



**FACULTAD DE MEDICINA  
INSTITUTO DE MEDICINA LEGAL**

**LINAJES DEL CROMOSOMA Y HUMANO:  
APLICACIONES GENÉTICO-POBLACIONALES  
Y FORENSES**



**Tesis doctoral**

**Alejandro J. Blanco Verea**

**2008**

**UNIVERSIDADE DE SANTIAGO DE COMPOSTELA**  
**FACULDADE DE MEDICINA**  
**INSTITUTO DE MEDICINA LEGAL**



**Linajes del cromosoma Y humano: aplicaciones genético-poblacionales  
y forenses**

Memoria que presenta, para optar al grado de doctor

**Alejandro José Blanco Verea**

En Santiago de Compostela, Julio de 2008



El Doctor *D. Ángel Carracedo Álvarez*, Catedrático de Medicina Legal de la Facultad de Medicina de la Universidad de Santiago de Compostela, la Doctora *D<sup>a</sup> María Brión Martínez*, investigadora del Complejo Hospitalario Universitario de Santiago (CHUS).

CERTIFICAN:

Que la presente memoria que lleva por título “**Linajes del cromosoma Y humano: aplicaciones genético-poblacionales y forenses**”, del licenciado en Ciencias Biológicas por la Universidad de Santiago de Compostela *Alejandro José Blanco Vereza*, ha sido realizada bajo nuestra dirección, considerándola en condiciones para optar al 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 30 de julio del 2008.

Fdo.: Prof. Dr. Ángel Carracedo Álvarez

Fdo.: Dra. María Brión Martínez



## Agradecimientos

A la primera persona que quiero mostrarle mis agradecimientos es a Ángel, por darme la oportunidad de formar parte de este gran grupo de investigación y por enseñarme lo que es realmente implicarse en el trabajo.

Agradecer especialmente a María todo el tiempo que ha dedicado en enseñarme, en hacerme ver y entender todo lo que para mí no tenía sentido. Agradecerle también el hecho de estar siempre ahí para resolver mis dudas y sobre todo darle las gracias por tener tanta paciencia conmigo.

A Maviky por el apoyo y ayuda prestados durante todos estos años de trabajo y por sus valiosos consejos en el trabajo de rutina.

Agradecimientos también a Paula, Toño, Vane, Alejandro y Chris porque cada uno de ellos me ha enseñado y ayudado de un modo u otro.

A Meli, Fonde, Nuria, Raquel, Eva, Gloria, Manuel, Francesca, María C, Anita, Cata, María D, Angela y Luisma por echar una mano siempre que se les necesita. Realmente podría hacerle unos agradecimientos particulares a cada uno.

A los más nuevos: Alberto, Rocío, Ana F., María Mayo y Ricardo por esa energía renovadora que transmiten cada día en el laboratorio.

Agradecimientos a Bea S., Ana Fernández, Bea Q. (que también me ayudaron mucho en mis comienzos) y a un gran número de investigadores de otros laboratorios nacionales e internacionales que han pasado por nuestro departamento permitiéndome compartir conocimientos y observar a través de ellos como es la investigación en otros laboratorios.

A mis amigos debo agradecerles que entendieran el poco tiempo que les he podido dedicar últimamente, también el escucharme y ayudarme cuando he tenido algún problema.

A Inma agradecerle el empuje alentador que me ha transmitido en las últimas fases de la elaboración de esta tesis.

A mis tíos Carlos y Esperanza por apoyarme y animarme continuamente durante todo este tiempo.

Agradecerles a mis primos Carlota y Nico que tuvieron tanta paciencia conmigo y el hecho de aguantar mis pesadas charlas.

A mi abuelo agradecerle el mostrarme que el esfuerzo siempre tiene su recompensa.

Y por último a mi madre y a mi padre todo mi agradecimiento por lo que han hecho por mí durante la realización de esta tesis y por estar siempre a mi lado.

**A MIS PADRES**





# Índice

Abreviaturas.....	XIII
<b>I.Introducción.....</b>	<b>19</b>
<b>1. El genoma humano .....</b>	<b>24</b>
<b>2. El cromosoma Y .....</b>	<b>25</b>
2.1. Definición.....	25
2.2. Estructura .....	26
2.3. Herencia.....	30
<b>3. Polimorfismos del cromosoma Y .....</b>	<b>31</b>
3.1. Polimorfismos binarios (SNPs).....	33
3.1.1. Definición .....	33
3.1.2. Búsqueda de SNPs.....	33
3.1.3. Métodos de genotipado de SNPs .....	34
3.1.4. Haplogrupos .....	39
3.1.5. Aplicaciones poblacionales .....	48
3.1.6. Aplicaciones forenses .....	58
3.2. Microsatélites o STRs .....	59
3.2.1. Definición .....	59
3.2.2. Tasas de mutación.....	60
3.2.3. Métodos de detección .....	62
3.2.4. Aplicaciones poblacionales.....	65
3.2.5. Aplicaciones forenses .....	66
<b>II. Justificación y objetivos .....</b>	<b>71</b>
<b>III. Material y métodos.....</b>	<b>77</b>
<b>1. Muestras.....</b>	<b>79</b>

