se eliminan los ddNTPs no incorporados más no los primers extendidos; esta

purificación se lleva a cabo por medio de la enzima SAP (*Shrimp Alkaline Phosphatase*), la cual desfosforila el extremo 5' de los ddNTPs y los inactiva para no causar interferencia en el análisis de datos en la electroforesis capilar.

Esta reacción enzimática se realizó de la siguiente manera: todo el volumen de reacción de SNaPshot (5-10 μL) y 1 μL de USB® SAP (Affymetrix, Inc.), a 37°C por 60 minutos seguidos de 15 minutos a 85°C, en un termociclador GeneAmp® PCR System 9700 (AB).

3.5.3 Electroforesis Capilar

Una vez finalizada la purificación de los productos de mini-secuencia, se procede con la separación de los fragmentos fluoro-marcados por medio de electroforesis capilar. La preparación para la separación consta de una mezcla de 0.3 μL de marcador interno de tamaño GeneScan-LIZ120™, 9.5 μL de formamida HiDi™ y 2 μL de muestra purificada. Una vez preparada la muestra, se llevó a cabo el corrido de las muestras en una secuenciador ABI PRISM 3130xl® Genetic Analyzer (AB) utilizando el polímero de corrido POP4®. Las condiciones de electroforesis fueron las del módulo que por defecto trae el secuenciador.

Los resultados de la electroforesis capilar fueron analizados en el software informático Genemapper™ ID 3.2.1 para el cual se generaron ventanas de aceptación (BINs) con la herramienta Genemapper Manager de Genemapper™ ID para el tallaje y asignación automática de los alelos SNP.

Marcador	Primer de mini-secuenciación	Anillamiento	Cambio de Base	CF (μM)	Multiplex	Diseño
SRY-1532	TGTATCTGACTTTTTCACACAGT	Fwd	A > G	0.15	1	Whitfield y cols (1995)
M213	TCAGAACTTAAACATCTCGTTAC	Rev	T > C	0.35	1	Brión y cols. (2004)
M173	CTTACAATCAAGGGCATTAGAAC	Fwd	A > C	0.3	1	Brión y cols. (2004)
M9	GGCCTAAGATGGTTGAAT	Fwd	C > G	0.2	1	Brión y cols. (2004)
M22	GCCATTCTGGTGGCTCT	Fwd	A > G	0.05	1	Brión y cols. (2004)
M70	TAGGATTCTGTTGTTGTTAGTCTTAG	Rev	A > C	0.1	1	Brión y cols. (2004)
TAT	CTCTGAAATATTAATAAAACAAC	Rev	T > C	0.4	1	Brión y cols. (2004)
92R7	GCATGAACACAAAAGACGTAGAAG	Rev	G > A	0.2	1	Brión y cols. (2004)
M173	CTTACAATCAAGGGCATTAGAAC	Fwd	A > C	0.3	1	Brión y cols. (2004)
P25	TCTGCCTGAAACCTGCCTG	Fwd	C > A	0.1	1	Brión y cols. (2004)
M269	GGAATGATCAGGGTTGGTTAAT	Fwd	T > C	0.4	1	Brión y cols. (2004)
M201	GATCTAATAATCCAGTATCAACTGAGG	Fwd	G > T	0.15	2	Brión y cols. (2004)
M170	ACACAACCCACTGAAAAAAA	Fwd	A > C	0.2	2	Brión y cols. (2004)
M26	CATAGGCCATTCAGTGTCTCTG	Rev	G > A	0.1	2	Brión y cols. (2004)
M62	CAATGTTTGTGGCCATGGA	Rev	T > C	0.4	2	Brión y cols. (2004)
M172	CCCCAAACCAATTTTGTGCTT	Fwd	T > G	0.05	2	Brión y cols. (2004)
M96	GTAAGTGGAAAACAGTCTCTCATAATA	Fwd	G > C	0.725	3	Brión y cols. (2004)
M35	AGTCTCTGCCTGTGTC	Rev	G > C	0.34	3	Brión y cols. (2004)
M78	ACACTTAACAAAAGATACTTCTTTC	Fwd	C > T	1.21	3	Brión y cols. (2004)
M81	TAAATTTTGTCTTTTGTGAA	Fwd	C > T	1.21	3	Brión y cols. (2004)
M123	TTCTAGGTATTCAGGCGATG	Rev	G > A	1.55	3	Brión y cols. (2004)
M34	TTGCAGACACACCACATGTG	Rev	G > T	0.52	3	Brión y cols. (2004)
M3	GGGTACCTCTGGGACTGA	Fwd	C > T	0.25	5	Blanco-Verea y cols. (2010)
M19	CAAAGTATTTTGTGAAGACTGTTGTA	Fwd	T > A	0.25	5	Blanco-Verea y cols. (2010)
M194	ACTTTAAACATACAGGGAGTGTITTT	Fwd	T > C	0.15	5	Blanco-Verea y cols. (2010)
M199	GAAATGTTAAATGGCTTACACTTG	Fwd	C > G	0.3	5	Blanco-Verea y cols. (2010)
M242	AAAAAGGTGACCAAGGTGCT	Fwd	C > T	0.15	5	Blanco-Verea y cols. (2010)
M167	AAGCCACAGGGTGC	Fwd	C > T	0.4	6	Ramos-Luis y cols (2009)
M222	TCTGACAACATTCCTAATCCCTCATCCGA	Fwd	G > A	0.5	6	Ramos-Luis y cols (2009)
U152	GAAAGTCTGACAACATATACATTACTTTGAGAAGTATGG	Rev	G > A	1.17	6	Ramos-Luis y cols (2009)
U106	AAGTCTGACAATTCAGAGAAGAAGCAATTGAACCC	Rev	C > T	0.5	6	Ramos-Luis y cols (2009)
U198	ACGTCGTGAAAGTCTGACAATTCATCATGGTATAACTG	Fwd	G > A	0.5	6	Ramos-Luis y cols (2009)

Tabla 6. Secuencia de las sondas empleadas en la extensión de una sola base con el kit SNaPshot de 32 Y-SNPs. La tercera columna indica la posición de anillamiento de la sonda, Fwd (Forward) o Rev (Reverse). La cuarta columna indica los alelos interrogados por el ensayo. La quinta columna, indica la concentración final de la sonda en el respectivo multiplex de la sexta columna.

3.6 Análisis Estadísticos

Para los análisis de todos los Y-STRs amplificados en la presente tesis (*Yfiler*[®], Powerplex Y23 y RM-Y-STRs), se estimaron las frecuencias haplotípicas y diversidad génicas, por medio del software Arlequin v3.51.2 (Excoffier and Lischer 2010). Así mismo, se calculó la capacidad de discriminación haplotípica (DC), dividiendo el número de diferentes haplotipos encontrado por cada departamento por el tamaño de la muestra. Con el fin de estimar los valores reales de diversidad del sistema microsatélite DYS389I, se restaron el número de repeticiones a partir del sistema DYS389II (*i.e.* DYS389I (Real) = DYS389II – DYS389I (genotipado)). Para la identificación de alelos intermedios en la población boliviana, se consultó su frecuencia en <http://www.yhrd.org>.

Para determinar, el porcentaje de varianza y la posible estructura genética presente en la población boliviana, se usaron diferentes arreglos para su determinación, como se explica a continuación; como se mencionó en el apartado de muestras, la muestra global se dividió en los 5 departamentos muestreados como también en las 3 zona geográficas (Andina, Sub-Andina y Llanos), de esta manera el análisis de varianza (AMOVA) fue ejecutado bajo estos 2 distintos arreglos. Para el cálculo de la AMOVA, se eligió la opción por distancias R_{ST} , debido a la naturaleza de los polimorfismos.

Adicionalmente, para el análisis entre poblaciones y la población de Bolivia, se determinaron las frecuencias de haplogrupos de varias poblaciones de América del Sur disponibles y a partir de estos patrones de frecuencias se realizó un análisis de componentes principales (PCA), que permite observar las relaciones entre la población de Bolivia y las de su alrededor. Este análisis fue realizado por medio del paquete estadístico R (<http://www.-r-project.org>).

Por último, se determinó el nivel de mestizaje promedio de la población boliviana a través de la frecuencia de Y-SNPs disponible para diferentes poblaciones mundiales, analizadas por (Simms *et al.* 2012); este análisis se basó en el método de mínimos cuadrados propuesto por (Long *et al.* 1991), el cual consiste en la comparación de los principales haplogrupos de Y-SNPs presentes en diferentes poblaciones y agrupados en grupos parentales de África, Europa, este de Asia, sur de Asia y de grupos nativo Americanos; el algoritmo permite estimar el nivel de mestizaje, en este caso de la población de Bolivia, a partir de los grupos parentales. Este análisis se realizó por medio del Software informático SPSS v17.0



IV. Resultados y Discusión





4.1 Variabilidad genética del Yfiler® en población boliviana

4.1.1 Diversidad molecular de los Y-STRs

El conjunto de muestras analizadas ($n=223$) con *Yfiler* y ($n = 226$) con Y-SNPs, están publicadas en el portal <http://www.yhrd.org> bajo el código de acceso [YA003827] [Bolivia, Boliviano]. Las muestras analizadas con *Yfiler* sumaron un total de 202 haplotipos y un total de 128 haplotipos con el haplotipo mínimo. (Ver Tabla S1 disponible *on line*); Para los haplotipos de *Yfiler*, 15 distintos haplotipos fueron compartidos por dos individuos y dos distintos haplotipos compartidos por cuatro individuos. En el caso de los haplotipos mínimos, un solo haplotipo fue compartido por 22 individuos, otro más por 17 y también uno más por 5; dos haplotipos también fueron compartidos por 7 individuos, 3 haplotipos por 4 individuos, 9 haplotipos por 3 individuos y por último 15 haplotipos por dos individuos. Por departamentos, se obtuvieron un total de 17, 32, 25, 24 y 60 haplotipos mínimos para Pando, Beni, Santa Cruz, Cochabamba and La Paz, respectivamente (Tabla 7), igualmente 27 haplotipos fueron compartidos entre los departamentos. Entre los haplotipos mínimos más frecuentes se encuentra el haplotipo, 13-14-17-23-10-16-14 con una frecuencia de 16.8%, 7.7% y 7.1% en los departamentos de La Paz, Santa Cruz and Cochabamba, respectivamente. Como es de esperar los índices de diversidad fueron mayores para los haplotipos de *Yfiler* que para los del haplotipo mínimo. Desde un enfoque forense este es un hallazgo relevante, ya que la capacidad de discriminación entre individuos en la población de Bolivia, asciende a un valor de 0.91 para las muestras tipadas con *Yfiler* pero se disminuye drásticamente a 0.57 solo considerando el haplotipo mínimo (Tabla 7).

Analizando la distribución de los índices de diversidad entre los departamentos, Santa Cruz (Región Llanos), muestra los mayores valores respecto a los demás departamentos, tanto para el haplotipo mínimo como para los haplotipos con *Yfiler* (con la excepción en este último para Beni (Region Llanos) en la diversidad génica como para el promedio de las diferencias *pairwise*); Igualmente cuando son analizados estos índices de diversidad entre las regiones eco-geográficas, la región de los Llanos es la que exhibe los mayores valores de diversidad tanto para los haplotipos mínimos como para los de *Yfiler* (con excepción de este último en la diversidad haplotípica para la región Sub-andina)

Respecto a los alelos intermedios se reporta un individuo de Santa Cruz con un alelo 14.2 en el *locus* DYS19; es de resaltar que este alelo intermedio ha sido reportado anteriormente

al menos 4 veces en individuos de poblaciones americanas (<http://www.yhrd.org>). Así mismo, 3 individuos de Cochabamba presentaron alelos intermedio en el *locus* DYS458, dos de ellos con el alelo 17.2 (Individuos pertenecientes a los haplogrupos J* y I*) y el tercero con un alelo 20.2 perteneciente al haplogrupo R1b1*, (que en el presente estudio se refiere a R1b1(xM269)). Así mismo, para este mismo departamento, se reportó un individuo (Hg Q*) con un alelo nulo para el sistema DYS439, alelo que ha sido reportado con anterioridad en al menos nueve poblaciones, (nativo americanas, euroasiáticas grupos del este de Asia, como también un único individuo africano-americano, consultado en <http://www.yhrd.org>)

Población	N	K	DC	UH	HD	h	M
Yfiler							
Pando	21	20	0.95	20	0.9952 ± 0.0165	0.610417 ± 0.325057	9.766667 ± 4.659243
Beni	41	38	0.93	36	0.9963 ± 0.0063	0.652287 ± 0.337224	10.436585 ± 4.857652
Santa Cruz	26	26	1.00	23	1.0000 ± 0.0107	0.649808 ± 0.341165	10.396923 ± 4.899982
Cochabamba	28	28	1.00	27	1.0000 ± 0.0095	0.610946 ± 0.321076	9.775132 ± 4.614161
La Paz	107	93	0.87	87	0.9965 ± 0.0021	0.599442 ± 0.306762	9.591077 ± 4.431930
Llanos	88	82	0.93	81	0.9984 ± 0.0020	0.642682 ± 0.328070	10.282915 ± 4.737945
Andina	107	93	0.47	87	0.9965 ± 0.0021	0.593480 ± 0.303917	9.495680 ± 4.390827
Sub-Andina	28	28	2.32	27	1.0000 ± 0.0095	0.604993 ± 0.318155	9.679894 ± 4.572186
Todo Bolivia	223	202	0.91	196	0.9989 ± 0.0006	0.624785 ± 0.317280	9.996566 ± 4.588061
Haplotipo mínimo							
Pando	21	17	0.81	10	0.9810 ± 0.0197	0.576190 ± 0.334459	4.033333 ± 2.097376
Beni	41	32	0.78	21	0.9890 ± 0.0074	0.598955 ± 0.337344	4.192683 ± 2.125982
Santa Cruz	26	25	0.96	16	0.9969 ± 0.0117	0.640879 ± 0.363256	4.486154 ± 2.282552
Cochabamba	28	24	0.86	14	0.9815 ± 0.0181	0.561224 ± 0.322740	3.928571 ± 2.029159
La Paz	107	60	0.56	45	0.9552 ± 0.0120	0.582890 ± 0.324388	4.080233 ± 2.050375
Llanos	88	66	0.75	52	0.9916 ± 0.0034	0.610651 ± 0.338452	4.274556 ± 2.138451
Andina	107	60	0.56	45	0.9552 ± 0.0120	0.582890 ± 0.324388	4.080233 ± 2.050375
Sub-Andina	28	24	0.86	14	0.9815 ± 0.0181	3.928571 ± 2.029159	0.561224 ± 0.322740
Todo Bolivia	223	128	0.57	128	0.9806 ± 0.0043	0.594895 ± 0.328544	4.164263 ± 2.078542
Y-SNPs							
Pando	22	-	-	-	0.6857 ± 0.0892	0.099471 ± 0.061342	2.685714 ± 1.482775
Beni	41	-	-	-	0.6585 ± 0.0479	0.093586 ± 0.057017	2.526829 ± 1.385439
Santa Cruz	26	-	-	-	0.7908 ± 0.0674	0.122963 ± 0.072505	3.320000 ± 1.756627
Cochabamba	28	-	-	-	0.7328 ± 0.0689	0.105232 ± 0.063484	2.841270 ± 1.538839
La Paz	109	-	-	-	0.6928 ± 0.0288	0.072298 ± 0.045729	1.952037 ± 1.114639
Llanos	89	-	-	-	0.7965 ± 0.0293	0.102048 ± 0.057705	3.673720 ± 1.874895
Andina	109	-	-	-	0.7161 ± 0.0327	0.067576 ± 0.040865	2.432728 ± 1.328156
Sub-Andina	28	-	-	-	0.7354 ± 0.0701	0.076940 ± 0.046623	2.769841 ± 1.506807

Tabla 7. Índices moleculares de diversidad en población Boliviana analizada. N = Tamaño muestral; K = Número de distintos haplotipos; DC = Capacidad de discriminación; UH = número de haplotipos únicos; HD = Diversidad Haplotípica; h = diversidad génica entre los *loci*; M = promedio de las diferencias *pairwise*

4.1.2 Diversidad de los haplogrupos de Y-SNPs

Para la población boliviana analizada, el componente europeo esta principalmente representado por los haplogrupos R1b1-P25 (incluyendo el sub-haplogrupo R1b1b2-M269), I-M170, I2a2-M26, G-M201, J-M304 (que incluye los sub-haplogrupos J1a-M62 y J2-M172) que corresponden a un 65.5% y un componente nativo americano representado principalmente por los haplogrupos haplogroups P*, Q (que incluyen el sub-haplogrupo Q1a3a-M3) con un 28%, que corresponden en total al 94% del total de la muestra (**Figura**

12). El 6% restante, explicado por los haplogrupos BR-SRY10831.1, E-M96 (que incluye los sub-haplogrupos, E1b1b1-M35, E1b1b1a-M78 y E1b1b1b-M81), se encuentra normalmente distribuido a lo largo de la región andina a Llanos. Así mismo, se resalta que el 26.6% de estos haplotipos de ancestralidad africana reciente, se encontraron en la pequeña región ‘Afro-Boliviana’ de Bolivia en Tocaña (Región Andina – Departamento de la Paz – Provincia de Nor Yungas)

Tal vez, las diferencias más importantes entre las regiones es la frecuencia del haplogrupo R1b1-P25; este haplogrupo alcanza una frecuencia de 32% en la región de los Llanos y 71% en la región sub-andina. Sin embargo, el hallazgo más “chocante” es que el haplogrupo R1b1*(xM269) representado por un 46% en las regiones sub-andina y andina, pero se encuentra completamente ausente en la región de los Llanos (**Figura 12**).

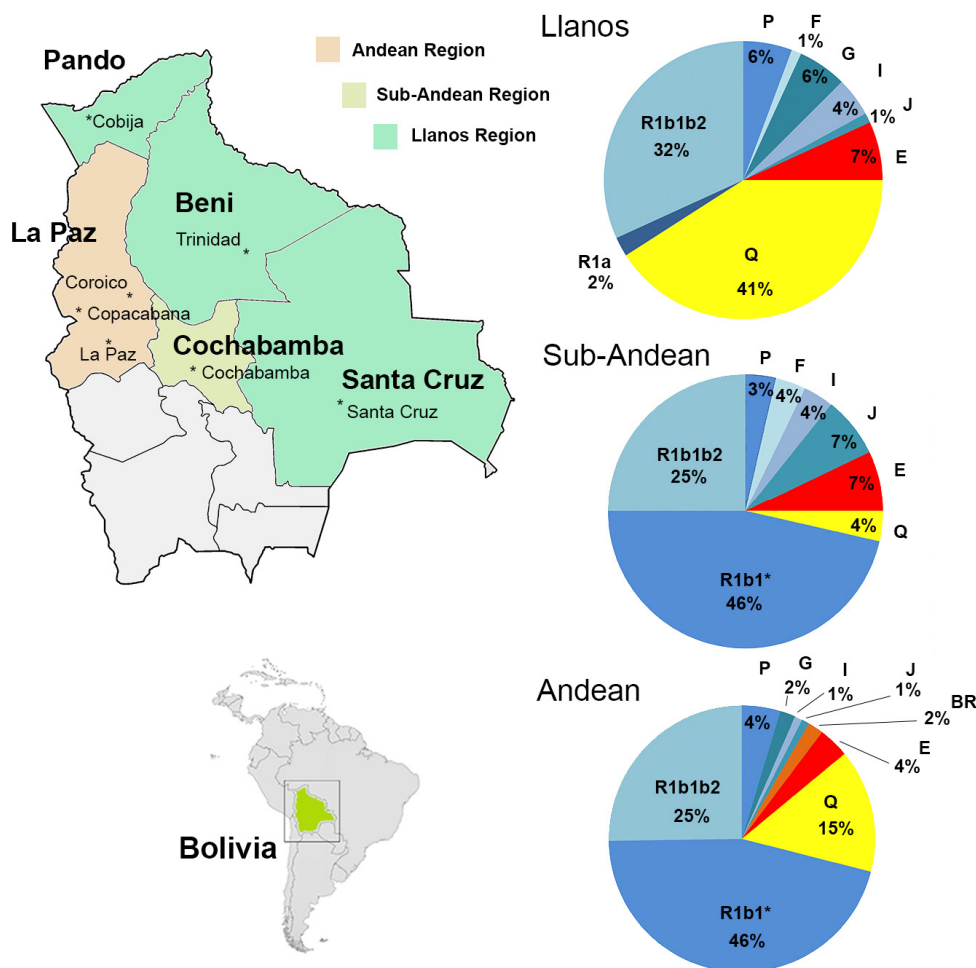


Figura 12. Mapa en el que se enseña los lugares de muestreo, señalados con un asterisco y los departamentos de Bolivia; los departamentos fueron agrupados en 3 regiones eco-geográficas (Andina, sub-Andina y Llanos); las frecuencias de los haplogrupos de cromosoma Y son resumidos en los gráficos circulares. La información completa de haplogrupos y haplotipos se muestra en la Tabla S1 disponible *on line*.

En general la región andina y sub-andina muestran un componente principalmente europeo que representan un 77% y un 86% del total de la muestra, mientras que la región de los Llanos, la que mayor componente representa solo un 46% del total de la muestra. Esta última región es la que mayor componente nativo americano presenta dada la de frecuencia de los haplogrupos nativo americanos representados por: 34% Q1a3a*-M3, 7% Q*-M242, y 6% P*-92R7; el porcentaje restante corresponde a los haplogrupos de componente típico europeo (R1b1b2-M269) (**Figura 12**).

Finalmente, la región de los Llanos muestra el valor diversidad de haplogrupo mas alto (0.7965 ± 0.0293), seguido de las regiones andina y sub andina (0.7354 ± 0.0701 , 0.7161 ± 0.0327 ; respectivamente).

4.1.3 Análisis de varianza molecular (AMOVA) de los Y-STRs

Los análisis de varianza (AMOVA) fueron estimados para los haplotipos *Yfiler* como para los haplotipos mínimos. Como es de esperar la varianza ocurre mayormente dentro de las poblaciones. La varianza entre las poblaciones estimada para los haplotipos mínimos fue levemente mayor que los estimados para los perfiles *Yfiler* (Tabla 8). Este efecto ha se ha descrito con anterioridad (Brisighelli *et al.* 2012), y puede ser debido al hecho de que los perfiles de *Yfiler* aumentan la cantidad de haplotipos únicos en toda la muestra y por lo tanto aumentando la diversidad haplotípica a su máximo valor teórico (Tabla7).

Los haplotipos de cromosoma Y analizados en la presente tesis y adicionalmente agrupado en departamentos aumentan la varianza entre la población de 4.94 (usando el haplotipo mínimo) a 5.51% (usando el haplotipo *Yfiler*). Así mismo, cuando se incorporan más perfiles de Bolivia, descritos en la literatura, el análisis tiene un aumento en la varianza substancial de 9.74 (usando el haplotipo mínimo) a 11.15% (usando el haplotipo *Yfiler*).