<b>2. Extracción de ADN</b> .....	79
<b>2.1. Extracción a partir de sangre total</b> .....	79
2.1.3. Protocolo Fenol-Cloroformo-Alcohol Isoamílico .....	79
2.1.2. Protocolo resina quelante Chelex <sup>®</sup> 100 Resin.....	81
2.1.3. Protocolo QIAamp <sup>®</sup> DNA blood mini kit de QIAGEN .....	81
<b>2.2. Extracción a partir de mancha de sangre</b> .....	82
2.2.1. Protocolo Fenol-Cloroformo-Alcohol Isoamílico .....	82
2.2.2. Protocolo resina quelante Chelex <sup>®</sup> 100 Resin.....	83
<b>3. Cuantificación</b> .....	83
<b>3.1. Cuantificación fluorimétrica de ADN</b> .....	83
<b>3.2. Cuantificación espectrofotométrica del ADN</b> .....	84
<b>3.3. Quantifiler<sup>™</sup> Human DNA Quantification Kit y</b> <b>Quantifiler<sup>™</sup> Y Human Male DNA Quantification Kit</b> .....	84
<b>4. Amplificación</b> .....	86
<b>5. Purificación</b> .....	86
<b>5.1 Protocolo CENTRICON<sup>®</sup> Millipore</b> .....	86
<b>6. Detección de marcadores de cromosoma Y</b> .....	87
<b>6.1. Detección de STRs</b> .....	87
<b>6.2. Detección de SNPs</b> .....	87
<b>7. Análisis estadísticos</b> .....	88
<b>IV. Resultados</b> .....	91
<b>Bloque I. Dinámica evolutiva de los STRs y SNPs del cromosoma Y</b> .....	93
<b>Artículo 1: “Duplicaciones de los loci P25 y 92R7 específicos del</b> <b>cromosoma Y y sus implicaciones forenses”</b> .....	95
<b>Artículo 2: “Evidencias filogenéticas para múltiples eventos de</b> <b>duplicación independiente en el locus DYS19”</b> .....	107

---

<b>Bloque II. Utilidad del desarrollo de multiplexes en genética poblacional y forense</b> .....	115
<b>Artículo 3:</b> “Análisis jerárquico de 30 SNPs del cromosoma Y en poblaciones europeas” .....	117
<b>Artículo 4:</b> “Multiplex de Y-SNPs para el análisis de los haplogrupos principales útil en la predicción del origen geográfico de linajes masculinos” .....	125
<b>Bloque III. Estudios poblacionales basados en la variabilidad de los polimorfismos del cromosoma Y</b> .....	137
<b>Artículo 5:</b> “El legado genético masculino de El Salvador” .....	139
<b>Artículo 6:</b> “El componente genético masculino de dos poblaciones del Sur-Oeste colombiano” .....	147
<b>Artículo 7:</b> “Linajes del cromosoma Y en población nativa del Sur de América” .....	155
<b>Artículo 8:</b> “Moros y sarracenos en Europa: estima del legado masculino medieval del Norte de África en el Sur de Europa” .....	181
<b>V. Conclusiones</b> .....	193
<b>VI. Bibliografía</b> .....	199
<b>Glosario</b> .....	227