Igualmente, la AMOVA fue estimada para las 3 regiones eco-geográficas analizadas en la presente tesis (Tabla 8).

Este análisis demuestra que la geografía (en términos de altitud) tiene un impacto significativo (varianza entre las poblaciones, 3.75 (usando el haplotipo *Yfiler*) y 5.04% (usando el haplotipo mínimo), sin embargo es mucho menor cuando es analizada la muestra por departamentos.

	Pop.	N	17 Y-STRs		7 Y-STRs	
			Among	Within	Among	Within
Bolivia (Presente estudio [regiones])	3	223	3.75	96.25	5.04	94.96
Bolivia (Presente estudio [departamentos])	5	223	5.51	94.49	4.94	95.06
Bolivia (Presente estudio + literatura)	10	521	8.94	91.06	8.05	91.95
América (todas las poblaciones)	16	1660	14.60	85.40	19.21	80.79
América menos Colombia	15	1355	17.99	82.01	21.32	78.68
América menos Argentina	14	1582	14.00	86.00	18.73	81.27
América menos Chile	14	1268	16.03	83.97	22.44	77.56
América menos Brazil	15	1296	15.39	84.61	19.11	80.89
América menos Bolivia	6	1139	4.86	95.14	7.28	92.72

Tabla 8. AMOVA estimada para las distintas poblaciones Bolivianas de América del sur. El análisis se estimó usando el grupo de haplotipos por *Yfiler* (17 YSTRs) como con el haplotipo mínimo (9 YSTRs). Diferentes análisis de AMOVA fueron estimados (i) Bolivianos analizados en el presente estudio agrupados en las 3 regiones eco-geográficas (Andina, Sub-Andina y Llanos), (ii) Bolivianos analizados en el presente estudio agrupados por los 5 departamentos (ver **Figura 1**) y (iii) Bolivianos analizados en el presente estudio agrupados por los 5 departamentos más otras 3 poblaciones Bolivianas disponibles en la literatura, sumando en total 8 poblaciones (descritas a continuación). La columna “Pop” Indica el número de grupos poblacionales considerados, N se refiere a número total de la muestra. La significancia estadística fue evaluada usando el procedimiento de permutación implementado en Arlequin (10,000 permutaciones); Todos los valores de *P* fueron estadísticamente significantes. Los haplotipos de cromosoma Y fueron tomados de la literatura, Gayà-Vidal et al. (2011) (Bolivia: Aymara, *n* = 59; Quechua *n* = 55), Tirado et al. (2009) (Bolivia: departamento de El Beni, *n* = 87), Vullo et al. (2014) (Bolivia: departamento Chuquisaca, *n* = 38, departamento de La Paz, *n* = 57) Romero et al. (2008) (Colombia: Caribe, *n* = 305), Toscanini et al. (2006) (Argentina: Collas, *n* = 29; Toba, *n* = 49), Roby et al. (2009) (Chile: Santiago de Chile, *n* = 196; Iquique, *n* = 196), y Caine et al. (2010) más Schwengber et al. (2009) (South Brazil: Rio Grande do Sul and Santa Catarina, *n* = 364).

Un total de 1,437 perfiles completos de *Yfiler* fueron recuperados de la literatura, en las que se contaban principalmente, poblaciones de América del Sur y de Bolivia (Iannacone et al. 2005, Toscanini et al. 2006, Lee et al. 2007, Borjas et al. 2008, López-Parra et al. 2008, Romero et al. 2008, Gonzalez-Andrade et al. 2009, Roby et al. 2009, Schwengber et al. 2009, Tirado et al. 2009, Blanco-Verea et al. 2010, Caine et al. 2010, Marino and Furfuro 2010, Gayà-Vidal et al. 2011), con el fin de determinar los patrones de variación de los Bolivianos en una magnitud continental. La AMOVA que consideraba todas las poblaciones de América del Sur con perfiles de *Yfiler* arrojó valores considerables de varianza entre las poblaciones (14.60 [*Yfiler*]- 19.21% [haplotipo mínimo]). Cuando se eliminó una población a la vez por análisis de AMOVA, se obtienen valores muy altos de varianza entre las poblaciones, como es el caso cuando se elimina Colombia (17.99 [*Yfiler*]- 21.32% [haplotipo mínimo]), Chile (16.03 [*Yfiler*] - 22.44% [haplotipo mínimo]) y

Argentina (14.00 [Yfiler] - 18.73% [haplotipo mínimo]). Aun así, cuando la muestra de Bolivia es eliminada de análisis de AMOVA la varianza entre las poblaciones disminuye considerablemente (4.86 [Yfiler] - 7.28% [haplotipo mínimo]), sugiriendo que los haplotipos de la población Boliviana, contribuyen de manera importante a la estratificación de la muestra.

4.1.4 Patrones espaciales de Y-STRs y de la variación en Y-SNPs

Un diagrama de componentes principales (PCA) basado en las frecuencias de los haplogrupos fue generado para facilitar la visualización de las relaciones genéticas a nivel poblacional entre Bolivia y otras poblaciones sur americanas y continentales (Alves *et al.* 2007). La componente 1, PC1 (que explica el 42.7% de la variabilidad) y PC2 (32.4%) muestra tres grupos claramente diferenciados, de poblaciones Africanas, del Este de Asia y Europeas (**Figure 13**). Es de resaltar, que poblaciones con un conocido nivel de ancestralidad nativo americana como los Aymara y los quechuas de Bolivia o los Colla y Diaguitas de Argentina aparecen en PC1 y en PC2 más cercanas a las poblaciones europeas, que a poblaciones del Este de Asia, (lo que se esperaría debido a la ancestralidad demográfica histórica) y este hecho tienen sin lugar a dudas se debe a que muchas de estas poblaciones consideraras tienen un importante porcentaje de mestizaje (ancestralidad) europeo.

De este mismo gráfico se puede resaltar que las regiones andina y sub-andina, analizadas en la presente tesis, están más relacionadas con los europeos, como es esperado por su alto componente de ancestralidad de haplogrupos europeos en sus cromosomas (PC1 y PC2, **Figura 13**). Finalmente es en la PC3 (20.4%) donde las poblaciones de alto componente nativo americano en el cromosoma Y donde se diferencian de los grupos africanos, europeos y del este de Asia. Las regiones andina y sub-andina del presente estudio, se localizan más cerca del polo Europa-Este de Asia, mientras que la región Llanos está más cercana a las otras poblaciones nativo americanas del gráfico (PC1 vs. PC3) como son la Aymara y la Quechua.

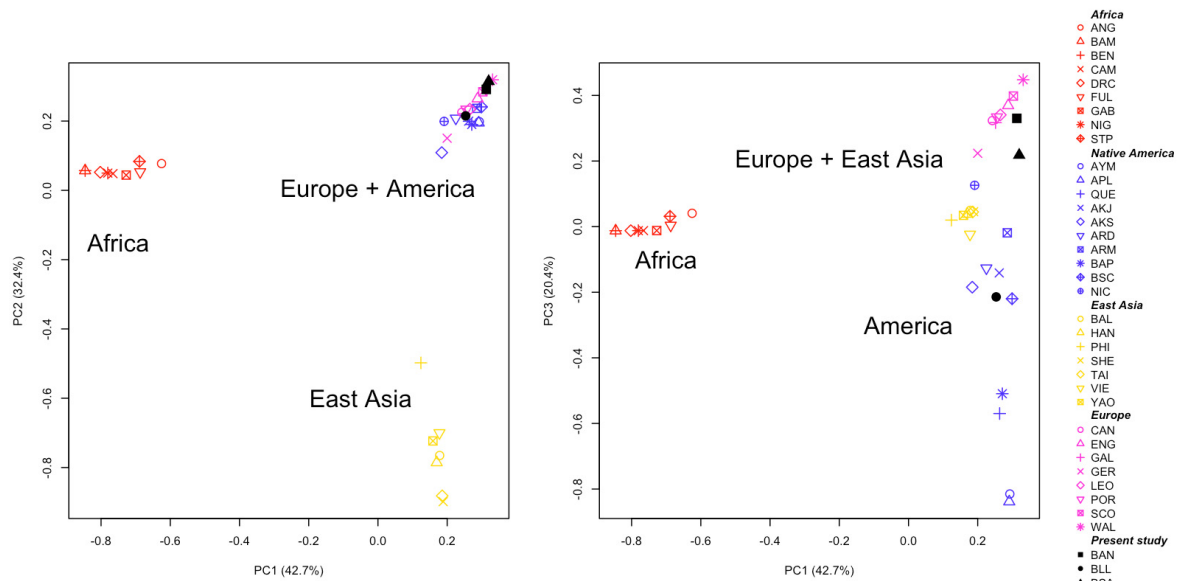


Figura 13. Análisis de componentes Principales (PCA) de las frecuencias de haplogrupos de Cromosoma Y; los ejes explican el porcentaje de varianza explicado por cada componente. Códigos de las poblaciones: África = Fon (BEN) (Luis *et al.* 2004, Simms *et al.* 2011); Fulbe (FUL) (Rosa *et al.* 2007); São Tomé de Príncipe (STP) (Gonçalves *et al.* 2007); Bakaka, Ewondo, Ngoumba (CAM), Makina and Nzebi (GAB), Tiv y Igala (NIG) (Montano *et al.* 2011); Angola (ANG) (Beleza *et al.* 2005); Bamelike (BAM) (Luis *et al.* 2004); República democrática del Congo (DRC) (de Filippo *et al.* 2011); Nativos Americanos = Nicaragua (NIC) (Nuñez *et al.* 2010); Bolivia - Aymara (AYM) y Bolivia - Quechua (QUE) (Gayà-Vidal *et al.* 2011); Andina - La Paz (BAP), y Sub-Andina - Chuquisaca (BSC) (Vullo *et al.* 2014); Apalai (APL) (McLean *et al.* 2005); Diaguitas (ARD), Kolla - Jujuy (AKJ), Kolla - Salta (AKS), y Mapuche (ARM) (Blanco-Verea *et al.* 2010); Este de Asia = Bali (Indonesia) (BAL) (Karafet *et al.* 2005); Filipinas (PHI), Taiwán (TAI), She (SHE), Yao (YAO), y Vietnam (VIE) (Karafet *et al.* 2010); Han (HAN) (Zhong *et al.* 2011); Europa = Cantabria (CAN), León (LEO) (Flores *et al.* 2004); Inglaterra (ENG), Gales (WAL), Escocia (SCO) (Campbell 2007); Galicia (GAL) (Brión *et al.* 2004); Alemania (GER) (Kayser *et al.* 2005); Portugal (POR) (Beleza *et al.* 2006); Presente estudio= Region Andina (BAN), Sub-Andina Region (BSA), y Llanos (BLL).

4.1.5 Análisis de componentes de ancestralidad

El análisis de ancestralidad, basado en los componentes mayoritarios de haplogrupos de Y-SNPs encontrados en poblaciones continentales fue usado para determinar la proporción de ancestralidad en la población de Bolivia. (Simms *et al.* 2012) (**Tabla 9**). La mayor proporción de ancestralidad, para la región de los Llanos es nativa americana (80%), seguida de una ancestralidad europea (19%); Así mismo, la región sub-andina muestra un alto nivel de *mestizaje* con Europa (94%) y una disminuida proporción nativa americana (6%). El caso de la región andina, es un caso intermedio entre las dos regiones mencionadas anteriormente, donde la composición de ancestralidad es de 76% europea y 24% nativo americana.

Grupos parentales	Llanos	Sub-Andina	Andina
África	0.000 ± 0.258	0.000 ± 0.870	0.000 ± 0.927
Europa	0.194 ± 0.236	0.945 ± 0.797	0.757 ± 0.849
Nativo Americano	0.799 ± 0.054	0.055 ± 0.182	0.243 ± 0.194
Este de Asia	0.000 ± 0.035	0.000 ± 0.119	0.000 ± 0.127
Sur de Asia	0.000 ± 0.071	0.000 ± 0.241	0.000 ± 0.256

Tabla 9. Proporciones de ancestralidad generadas para las regiones eco-geográficas de Bolivia. Estas proporciones fueron estimadas usando los conjuntos de Y-SNPs parentales de Simms et al. (Simms *et al.* 2012)

Adicionalmente, aunque se encontraron haplogrupos típicos africanos (E y BR) que están representados en un 6% en la muestra total, el análisis de mestizaje, no es capaz de captar esta componente en la población Boliviana. Finalmente, el análisis muestra que no hay contribución de ancestralidades del Este y Sur de Asia para la población.

4.1.6 Análisis del sesgo por género entre los polimorfismos de cromosoma Y, ADN mitocondrial y AIMs

Para el total de los 223 Y-SNPs y Y-STRs tipados en la muestra de la población Boliviana, 152 haplotipos de la región hipervariable I (HVS-I) ADN mitocondrial fueron descritos como también 46-AIM-indels (**Table S2**) y adicionalmente 6 muestras fueron tipadas por el panel de ancestralidad LACE (Galanter *et al.* 2012). Este conjunto de datos son un escenario ideal para la evaluación de la contribución de cada género de distinto grupo ancestral a los bolivianos de hoy en día (**Figura 14**).

La ancestralidad predominante en los bolivianos es principalmente nativo americana (94.7%) y la restante es europea (5.3%); esta ancestralidad europea está concentrada mayormente en la region de los Llanos donde su valor es de 9.6% del total de esta región.

Desde la óptica mitocondrial la ancestralidad boliviana, es aún más nativo americana (98.7%), con una pequeña proporción africana (1.3%); la mayoría de este componente de ancestralidad está concentrado en la región sub-andina donde su valor es del 11.1% del total de esta región.

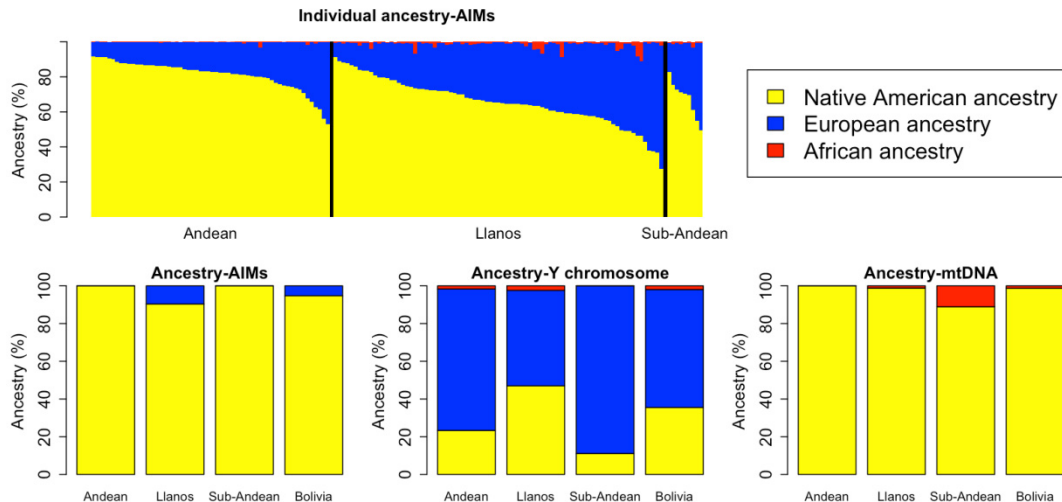


Figura 14. Proporciones de ancestralidad estimadas desde distintos marcadores como son, AIMs-Indels, ADNmt, y cromosoma Y. Las estimaciones para los AIMs y para el ADNmt fueron tomadas de (Taboada-Echalar *et al.* 2013), y fueron usadas para generar esta figura que tiene el objetivo de comparar estos valores con los obtenidos para Cromosoma Y. Todos los individuos considerados para esta figura, fueron tipados para los marcadores AIMs y los marcadores uniparentales (Ver Tabla S2 disponible *on line*).

Por el contrario y como se ha analizado anteriormente, el cromosoma Y muestra principalmente un componente Europeo (65%) y el componente restante es nativo americano (29%) y africano (6%). La región de los Llanos como ya se ha demostrado, es la que tiene la mayor ancestralidad nativo americana de cromosoma Y; Teniendo en cuenta que el componente nativo americano está determinado por los haplogrupos Q1a3a*-M3, Q*-M242, and P*-92R7, la región sub-andina es donde la ancestralidad europea predomina.

Región	Muestra	HG Cr-Y	HG ADNmt	Panel LACE 446-AIM (%)			Panel AIM-Indels (%)		
				N. Americano	Africana	Europea	N. Americano	Africana	Europea
Andean	337COROICO	R1b1(xM269)	A2	100	0	0	90	0	10
Andean	422COPACABANA	R1b1(xM269)	B4	100	0	0	82	0	18
Andean	409COPACABANA	R1b1(xM269)	B4	100	0	0	90	0	10
Andean	433COPACABANA	R1b1(xM269)	D4h3a4	100	0	0	75	0	25
Andean	414COPACABANA	R1b1(xM269)	B4	100	0	0	78	0	22
Andean	418COPACABANA	R1b1(xM269)	B4	98	0	20	80	0	20

Tabla 10. Sesgo por género entre los polimorfismos de cromosoma Y, ADNmt, AIM-Indels y el Panel LACE en Bolivia. La tabla muestra los 6 individuos de la region andina tipados para el panel de AIM-Indels y ADNmt por (Taboada-Echalar *et al.* 2013), para el Panel LACE analizados en (Galanter *et al.* 2012) y los Y-SNPs analizados en la presente tesis. Los valores de porcentajes de los AIM-Indels, que estimar la componente principal de ancestralidad continental, fueron redondeando al valor promedio.

En la tabla 10 se muestra a modo de resumen, las estimas de ancestralidad de 6 muestras pertenecientes al departamento de la Paz, (Región Andina) que fueron tipadas para el panel de 46 AIM-Indels, el Panel LACE y para los marcadores uniparentales. El panel LACE contiene más SNPs (446 en total) y esto se debe traducirse en estimas de ancestralidad mas exactas que se aproximan mejor a la ancestralidad genómica global (Pardo-Seco *et al.* 2014). El resultado del panel LACE indica que todas las muestras tienen virtualmente un

100% de ancestralidad nativo americana (0.9966 ± 0.0082); por el otro lado, el panel de AIM-Indels, parece infra estimar la ancestralidad nativo americana con una alta desviación estándar (0.8250 ± 0.0625) y favorece un componente ancestral europeo (0.1800 ± 0.0678), aun así las diferencias son pequeñas. La ancestralidad europea estimada por los AIM-Indels no necesariamente explica la ascendencia europea; se ha demostrado que este efecto puede estar dado por la alta heterogeneidad en las estimas de ancestralidad obtenidas a partir de una cantidad baja de SNPs (Pardo-Seco *et al.* 2014). En este caso en particular el panel LACE contiene un mayor número SNPs que permite aproximarse mejor a la ancestralidad genómica global. La información del cromosoma Y (SNPs and STRs), claramente identifica estos individuos a un haplogrupo típicamente europeo R1b1 (xM269) (que es uno de los predominantes en la región andina, **Figura 12**), mientras tanto el haplogrupo asignado por el ADNmt para todos ellos es típico de ancestralidad nativo americano.

4.1.7 Importancia del estudio de la variabilidad del cromosoma Y en Bolivia.

Los resultados anteriores nos permiten concluir que la ancestralidad predominante de la población Bolivia analizada es Europea (62.5%), con un notable componente Nativo americano (35.5%). Es por lo tanto claro, que la contribución Europea al *pool* de cromosomas Y en Bolivia es alto, como ya se ha descrito en otras poblaciones Sur y Centro Americanas (Lovo-Gómez *et al.* 2007, Salas *et al.* 2008, Acosta *et al.* 2009, de Azevedo *et al.* 2009). Igualmente, a diferencia del gran impacto que ha tenido la influencia Africana en otras poblaciones de Sur y Centro América, como sucede en el caso de Colombia y el Caribe (Salas *et al.* 2005, Salas *et al.* 2005, Mendizabal *et al.* 2008, Salas *et al.* 2008, Nuñez *et al.* 2010, Marcheco-Teruel *et al.* 2014), la ancestralidad africana en Bolivia es muy baja (2%).

Así mismo, la variación de la ancestralidad continental varía sustancialmente en las regiones eco-geográficas, como es el caso más notable entre la region de los Llanos vs Andina. La región Llanos es la región con el componente nativo americano más alto (47%), en comparación con la región Andina (23.5%) y Sub-Andina (11.1%). Los análisis de ancestralidad de mostraron una vez más el contraste de las regiones Llanos, Andina y Sub-Andina, mostrando altos niveles de ancestralidad Europea en altas altitudes *versus* altos niveles de ancestralidad nativo americana en bajas altitudes. El haplogrupo Q es el que representa a mayor grado el componente nativo americano en los Bolivianos, como también

se ha descrito para otras poblaciones Sur Americanas (Roewer *et al.* 2013). Respecto a los índices de diversidad molecular de los marcadores Y-STRs y Y-SNPs, la región de los Llanos, fue la que en general mostró mayores valores. Adicionalmente el análisis de AMOVA de Y-STRs demostró, que cuando la población de Bolivia es eliminada del grupo de otras poblaciones Suramericanas, esta población mostraba una estratificación notable. Así mismo, esta estratificación no puede ser apreciada en este grupo sub-continental, cuando el análisis de basa en el haplotipo mínimo, lo que permite concluir que el haplotipo mínimo no aporta variación importante entre las poblaciones, esto es principalmente debido a la gran homogeneidad que presentan estos 9 *loci* entre las poblaciones analizadas.

El análisis comparativo de la ancestralidad de los haplotipos de cromosoma Y con las ancestralidades estimadas para AIM-Indels y ADNmt (Heinz *et al.* 2013, Taboada-Echalar *et al.* 2013) mostró una concordancia, que permite afirmar con certeza que el componente ancestral Nativo Americano es muy importante en Bolivia. Sin embargo, es el cromosoma Y el que revela la ancestralidad Europea predominante (62.5%; especialmente en las regiones sub-andina [88.9%] y andina [75%]) en comparación con los autosomas (5.3%) y el ADNmt (0%). Por lo tanto, estos datos permiten indicar con veracidad que existe un claro sesgo de género en Bolivia, como ya se había sugerido anteriormente por (Taboada-Echalar *et al.* 2013); este sesgo de géneros es explicado por el predominante impacto masculino que tuvieron los conquistadores españoles en Bolivia durante los años coloniales, un efecto frecuente en otras poblaciones americanas (Salas *et al.* 2008).

Así mismo el presente estudio da soporte a la naturaleza pluri-étnica de la población boliviana en las distintas regiones del país. Sin embargo el escenario demográfico se torna aún más complejo cuando se comparan los resultados del presente trabajo con los resultados obtenidos por otros autores en las mismas regiones. Por ejemplo mientras las estimas de la ancestralidad del presente estudio fueron calculadas en un 19% para el componente nativo americano en región andina, la muestra de La Paz (localizada igualmente en la región andina) de Vullo *et al.* (2014), muestra un 70% de ancestralidad Nativo Americana. Como se ha demostrado en otras poblaciones de América del Sur (i.e. Argentina (Toscanini *et al.* (2007, 2008, 2010, 2012)) también existe una notable estratificación poblacional entre los bolivianos, lo que justificaría considerar la posibilidad de establecer bases de datos de referencia para estudios forenses y clínicos, con muestreos más detallados de la población boliviana.

4.2. Valoración de la variabilidad de los *Rapidly Mutating* Y-STRs en Bolivia en el contexto de las poblaciones mundiales

En este bloque, se analiza la variabilidad de los *Rapidly Mutating* Y-STRs en 111 poblaciones que representan los grandes grupos continentales. Luego, se analiza pormenorizadamente esta variabilidad en una muestra de población boliviana en el contexto de otros grupos nativos americanos también genotipados para este conjunto de marcadores.

4.2.1 Diversidad Global de los haplotipos de RM-YSTRs en diferenciación de los linajes masculinos

En este estudio colaborativo, participaron 52 laboratorios (Laboratorios participantes ver (Ballantyne *et al.* 2014)), entre ellos la unidad de genética forense de la Universidad de Santiago de Compostela, donde se analizaron 111 poblaciones con este nuevo conjunto de marcadores, agrupados en 17 grupos regionales, y entre ellos se encuentra el grupo regional denominado Nativo Americano en el que se incluyen las muestras de la población de Bolivia, tipadas en el bloque anterior pertenecientes a los haplogrupos Q1a3a*-M3, Q*-M242 y P*-92R7 y que se ($n=99$).

A nivel global el desempeño de este multiplex de 13 RM-YSTRs pudo generar 12,156 haplotipos únicos (de 12,272 muestras) para las poblaciones analizadas; la diversidad haplotípica calculada para todas poblaciones se estimó en 0.9999985 (varianza muestral = 2.85×10^{-8}); los 102 haplotipos compartidos entre las muestras, pertenecían a individuos pertenecientes a las mismas zonas geográficas y en el caso de 6 de estos haplotipos compartidos, se trataba de individuos de la misma población muestreada; entre ellas se encuentran dos individuos de las poblaciones de Chubut y Rio Negro (Argentina), dos individuos Inuit de Groenlandia, Republica Checa (2 individuos), Colonia y Varsovia (2 individuos) y dos individuos de Bután, respectivamente. Con lo anterior se puede calcular la proporción de haplotipos compartidos entre poblaciones y esta fue muy baja, (0.05%) y también se calculó la proporción de haplotipos compartidos dentro de las poblaciones y el valor fue un poco mayor (0.8%).

4.2.2 Análisis global de la sub-estructura geográfica con los *Rapidly Mutating* Y-STRs

El efecto de sub-estructura se evaluó, mediante el estadístico θ (*Theta*) que para la mayoría de las poblaciones analizadas fue muy bajo, indicativo de poca sub estructura poblacional; así mismo para todas las poblaciones continentales se estimó una AMOVA basada en valores F_{ST} , la cual demostró que el 99.98% de la variación haplotípica esta explicada

dentro las poblaciones, un valor muy alto para este tipo de marcadores de Cromosoma Y, también un valor de 0.02% entre las poblaciones y 0.002% entre los grupos regionales. El valor F_{ST} Global fue calculado en 0.00017 y para los grupos regionales el promedio de *pairwise* F_{ST} fue tan solo de 0.000127. Así mismo el valor promedio estimado de *pairwise* F_{ST} para todas las comparaciones poblaciones fue de 0.000826.

Para confirmar la ausencia de Sub estructura, e ilustrar la magnitud de las diferencias haplotípicas entre las poblaciones y las áreas geográficas, se realizó un gráfico de *MDS* (Escalamiento Multi-Dimensional) con los valores F_{ST} linearizados en distancias Slatkin (**Figura 15**) y este refuerza los resultados de la AMOVA, ya que no se observa ningún tipo de agrupamiento específico de las poblaciones e incluso poblaciones pertenecientes al mismo grupo geográfico no forman *clústeres*; el efecto que pudiera tener el tamaño de la muestra en las relaciones entre las poblaciones fue descartado (**Figura 15** - Gráfico superior izquierda), ya que no se observa ningún patrón de agrupamiento; todo esto permite concluir que el alto nivel de mutación de los marcadores RM-YSTRs, desaparece y elimina cualquier tipo de historia compartida entre las poblaciones, efecto explicado por el número de haplotipos únicos en las distintas poblaciones. Igualmente como consecuencia práctica de lo anterior, la necesidad de aplicar correcciones de sub-estructura, por ejemplo en el campo forense, no es necesaria como es frecuente en algunos otros marcadores de ADN.

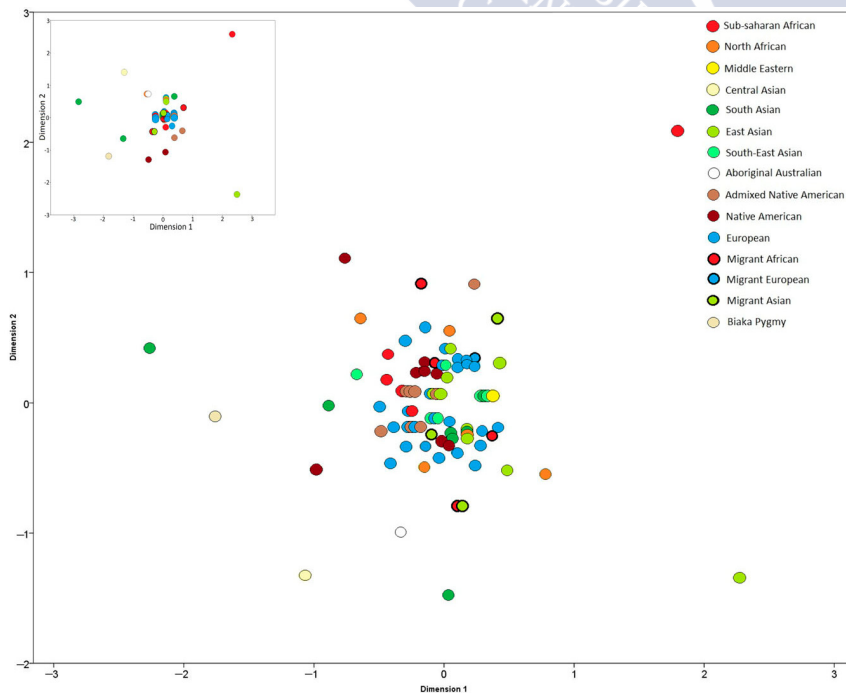


Figura 15. Gráfico de 2 dimensiones del análisis de *MDS* con los valores de F_{ST} linearizados por el método de Slatkin para los haplotipos de RM-YSTRs en la muestras de todos los 111 grupos poblacionales ($n=12,272$) (Stress = 0.07462); El gráfico pequeño, (Superior izquierda) muestra el mismo análisis de *MDS* para todos los grupos poblacionales con un tamaño de muestra de 100 individuos por población (Stress = 0.02416).

4.2.3 Mejoramiento en la diferenciación masculina con RM-YSTRs

La utilidad de estos marcadores en futuras aplicaciones forenses, antropológicas y en estudios genealógicos, puede ser evaluada al ser comparada con las técnicas actuales de identificación; este análisis comparativo se realizó con el conjunto de marcadores de RM-YSTRs y el *Yfiler*. El análisis permite demostrar que los *Rapidly Mutating* tienen un desempeño mejor a nivel global, por ejemplo en términos de la diversidad haplotípica global estimada para el grupo de 7,784 individuos de 65 poblaciones tipadas para ambos múltiplexes, hubo un aumento global de 0.99995 con *Yfiler* a 0.999997 con RM-YSTRs y en el número de haplotipos únicos aumentaron de 6,469 a 7,647. Así mismo se puede observar en la Tabla 11, como el desempeño de los RM-YSTRs es superior en los diferentes grupos regionales, donde a nivel global la diversidad haplotípica tiene un aumento en promedio de 0.0026 y el número de haplotipos únicos, o *singletons*, tiene un aumento en promedio de 77 haplotipos. Igualmente existen diferencias significativas entre los 2 grupos de marcadores evaluados, en términos de haplotipos compartidos tanto a nivel regional, entre las poblaciones y dentro de las poblaciones. (Figura 16). Globalmente, el panel *Yfiler* mostró 506 haplotipos compartidos en 1,318 individuos para todos los grupos poblacionales analizados, *i.e.* 65 poblaciones, en comparación con tan solo 66 haplotipos de RM-YSTRs entre 70 individuos.

Grupo	Individuos	Poblaciones	Haplotipos <i>Yfiler</i>	Haplotipos RM-YSTRs	Diversidad Haplotípica <i>Yfiler</i>	Diversidad Haplotípica RM-YSTRs	Porcentaje promedio de distintos loci <i>Yfiler</i>	Porcentaje promedio de distintos loci RM-YSTRs
Norte África	193	2	173	189	0.99792	0.99973	59%	73%
Asia Central	83	1	67	77	0.99060	0.99824	63%	77%
Asia Sur	497	6	450	483	0.99951	0.99988	68%	83%
Este de Asia	633	5	580	630	0.99952	0.99999	64%	82%
Sudeste Asia	200	2	175	198	0.99759	0.99990	66%	81%
Aborígen Australia	100	1	74	96	0.99152	0.99919	68%	80%
Nativo Americano	279	8	233	275	0.99761	0.99988	61%	82%
Nativo Americano (Mezcla)	458	6	444	454	0.99986	0.99996	67%	81%
Europeo	4,041	25	3,696	4,025	0.99991	0.999998	66%	82%
Emigrante África Sub-Sahariana	442	3	407	439	0.99909	0.99997	63%	81%
Emigrante Europeo	552	3	541	551	0.99993	0.99999	63%	81%
Emigrante Asiático	205	2	176	203	0.99737	0.99990	66%	83%
Pigmeos Biaka	101	1	83	94	0.99505	0.99822	65%	83%
Global	7,784	65	6,975	7,714	0.99995	0.999997	65%	81%

Tabla 11. Comparación de los haplotipos de *Yfiler* y RM-YSTRs en el conjunto global de 7,784 individuos pertenecientes a 65 poblaciones, agrupadas en grupos regionales.

Igualmente, llama la atención que el panel de RM-YSTRs detectó en 29 poblaciones de las 65 (44.6%), coincidencias de haplotipos dentro de las mismas poblaciones, como es el claro ejemplo de la población Argentina de Wichis de la provincia de Salta la cual presenta una proporción del 11% en haplotipos compartidos (Figura 16); de la misma manera, a nivel

global tan solo el 6.2% (4 poblaciones) de todos los grupos poblacionales analizados, mostraron coincidencias de haplotipos para el panel de RM YSTRs fuera de sí mismas, como fue el caso de coincidencias de un haplotipo por dos individuos de la población de República Checa y una coincidencia entre un haplotipo de un individuo de Colonia (Alemania) y Varsovia (Polonia). Este hallazgo indica que el 93.8% de los grupos poblacionales analizados (61 poblaciones) no muestran haplotipos compartidos con poblaciones fuera de sí misma, que en comparación con *Yfiler* este valor asciende a 13 poblaciones, es decir el 80% de la población (**Figura 17**). Lo anterior puede traducirse en que el conjunto de marcadores RM-YSTRs tiene un gran poder de discriminación que permite una diferenciación de linajes masculinos muy alta, incluso en poblaciones regionalmente cercanas, como las muestreadas en Europa donde tan solo se encontraron 4 coincidencias, comentadas anteriormente, en contraste con el panel de *Yfiler* (**Figura 17 – Recuadros inferior izquierda**).

Dado este panorama, en el que se observó que el número de coincidencias de los haplotipos entre las poblaciones y fuera de ellas es relativamente bajo para el grupo de 13 RM-YSTRs, se procedió a estimar a nivel global a través de un análisis de redes en 1000 individuos escogidos aleatoriamente del grupo total de 7,784, como se afectaría el agrupamiento geográfico de los distintos paneles, analizando solo 10 *loci* RM-YSTRs (solo los RM-YSTRs de una sola copia por marcador (Ver materiales y Métodos - apartado 3.4.3) y los 15 *loci* para el *Yfiler* (excluyendo el sistema DYS385a/b). Los resultados de la **figura 18** muestran como, por ejemplo, el panel *Yfiler* muestra *clústeres* para distintos grupos regionales, efecto que no es presentado por los RM-YSTRs, con excepción de los haplotipos de los grupos regionales del norte de África. Esto corrobora lo que anteriormente se había demostrado en los análisis de AMOVA y el análisis de *MDS*, que indican que hay ausencia de sub-estructura poblacional para el panel de RM-YSTRs. Por lo tanto para estos marcadores, la necesidad de bases de datos de referencia, para la estima de las frecuencias haplotípicas con fines forenses, es notablemente reducida en comparación con el panel *Yfiler*.

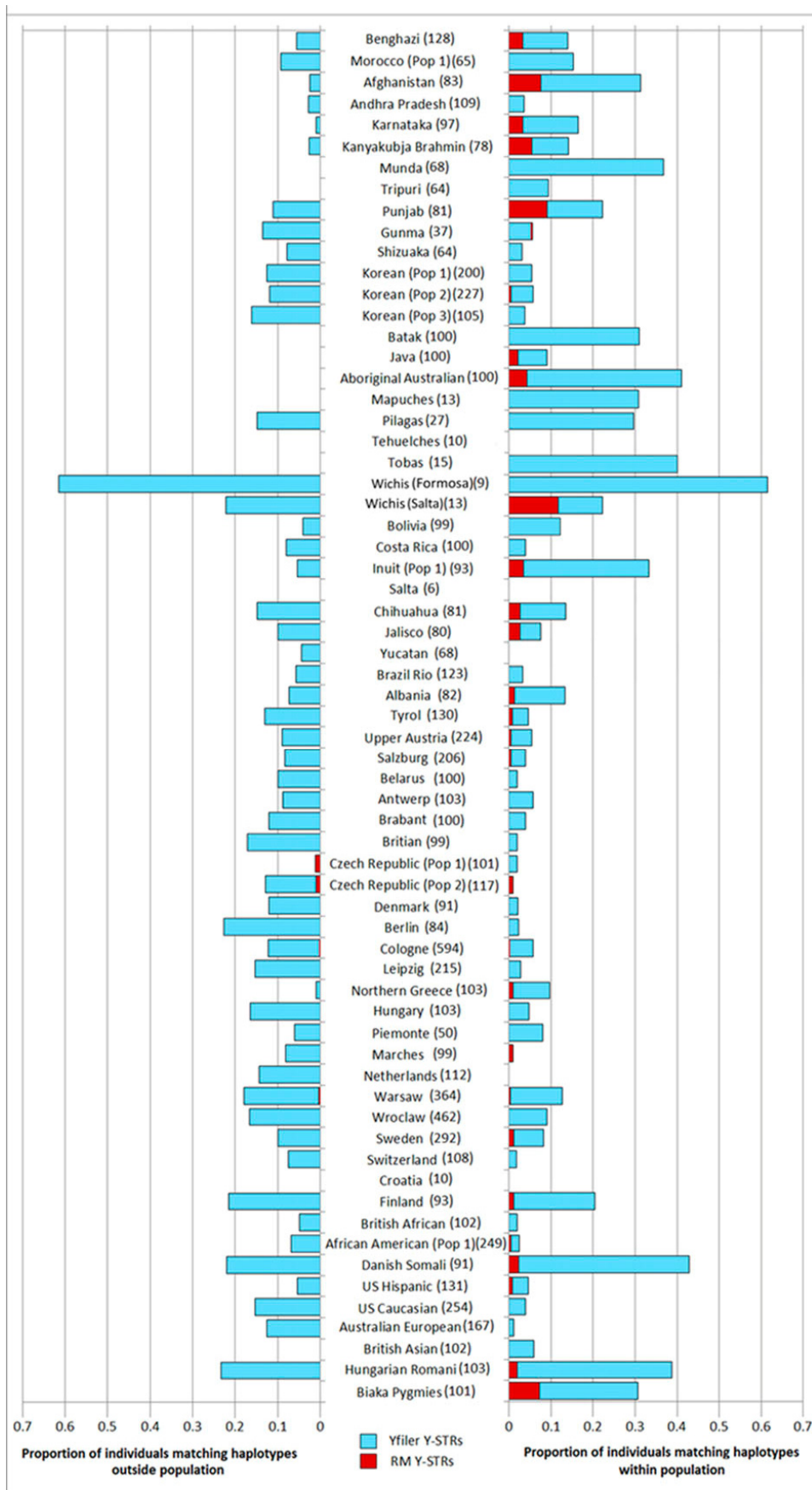


Figura 16. Proporción de individuos con haplotipos compartidos dentro de las poblaciones (Izquierda) y entre las poblaciones (derecha) Para Yfiler (Barras azul Claro) y para RM-YSTRs (Barras rojas).

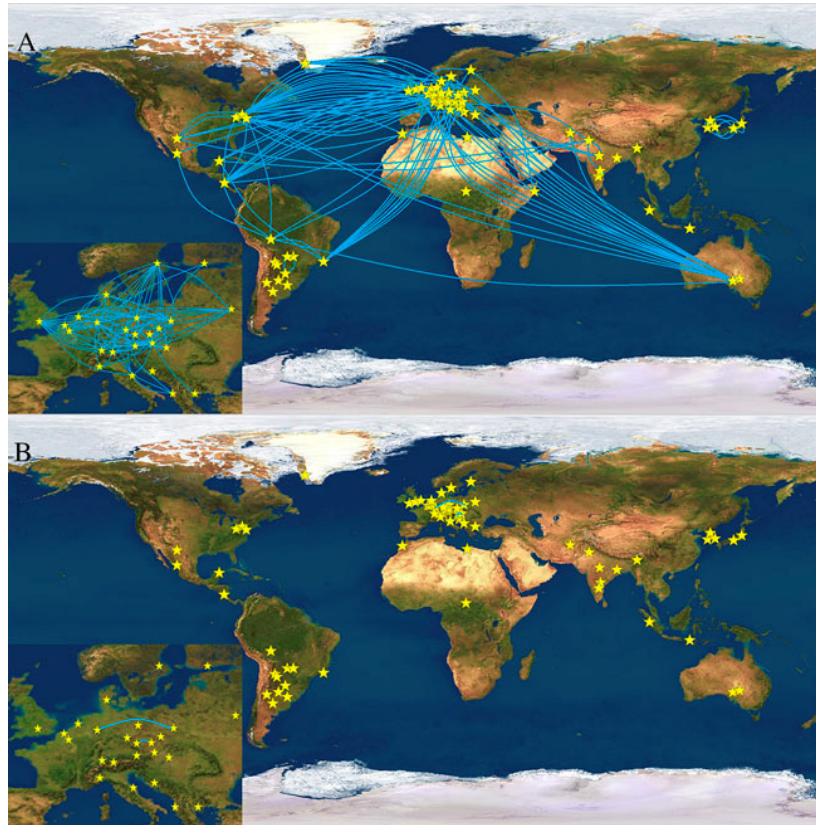


Figura 17. Representación geográfica de las coincidencias de haplotipos entre las poblaciones analizadas para el panel de RM-YSTRs y el panel *Yfiler*. El panel A) muestra los haplotipos con coincidencias para *Yfiler* y el panel B) para los RM-YSTRs. Las imágenes inferiores izquierda, muestran una ampliación de las poblaciones europeas.

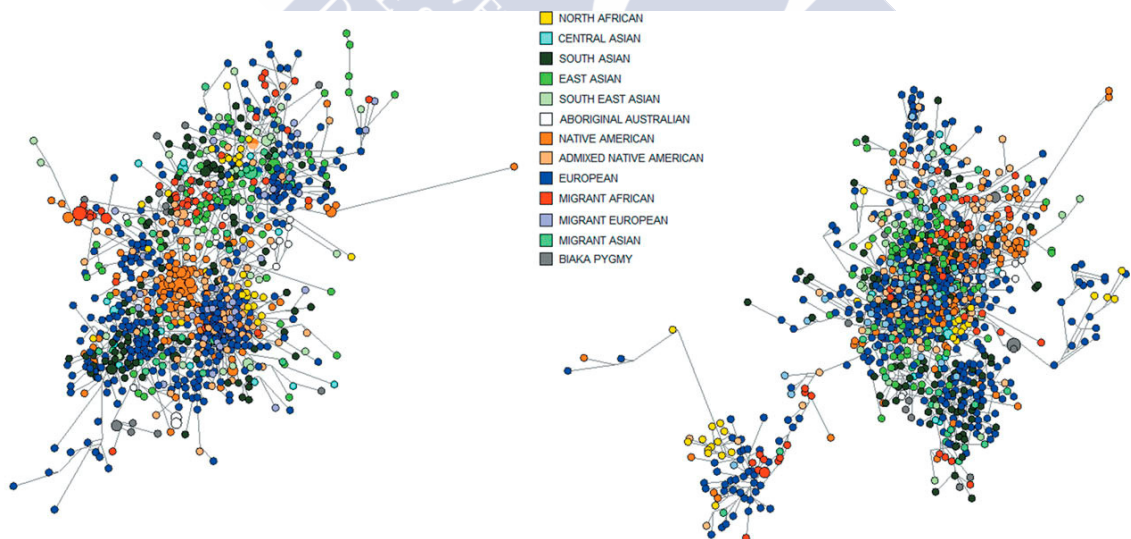


Figura 18. Redes *Weighted median-joining* estimadas para los Y-STRs de una sola copia en 1000 individuos escogidos aleatoriamente del grupo total de muestras 7,784, pertenecientes a todos los grupos poblacionales tipados (indicados en la leyenda). A) Red basada en los haplotipos de *Yfiler* con 15 loci (DYS385a/b, excluido) y B) Red basada en los haplotipos de 10-loci del panel RM-YSTRs (excluyendo los sistemas multicopia DYF399S1, DYF403S1 y DYF404S1).

4.2.4 Habilidad de los RM-YSTRs para detectar relaciones paternas no visibles.

Como se observó en la **Figura 16**, el panel de RM-YSTRs permite detectar en muchos casos (44.6%) coincidencias entre las mismas poblaciones, esto hace pensar que este panel puede tener la habilidad de descifrar y ver relaciones parentales que no son visibles; es por esto se simuló una situación para evaluar el poder de detección de relaciones paternas no visibles. Esta consistió en formar conjuntos de poblaciones en tres grupos, Rural (con una tasa considerable de eventos poblacionales endogámicos), Urbana (Poblaciones con una tasa considerable de apareamientos aleatorios) y endogámica (Poblaciones que por creencias y costumbres tiene un alto nivel de endogamia); estos grupos se evaluaron con el panel RM-YSTRs y los resultados mostraron evidentemente que el conjunto de poblaciones agrupadas en la clase endogámica, mostraron una reducción en la diversidad haplotípica y una menor proporción de haplotipos únicos que los grupos rurales y urbanos. Esto por lo tanto permite concluir que este panel permite observar relaciones parentales que no son visibles y que en comparación con el *Yfiler* bajo el mismo conjunto de grupos, los RM-YSTRs muestran mayor diversidad haplotípica y menor número de haplotipos no únicos en comparación con *Yfiler*.