# Abreviaturas



## Abreviaturas

<b>A</b>	Adenina
<b>ADN/DNA</b>	Ácido desoxiribonucleico
<b>ADNmit</b>	Ácido desoxiribonucleico mitocondrial
<b>ACP</b>	Análisis de Componentes Principales
<b>ARN</b>	Ácido ribonucleico
<b>AMOVA</b>	Análisis de la Varianza Molecular
<b>Aprox.</b>	Aproximadamente
<b>ASOs</b>	Allele-Specific Oligonucleotides
<b>bp</b>	Pares de bases
<b>BSA</b>	Bovine Serum Albumina
<b>C</b>	Citosina
<b>CCD</b>	Charge-Coupled Device (sensor-dispositivo de cargas interconectadas)
<b>CE</b>	Electroforesis Capilar
<b>cm/cm<sup>2</sup></b>	Centímetro/centímetro cuadrado
<b>DHPLC</b>	Denaturing High-Performance Liquid Chromatography
<b>dNTP</b>	Desoxinucleótido trifosfato
<b>ddNTP</b>	Didesoxinucleótido trifosfato
<b>EDTA</b>	Ácido Etilen-Diaminotetraacético
<b>FRET</b>	Fluorescence Resonance Energy Transfer
<b>G</b>	Guanina
<b>gr</b>	Gramo
<b>h.</b>	Hora
<b>Hg</b>	Haplogrupo
<b>ISFH</b>	International Society of Forensic Haemogenetics
<b>ISFG</b>	International Society of Forensic Genetics
<b>Kb</b>	Kilobases
<b>L</b>	Litro
<b>LINEs</b>	Long Interspersed Nuclear Elements
<b>m</b>	Metro
<b>M</b>	Molar



<b>Mb</b>	Megabases
<b>MALDITOF</b>	Matrix Assisted Laser Desorption Ionization - Time of Flight
<b>MDS</b>	Multidimensional Scaling
<b>mg</b>	Miligramo
<b>min.</b>	Minuto
<b>ml</b>	Mililitro
<b>mM</b>	Milimolar
<b>MSY</b>	Male-Specific Y
<b>ng</b>	Nanogramo
<b>nm</b>	Nanómetros
<b>NIST</b>	National Institute of Standards and Technology
<b>NRPY</b>	Non-recombining portion Y
<b>NRV</b>	Non-Recombining Y
<b>OLA</b>	Oligonucleotide Ligation
<b>PAR</b>	Pseudoautosomal Region
<b>PCR</b>	Polimerase Chain Reaction
<b>PSV</b>	Paralogous Sequence Variants
<b>RFLPs</b>	Restriction Fragment Length Polymorphism
<b>r.p.m.</b>	Revoluciones por minuto
<b>SAP</b>	Shrimp Alkaline Phosphatase
<b>SDS</b>	Dodecil Sulfato de Sodio
<b>seg.</b>	Segundos
<b>SINEs</b>	Short Interspersed Nuclear Elements
<b>SNPs</b>	Single Nucleotide Polymorphisms
<b>SSCP</b>	Single-Strand Conformation Polymorphism
<b>SSRs</b>	Simple Sequence Repeats
<b>STRs</b>	Short Tandem Repeats
<b>T</b>	Timina
<b>T<sup>a</sup></b>	Temperatura
<b>Taq</b>	<i>Termus aquaticus</i>
<b>TE</b>	TrisHCl + EDTA
<b>TMRCA</b>	Time Most Recent Common Ancestral

<b>U</b>	Unidades
<b>UEPs</b>	Unique Event Polymorphisms
<b>V</b>	Voltios
<b>Vol.</b>	Volumen
<b>VNTR</b>	Variable Number of Tandem Repeat
<b>YAP</b>	Y-chromosome Alu Polymorphism
<b>YCC</b>	Y Chromosome Consortium
<b>y cols.</b>	Y colaboradores
<b>YHRD</b>	Y chromosome Haplotype Referente Database
<b>Y-STR</b>	Repeticiones en tandem cortas del cromosoma Y
<b>µg</b>	Microgramo
<b>µl</b>	Microlitro
<b>µM</b>	Micromolar
<b>°C</b>	Grados centígrados



# I. Introducción



Hoy en día la genética ha tomado una gran relevancia tanto en el mundo científico como en la sociedad actual, a nadie se le escapa la trascendencia que para la humanidad tienen los avances de la genética, pero entender todos los conceptos que conllevan estos avances, así como emplearlos correctamente, provoca en ocasiones debates éticos, religiosos, jurídicos o morales en torno a esta disciplina.

La genética es una ciencia relativamente nueva. Fue *W. Batenson* en 1903 quien bautizó con este nombre a esta ciencia naciente después de redescubrir en 1900 los trabajos de *Gregor Mendel* de 1866. Este monje agustino es considerado el padre de la genética clásica por su trabajo en el que plantea las tres leyes de la herencia mendeliana. Sus conclusiones fueron el fruto del estudio de hibridaciones entre plantas del guisante en un pequeño huerto de 7m x 35m. Mendel ideó una simbología que le permitió representar y entender los mecanismos que hacen posible la transmisión de las características hereditarias de padres a hijos. Otros muchos científicos han contribuido al desarrollo de los principios básicos de las leyes de la herencia generalizando y ampliando los planteamientos mendelianos a un gran número de organismos vivos.