4.2.5 Bolivia y su aporte a la variabilidad de los RM-YSTRs en el contexto regional.

En la presente sección analizaremos la variabilidad de la población de Bolivia, en el grupo regional nativo americano y como esta se contextualiza en el escenario continental.

4.2.5.1 Bolivia, el grupo regional nativo americano y la Sub-estructura geográfica con los RM-YSTRs

Para el análisis de sub-estructura de la población de Bolivia como para las demás poblaciones incluidas en el grupo regional de nativos americanos se estimó el valor θ (*Theta*); este valor fue muy bajo, lo que indica baja o poca subestructura poblacional para el conjunto de RM-YSTRs, en cada una de las poblaciones del grupo, así como para la muestra de Bolivia (Bolivia $\theta = 0$, ver Tabla 12). Este comportamiento de ausencia de sub-estructura es confirmado por el análisis de *MDS* que fue ejecutado para todos los grupos regionales (**Figura 15**); en este se puede observar que no existe ningún tipo de agrupamiento de las poblaciones del grupo nativo americano específico, y no se evidencia ningún patrón geográfico que sugiera sub-estructura; Así mismo este análisis se ejecutó de nuevo para todas poblaciones con el número de muestras fijado en 100 para todas las poblaciones (Gráfico superior izquierda - **Figura 15**) y el comportamiento fue el mismo,

indicando que el número de muestras no influye en los análisis de sub-estructura, como tampoco en el grupo regional nativo americano.

Grupo	Población	Número Individuos	Número Haplotipos	Diversidad Haplotípica	θ	Número promedio de diferencias en <i>locus</i>
Nativos Americanos	Tehuelches (Argentina)	10	10	1.0	0	16.16
	Bolivia	99	99	1.0	0	16.78
	Mocovies (Argentina)	7	7	1.0	0	15.38
	Pilagás (Argentina)	31	31	1.0	0	14.91
	Tobas (Argentina)	15	15	1.0	0	14.62
	Mapuches (Argentina)	13	13	1.0	0	16.58
	Wichis (Salta, Provincia Argentina)	9	8	0.97222	0.01818182	15.08
	Wichis (Formosa, Provincia Argentina)	14	14	1.0	0	11.38
	Inuit (Groenlandia Población 1)	95	92	0.99910	0.00137764	16.90
	Inuit (Groenlandia Población 2)	72	69	0.99883	0.00099431	17.01

Tabla 12. Características de los haplotipos de RM-YSTRs en 365 muestras pertenecientes al grupo regional Nativo Americano.

4.2.5.2 Bolivia en el grupo regional nativo americano en el mejoramiento en la diferenciación masculina con RM-YSTRs

Con el fin de evaluar la utilidad de los marcadores RM-YSTRs y su desempeño en el área forense, se realizó una comparativa entre los grupos regionales, entre ellos el nativo americano en el que se encuentra la población de Bolivia, y se analizaron los niveles de diversidad haplotípica obtenidos para estos marcadores y con *Yfiler* (tipado en las mismas muestras); La diversidad haplotípica calculada para los RM-YSTRs en las muestras de Bolivia ascendió al valor más alto posible (1), por el contrario el estimado para el *Yfiler* fue de 0.998763. Este valor obtenido indica que con los RM-YSTRs, en el 100% de los casos se obtienen haplotipos distintos sobre los tipados con *Yfiler* (93 haplotipos) (Tabla 13);

Grupo	Población	Número Individuos	Número Haplotipos <i>Yfiler</i>	Número Haplotipos RM-YSTRs	Diversidad Haplotípica <i>Yfiler</i>	Diversidad Haplotípica RM-YSTRs	Promedio % de diferentes loci <i>Yfiler</i>	Promedio % de diferentes loci RM-YSTRs
Nativos Americanos	Tehuelches (Argentina)	10	10	10	1	1.0	60%	77%
	Bolivia	99	93	99	0.998763	1.0	56%	80%
	Pilagas (Argentina)	27	23	27	0.981481	1.0	45%	71%
	Tobas (Argentina)	15	12	15	0.971429	1.0	36%	70%
	Mapuches (Argentina)	13	11	13	0.974359	1.0	51%	79%
	Wichis (Salta Provincia, Argentina)	9	8	8	0.972222	0.972222	40%	72%
	Wichis (Formosa Provincia, Argentina)	13	7	13	0.794872	1.0	20%	54%
	Inuit (Groenlandia Población 1)	93	72	90	0.990182	0.999065	66%	80%

Tabla 13. Comparativa de las características haplotípicas para muestras tipadas con *Yfiler* y RM-YSTRs en el grupo regional Nativo Americano, en 279 individuos pertenecientes a 8 poblaciones.

Igualmente el grupo regional nativo americano compuesto de 279 individuos en 8 poblaciones tipados para ambos marcados, mostró un total de 233 haplotipos únicos con *Yfiler* y un promedio de diversidad haplotípica de 0.99761, en comparación de un total de 275 haplotipos con los RM-YSTRs y una diversidad haplotípica promedio de 0.99988, lo que implica una mayor capacidad de discriminación entre individuos del grupo (Tabla 11).

También se detectaron diferencias significativas, entre *Yfiler* y RM-YSTRs a nivel de haplotipos compartidos entre los individuos de los grupos, entre las poblaciones y dentro de las poblaciones (**Figura 16**). Entre las poblaciones y fuera de ellas, los nativos de Bolivia, mostraron solo coincidencias de haplotipos para el panel de *Yfiler* (~11% y ~5% respectivamente, **Figura 16**), contándose entre las 35 poblaciones totales (53.8%) que no mostraron coincidencias entre y fuera de las poblaciones para el panel de RM-YSTRs; para el grupo regional de los nativos americanos, los individuos del grupo étnico Wichis de la provincia de Salta fueron los que mayor porcentaje de coincidencias de haplotipos para el panel de RM-YSTRs registraron (11%) y los Wichis de la provincia de Formosa los que mayor coincidencias de haplotipos para el panel *Yfiler* registraron (62%), esto debido en gran parte por los pequeños tamaños muestrales de estas 2 etnias ($n=13$ y $n=9$, respectivamente).

En la **figura 17** (Panel superior A), se observan las coincidencias de haplotipos de compartidos por *Yfiler* en la población nativa de Bolivia con poblaciones fuera de ellas; en total se registran 7 coincidencias, entre ellas: con un individuo de México, uno de Costa

Rica, dos con emigrantes europeos uno de Estados Unidos y otro de Australia y dos con 2 países europeos. El resto del grupo nativo americano no muestra muchas coincidencias, a excepción de los grupos étnicos de Argentina en las que se cuentan 3 coincidencias.

Igualmente, como se mostró para las 65 poblaciones analizadas en el presente estudio, el número de haplotipos compartidos de RM-YSTRs en el grupo regional nativo americano, no se registraron coincidencias fuera del grupo.

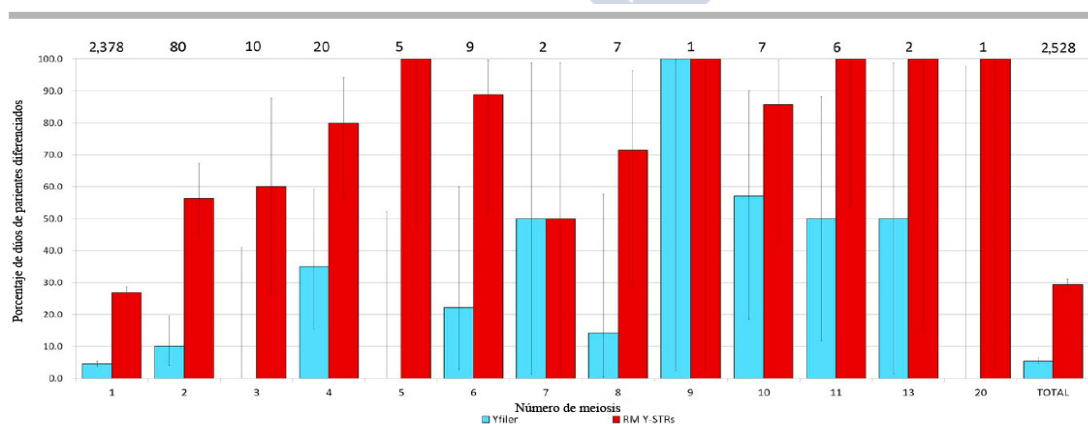
Respecto a los análisis de redes realizados para todos los grupos regionales, en la **Figura 18** puede observarse como para el panel de 17 Y-STRs, el grupo regional nativo americano forma un agrupamiento o clúster hacia el interior de la red (**Figura 18A** círculos color naranja), indicando una relación geográfica y haplotípica entre los diferentes grupos étnicos evaluados; sin embargo cuando se analizan solo los marcadores de una sola copia del panel RM-YSTRs, (**Figura 18B**) no se presentan agrupamiento o clústeres algunos, lo que de nuevo indica y confirma la ausencia de sub-estructura, como se observa en el gráfico de MDS para este mismo grupo regional.

4.2.6 Capacidad de diferenciación entre individuos masculinos emparentados.

Por último se realizó un ensayo complementario que evaluó el poder de diferenciación de los marcadores RM-YSTRs en individuos emparentados; se tiparon en total 2,372 dúos paternalmente relacionados divididos en 2,339 dúos padre-hijo, 30 dúos hermano-hermano y 3 dúos sobrino-tío que previamente fueron confirmados por análisis autosómicos, adicionando 156 pares anteriormente estudiados por (Ballantyne *et al.* 2012). El análisis conjunto de estos dúos confirmados con los marcadores RM-YSTRs, permitió como resultado diferenciar al menos por una diferencia alélica, entendida como una mutación, al menos un *locus* en 742 del total de 2,528 dúos emparentados de una a 20 generaciones, lo que en comparación con *Yfiler* tan solo ocurrió en 118 casos de los 2,161. Esto se traduce, en que el conjunto de RM-YSTRs pudo diferenciar a padres de sus hijos en un 26.9% de los casos en comparación de tan solo 4.5% con *Yfiler* y a hermanos unos de otros en el 56.3% de los casos respecto a solo un 10% con *Yfiler*. Finalmente se puede afirmar que entre mayor sea la distancia entre los miembros relacionados entre sí, mayor oportunidad habrá de mutación y por lo tanto mayor probabilidad de obtener haplotipos distintos de RM-YSTRs en la misma línea de parientes relacionados (**Figura 19**).

4.2.7 Importancia del estudio de la variabilidad con RM-YSTRs en Bolivia y el grupo nativo americano.

Los resultados que se obtienen con este panel de marcadores para la población de Bolivia indiscutiblemente tienen una importancia a nivel de las relaciones poblacionales entre los individuos; como se ha podido demostrar este panel permite discriminar en el 100% de los casos a todos los individuos tipados llegando al valor de diversidad haplotípica más alto posible, lo que sin lugar a dudas hace pensar que la aplicación de esta herramienta en la diferenciación de individuos relacionados puede ser muy efectiva. A diferencia de algunos grupos de las poblaciones nativo americanas (*i.e* Wichis, Argentina) que muestran alto nivel de coincidencia de haplotipos en la misma población, la población de Bolivia no mostró coincidencias haplotípicas dentro de la población, lo que puede sugerir que la población nativa de Bolivia no tiene ningún comportamiento endogámico y además no presenta sub-estructura poblacional. Así mismo el hecho que se hayan observado coincidencias con otras poblaciones mundiales con *Yfiler* (7 en total, **Figura 17**) es indicativo, de cómo se ha mencionado en el bloque anterior, que esta población con este panel de marcadores muestra un nivel ancestralidad mezclado con un componente europeo, que con el panel RM-YSTRs no es observada dado el alto nivel de mutación de los marcadores que no permiten establecer alguna estructura poblacional. Por otro lado el uso de este nuevo panel en aplicaciones futuras, permitiría ahondar en la diferencia de los dos grandes grupos nativos de Bolivia, los quechuas y Aymaras, que comparten historias similares y de esta manera entender más claramente su contribución a las poblaciones



actuales y su diferencia *per se*.

Figura 19. Diferenciación empírica en 2,528 dúos de parientes separados por 20 meiosis para *Yfiler* (Barras azules) y RM-YSTRs (Barras rojas). En esta figura se combina los 2,372 dúos de parientes del presente estudio

con los 156 analizados anteriormente por (Ballantyne *et al.* 2012). Los valores que se encuentran en la parte superior representan los números absolutos de los pares de parientes en los que la tasa de estimación está basada; las barras de error representan los intervalos de confianza binomial.

4.3. Valoración de la variabilidad del PowerPlex® Y23 (PPY23) en Bolivia en el contexto de las poblaciones mundiales.

En este bloque analizaremos los resultados del ejercicio colaborativo, con la participación de 93 laboratorios, donde se analizaron 129 poblaciones pertenecientes a 51 países, agrupados según su residencia continental (5 grupos) y según su ancestralidad continental (5 grupos); todas las muestras fueron anteriormente tipadas con *Yfiler* y en el presente estudio los laboratorios tiparon las mismas muestras para el panel PPY23; En este estudio participó la unidad de genética forense de la Universidad de Santiago de Compostela. Para el panel de PPY23 se tiparon un sub grupo de muestras de la población de Bolivia (previamente tipadas con *Yfiler*, $n=100$), compuestas por muestras clasificadas como [Mestizo] ($n=44$) y [Native American] ($n=56$), las cuales también previamente fueron tipadas para Y-SNPs que permitió la determinación de los haplogrupos Europeos y Nativo americanos (Ver apartado 4.1.2).

4.3.1 Generalidades del análisis comparativo de parámetros forenses de discriminación en PPY23 con *Yfiler*

Para todas las muestras de las diferentes poblaciones tipadas con *Yfiler* y con el panel PPY23 se evaluaron parámetros de diversidad genética como, la diversidad haplotípica, la capacidad de discriminación (DC) y otros parámetros forenses como la probabilidad de coincidencia (*random match probability*). Con el nuevo panel de marcadores tipado con PPY23 fueron observados 18,860 distintos haplotipos de acuerdo a la organización de residencia continental (Tabla 14) de las 19,630 muestras analizadas; de todos estos haplotipos, 18,237 haplotipos fueron únicos (92.9%) y el haplotipo más frecuente a lo largo de todas las poblaciones, fue detectado 11 veces en tres distintas poblaciones. Bajo esta organización de meta-poblaciones por residencia continental, Asia mostró el valor de DC mas alto (>0.97) seguido por Europa y Latinoamérica (DC ~ 0.96) y por último África con el valor más bajo de los grupos (DC ~ 0.85). De igual manera, bajo la organización por ancestralidad continental, los valores de DC (>0.96) para los grupos Asiáticos, Europeos y de Americanos *mixed* fueron similares (Tabla 14); aun así, para el grupo ancestral Nativo Americano, hay una disminución importante en el valor DC (0.83) pero un aumento notable en el grupo ancestral africano (0.94). Respecto a los haplotipos obtenidos con PPY23, se puede resaltar que 42 de las 129 poblaciones analizadas, que representan 32.6% del total,

presentaron haplotipos únicos de PPY23, entre ellas siete poblaciones asiáticas, 23 Europeas, seis latino-americanas y seis norte-americanas.

PPY23	Agrupamiento por residencia continental						Agrupamiento por ancestralidad continental				
	General	África	Asia	Europa	Latino-América	Norte-América	Africana	Asiática	Europea	Nativo Americana	Americana M _{ixed}
# Haplotipos observados	19630 haplotipos 129 poblaciones	445 haplotipos 6 poblaciones	3458 haplotipos 22 poblaciones	11968 haplotipos 72 poblaciones	1183 haplotipos 14 poblaciones	2576 haplotipos 15 poblaciones	1294 haplotipos 10 poblaciones	3976 haplotipos 28 poblaciones	12585 haplotipos 76 poblaciones	558 haplotipos 5 poblaciones	1217 haplotipos 10 poblaciones
n=1 (Únicos)	18237 (92.9%)	337 (75.7%)	3293 (95.2)	11185 (93.5%)	1094 (92.5%)	2378 (92.3%)	1170 (90.4%)	3760 (94.6%)	11801 (93.8%)	396 (71.0%)	1159 (95.2%)
n=2	531	27	65	314	38	67	35	79	323	52	24
n=3	64	6	9	32	3	12	6	14	31	10	2
n=4	16	4	2	7	1	3	4	4	5	3	1
n=5	6	1	-	3	-	2	1	-	3	2	-
n=6	2	-	-	1	-	-	-	-	-	1	-
n=7	2	1	-	-	-	-	1	-	-	-	-
n=8	1	1	-	-	-	-	1	-	-	-	-
n=9	-	-	-	-	-	-	-	-	-	-	-
n=10	-	-	-	1	-	-	-	-	1	-	-
n=11	1	-	-	-	-	-	-	-	-	-	-
HD	0.999995	0.998704	0.999983	0.999992	0.999924	0.999953	0.999837	0.999982	0.999993	0.999131	0.999951
MP	5.63×10^{-5}	3.54×10^{-3}	3.07×10^{-4}	9.11×10^{-5}	9.21×10^{-5}	4.35×10^{-4}	9.35×10^{-4}	2.70×10^{-4}	8.60×10^{-5}	2.66×10^{-3}	8.70×10^{-4}
MP como 1 en ...	17760	283	3262	10975	1086	2298	1070	3706	11623	377	1150
# haplotipos distintos	18860	377	3369	11543	1136	2463	1218	3857	12164	464	1186
# haplotipos específicos	-	376	3367	11516	1127	2446	1213	3848	12136	458	1176
Población	-	376	3367	11516	1127	2446	1213	3848	12136	458	1176
DC	0.9608	0.8472	0.9743	0.9645	0.9603	0.9561	0.9413	0.9701	0.9665	0.8315	0.9745

HD, Diversidad haplotípica; MP, Probabilidad de Coincidencia (Matching Probability); DC, Capacidad de discriminación.

Tabla 14. Número de distintos haplotipos obtenidos con el panel PPY23, en todas poblaciones analizadas en conjunto (Columna General) y por los grupos definidos por residencia continental y ancestralidad continental

En comparativa con el panel de marcadores de *Yfiler*, los valores de diversidad, el número de haplotipos únicos y el número de distintos haplotipos aumenta considerablemente, en forma linear (Tabla 15). Se puede observar que a nivel global (Columnas General –Tablas 14, 15) el aumento en la proporción de haplotipos únicos incrementa de 77.8% a 92.9% en *Yfiler* y PPY23. Así mismo los índices, de capacidad de discriminación (DC) y de diversidad haplotípica tienen un incremento notable de 85.7 % a 96.1%, y 0.999962 a 0.999995 con *Yfiler* y PPY23, respectivamente. Con esto se puede concluir que al aumentar el número de marcadores, hay un incremento notable de todos los parámetros forenses de discriminación, como también es demostrado cuando el panel de PPY23 es comparado con los paneles de marcadores incluidos, en el haplotipo mínimo (MHT), el de SWGDAM, y el PowerPlex® 12.

Los parámetros forenses también fueron evaluados respecto a su variación cuando solo se analizan los *loci* con tamaños menores a 220 pares de bases (pb), que permitirían simular una situación de muestras degradadas, incluidos en *Yfiler* (DYS456, DYS389I, DYS458, DYS19, DYS393, DYS391, GATAH4, y DYS437) y en PPY23 (DYS576, DYS389I, DYS391, DYS481, DYS570, DYS635, DYS393, y DYS458); la comparativa muestra que existe un aumento muy importante en los niveles de capacidad de discriminación donde este valor (Tabla 16), es el casi el doble como también en el número de distintos haplotipos. Se puede afirmar que este cambio es evidente por la incorporación de los

marcadores específicos de PPY23, DYS576, DYS481, DYS570, y DYS635, que han mostrado tener altos niveles de diversidad génica.

Yfiler	Agrupamiento por residencia continental						Agrupamiento por ancestralidad continental				
	General 19630 haplotipos 129 poblaciones	África 445 haplotipos 6 poblaciones	Asia 3458 haplotipos 22 poblaciones	Europa 11968 haplotipos 72 poblaciones	Latino-América 1183 haplotipos 14 poblaciones	Norte-América 2576 haplotipos 15 poblaciones	Africana 1294 haplotipos 10 poblaciones	Asiática 3976 haplotipos 28 poblaciones	Europea 12585 haplotipos 76 poblaciones	Nativo Americana 558 haplotipos 5 poblaciones	Americana Mixed 1217 haplotipos 10 poblaciones
n=1 (Únicos)	15263 (77.8%)	305 (68.5%)	3058 (88.4%)	9121 (76.2%)	1044 (88.3%)	2203 (85.5%)	1107 (85.5%)	3456 (86.9%)	9549 (75.9%)	342 (61.3%)	1109 (91.1%)
n=2	1064	32	148	675	51	93	52	175	728	44	38
n=3	256	4	16	181	8	24	5	19	191	12	8
n=4	94	7	7	60	2	11	8	12	63	8	2
n=5	63	2	4	31	1	7	2	6	36	6	
n=6	21	1	-	15		3	1	-	14	2	
n=7	12	-	-	11		-	-	-	16	-	
n=8	12	1	1	5		-	1	3	4	-	
n=9	9	-	-	4		2	-	-	3	2	
n=10	4	-	-	3		-	-	-	3	-	
n=11	5	-	-	6		-	-	1	7	-	
n=12	3	1	-	2		-	1	-	2	-	
n=13	2	-	-	-		-	-	-	-	-	
n=14	1	-	-	2		-	-	-	-	-	
n=15	2	-	-	2		-	-	-	4	-	
n=16	-	-	-	2		-	-	-	-	-	
n=17	2	-	-	-		-	-	-	1	-	
n=18	1	-	-	2		-	-	-	-	-	
n=19	-	-	-	1		-	-	-	1	-	
n=20	2	-	-	1		-	-	-	1	-	
n=21	1	-	-	-		-	-	-	2	-	
n=22	1	-	-	-		-	-	-	-	-	
n=23	-	-	-	-		-	-	-	-	-	
n=24	-	-	-	-		-	-	-	-	-	
n=25	1	-	-	-		-	-	-	-	-	
n=26	-	-	-	-		-	-	-	-	-	
n=27	-	-	-	-		-	-	-	-	-	
n=28	-	-	-	-		-	-	-	-	-	
n=29	-	-	-	-		-	-	-	-	-	
n=30	-	-	-	-		-	-	-	-	-	
n < (30, 40)	1	-	-	1		-	-	-	1	-	
HD	0.999962	0.997824	0.999949	0.999930	0.999861	0.999874	0.999708	0.999936	0.999934	0.998134	0.999900
MP	8.94×10^{-5}	4.42×10^{-3}	3.40×10^{-4}	1.53×10^{-4}	9.84×10^{-4}	5.14×10^{-4}	1.06×10^{-3}	3.15×10^{-4}	1.46×10^{-4}	3.65×10^{-3}	9.22×10^{-4}
MP como 1 en ...	11189	227	2939	6527	1017	1945	940	3174	6856	274	1086
# haplotipos distintos	16820	353	3234	10125	1106	2343	1177	3672	10626	416	1157
# haplotipos específicos Población	-	345	3202	9834	1007	2116	1119	3631	10425	384	1046
DC	0.8569	0.7933	0.9352	0.8460	0.9349	0.9095	0.9096	0.9235	0.8443	0.7455	0.9507

HD, Diversidad haplotípica; MP, Probabilidad de Coincidencia (Matching Probability); DC, Capacidad de discriminación.

Tabla 15. Número de distintos haplotipos obtenidos con el panel Yfiler, en todas poblaciones analizadas en conjunto (Columna General) y por los grupos definidos por residencia continental y ancestralidad continental

	Yfiler >220 pb ^a	PPY23 >220 pb ^b
HD	0.998569	0.999695
MP	1.48×10^{-3}	3.56×10^{-4}
MP como 1 en ...	675	2809
# Haplotipos distintos	6418	11,702
DC	0.3269	0.5961

Tabla 16. Parámetros forenses estimados para Y-STRs de tamaño menor a 220 pb en los paneles Yfiler y PPY23. HD, Diversidad haplotípica; MP, Probabilidad de coincidencia (Match probability); DC, Capacidad de discriminación. ^a Marcadores incluidos DYS456, DYS389I, DYS458, DYS19, DYS393, DYS391, GATAH4, DYS437. ^b Marcadores incluidos DYS576, DYS389I, DYS391, DYS481, DYS570, DYS635, DYS393, DYS458.

4.3.2 Análisis comparativo de la estructura poblacional con los paneles Yfiler y PPY23

Las distancias genéticas (R_{ST}) estimadas por análisis de varianza molecular AMOVA, para los diferentes grupos de residencia continental, fueron evaluadas para ambos paneles; las distancias genéticas muestran que el panel PPY23 distancia el grupo de poblaciones de África y las demás meta-poblaciones no Africanas con valores altos ($R_{ST} > 0.2$) (Tabla 17), sin embargo las distancias entre los demás grupos fueron más pequeñas pero significativas, como es el caso de las poblaciones de Norte-América y Latino-América ($R_{ST} = 0.009$).

Panel	Continente	África	Asia	Europa	Latino-América	Norte-América
Yfiler	África		+	+	+	+
	Asia	0.22740		+	+	+
	Europa	0.22823	0.10203		+	+
	Latino-América	0.23016	0.08182	0.02695		+
	Norte-América	0.1755	0.07394	0.01914	0.0081	
PPY23	África		+	+	+	+
	Asia	0.26082		+	+	+
	Europa	0.28376	0.06978		+	+
	Latino-América	0.28108	0.05315	0.01653		+
	Norte-América	0.21391	0.04687	0.01842	0.00895	

Tabla 17. Distancias genéticas *pairwise* estimadas en valores R_{ST} (debajo de la diagonal) para los paneles *Yfiler* y *PPY23*, obtenidos por análisis de AMOVA; Las significancias de las distancias genéticas están arriba de la diagonal y representan valores $p > 10^4$ indicados con (+).

Este mismo análisis de AMOVA mostró que la variación haplotípica para el panel *PPY23* principalmente se encuentra dentro de las poblaciones (*within populations*) con un 85.1%, en un 9.1% entre las poblaciones dentro de las meta-poblaciones definida por la residencia continental (*among populations within meta-populations*), y un 5.8% entre (*among meta-populations*) las meta-poblaciones; este mismo análisis de AMOVA estimado para *Yfiler* muestra que la principal componente de variación está igualmente dentro de las poblaciones en un 83.7%, en un 9.43% entre las poblaciones dentro de las meta-poblaciones y en un 6.86% entre las meta-poblaciones.

De lo anterior se observa que a medida que aumenta el número de marcadores Y-STRs tipados, las distancias entre las meta-poblaciones disminuye; igualmente la tendencia muestra que la proporción de variación tanto entre las poblaciones y dentro de las meta-poblaciones aumenta cuando se incrementa el número de marcadores tipados, mientras que la variación entre las población disminuye.

Finalmente el análisis de escalamiento multidimensional generado a partir de los valores de R_{ST} linearizados para las 129 poblaciones, distingue claramente las poblaciones africanas de las no-africanas e igualmente confirma el hallazgo reportado anteriormente por (Roewer *et al.* 2005), donde existe una distribución diferencial de variación haplotípica entre las poblaciones del este y oeste de Europa (**Figura 20**).

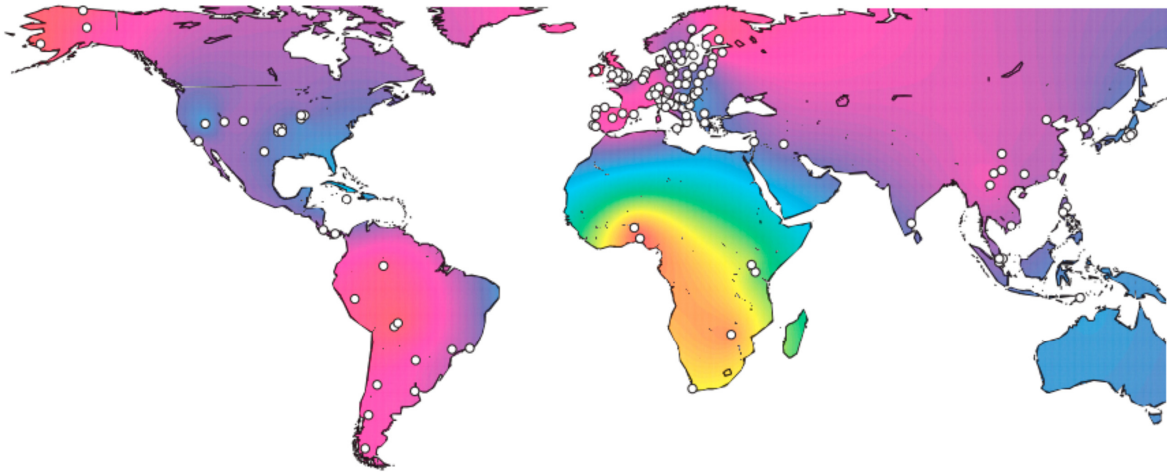


Figura 20. Estructura poblacional determinada por el panel PPY23; el mapa expresa las dos primeras componentes del MDS basado en las distancias R_{ST} . Los círculos en blanco representan las poblaciones muestreadas; la escala de colores tipo-arcoíris de amarillo a magenta explican los extremos de la escala.

4.3.3 Las poblaciones Latino-Americanas y Nativo americanas y su efecto en el análisis haplotípico con Yfiler y PPY23.

En esta sección analizaremos la contribución a la variabilidad haplotípica y a la subestructura que aportan los conjuntos de poblaciones en la agrupación por residencia continental y la agrupación por ancestralidad continental en los que se incluyen las muestras de Bolivia. El primer grupo agrupa a Bolivia con el conjunto de poblaciones por residencia continental Latino-Americano que contiene 14 poblaciones que se describen a continuación: Bolivia [Mestizo], Bolivia [Nativo Americano], Buenos Aires, Argentina [Europeo], Costa Rica [Mestizo], Formosa, Argentina [Europeo], Jamaica [Jamaiquino], Mendoza, Argentina [Europeo], Neuquén, Argentina [Europeo], Panamá [Panameño], Perú [Peruano], Rio de Janeiro, Brasil [Brasileño *Admixed*], Santa Cruz, Argentina [Europeo], São Gabriel de Cachoeira, Brasil [Nativo Americano] y São Paulo, Brasil [Brasileño *Admixed*]; así mismo, la agrupación por ancestralidad continental incluye cinco poblaciones nativo americanas en total entre ellas: Alaska central (Athapaskan), Norte de Alaska (Inupiat), São Gabriel de Cachoeira (Brasil), Alaska Occidental (Yupik) y la población nativa de Bolivia.

En la tabla 14 (panel izquierdo) se observa que el grupo Latino-Americano (Agrupamiento por residencia continental) analizado con PPY23 muestra en total 1183 haplotipos/individuo, de los cuales 1094 (92.5%) son únicos y 1127 son específicos de la población; respecto los parámetros forenses de variabilidad, la capacidad de discriminación de este panel para este grupo tiene un valor 0.96 y la probabilidad de coincidencia

(*matching probability*) indica para esta población, 1 en 1086 tiene un haplotipo distinto; también los parámetros de diversidad para el panel son altas, con valores de 0.999924; esto permite concluir que este tipo de agrupamiento explica buena parte de variabilidad del análisis y también muestra que este grupo, analizado con este panel de marcadores, permite obtener uno de los niveles más altos de diversidad. De la misma manera, en esta tabla 14 (panel derecho) se observa que el grupo nativo americano, en las que se encuentran las muestras nativas de Bolivia, muestra una variabilidad que es menor en comparación con los demás grupos, solo un 71% de haplotipos únicos (396 haplotipos), lo que evidentemente tiene un efecto en los parámetros de diversidad forense como la capacidad de discriminación que muestra un valor bajo (0.83) y en la probabilidad de coincidencia. Esta baja variabilidad podría estar relacionada con una alta homogeneidad haplotípica de este grupo al ser tipadas con el panel PPY23.

Igualmente la tabla 15 muestra la variabilidad en los dos grupos descritos anteriormente pero analizados con el panel *Yfiler*; en el grupo de agrupamiento por residencia continental el conjunto de poblaciones Latino-Americanas, muestra una disminución en el porcentaje de haplotipos únicos (88.3%), que en comparación con PPY23, lo que se traduce en una reducción en la diferenciación de 50 haplotipos. Igualmente la capacidad de discriminación es menor (0.93) pero en promedio es una de las más altas con respecto a los demás grupos. Respecto al agrupamiento por ancestralidad continental el cambio en la variabilidad es más notable, donde el grupo nativo americano de nuevo muestra los niveles más bajos de haplotipos únicos, y en comparación con el panel PPY23 hay una disminución del 10% en este último valor como también en la capacidad de discriminación; todo esto parece indicar que el grupo nativo americano, analizado con los dos distintos paneles muestra una homogeneidad y q para lograr una mejor resolución, en términos de mayor número de haplotipos únicos, será necesario incrementar el número de marcadores Y-STRs.

4.3.4 Estructura poblacional de las poblaciones Latino-Americanas

Para el grupo de residencia continental analizado con PPY23 se estimaron las distancias genéticas a partir de una AMOVA con valores R_{ST} (Tabla 18); el análisis de las distancias muestra que la población de Jamaica es aquella que mayor distancias presenta con las demás poblaciones y con los mayores valores para la poblaciones Bolivia (nativa y mestiza) y nativa de São Gabriel de Cachoeira, todas estadísticamente significativas ($p > 10^{-4}$) (0.3534, 0.3234, y 0.3282, respectivamente)

		Latino-América													
Población		Bolivia [Mestizo]	Bolivia [Nativo Americano]	Buenos Aires, Argentina	Costa Rica	Formosa, Argentina	Jamaica	Mendoza, Argentina	Neuquén, Argentina	Panamá	Perú	Río de Janeiro, Brasil	Santa Cruz, Argentina	São Gabriel de Cachoeira, Brasil	São Paulo, Brasil
Latino-América	Bolivia [Mestizo]		0.13038	+	+	+	+	+	+	+	0.00663	+	+	0.00020	+
	Bolivia [Nativo Americano]	0.0247		+	+	+	+	+	+	+	+	+	+	+	+
	Buenos Aires, Argentina	0.1055	0.1432		0.04920	0.83932	+	0.38135	0.45847	0.00069	+	0.01911	0.12326	+	0.23097
	Costa Rica	0.1039	0.1501	0.0081		0.06336	+	0.15484	0.22928	0.00010	+	0.09059	0.19008	+	0.01683
	Formosa, Argentina	0.1105	0.1638	0.0014	0.0031		+	0.97327	0.63142	0.00208	+	0.15325	0.61826	+	0.51322
	Jamaica	0.3234	0.3534	0.2265	0.1951	0.1967		+	+	+	+	+	+	+	+
	Mendoza, Argentina	0.1135	0.1613	0 ^a	0.0004	0 ^a	0.2102		0.49965	0.00030	+	0.45055	0.77101	+	0.38719
	Neuquén, Argentina	0.1171	0.1676	0 ^a	0.0008	0 ^a	0.2219	0 ^a		0.10900	0.00020	0.16751	0.41154	+	0.63825
	Panamá	0.1121	0.1548	0.0299	0.0231	0.0190	0.1265	0.0260	0.0168		+	0.00030	0.00089	+	0.00861
	Perú	0.0275	0.0609	0.0353	0.0330	0.0411	0.2604	0.0441	0.0350	0.0563		+	+	+	+
	Río de Janeiro, Brasil	0.1261	0.1726	0.0166	0.0028	0.0009	0.1421	0.0051	0.0054	0.0129	0.0546		0.13118	+	0.30086
	Santa Cruz, Argentina	0.1263	0.1843	0.0022	0.0108	0 ^a	0.2535	0 ^a	0 ^a	0.0423	0.0520	0.0185		+	0.05524
	São Gabriel de Cachoeira, Brasil	0.0930	0.1070	0.0740	0.0811	0.0791	0.3282	0.0809	0.0772	0.1173	0.0571	0.0985	0.0944		+
	São Paulo, Brasil	0.1205	0.1693	0.0201	0.0087	0 ^a	0.1428	0.0119	0.0085	0.0079	0.0524	0 ^a	0.0297	0.0928	

Tabla 18. Distancias genéticas *pairwise* estimadas en valores R_{ST} (debajo de la diagonal) para el panel PPY23, obtenidos por análisis de AMOVA para las poblaciones Latino-Americanas; Las significancias de las distancias genéticas están arriba de la diagonal y representan valores $p > 10^4$ indicados con (+). ^a Debido a la aproximación numérica el valor R_{ST} era negativo en este caso indicó la distancia con cero

Respecto a las poblaciones de Bolivia analizadas se puede resaltar que la población nativa, mostró distancias significativas con todas las poblaciones a excepción de la población de mestizos de Bolivia (R_{ST} 0.06 – 0.3534), incluso a nivel global la tendencia es que esta población presentará mayor distancias que la población mestiza; por otro lado la población mestiza de Bolivia solo mostró distancias cercanas con la población de Perú y de São Gabriel de Cachoeira en el grupo de poblaciones Latino-Americanas. A nivel global solo la población mestiza de Bolivia mostró distancias cercanas con las poblaciones de Los Ángeles, USA [Mexicano], Norte de Alaska, USA [Inupiat], y Alaska occidental, USA [Yupik], poblaciones con gran número de individuos nativos americanos, que permiten entender el porqué de la cercanía con la población Boliviana.

Finalmente, llama la atención que estas dos poblaciones al ser analizadas a nivel global por el análisis de escalamiento multidimensional considerando las dimensiones 3 y 4, ambas poblaciones Bolivianas presentaron un claro patrón de diferenciación respecto a todo el continente como también fue el caso algunas poblaciones del norte de América y el este asiático (**Figura 21**).

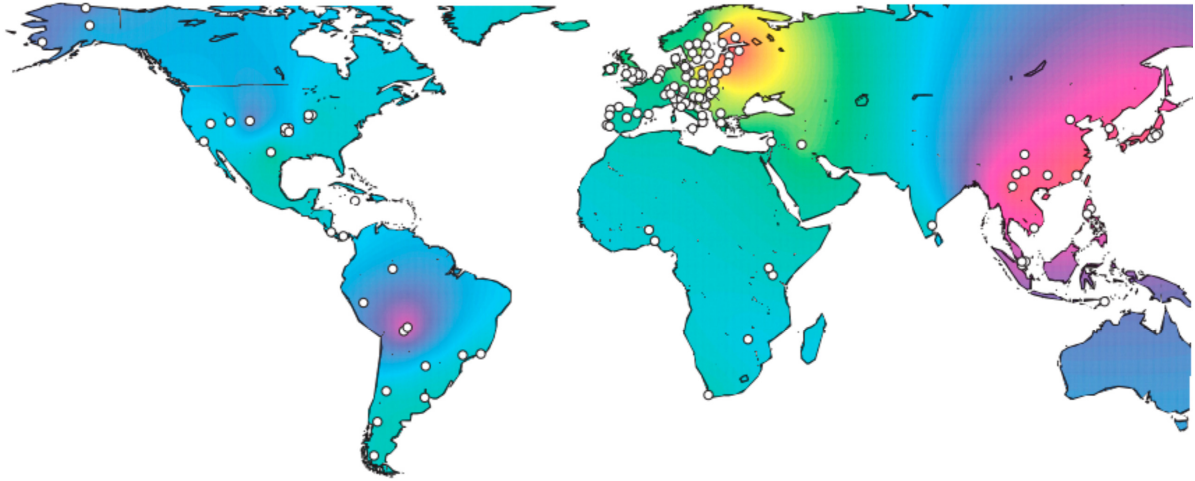


Figura 21. Estructura poblacional determinada por el panel PPY23; el mapa expresa las cuatro componentes del MDS basado en las distancias R_{ST} . Los círculos en blanco representan las poblaciones muestreadas; la escala de colores tipo-arcoíris de amarillo a magenta explican los extremos de la escala.



V. Consideraciones finales



Por medio de los diferentes análisis genéticos de cromosoma Y llevados a cabo en la población Boliviana en la presente tesis, podemos considerar que esta población tiene un componente notable de ancestralidades compartidas con grupos europeos y nativo americanos. Así, en los datos se puede observar la presencia de distintos haplogrupos de origen europeo (Resultados-Bloque 1 - 62.5%) y de haplogrupos nativos americanos (35.5% - principalmente en bajas altitudes). Estos porcentajes varían además significativamente respecto a la altitud.

Los resultados permiten además confirmar, a grandes rasgos, que esta población, y de manera similar a otras poblaciones de América del sur, contienen información que es fundamental para el entendimiento de cómo se llevó a cabo el proceso de colonización del continente. La presencia de ancestralidad nativo americana atestigua que las actuales poblaciones indígenas de América, y en particular la población boliviana, están relacionadas con poblaciones del noreste de Asia. Estudios recientes basados en el análisis de 'redes' (Battaglia *et al.* 2013) en poblaciones nativo americanas, y por medio del análisis de microsátelites de cromosoma Y en individuos pertenecientes al haplogrupo M3-Q1a3a* ($n=436$), demuestran que los patrones filogenéticos de tipo *star-like* para este haplogrupo, explican que estas poblaciones encajan bien con un modelo de entrada del hombre anatómicamente moderno reciente en América y una amplia y rápida expansión de este linaje en todo el continente. Lo anterior permite confirmar varios de los hallazgos reportados en la presente tesis, ya que el componente nativo americano de la población de Bolivia está compuesto principalmente por este haplogrupo. Estos hallazgos permiten dar soporte a la presencia del legado paterno que existió antes del periodo colonial principalmente heredado probablemente del imperio Inca y que se extendió en una amplia región andina de América del sur.

Igualmente, la comparativa efectuada con otros marcadores de linaje, como el ADNmt, y marcadores de ancestralidad autosómica (AIM-Indels), permiten confirmar que este componente nativo en la población boliviana es todavía más alto en estos otros marcadores (Figura 14). En general, los resultados atestiguan la naturaleza pluri-étnica de la población Boliviana y como esta varía en las diferentes regiones del país.

Desde el enfoque de la identificación humana, los marcados analizados en la presente tesis permiten hacer inferencias de distinta naturaleza. Primero, los datos aportados en los estudios realizados son de interés para la investigación criminal, dado que las bases de

datos de la población Boliviana eran hasta este momento muy escasas. Los marcadores Y-STRs (*Yfiler*) permiten hacer inferencias respecto a la probabilidad de encontrar una coincidencia haplotípica en la población de referencia, lo que se traduce en la posibilidad de discriminar individuos masculinos que puedan estar relacionados con un delito. Esto es especialmente importante en casos de las agresiones sexuales.

Segundo, los marcadores Y-SNPs permiten establecer el linaje geográfico perteneciente a una muestra o individuo en cuestión, permitiendo así formular hipótesis sobre el posible origen geográfico del perfil genético de cromosoma Y. En términos de identificación, este hecho indica que el cromosoma Y podría ser una herramienta útil para guiar investigaciones policiales y/o judiciales o establecer parentesco geográfico con muestras de referencia. En contrapartida, el hecho de que Bolivia es una población con una compleja historia e mestizaje, existe la gran desventaja de que en muchos casos el origen geográfico (continental) del haplogrupo inferido a través del cromosoma Y, diste mucho de la ancestralidad inferidas a través de las características fenotípicas del individuo. No hay que olvidar en este contexto, que el cromosoma Y representa en términos de ancestralidad un solo locus del genoma humano, y que por lo tanto su capacidad predictiva de ancestralidad es mucho más limitada en un contexto de mezcla poblacional. En el contexto americano, hay que tener en cuenta además que existe un importante sesgo de género en relación a las ancestralidades. Por ejemplo, tal y como se ha visto en los estudios presentados, la ancestralidad europea predomina en el cromosoma Y, mientras que la ancestralidad nativa predomina en el ADN mitocondrial. El ADN autosómico, presenta una ancestralidad promedio, aunque en el caso de Bolivia, predomina el componente nativo americano.

Paralelamente, como ocurre en otras poblaciones de América del sur, en esta tesis se ha demostrado que Bolivia tiene una notable estratificación poblacional (Apartado 4.1.3) y es por esto que se sugiere y se justifica la creación de bases de datos de referencia en el área forense que representen regiones o lugares (o etnicidades) concretos. Este último objetivo se ha alcanzado en esta tesis, ya que el tipado de Y-STRs (*Yfiler*) para la población Boliviana se ha hecho disponible *on line* en la base de datos de cromosoma Y de <http://www.yhrd.org> bajo el código de acceso [YA003827] [Bolivia, Boliviano], lo que permite consultar los perfiles genéticos más frecuentes de la región y puede ser usada por el sistema de administración de justicia o las instituciones de identificación humana con el fin de buscar coincidencias o exclusiones de haplotipos en un caso dado.

El estudio de estos polimorfismos presenta también ciertas limitaciones en la discriminación de individuos pertenecientes a poblaciones de tamaño efectivo poblacional pequeños, y que por lo tanto a menudo presentan altos niveles de endogamia (es decir, los miembros de estos grupos están emparentados en algún grado genealógicamente cercano). En estas poblaciones, las coincidencias de haplotipos son muy altas, por lo que el análisis de cromosoma Y podría ser incluso más limitado como herramienta de exclusión/inclusión en la casuística forense. Este efecto puede verse de alguna manera en el grupo de muestras tipadas con *Yfiler* en el bloque 2 de resultados (Apartado 4.2.5.1) en el que se observa que para las 99 muestras nativo americanas, solo 93 tuvieron haplotipos únicos, que en términos de identificación se traduce en un valor de capacidad de discriminación de ~ 0.94 . Es por esta razón que en la mayor parte de los casos es preferible usar polimorfismos autosómicos debido a que estos sí permiten individualizar y discriminar con alta probabilidad (usualmente >0.9999).

Por lo tanto, y siguiendo con las inferencias que permiten realizar los marcadores analizados en la presente tesis, uno de los motivos por el cual se ha diseñado el panel de marcadores RM-YSTRs, ha sido el poder manejar las limitaciones del *Yfiler* en relación al poder de discriminación en grupos de personas o poblaciones estrechamente relacionadas. El análisis de los RM-YSTRs ha demostrado en el estudio de las distintas poblaciones mundiales y en particular la muestra de Bolivia analizada en la presente tesis su gran poder de discriminación en comparación con paneles de Y-STRs previos, ya que el nivel de individualización se incrementa notablemente cuando se compara con los mismos resultados de *Yfiler*. En términos de poder de discriminación las 99 muestras bolivianas analizadas presentan haplotipos distintos, lo que se traduce en el valor más alto de discriminación posible, una constante en muchas otras poblaciones mundiales analizadas ($\sim 60\%$ de las poblaciones analizadas a nivel mundial muestran el máximo poder de discriminación en relación a los tamaños muestrales analizados). Así mismo, tal y como cabría esperar, por el nivel de mutación elevada de los marcadores, no se observa una estructura poblacional bien definida en los distintos grupos poblacionales y esto desde el punto de vista forense podría ser muy favorable, ya que una única base de datos de referencia podría ser suficiente para manejar la casuística del país. Por último, la habilidad de identificación de relaciones paternas no visibles (ej. en poblaciones con historial de alto nivel de endogamia) y la habilidad de diferenciar padres de hijos y hermanos entre sí, hace que este conjunto de marcadores con tasa de mutación alta, sea apropiado en muchas

situaciones de interés forense que requieren individualizar personas y no el establecimiento de la relación de parentesco, lo que convierte a estos marcadores en una herramienta muy útil a la hora de excluir o incluir sospechosos e inocentes en un caso judicial.

Existe sin embargo una importante desventaja en el área forense para los RM-YSTRs y es que por su alto nivel de mutación, de momento no es apropiado y útil incluir estos marcadores en la rutina para la resolver casos de paternidad o búsqueda de miembros familiares emparentados por vía paterna. Esta desventaja podría ser manejada analizando posibles factores de corrección que permitan compensar la probabilidad en la ocurrencia de estas mutaciones. Igualmente sería necesario establecer un umbral de mutaciones que permitiera excluir una paternidad con ciertas garantías. Así, igual que para los Y-STRs convencionales usados en la rutina (*Yfiler*) se ha determinado un umbral de tres mutaciones para declarar una exclusión (Kayser and Sajantila 2001), sería conveniente establecer un umbral para los marcadores RM-YSTRs. Es de sentido común, que el umbral debería superar en este caso los tres eventos mutacionales ya que como se ha demostrado en el presente estudio a nivel global más del 3% de las paternidades padre-hijo confirmadas presentan tres o cuatro mutaciones en los sistemas RM-YSTRs. Por lo tanto, a la espera de solventar las limitaciones de los RM-YSTRs y entender mejor la naturaleza mutacional de los mismos, el uso de los marcadores de *Yfiler* es de momento la mejor solución para la mayor parte de los casos de parentesco y de paternidad.

En nuestro análisis de marcadores incluidos en la presente tesis, se presentan resultados de una alternativa que permite solucionar algunas de las desventajas mencionadas para los dos conjuntos de marcadores que se han comentado hasta el momento. Esta solución combina la versatilidad de los marcadores *Yfiler* que se usan en la rutina comúnmente y la capacidad de individualización de los RM-YSTRs, las cuales se han tenido en cuenta en la elaboración del conjunto de marcadores del PowerPlex[®] Y23. Esto ha sido posible por la implementación de los 17 marcadores del *Yfiler* mas seis marcadores de alta diversidad génica que incluyen dos sistemas RM-YSTRs (DYS570 y DYS576), presentes en multiplex uno del multiplex de 13 loci de RM-YSTRs. Como se ha demostrado en la presente tesis tanto en las muestras del grupo nativo americano y del grupo Latino-Americano, el número de haplotipos únicos obtenidos con el PPY23 aumenta de un 61.3% a un 71.0% y de 88.3% a un 92.5%, respecto a *Yfiler*, en ambos grupos respectivamente. Igualmente las capacidades de discriminación aumentan (Ver apartado 4.3.3), lo que permite concluir que este panel tiene una utilidad forense importante. Respecto al desempeño de este panel en

muestras degradadas, el análisis de los marcadores con amplicones menores a >220 pb (Apartado 4.3.1), indica que aunque los resultados son muy variables debido a que los marcadores de este rango de tamaño incluyen los de mayor tasas de mutación (*i.e.* RM-YSTRs), puede decirse que este panel puede ser útil para casos en los que las muestras tengas comprometido el ADN (degradado o de bajo peso molecular).

Por último, los análisis de distancias génicas llevados a cabo con el PPY23 muestran que bajo la organización de meta-poblaciones, las poblaciones de África están claramente diferenciadas de los demás grupos de poblaciones, y que los demás grupos poblacionales muestran pequeñas distancias entre sí, como es el caso de las poblaciones de Norte-américa y Latino-américa (incluida Bolivia). Finalmente para la implementación de estos marcadores en la rutina forense debe tenerse en cuenta el uso obligatorio de aproximaciones de verosimilitud que tenga en cuenta las frecuencias alélicas, las tasas de mutación y el presunto nivel de parentesco que se desea determinar.





VI. Conclusiones



A) Análisis de la población Boliviana a partir de polimorfismos SNPs en cromosoma Y

- La distribución de las frecuencias de haplogrupos de Cromosoma Y en Bolivia muestra claramente que esta población tiene una importante proporción de componente europeo y nativo americano en un menor grado, y esta distribución está determinada por la geografía de la región estudiada, donde en mayores altitudes (Altiplano Andino y Sub-Andino) el componente europeo es el más predominante y a bajas altitudes (Llanos amazónicos) el componente nativo es mayor.

B) Análisis comparativo de la población Boliviana de marcadores uniparentales y autosómicos (AIM-Indels)

- El análisis conjunto de los linajes uniparentales en la población Boliviana muestra claramente que existe un notable sesgo por género, que consiste fundamentalmente en que el ADN mitocondrial está representado mayoritariamente por linajes nativos americanos, mientras que en el cromosoma Y existe una mayor proporción de linajes europeos (62.5%), en comparación con los nativos americanos (35.5%) y los africanos (2%).
- El análisis comparativo de las proporciones de ancestralidad autosómica por medio de los AIM-Indels permite confirmar que la población de Bolivia presenta una proporción ancestral especialmente nativa americana (más del 90% en promedio) y que la proporción europea es muy baja, lo que nos permite concluir que el componente europeo está fundamentalmente representado en el cromosoma Y.

C) Estudios de los microsatélites de Cromosoma Y con *Yfiler* en la población Boliviana

- La diversidad haplotípica estimada para todos los *loci* en el conjunto de muestras de Bolivia resultó en una alta capacidad de discriminación y alta diversidad haplotípica. Este hecho es particularmente importante en la casuística forense que se pueda llevar a cabo en Bolivia.

- La distribución de la varianza cuando se consideran los Y-STRs de *Yfiler* en Bolivia y las demás poblaciones de América del sur muestra que Bolivia tiene un efecto importante en la estratificación del subcontinente americano.

D) Estudios de los microsatélites de Cromosoma Y con RM-YSTRs en una muestra nativa de la población Boliviana

- Los resultados de los marcadores de nueva generación RM-YSTRs en población nativa Boliviana mostraron que este nuevo conjunto de marcadores permite aumentar considerablemente el nivel de discriminación haplotípica entre individuos debido a sus altas tasas de mutación.
- El nivel de diversidad haplotípica de cromosoma Y estimado con los RM-YSTRs supera al obtenido con *Yfiler* en los bolivianos estudiados. Así, por ejemplo, se han observado haplotipos repetidos en individuos bolivianos (genealógicamente) no relacionados para el *Yfiler*, mientras que todos los haplotipos de RM-YSTRs fueron únicos, indicando un mayor poder de discriminación para estos últimos.
- De manera genérica, se puede decir que la ausencia de coincidencias haplotípicas entre la población de Bolivia con el grupo nativo americano y con las demás poblaciones mundiales (y que representan ampliamente la variabilidad intracontinental mundial), demuestra el poder de los RM-YSTRS en la diferenciación de los linajes masculinos.
- La implementación de estos marcadores en la rutina forense para estudios de parentesco y paternidad de momento no es aconsejable hasta que no se establezca con mayor precisión el número umbral de eventos mutacionales que permita definir con una confianza criterios amplios de inclusión y exclusión.

E) Estudios de los microsatélites de Cromosoma Y con PowerPlex® Y23 en una muestra de la población Boliviana

- El panel de PowerPlex® Y23 analizado para las muestras de población Boliviana incluidas en el grupo nativo Americano y Latino-Americano muestra una mayor diferenciación en términos de número de haplotipos y de capacidad de discriminación que cuando son analizadas con *Yfiler*.
- La implementación de estos marcadores de rutina implican obligatoriamente incluir aproximaciones de verosimilitud que tenga en cuenta las frecuencias alélicas, las tasas de mutación y el presunto nivel de parentesco patrilineal que se desea determinar.







VII. Bibliografía



- Acosta, M. A., A. Blanco-Verea, M. V. Lareu, M. Brion and A. Carracedo (2009). "The genetic male component of two South-Western Colombian populations." Forensic Science International: Genetics **3**(2): e59-e61.
- Alvarez-Iglesias, V., F. Barros, A. Carracedo and A. Salas (2008). "Minisequencing mitochondrial DNA pathogenic mutations." BMC Med Genet **9**: 26.
- Alves, C., V. Gomes, M. J. Prata, A. Amorim and L. Gusmão (2007). "Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFISTR YFiler) in Portugal." Forensic Sci. Int. **171**(2): 250-255.
- Ballantyne, K. N., M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, Y. Choi, K. van Duijn, M. Vermeulen, S. Brauer, R. Decorte, M. Poetsch, N. von Wurmb-Schwark, P. de Knijff, D. Labuda, H. Vezina, H. Knoblauch, R. Lessig, L. Roewer, R. Ploski, T. Dobosz, L. Henke, J. Henke, M. R. Furtado and M. Kayser (2010). "Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications." Am J Hum Genet **87**(3): 341-353.
- Ballantyne, K. N., V. Keerl, A. Wollstein, Y. Choi, S. B. Zuniga, A. Ralf, M. Vermeulen, P. de Knijff and M. Kayser (2012). "A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages." Forensic Sci Int Genet **6**(2): 208-218.
- Ballantyne, K. N., A. Ralf, R. Aboukhalid, N. M. Achakzai, M. J. Anjos, Q. Ayub, J. Balazic, J. Ballantyne, D. J. Ballard, B. Berger, C. Bobillo, M. Bouabdellah, H. Burri, T. Capal, S. Caratti, J. Cardenas, F. Cartault, E. F. Carvalho, M. Carvalho, B. Cheng, M. D. Coble, D. Comas, D. Corach, M. E. D'Amato, S. Davison, P. de Knijff, M. C. De Ungria, R. Decorte, T. Dobosz, B. M. Dupuy, S. Elmrghni, M. Gliwinski, S. C. Gomes, L. Grol, C. Haas, E. Hanson, J. Henke, L. Henke, F. Herrera-Rodriguez, C. R. Hill, G. Holmlund, K. Honda, U. D. Immel, S. Inokuchi, M. A. Jobling, M. Kaddura, J. S. Kim, S. H. Kim, W. Kim, T. E. King, E. Klausriegler, D. Kling, L. Kovacevic, L. Kovatsi, P. Krajewski, S. Kravchenko, M. H. Larmuseau, E. Y. Lee, R. Lessig, L. A. Livshits, D. Marjanovic, M. Minarik, N. Mizuno, H. Moreira, N. Morling, M. Mukherjee, P. Munier, J. Nagaraju, F. Neuhuber, S. Nie, P. Nilasitsataporn, T. Nishi, H. H. Oh, J. Olofsson, V. Onofri, J. U. Palo, H. Pamjav, W. Parson, M. Petlach, C. Phillips, R. Ploski, S. P. Prasad, D. Primorac, G. A. Purnomo, J. Purps, H. Rangel-Villalobos, K. Rebala, B. Rerkamnuaychoke, D. R. Gonzalez, C. Robino, L. Roewer, A. Rosa, A. Sajantila, A. Sala, J. M. Salvador, P. Sanz, C. Schmitt, A. K. Sharma, D. A. Silva, K. J. Shin, T. Sijen, M. Sirker, D. Sivakova, V. Skaro, C. Solano-Matamoros, L. Souto, V. Stenzl, H. Sudoyo, D. S. Court, A. Tagliabracci, D. Taylor, A. Tillmar, I. S. Tsybovsky, C. Tyler-Smith, K. J. van der Gaag, D. Vanek, A. Volgyi, D. Ward, P. Willemse, E. P. Yap, R. Y. Yong, I. Z. Pajnic and M. Kayser (2014). "Towards male individualization with rapidly mutating Y-chromosomal STRs." Hum Mutat.
- Battaglia, V., V. Grugni, U. A. Perego, N. Angerhofer, J. E. Gomez-Palmieri, S. R. Woodward, A. Achilli, N. Myres, A. Torroni and O. Semino (2013). "The First Peopling of South America: New Evidence from Y-Chromosome Haplogroup Q." PLoS ONE **8**(8): e71390.

- Beleza, S., L. Gusmão, A. Amorim, Á. Carracedo and A. Salas (2005). "The genetic legacy of western Bantu migrations." Hum. Genet. **117**(4): 366-375.
- Beleza, S., L. Gusmão, A. Lopes, C. Alves, I. Gomes, M. Giouzeli, F. Calafell, Á. Carracedo and A. Amorim (2006). "Micro-phylogeographic and demographic history of Portuguese male lineages." Ann. Hum. Genet. **70**(Pt 2): 181-194.
- Blanco-Verea, A., M. Brion, P. Sanchez-Diz, J. C. Jaime, M. V. Lareu and A. Carracedo (2006). "Analysis of Y chromosome lineages in native South American population." International Congress Series **1288**: 222-224.
- Blanco-Verea, A., J. C. Jaime, M. Brión and A. Carracedo (2010). "Y-chromosome lineages in native South American population." Forensic Science International: Genetics **4**(3): 187-193.
- Borjas, L., L. P. Bernal, M. A. Chiurillo, F. Tovar, W. Zabala, N. Lander and J. L. Ramirez (2008). "Usefulness of 12 Y-STRs for forensic genetics evaluation in two populations from Venezuela." Leg Med (Tokyo) **10**(2): 107-112.
- Brinkmann, B., M. Klintschar, F. Neuhuber, J. Huhne and B. Rolf (1998). "Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat." Am J Hum Genet **62**(6): 1408-1415.
- Brión, M., B. Quintáns, M. Zarrabeitia, A. González-Neira, A. Salas, V. Lareu, C. Tyler-Smith and Á. Carracedo (2004). "Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis." Gene **329**: 17-25.
- Brión, M., J. J. Sanchez, K. Balogh, C. Thacker, A. Blanco-Verea, C. Børsting, B. Stradmann-Bellinghausen, M. Bogus, D. Syndercombe-Court, P. M. Schneider, A. Carracedo and N. Morling (2006). "Analysis of 29 Y-chromosome SNPs in a single multiplex useful to predict the geographic origin of male lineages." International Congress Series **1288**: 13-15.
- Brion, M., B. Sobrino, A. Blanco-Verea, M. V. Lareu and A. Carracedo (2004). "Hierarchical analysis of 30 Y-chromosome SNPs in European populations." Int J Legal Med **119**: 10-15.
- Brisighelli, F., A. Blanco-Verea, I. Boschi, P. Garagnani, V. L. Pascali, Á. Carracedo, C. Capelli and A. Salas (2012). "Patterns of Y-STR variation in Italy." Forensic Sci. Int. Genet. **6**: 834-839.
- Burton, M. L., C. C. Moore, J. W. M. Whiting, A. K. Romney, D. F. Aberle, J. A. Barcelo, M. M. Dow, J. I. Guyer, D. B. Kronenfeld, J. E. Levy and J. Linnekin (1996). "Regions Based on Social Structure." Current Anthropology **37**(1): 87-123.
- Butler, J. M. (2005). "Forensic DNA Typing, Second Edition: Biology, Technology, and Genetics of STR Markers."

- Butler, J. M. (2006). "Genetics and genomics of core short tandem repeat loci used in human identity testing." J Forensic Sci **51**(2): 253-265.
- Caine, L. M., M. M. de Pancorbo and F. Pinheiro (2010). "Y-chromosomal STR haplotype diversity in males from Santa Catarina, Brazil." J Forensic Leg Med **17**(2): 92-95.
- Calderon, S., D. Perez-Benedico, L. Mesa, D. Guyton, D. J. Rowold and R. J. Herrera (2013). "Phylogenetic and forensic studies of the Southeast Florida Hispanic population using the next-generation forensic PowerPlex(R) Y23 STR marker system." Leg Med (Tokyo) **15**(6): 289-292.
- Campbell, K. D. (2007). "Geographic patterns of haplogroup R1b in the British Isles." JOGG **3**: 1-13.
- Cinnioglu, C., R. King, T. Kivisild, E. Kalfoglu, S. Atasoy, G. L. Cavalleri, A. S. Lillie, C. C. Roseman, A. A. Lin, K. Prince, P. J. Oefner, P. Shen, O. Semino, L. L. Cavalli-Sforza and P. A. Underhill (2004). "Excavating Y-chromosome haplotype strata in Anatolia." Hum Genet **114**(2): 127-148.
- Coble, M. D., C. R. Hill and J. M. Butler (2013). "Haplotype data for 23 Y-chromosome markers in four U.S. population groups." Forensic Sci Int Genet **7**(3): e66-68.
- Cruciani, F., P. Santolamazza, P. Shen, V. Macaulay, P. Moral, A. Olckers, D. Modiano, S. Holmes, G. Destro-Bisol, V. Coia, D. C. Wallace, P. J. Oefner, A. Torroni, L. L. Cavalli-Sforza, R. Scozzari and P. A. Underhill (2002). "A back migration from Asia to Sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes." Am J Hum Genet **70**: 1197-1214.
- Davis, C., J. Ge, C. Sprecher, A. Chidambaram, J. Thompson, M. Ewing, P. Fulmer, D. Rabbach, D. Storts and B. Budowle (2013). "Prototype PowerPlex(R) Y23 System: A concordance study." Forensic Sci Int Genet **7**(1): 204-208.
- de Azevedo, D. A., L. A. F. da Silva, L. Gusmão and E. F. de Carvalho (2009). "Analysis of Y chromosome SNPs in Alagoas, Northeastern Brazil." Forensic Science International: Genetics Supplement Series **2**(1): 421-422.
- de Filippo, C., C. Barbieri, M. Whitten, S. W. Mpoloka, E. D. Gunnarsdottir, K. Bostoen, T. Nyambe, K. Beyer, H. Schreiber, P. de Knijff, D. Luiselli, M. Stoneking and B. Pakendorf (2011). "Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups." Mol. Biol. Evol. **28**(3): 1255-1269.
- Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin and N. B. Freimer (1994). "Mutational processes of simple-sequence repeat loci in human populations." Proc Natl Acad Sci U S A **91**(8): 3166-3170.

- Dupuy, B. M., M. Stenersen, T. Egeland and B. Olaisen (2004). "Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci." Hum Mutat **23**(2): 117-124.
- Durrin, L. K., J. L. Weber and J. Gorski (1984). "Chromatin structure, transcription, and methylation of the prolactin gene domain in pituitary tumors of Fischer 344 rats." J Biol Chem **259**(11): 7086-7093.
- Eltis, D. and D. Richardson (2010). "Atlas of the transatlantic slave trade." Yale University Press.
- Excoffier, L. and H. E. Lischer (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." Mol Ecol Resour **10**(3): 564-567.
- Flores, C., N. Maca-Meyer, A. M. González, P. J. Oefner, P. Shen, J. A. Perez, A. Rojas, J. M. Larruga and P. A. Underhill (2004). "Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography." Eur. J. Hum. Genet. **12**(10): 855-863.
- Foot, S., D. Vollrath, A. Hilton and D. C. Page (1992). "The Human Y chromosome: overlapping DNA clones spanning the euchromatic region." Science **258**: 60-66.
- Galanter, J. M., J. C. Fernández-López, C. R. Gignoux, J. Barnholtz-Sloan, C. Fernandez-Rozadilla, M. Via, A. Hidalgo-Miranda, A. V. Contreras, L. U. Figueroa, P. Raska, G. Jimenez-Sánchez, I. Silva Zolezzi, M. Torres, C. R. Ponte, Y. Ruiz, A. Salas, E. Nguyen, C. Eng, L. Borjas, W. Zabala, G. Barreto, F. Rondon González, A. Ibarra, P. Taboada, L. Porras, F. Moreno, A. Bigham, G. Gutierrez, T. Brutsaert, F. Leon-Velarde, L. G. Moore, E. Vargas, M. Cruz, J. Escobedo, J. Rodriguez-Santana, W. Rodriguez-Cintron, R. Chapela, J. G. Ford, C. Bustamante, D. Seminara, M. Shriver, E. Ziv, E. González Burchard, R. Haile, E. Parra and Á. Carracedo (2012). "Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas." PLoS Genet **8**(3): e1002554.
- Gayà-Vidal, M., P. Moral, N. Saenz-Ruales, P. Gerbault, L. Tonasso, M. Villena, R. Vasquez, C. M. Bravi and J.-M. Dugoujon (2011). "mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations." American Journal of Physical Anthropology **145**(2): 215-230.
- Geppert, M., M. Baeta, C. Núñez, B. Martínez-Jarreta, S. Zweynert, O. W. V. Cruz, F. González-Andrade, J. González-Solorzano, M. Nagy and L. Roewer (2011). "Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America." Forensic Science International: Genetics **5**(2): 100-104.

- Gonçalves, R., H. Spinola and A. Brehm (2007). "Y-chromosome lineages in São Tomé e Príncipe islands: evidence of European influence." Am. J. Hum. Biol. **19**(3): 422-428.
- Gonzalez-Andrade, F., L. Roewer, S. Willuweit, D. Sanchez and B. Martinez-Jarreta (2009). "Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis." Forensic Sci Int Genet **3**(3): e83-91.
- Gusmão, L., P. Sanchez-Diz, F. Calafell, P. Martin, C. A. Alonso, F. Alvarez-Fernandez, C. Alves, L. Borjas-Fajardo, W. R. Bozzo, M. L. Bravo, J. J. Builes, J. Capilla, M. Carvalho, C. Castillo, C. I. Catanesi, D. Corach, A. M. Di Lonardo, R. Espinheira, E. Fagundes de Carvalho, M. J. Farfan, H. P. Figueiredo, I. Gomes, M. M. Lojo, M. Marino, M. F. Pinheiro, M. L. Pontes, V. Prieto, E. Ramos-Luis, J. A. Riancho, A. C. Souza Goes, O. A. Santapa, D. R. Sumita, G. Vallejo, L. Vidal Rioja, M. C. Vide, C. I. Vieira da Silva, M. R. Whittle, W. Zabala, M. T. Zarrabeitia, A. Alonso, A. Carracedo and A. Amorim (2005). "Mutation rates at Y chromosome specific microsatellites." Hum Mutat **26**(6): 520-528.
- Hanson, E. K. and J. Ballantyne (2004). "A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework." J Forensic Sci **49**(1): 40-51.
- Hanson, E. K. and J. Ballantyne (2007). "An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system." PLoS One **2**(8): e688.
- Hedman, M., V. Pimenoff, M. Lukka, P. Sistonen and A. Sajantila (2004). "Analysis of 16 Y STR loci in the Finnish population reveals a local reduction in the diversity of male lineages." Forensic Sci Int **142**(1): 37-43.
- Heinz, T., V. Alvarez-Iglesias, J. Pardo-Seco, P. Taboada-Echalar, A. Gomez-Carballa, A. Torres-Balanza, O. Rocabado, A. Carracedo, C. Vullo and A. Salas (2013). "Ancestry analysis reveals a predominant Native American component with moderate European admixture in Bolivians." Forensic Sci Int Genet **7**(5): 537-542.
- Heyer, E., J. Puymirat, P. Dieltjes, E. Bakker and P. Knijff (1997). "Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees." Human Molecular Genetics **6**(5): 799-803.
- Iannacone, G. C., R. Y. Tito, P. W. Lopez, M. E. Medina and B. Lizarraga (2005). "Y-chromosomal haplotypes for the PowerPlex Y for twelve STRs in a Peruvian population sample." J Forensic Sci **50**(1): 239-242.
- Jobling, M. A. (2001). "In the name of the father: surnames and genetics." Trends Genet **17**(6): 353-357.
- Jobling, M. A. (2001). "Y-chromosomal SNP haplotype diversity in forensic analysis." Forensic Sci Int **118**(2-3): 158-162.

- Jobling, M. A. and P. Gill (2004). "Encoded evidence: DNA in forensic analysis." Nat Rev Genet **5**(10): 739-751.
- Jobling, M. A., M. Hurles and C. Tyler-Smith (2004). "Human evolutionary genetics : origins, peoples & disease." New York, NY : Garland Science.
- Jobling, M. A. and T. E. King (2004). "The distribution of Y-chromosomal haplotypes: forensic implications." International Congress Series **1261**(0): 70-72.
- Jobling, M. A., A. Pandya and C. Tyler-Smith (1997). "The Y chromosome in forensic analysis and paternity testing." Int J Legal Med **110**(3): 118-124.
- Jobling, M. A. and C. Tyler-Smith (1995). "Fathers and sons: the Y chromosome and human evolution." Trends Genet **11**(11): 449-456.
- Jobling, M. A. and C. Tyler-Smith (2003). "The human Y chromosome: an evolutionary marker comes of age." Nat Rev Genet **4**(8): 598-612.
- Karafet, T. M., B. Hallmark, M. P. Cox, H. Sudoyo, S. Downey, J. S. Lansing and M. F. Hammer (2010). "Major East-West division underlies Y chromosome stratification across Indonesia." Mol. Biol. Evol. **27**(8): 1833-1844.
- Karafet, T. M., J. S. Lansing, A. J. Redd, S. Reznikova, J. C. Watkins, S. P. Surata, W. A. Arthawiguna, L. Mayer, M. Bamshad, L. B. Jorde and M. F. Hammer (2005). "Balinese Y-chromosome perspective on the peopling of Indonesia: genetic contributions from pre-neolithic hunter-gatherers, Austronesian farmers, and Indian traders." Hum Biol. **77**(1): 93-114.
- Karafet, T. M., F. L. Mendez, M. B. Meilerman, P. A. Underhill, S. L. Zegura and M. F. Hammer (2008). "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree." Genome Res **18**(5): 830-838.
- Karafet, T. M., S. L. Zegura, O. Posukh, L. Osipova, A. Bergen, J. Long, D. Goldman, W. Klitz, S. Harihara, P. de Knijff, V. Wiebe, R. C. Griffiths, A. R. Templeton and M. F. Hammer (1999). "Ancestral Asian source(s) of new world Y-chromosome founder haplotypes." Am J Hum Genet **64**(3): 817-831.
- Kayser, M., S. Brauer, G. Weiss, W. Schiefenovel, P. A. Underhill and M. Stoneking (2001). "Independent histories of human Y chromosomes from Melanesia and Australia." Am J Hum Genet **68**(1): 173-190.
- Kayser, M., R. Kittler, A. Erler, M. Hedman, A. C. Lee, A. Mohyuddin, S. Q. Mehdi, Z. Rosser, M. Stoneking, M. A. Jobling, A. Sajantila and C. Tyler-Smith (2004). "A comprehensive survey of human Y-chromosomal microsatellites." Am J Hum Genet **74**(6): 1183-1197.

- Kayser, M., O. Lao, K. Anslinger, C. Augustin, G. Bargel, J. Edelmann, S. Elias, M. Heinrich, J. Henke, L. Henke, C. Hohoff, A. Illing, A. Jonkisz, P. Kuzniar, A. Lebioda, R. Lessig, S. Lewicki, A. Maciejewska, D. M. Monies, R. Pawlowski, M. Poetsch, D. Schmid, U. Schmidt, P. M. Schneider, B. Stradmann-Bellinghausen, R. Szibor, R. Wegener, M. Wozniak, M. Zoledziewska, L. Roewer, T. Dobosz and R. Ploski (2005). "Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis." Hum. Genet. **117**(5): 428-443.
- Kayser, M., L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Kruger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. Knijff, M. Stoneking and A. Sajantila (2000). "Characteristics and frequency of germline mutations at microsatellite loci from the Human Y chromosome, as revealed by direct observation in father/son pairs." American journal of Human Genetics **66**: 1580-1588.
- Kayser, M. and A. Sajantila (2001). "Mutations at Y-STR loci: implications for paternity testing and forensic analysis." Forensic Sci Int **118**(2-3): 116-121.
- Kemp, B. M., R. S. Malhi, J. McDonough, D. A. Bolnick, J. A. Eshleman, O. Rickards, C. Martinez-Labarga, J. R. Johnson, J. G. Lorenz, E. J. Dixon, T. E. Fifield, T. H. Heaton, R. Worl and D. G. Smith (2007). "Genetic analysis of early holocene skeletal remains from Alaska and its implications for the settlement of the Americas." Am J Phys Anthropol **132**(4): 605-621.
- Kovatsi, L., J. L. Saunier and J. A. Irwin (2009). "Population genetics of Y-chromosome STRs in a population of Northern Greeks." Forensic Sci Int Genet **4**(1): e21-22.
- Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J.

- Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y. J. Chen (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
- Lee, J., A. A. Ewis, M. E. Hurler, H. Kashiwazaki, T. Shinka and Y. Nakahori (2007). "Y chromosomal STRs haplotypes in two populations from Bolivia." Legal medicine (Tokyo, Japan) **9**(1): 43-47.
- Lell, J. T., R. I. Sukernik, Y. B. Starikovskaya, B. Su, L. Jin, T. G. Schurr, P. A. Underhill and D. C. Wallace (2002). "The dual origin and Siberian affinities of Native American Y chromosomes." Am J Hum Genet **70**(1): 192-206.
- Lewis, M. (2009). "Ethnologue. Languages of the world." Dallas, Texas: SIR International.
- Lipski, J. (2006). "Afro-Bolivian language today: the oldest surviving Afro-Hispanic speech community." Afro-Hispanic Review: 25.
- Lipski, J. (2006). "Afro-Bolivian Spanish and Helvécia Portuguese: semi-creole parallels." Papia **16**: 96-116.
- Lipski, J. (2007). "Afro-Bolivian Spanish: The Survival of a True Creole Prototype." John Benjamins Publishing.
- Lizcano, F. F. (2005). "Composición Étnica de las Tres Áreas Culturales del Continente Americano al Comienzo del Siglo XXI." Convergencia **2**(38): 185-232.
- Long, J. C., R. C. Williams, J. E. McAuley, R. Medis, R. Partel, W. M. Tregellas, S. F. South, A. E. Rea, S. B. McCormick and U. Iwaniec (1991). "Genetic variation in Arizona Mexican Americans: Estimation and interpretation of admixture proportions." American Journal of Physical Anthropology **84**(2): 141-157.

- López-Parra, A. M., M. Tirado, C. Baeza, F. Bert, A. Corella, A. Pérez-Pérez, C. Gamba, E. Fernández and E. Arroyo-Pardo (2008). "Genetic structure of the population of Beni department (North Bolivia)." Forensic Sci. Int.: Genet. Suppl. Series **1**(1): 348-349.
- Lovo-Gómez, J., A. Blanco-Verea, M. V. Lareu, M. Brión and A. Carracedo (2007). "The genetic male legacy from El Salvador." Forensic Science International **171**(2): 198-203.
- Luis, J. R., D. J. Rowold, M. Regueiro, B. Caeiro, C. Cinnioglu, C. Roseman, P. A. Underhill, L. L. Cavalli-Sforza and R. J. Herrera (2004). "The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations." Am. J. Hum. Genet. **74**(3): 532-544.
- Marcheco-Teruel, B., E. J. Parra, E. Fuentes-Smith, A. Salas, H. N. Buttenschon, D. Demontis, M. Torres-Espanol, L. C. Marin-Padron, E. J. Gomez-Cabezas, V. Alvarez-Iglesias, A. Mosquera-Miguel, A. Martinez-Fuentes, A. Carracedo, A. D. Borglum and O. Mors (2014). "Cuba: exploring the history of admixture and the genetic basis of pigmentation using autosomal and uniparental markers." PLoS Genet **10**(7): e1004488.
- Marino, M. and S. Furfuro (2010). "Genetic population data of 12 Y-chromosome STRs loci in Mendoza population (Argentina)." Forensic Sci Int Genet **4**(4): e89-93.
- Maroñas, O., C. Phillips, J. Söchtig, A. Gomez-Tato, R. Cruz, J. Alvarez-Dios, M. C. de Cal, Y. Ruiz, M. Fondevila, Á. Carracedo and M. V. Lareu (2014). "Development of a forensic skin colour predictive test." Forensic Science International: Genetics **13**(0): 34-44.
- McLean, D. C., Jr., I. Spruill, G. Argyropoulos, G. P. Page, M. D. Shriver and W. T. Garvey (2005). "Mitochondrial DNA (mtDNA) haplotypes reveal maternal population genetic affinities of Sea Island Gullah-speaking African Americans." Am. J. Phys. Anthropol. **127**(4): 427-438.
- Mendizabal, I., K. Sandoval, G. Berniell-Lee, F. Calafell, A. Salas, A. Martinez-Fuentes and D. Comas (2008). "Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba." BMC Evol Biol **8**: 213.
- Mohyuddin, A., Q. Ayub, S. Siddiqi, D. R. Carvalho-Silva, K. Mazhar, S. Rehman, S. Firasat, A. Dar, C. Tyler-Smith and S. Q. Mehdi (2004). "Genetic instability in EBV-transformed lymphoblastoid cell lines." Biochim Biophys Acta **1670**(1): 81-83.
- Montano, V., G. Ferri, V. Marcari, C. Batini, O. Anyaele, G. Destro-Bisol and D. Comas (2011). "The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa." Mol. Ecol. **20**(13): 2693-2708.
- Mosquera-Miguel, A., V. Alvarez-Iglesias, M. Cerezo, M. V. Lareu, A. Carracedo and A. Salas (2009). "Testing the performance of mtSNP minisequencing in forensic samples." Forensic Sci Int Genet **3**(4): 261-264.

- Mulero, J. J., C. W. Chang, L. M. Calandro, R. L. Green, Y. Li, C. L. Johnson and L. K. Hennessy (2006). "Development and validation of the AmpFISTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system." J Forensic Sci **51**(1): 64-75.
- Núñez, C., M. Baeta, C. Sosa, Y. Casalod, J. Ge, B. Budowle and B. Martínez-Jarreta (2010). "Reconstructing the population history of Nicaragua by means of mtDNA, Y-chromosome STRs, and autosomal STR markers." American Journal of Physical Anthropology **143**(4): 591-600.
- Ohta, T. and M. Kimura (1973). "A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population." Genet Res **22**: 201-204.
- Onofri, V., F. Alessandrini, C. Turchi, M. Pesaresi, L. Buscemi and A. Tagliabracci (2006). "Development of multiplex PCRs for evolutionary and forensic applications of 37 human SNPs." Forensic Sci Int **157**: 23-25.
- Oota, H., W. Settheetham-Ishida, D. Tiwawech, T. Ishida and M. Stoneking (2001). "Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence." Nat Genet **29**(1): 20-21.
- Pardo-Seco, J., F. Martín-Torres and A. Salas (2014). "Evaluating the accuracy of AIM panels at quantifying genome ancestry." BMC Genomics **30**(15(1)): 543.
- Perego, U. A., H. Lancioni, M. Tribaldos, N. Angerhofer, J. E. Ekins, A. Olivieri, S. R. Woodward, J. M. Pascale, R. Cooke, J. Motta and A. Achilli (2012). "Decrypting the Mitochondrial Gene Pool of Modern Panamanians." PLoS ONE **7**(6): e38337.
- Phillips, C., M. Fondevila and M. V. Lareu (2012). "A 34-plex autosomal SNP single base extension assay for ancestry investigations." Methods Mol Biol **830**: 109-126.
- Pinto, N., L. Gusmão and A. Amorim (2014). "Mutation and mutation rates at Y chromosome specific Short Tandem Repeat Polymorphisms (STRs): A reappraisal." Forensic Science International: Genetics **9**(0): 20-24.
- Promega Corporation Technical Manual PowerPlex® Y23 System Instructions for use of products DC2305 AND DC2320, <http://www.promega.com/>.
- Purps, J., S. Siegert, S. Willuweit, M. Nagy, C. Alves, R. Salazar, S. M. Angustia, L. H. Santos, K. Anslinger, B. Bayer, Q. Ayub, W. Wei, Y. Xue, C. Tyler-Smith, M. B. Bafalluy, B. Martínez-Jarreta, B. Egyed, B. Balitzki, S. Tschumi, D. Ballard, D. S. Court, X. Barrantes, G. Bassler, T. Wiest, B. Berger, H. Niederstatter, W. Parson, C. Davis, B. Budowle, H. Burri, U. Borer, C. Koller, E. F. Carvalho, P. M. Domingues, W. T. Chamoun, M. D. Coble, C. R. Hill, D. Corach, M. Caputo, M. E. D'Amato, S. Davison, R. Decorte, M. H. Larmuseau, C. Ottoni, O. Rickards, D. Lu, C. Jiang, T. Dobosz, A. Jonkisz, W. E. Frank, I. Furac, C. Gehrig, V. Castella, B. Grskovic, C. Haas, J. Wobst, G. Hadzic, K. Drobic, K. Honda, Y. Hou, D. Zhou, Y. Li, S. Hu, S.

- Chen, U. D. Immel, R. Lessig, Z. Jakovski, T. Ilievska, A. E. Klann, C. C. Garcia, P. de Knijff, T. Kraaijenbrink, A. Kondili, P. Miniati, M. Vouropoulou, L. Kovacevic, D. Marjanovic, I. Lindner, I. Mansour, M. Al-Azem, A. E. Andari, M. Marino, S. Furfuro, L. Locarno, P. Martin, G. M. Luque, A. Alonso, L. S. Miranda, H. Moreira, N. Mizuno, Y. Iwashima, R. S. Neto, T. L. Nogueira, R. Silva, M. Nastainczyk-Wulf, J. Edelmann, M. Kohl, S. Nie, X. Wang, B. Cheng, C. Nunez, M. M. Pancorbo, J. K. Olofsson, N. Morling, V. Onofri, A. Tagliabracci, H. Pamjav, A. Volgyi, G. Barany, R. Pawlowski, A. Maciejewska, S. Pelotti, W. Pepinski, M. Abreu-Glowacka, C. Phillips, J. Cardenas, D. Rey-Gonzalez, A. Salas, F. Brisighelli, C. Capelli, U. Toscanini, A. Piccinini, M. Piglionica, S. L. Baldassarra, R. Ploski, M. Konarzewska, E. Jastrzebska, C. Robino, A. Sajantila, J. U. Palo, E. Guevara, J. Salvador, M. C. Ungria, J. J. Rodriguez, U. Schmidt, N. Schlauderer, P. Saukko, P. M. Schneider, M. Sirker, K. J. Shin, Y. N. Oh, I. Skitsa, A. Ampati, T. G. Smith, L. S. Calvit, V. Stenzl, T. Capal, A. Tillmar, H. Nilsson, S. Turrina, D. De Leo, A. Verzeletti, V. Cortellini, J. H. Wetton, G. M. Gwynne, M. A. Jobling, M. R. Whittle, D. R. Sumita, P. Wolanska-Nowak, R. Y. Yong, M. Krawczak, M. Nothnagel and L. Roewer (2014). "A global analysis of Y-chromosomal haplotype diversity for 23 STR loci." *Forensic Sci Int Genet* **12c**: 12-23.
- Ramos-Luis, E., A. Blanco-Verea, M. Brión, V. Van Huffel, A. Carracedo and P. Sánchez-Diz (2009). "Phylogeography of French male lineages." *Forensic Science International: Genetics Supplement Series* **2(1)**: 439-441.
- Rasmussen, M., S. L. Anzick, M. R. Waters, P. Skoglund, M. DeGiorgio, T. W. Stafford, Jr., S. Rasmussen, I. Moltke, A. Albrechtsen, S. M. Doyle, G. D. Poznik, V. Gudmundsdottir, R. Yadav, A. S. Malaspinas, S. S. t. White, M. E. Allentoft, O. E. Cornejo, K. Tambets, A. Eriksson, P. D. Heintzman, M. Karmin, T. S. Korneliusen, D. J. Meltzer, T. L. Pierre, J. Stenderup, L. Saag, V. M. Warmuth, M. C. Lopes, R. S. Malhi, S. Brunak, T. Sicheritz-Ponten, I. Barnes, M. Collins, L. Orlando, F. Balloux, A. Manica, R. Gupta, M. Metspalu, C. D. Bustamante, M. Jakobsson, R. Nielsen and E. Willerslev (2014). "The genome of a Late Pleistocene human from a Clovis burial site in western Montana." *Nature* **506(7487)**: 225-229.
- Roby, R. K., S. D. Gonzalez, N. R. Phillips, J. V. Planz, J. L. Thomas, J. A. P. Astudillo, J. Ge, E. A. Morales, A. J. Eisenberg, R. Chakraborty, P. Bustos and B. Budowle (2009). "Autosomal STR allele frequencies and Y-STR and mtDNA haplotypes in Chilean sample populations." *Forensic Sci. Int.: Genet. Suppl. Series* **2(1)**: 532-533.
- Rodig, H., L. Roewer, A. Gross, T. Richter, P. de Knijff, M. Kayser and W. Brabetz (2008). "Evaluation of haplotype discrimination capacity of 35 Y-chromosomal short tandem repeat loci." *Forensic Sci Int* **174(2-3)**: 182-188.
- Roewer, L., P. J. Croucher, S. Willuweit, T. T. Lu, M. Kayser, R. Lessig, P. de Knijff, M. A. Jobling, C. Tyler-Smith and M. Krawczak (2005). "Signature of recent historical events in the European Y-chromosomal STR haplotype distribution." *Hum Genet* **116(4)**: 279-291.
- Roewer, L., M. Nothnagel, L. Gusmão, V. Gomes, M. González, D. Corach, A. Sala, E. Alechine, T. Palha, N. Santos, A. Ribeiro-dos-Santos, M. Geppert, S. Willuweit, M.

- Nagy, S. Zweynert, M. Baeta, C. Núñez, B. Martínez-Jarreta, F. González-Andrade, E. Fagundes de Carvalho, D. A. da Silva, J. J. Builes, D. Turbón, A. M. Lopez Parra, E. Arroyo-Pardo, U. Toscanini, L. Borjas, C. Barletta, E. Ewart, S. Santos and M. Krawczak (2013). "Continent-Wide Decoupling of Y-Chromosomal Genetic Variation from Language and Geography in Native South Americans." PLoS Genet **9**(4): e1003460.
- Romero, R. E., I. Briceno, P. Lizarazo Rdel, S. Willuweit, L. Roewer and A. Gomez (2008). "A Colombian Caribbean population study of 16 Y-chromosome STR loci." Forensic Sci Int Genet **2**(2): e5-8.
- Rosa, A., C. Ornelas, M. A. Jobling, A. Brehm and R. Villems (2007). "Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective." BMC Evol. Biol. **7**: 124.
- Salas, A., A. Acosta, V. Alvarez-Iglesias, M. Cerezo, C. Phillips, M. V. Lareu and A. Carracedo (2008). "The mtDNA ancestry of admixed Colombian populations." Am J Hum Biol **20**(5): 584-591.
- Salas, A., A. Carracedo, M. Richards and V. Macaulay (2005). "Charting the ancestry of African Americans." Am J Hum Genet **77**(4): 676-680.
- Salas, A., J. C. Jaime, V. Alvarez-Iglesias and A. Carracedo (2008). "Gender bias in the multiethnic genetic composition of central Argentina." J Hum Genet **53**(7): 662-674.
- Salas, A., B. Quintans and V. Alvarez-Iglesias (2005). "SNaPshot typing of mitochondrial DNA coding region variants." Methods Mol Biol **297**: 197-208.
- Salas, A., M. Richards, M.-V. Lareu, B. Sobrino, S. Silva, M. Matamoros, V. Macaulay and A. Carracedo (2005). "Shipwrecks and founder effects: Divergent demographic histories reflected in Caribbean mtDNA." American Journal of Physical Anthropology **128**(4): 855-860.
- Schwengber, S. P., T. Kommers, C. H. Matte, P. E. Raimann, B. A. Carvalho, F. P. Leite, M. A. Medeiros, L. F. Souza, C. S. Castro, F. G. Chassot and S. L. Bonatto (2009). "Population data of 17 Y-STR loci from Rio Grande do Sul state (South Brazil)." Forensic Sci Int Genet **4**(1): e31-33.
- Seielstad, M. T., E. Minch and L. L. Cavalli-Sforza (1998). "Genetic evidence for a higher female migration rate in humans." Nat Genet **20**(3): 278-280.
- Semino, O., A. S. Santachiara-Benerecetti, F. Falaschi, L. L. Cavalli-Sforza and P. A. Underhill (2002). "Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny." Am J Hum Genet **70**(1): 265-268.

- Sia, E. A., R. J. Kokoska, M. Dominska, P. Greenwell and T. D. Petes (1997). "Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes." Mol Cell Biol **17**(5): 2851-2858.
- Simms, T. M., E. Martínez, K. J. Herrera, M. R. Wright, O. A. Pérez, M. Hernández, E. C. Ramírez, Q. McCartney and R. J. Herrera (2011). "Paternal lineages signal distinct genetic contributions from British Loyalists and continental Africans among different Bahamian islands." Am. J. Phys. Anthropol. **146**(4): 594-608.
- Simms, T. M., M. R. Wright, M. Hernandez, O. A. Perez, E. C. Ramirez, E. Martinez and R. J. Herrera (2012). "Y-chromosomal diversity in Haiti and Jamaica: contrasting levels of sex-biased gene flow." Am J Phys Anthropol **148**(4): 618-631.
- Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L. G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, k. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S. F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S. P. Yang, R. H. Waterston, R. K. Wilson, S. Rozen and D. C. Page (2003). "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes." Nature **423**: 825-837.
- Taboada-Echalar, P., V. Álvarez-Iglesias, T. Heinz, L. Vidal-Bralo, A. Gómez-Carballa, L. Catelli, J. Pardo-Seco, A. Pastoriza, Á. Carracedo, A. Torres-Balanza, O. Rocabado, C. Vullo and A. Salas (2013). "The genetic legacy of the pre-Colonial period in contemporary Bolivians." PLoS One **8**(3): e58980.
- The HapMap Consortium (2003). "The International HapMap Project." Nature **426**(6968): 789-796.
- The HapMap Consortium (2005). "A haplotype map of the human genome." Nature **437**(7063): 1299-1320.
- Thorisson, G. A., and Stein, L.D (2003). "The SNP Consortium website: past, present and future." Nucleic Acids Research **31**: 124-127.
- Tirado, M., A. M. Lopez-Parra, C. Baeza, F. Bert, A. Corella, A. Perez-Perez, D. Turbon and E. Arroyo-Pardo (2009). "Y-chromosome haplotypes defined by 17 STRs included in AmpFISTR Yfiler PCR Amplification Kit in a multi ethnical population from El Beni Department (North Bolivia)." Leg Med (Tokyo) **11**(2): 101-103.
- Toscanini, U., M. García-Magariños, G. Berardi, T. Egeland, E. Raimondi and A. Salas (2012). "Evaluating methods to correct for population stratification when estimating paternity indexes." PLoS One **7**(11): e49832.

- Toscanini, U., L. Gusmão, G. Berardi, A. Amorim, A. Carracedo, A. Salas and E. Raimondi (2006). "Genetic variability of 17 Y chromosome STRs in two Native American populations from Argentina." International Congress Series **1288**(0): 154-155.
- Toscanini, U., L. Gusmão, G. Berardi, A. Amorim, A. Carracedo, A. Salas and E. Raimondi (2007). "Testing for genetic structure in different urban Argentinian populations." Forensic Sci. Int. **165**(1): 35-40.
- Toscanini, U., L. Gusmão, G. Berardi, A. Amorim, A. Carracedo, A. Salas and E. Raimondi (2008). "Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data." Forensic Sci. Int. Genet. **2**(4): 274-280.
- Toscanini, U., A. Salas, M. García-Magariños, L. Gusmão and E. Raimondi (2010). "Population stratification in Argentina strongly influences likelihood ratio estimates in paternity testing as revealed by a simulation-based approach." Int J Legal Med **124**(1): 63-69.
- Underhill, P. A., G. Passarino, A. A. Lin, P. Shen, M. Mirazon Lahr, R. A. Foley, P. J. Oefner and L. L. Cavalli-Sforza (2001). "The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations." Ann Hum Genet **65**(Pt 1): 43-62.
- Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K.

- Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.
- Vullo, C., V. Gomes, C. Romanini, A. M. Oliveira, O. Rocabado, J. Aquino, A. Amorim and L. Gusmao (2014). "Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations." Int J Legal Med.
- Wei, W., Q. Ayub, Y. Xue and C. Tyler-Smith (2013). "A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping." Forensic Sci Int Genet **7**(6): 568-572.
- Whitfield, L. S., J. E. Sulston and P. N. Goodfellow (1995). "Sequence variation of the human Y chromosome." Nature **378**(6555): 379-380.
- Wyman, A. R. and R. White (1980). "A highly polymorphic locus in human DNA." Proc Natl Acad Sci U S A **77**(11): 6754-6758.
- Y-Chromosome-Consortium (2002). "A nomenclature system for the tree of human Y-chromosomal binary haplogroups." Genome Res **12**(2): 339-348.
- Zegura, S. L., T. M. Karafet, L. A. Zhivotovsky and M. F. Hammer (2004). "High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas." Mol Biol Evol **21**(1): 164-175.
- Zerjal, T., Y. Xue, G. Bertorelle, R. S. Wells, W. Bao, S. Zhu, R. Qamar, Q. Ayub, A. Mohyuddin, S. Fu, P. Li, N. Yuldasheva, R. Ruzibakiev, J. Xu, Q. Shu, R. Du, H. Yang, M. E. Hurles, E. Robinson, T. Gerelsaikhan, B. Dashnyam, S. Q. Mehdi and C. Tyler-Smith (2003). "The genetic legacy of the Mongols." Am J Hum Genet **72**(3): 717-721.
- Zhivotovsky, L. A. and P. A. Underhill (2005). "On the evolutionary mutation rate at Y-chromosome STRs: comments on paper by Di Giacomo et al. (2004)." Hum Genet **116**(6): 529-532.
- Zhivotovsky, L. A., P. A. Underhill, C. Cinnioglu, M. Kayser, B. Morar, T. Kivisild, R. Scozzari, F. Cruciani, G. Destro-Bisol, G. Spedini, G. K. Chambers, R. J. Herrera, K.

K. Yong, D. Gresham, I. Tournev, M. W. Feldman and L. Kalaydjieva (2004). "The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time." Am J Hum Genet **74**(1): 50-61.

Zhong, H., H. Shi, X. B. Qi, Z. Y. Duan, P. P. Tan, L. Jin, B. Su and R. Z. Ma (2011). "Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route." Mol. Biol. Evol. **28**(1): 717-727.



VIII. Índice de Figuras

Figura 1.	(a) Representación esquemática de los Cromosomas sexuales X e Y; En los extremos del cromosoma Y, se encuentran las regiones pseudoautosómicas, PAR 1 y PAR 2, y el resto del cromosoma lo ocupa la MSY o NRY. (b) El Cromosoma Y está compuesto de dos grandes regiones, una eucromática y otra heterocromática (Tomado y Modificado de (Butler 2005)).	7
Figura 2.	(a) Diagrama Completo del Cromosoma Y (b) Detalle de la eucromatina de la región MSY y las clases de repeticiones. La barra 1 Mb indica la escala del diagrama (Tomado y modificado de (Skaletsky <i>et al.</i> 2003))	8
Figura 3.	Árbol filogenético de los 18 haplogrupos principales de Cromosoma Y (A-R); los eventos mutacionales se presentan en las ramas del árbol (Tomado y modificado de (Jobling and Tyler-Smith 2003))	13
Figura 4.	Mapa actual del árbol de máxima parsimonia, donde se resaltan los 20 clados o Haplogrupos mayores (A-T) (Tomado de (Karafet <i>et al.</i> 2008))	14
Figura 5.	Distribución geográfica global de los 18 haplogrupos principales de Cromosoma Y (Jobling y Tyler-Smith, 2003). La distribución demuestra una tendencia migración de los haplogrupos desde África hacia América y Oceanía. (Poblaciones tomadas de Jobling y Tyler-Smith, 2003).	16
Figura 6.	<i>Slippage</i> replicativo. a) Replicación normal sin pérdida ni ganancia de repetición. b) Inserción por ganancia de repetición en la hebra nascente. c) Deleción por pérdida de repetición en la hebra nascente.	20
Figura 7.	Capacidad de discriminación adquirida por los marcadores RM Y-STRs y <i>Yfiler</i> ® en un grupo de 8 regiones geográficas y globales del panel HGDP-CEPH; el grupo de marcadores de alta tasa mutacional muestra un mayor poder de discriminación en 7 de las 8 regiones. Tomado de (Ballantyne <i>et al.</i> 2012).	22
Figura 8.	Y-STRs usados en la rutina forense. Con el fin de asegurar la compatibilidad entre los distintos grupos de marcadores, se han establecido, el haplotipo mínimo (MHT) y el extendido que incorpora los sistemas DYS438 y el DYS439, que es el recomendado por el SWGDAM (Scientific Working Group on DNA Analysis Methods; www.SWGDAM.org).	25
Figura 9.	Marcadores incorporados en el kit PowerPlex® Y23 ® y su localización en el cromosoma Y.	26
Figura 10.	Posibles rutas de migración desde Siberia (Continente Euroasiático) hacia Norteamérica (Continente Americano); la ruta 1 . Establece la migración desde Siberia hacia Alaska por el puente terrestre de Beringia y luego a través del corredor libre de capas de hielo en Norteamérica, y la ruta 2 . Es la ruta costera que establece la migración por un corredor costero a través de Beringia; ambas rutas no se encontraban disponibles al mismo tiempo, por lo que pueden ser excluyentes una de la otra. Con asterisco (*) son indicados, los hallazgos arqueológicos con evidencia genética masculina disponible que dan soporte a la presencia humana en los periodos posteriores a 16 KYA. (Tomado y modificado de (Jobling <i>et al.</i> 2004))	30

Figura 11. Árbol Filogenético de los haplogrupos de cromosoma Y detectados por los 5 distintos múltiplexes usados para el tipado de Y-SNPs de acuerdo a la nomenclatura propuesta por (Karafet <i>et al.</i> 2008).....	48
Figura 12. Mapa en el que se enseña los lugares de muestreo, señalados con un asterisco y los departamentos de Bolivia; los departamentos fueron agrupados en 3 regiones eco-geográficas (Andina, sub-Andina y Llanos); las frecuencias de los haplogrupos de cromosoma Y son resumidos en los gráficos circulares. La información completa de haplogrupos y haplotipos se muestra en la Tabla S1 disponible <i>on line.</i>	59
Figura 13. Análisis de componentes Principales (PCA) de las frecuencias de haplogrupos de Cromosoma Y; los ejes explican el porcentaje de varianza explicado por cada componente. Códigos de las poblaciones: África = Fon (BEN) (Luis <i>et al.</i> 2004, Simms <i>et al.</i> 2011); Fulbe (FUL) (Rosa <i>et al.</i> 2007); São Tomé de Príncipe (STP) (Gonçalves <i>et al.</i> 2007); Bakaka, Ewondo, Ngoumba (CAM), Makina and Nzebi (GAB), Tiv y Igala (NIG) (Montano <i>et al.</i> 2011); Angola (ANG) (Beleza <i>et al.</i> 2005); Bamelike (BAM) (Luis <i>et al.</i> 2004); República democrática del Congo (DRC) (de Filippo <i>et al.</i> 2011); Nativos Americanos = Nicaragua (NIC) (Nuñez <i>et al.</i> 2010) ; Bolivia - Aymara (AYM) y Bolivia - Quechua (QUE) (Gayà-Vidal <i>et al.</i> 2011); Andina - La Paz (BAP), y Sub-Andina - Chuquisaca (BSC) (Vullo <i>et al.</i> 2014) ; Apalai (APL) (McLean <i>et al.</i> 2005); Diaguitas (ARD), Kolla - Jujuy (AKJ), Kolla - Salta (AKS), y Mapuche (ARM) (Blanco-Verea <i>et al.</i> 2010); Este de Asia = Bali (Indonesia) (BAL) (Karafet <i>et al.</i> 2005); Filipinas (PHI), Taiwán (TAI), She (SHE), Yao (YAO), y Vietnam (VIE) (Karafet <i>et al.</i> 2010); Han (HAN) (Zhong <i>et al.</i> 2011); Europa = Cantabria (CAN), León (LEO) (Flores <i>et al.</i> 2004); Inglaterra (ENG), Gales (WAL), Escocia (SCO) (Campbell 2007); Galicia (GAL) (Brión <i>et al.</i> 2004); Alemania (GER) (Kayser <i>et al.</i> 2005); Portugal (POR) (Beleza <i>et al.</i> 2006); Presente estudio= Region Andina (BAN), Sub-Andina Region (BSA), y Llanos (BLL).....	63
Figura 14. Proporciones de ancestralidad estimadas desde distintos marcadores como son, AIMs-Indels, ADNmt, y cromosoma Y. Las estimaciones para los AIMs y para el ADNmt fueron tomadas de (Taboada-Echalar <i>et al.</i> 2013), y fueron usadas para generar esta figura que tiene el objetivo de comparar estos valores con los obtenidos para Cromosoma Y. Todos los individuos considerados para esta figura, fueron tipados para los marcadores AIMs y los marcadores uniparentales (Ver Tabla S2 disponible <i>on line</i>).....	65
Figura 15. Gráfico de 2 dimensiones del análisis de MDS con los valores de F_{ST} linearizados por el método de Slatkin para los haplotipos de RM-YSTRs en la muestras de todos los 111 grupos poblacionales ($n=12,272$) (Stress = 0.07462); El gráfico pequeño, (Superior izquierda) muestra el mismo análisis de MDS para todos los grupos poblacionales con un tamaño de muestra de 100 individuos por población (Stress = 0.02416).....	69
Figura 16. Proporción de individuos con haplotipos compartidos dentro de las poblaciones (Izquierda) y entre las poblaciones (derecha) Para Yfiler (Barras azul Claro) y para RM-YSTRs (Barras rojas).....	72
Figura 17. Representación geográfica de las coincidencias de haplotipos entre las poblaciones analizadas para el panel de RM-YSTRs y el panel <i>Yfiler</i> . El panel A) muestral los	

haplotipos con coincidencias para <i>Yfiler</i> y el panel B) para los RM-YSTRs. Las imágenes inferiores izquierda, muestran una ampliación de las poblaciones europeas.....	73
Figura 18. Redes <i>Weighted median-joining</i> estimadas para los Y-STRs de una sola copia en 1000 individuos escogidos aleatoriamente del grupo total de muestras 7,784, pertenecientes a todos los grupos poblacionales tipados (indicados en la leyenda). A) Red basada en los haplotipos de <i>Yfiler</i> con 15 <i>loci</i> (DYS385a/b, excluido) y B) Red basada en los haplotipos de 10- <i>loci</i> del panel RM-YSTRs (excluyendo los sistemas multicopia DYF399S1, DYF403S1 y DYF404S1).....	73
Figura 19. Diferenciación empírica en 2,528 dúos de parientes separados por 20 meiosis para <i>Yfiler</i> (Barras azules) y RM-YSTRs (Barras rojas). En esta figura se combina los 2,372 dúos de parientes del presente estudio con los 156 analizados anteriormente por (Ballantyne <i>et al.</i> 2012). Los valores que se encuentran en la parte superior representan los números absolutos de los pares de parientes en los que la tasa de estimación está basada; las barras de error representan los intervalos de confianza binomial.....	78
Figura 20. Estructura poblacional determinada por el panel PPY23; el mapa expresa las dos primeras componentes del MDS basado en las distancias R_{ST} . Los círculos en blanco representan las poblaciones muestreadas; la escala de colores tipo-arcoíris de amarillo a magenta explican los extremos de la escala.....	83
Figura 21. Estructura poblacional determinada por el panel PPY23; el mapa expresa las cuatro componentes del MDS basado en las distancias R_{ST} . Los círculos en blanco representan las poblaciones muestreadas; la escala de colores tipo-arcoíris de amarillo a magenta explican los extremos de la escala.....	86



IX. Índice de Tablas

Tabla 1. Características de los Y-STRs analizados en el presente estudio. (Tomada y modificada de (Butler 2006)).....	43
Tabla 2. Concentración final de primers usados para cada reacción PCR multiplex, tomado de (Ballantyne <i>et al.</i> 2014).....	45
Tabla 3. Volumen de componentes para las tres distintas reacciones PCR multiplex; el volumen final de reacción es de 10 μ L, tomado de (Ballantyne <i>et al.</i> 2014).....	45
Tabla 4. Secuencia de primers empleados en la amplificación, diseñados por Whitfield y cols. (1995), Brión y cols. (2004), Blanco-Verea y cols. (2010) y Ramos-Luis y cols. (2009). La tercera columna indica el tamaño del amplicón; la cuarta columna indica los alelos esperados, la quinta columna indica la concentración final de la pareja de primers en el respectivo arreglo de multiplex de la sexta columna.	49
Tabla 5. Fluorocromos asociados a los ddNTPs incorporados durante la reacción de SNaPshot TM	50
Tabla 6. Secuencia de las sondas empleadas en la extensión de una sola base con el kit SNaPshot de 32 Y-SNPs. La tercera columna indica la posición de anillamiento de la sonda, Fwd (Forward) o Rev (Reverse). La cuarta columna indica los alelos interrogados por el ensayo. La quinta columna, indica la concentración final de la sonda en el respectivo multiplex de la sexta columna.	52
Tabla 7. Índices moleculares de diversidad en población Boliviana analizada. N = Tamaño muestral; K = Número de distintos haplotipos; DC = Capacidad de discriminación; UH = número de haplotipos únicos; HD = Diversidad Haplotípica; h = diversidad génica entre los <i>loci</i> ; M = promedio de las diferencias <i>pairwise</i>	58
Tabla 8. AMOVA estimada para las distintas poblaciones Bolivianas de América del sur. El análisis se estimó usando el grupo de haplotipos por <i>Yfiler</i> (17 YSTRs) como con el haplotipo mínimo (9 YSTRs). Diferentes análisis de AMOVA fueron estimados (i) Bolivianos analizados en el presente estudio agrupados en las 3 regiones eco-geográficas (Andina, Sub-Andina y Llanos), (ii) Bolivianos analizados en el presente estudio agrupados por los 5 departamentos (ver Figura 1) y (iii) Bolivianos analizados en el presente estudio agrupados por los 5 departamentos más otras 3 poblaciones Bolivianas disponibles en la literatura, sumando en total 8 poblaciones (descritas a continuación). La columna “Pop” Indica el número de grupos poblacionales considerados, N se refiere a número total de la muestra. La significancia estadística fue evaluada usando el procedimiento de permutación implementado en Arlequin (10,000 permutaciones); Todos los valores de P fueron estadísticamente significantes. Los haplotipos de cromosoma Y fueron tomados de la literatura, Gayà-Vidal et al. (2011) (Bolivia: Aymara, $n = 59$; Quechua $n = 55$), Tirado et al. (2009) (Bolivia: departamento de El Beni, $n = 87$), Vullo et al. (2014) (Bolivia: departamento Chuquisaca, $n = 38$, departamento de La Paz, $n = 57$) Romero et al. (2008) (Colombia: Caribe, $n = 305$), Toscanini et al. (2006) (Argentina: Collas, $n = 29$; Toba, $n = 49$), Roby et al. (2009) (Chile: Santiago de Chile, $n = 196$; Iquique, $n = 196$), y Caine et al. (2010) más Schwengber et al. (2009) (South Brazil: Rio Grande do Sul and Santa Catarina, $n = 364$).	61

Tabla 9. Proporciones de ancestralidad generadas para las regiones eco-geográficas de Bolivia. Estas proporciones fueron estimadas usando los conjuntos de Y-SNPs parentales de Simms et al. (Simms <i>et al.</i> 2012).....	64
Tabla 10. Sesgo por género entre los polimorfismos de cromosoma Y, ADNmt, AIM-Indels y el Panel LACE en Bolivia. La tabla muestra los 6 individuos de la region andina tipados para el panel de AIM-Indels y ADNmt por (Taboada-Echalar <i>et al.</i> 2013), para el Panel LACE analizados en (Galanter <i>et al.</i> 2012) y los Y-SNPs analizados en la presente tesis. Los valores de porcentajes de los AIM-Indels, que estimar la componente principal de ancestralidad continental, fueron redondeando al valor promedio.	65
Tabla 11. Comparación de los haplotipos de <i>Yfiler</i> y RM-YSTRs en el conjunto global de 7,784 individuos pertenecientes a 65 poblaciones, agrupadas en grupos regionales.	70
Tabla 12. Características de los haplotipos de RM-YSTRs en 365 muestras pertenecientes al grupo regional Nativo Americano.....	75
Tabla 13. Comparativa de las características haplotípicas para muestras tipadas con <i>Yfiler</i> y RM-YSTRs en el grupo regional Nativo Americano, en 279 individuos pertenecientes a 8 poblaciones.	76
Tabla 14. Número de distintos haplotipos obtenidos con el panel PPY23, en todas poblaciones analizadas en conjunto (Columna General) y por los grupos definidos por residencia continental y ancestralidad continental.....	80
Tabla 15. Número de distintos haplotipos obtenidos con el panel <i>Yfiler</i> , en todas poblaciones analizadas en conjunto (Columna General) y por los grupos definidos por residencia continental y ancestralidad continental.....	81
Tabla 16. Parámetros forenses estimados para Y-STRs de tamaño menor a 220 pb en los paneles <i>Yfiler</i> y PPY23. HD, Diversidad haplotípica; MP, Probabilidad de coincidencia (<i>Match probability</i>); DC, Capacidad de discriminación. ^a Marcadores incluidos DYS456, DYS389I, DYS458, DYS19, DYS393, DYS391, GATAH4, DYS437. ^b Marcadores incluidos DYS576, DYS389I, DYS391, DYS481, DYS570, DYS635, DYS393, DYS458.....	81
Tabla 17. Distancias genéticas <i>pairwise</i> estimadas en valores R_{ST} (debajo de la diagonal) para los paneles <i>Yfiler</i> y PPY23, obtenidos por análisis de AMOVA; Las significancias de las distancias genéticas están arriba de la diagonal y representan valores $p > 10^4$ indicados con (+).....	82
Tabla 18. Distancias genéticas <i>pairwise</i> estimadas en valores R_{ST} (debajo de la diagonal) para el panel PPY23, obtenidos por análisis de AMOVA para las poblaciones Latino-Americanas; Las significancias de las distancias genéticas están arriba de la diagonal y representan valores $p > 10^4$ indicados con (+). ^a Debido a la aproximación numérica el valor R_{ST} era negativo en este caso indicó la distancia con cero.....	85