La “Evolución de las especies” fue otra obra realmente innovadora y contemporánea a la de Mendel, escrita por *Charles Darwin* en 1859 y que planteaba el concepto de la evolución de las especies a través de la selección natural. Lo que Darwin no pudo explicar de forma satisfactoria fue la abundante variabilidad genética observable en los seres vivos. Este problema se resolvió, en parte, gracias a los planteamientos de la genética mendeliana, surgiendo así la genética evolutiva que a su vez derivó en la genética poblacional. Las bases teóricas de la genética de poblaciones fueron fijadas por *Ronald Fisher*, *J. B. S. Haldane* y *Sewal Wright* en los años 30, pero fue Wright el que planteó uno de los desarrollos teóricos más importantes en genética de poblaciones y en evolución al proponer un mecanismo denominado deriva genética. Este mecanismo favorece la presencia de determinadas combinaciones génicas por efecto de la selección natural en poblaciones pequeñas. *Theodosius Dobzhansky*, basándose en los datos que había hasta el momento y en sus propias observaciones, publicó un libro titulado “La genética y el origen de las especies” en 1937 que fue clave para la integración de la genética en la teoría de la selección natural y sentó las bases de la moderna teoría de la evolución que actualmente se conoce como “teoría sintética de

*la evolución*”, término acuñado por *Julian Huxley* en 1942 en su obra “*Evolution: The Modern synthesis*”.

La aparición de la genética molecular y la aparición de nuevas técnicas fueron las causas promotoras del desarrollo de una nueva teoría evolutiva: “*la teoría neutra de la selección*” o “*neutralismo*”, propuesta por *Motoo Kimura* a finales de los 60. Según ésta teoría, la selección natural no es el factor más importante para explicar la evolución a nivel del ADN, sino la tasa de mutación y la deriva genética. La aceptación de esta teoría implica que la evolución a nivel molecular tiene unas causas cualitativamente distintas de la evolución a nivel morfológico. Mientras que la primera estaría principalmente determinada por la mutación y deriva, la última lo estaría por la selección natural. Esta teoría deja el debate abierto entre seleccionistas y neutralistas en lo que respecta a la interpretación de los mecanismos evolutivos.

*Francis Crick* y *James Watson*, en 1953, consiguieron integrar todos los datos moleculares disponibles hasta ese momento, sobre todo los datos obtenidos con las técnicas de difracción de rayos X desarrolladas por *Rosalind Franklin* y *Maurice Wilkins*, que fueron los que más les ayudaron a desarrollar un modelo de la estructura del ADN, estableciéndose así como los principales responsables del nacimiento de la genética molecular. A partir de este momento multitud de campos de la ciencia van a incorporar esta nueva rama de la genética.

La ciencia forense es uno de esos campos que ha adoptado a la genética, en su caso, con el fin de identificar a los individuos en base al análisis de los polimorfismos del ADN.

El primer locus de ADN polimórfico fue descubierto por *Wyman* y *White*, en 1980, usando una sonda de ADN arbitraria. De esta manera observaron fragmentos de más de 15 longitudes diferentes en una pequeña muestra de individuos. Más tarde se encontraron otros loci hipervariables en la secuencia del gen de la insulina humana, en el oncogen “*ras*”, en el pseudogen de la zeta-globina y en el gen de la mioglobina por *Weller* y *cols.* en 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 el descubrimiento de los

primeros VNTRs se vio que éstos podían ser aplicados a la medicina forense y sustituir a los marcadores clásicos.

Aunque la Ciencia poseía las herramientas necesarias para el estudio del ADN, su aplicación en la resolución de casos judiciales no se produjo hasta 1985, cuando el Ministerio del Interior Británico solicitó la ayuda de *Alec J. Jeffreys*, profesor de genética de la Univesidad de Leicester, para identificar al agresor sexual y asesino de dos chicas en Narborough (Leicestershire) usando su 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, que permite la amplificación de secuencias específicas del ADN.

La PCR y el descubrimiento de los STRs (Short Tandem Repeats), un nuevo tipo de polimorfismo con altos niveles de diversidad y simplicidad de tipado, permiten llevar a cabo los análisis de identificación con gran rapidez por la posibilidad de amplificar simultáneamente varios STRs en la misma reacción de PCR (PCR Multiplex). En la actualidad, en el campo forense se contempla la posibilidad de uso de otro tipo de polimorfismo simple denominado SNP (Single Nucleotide Polymorphism) con un poder de discriminación inferior al de los STRs, pero que podrían complementar en algunos casos la información proporcionada por los STRs.

Hoy en día la genética poblacional también recurre al uso de estos polimorfismos para poder observar la diversidad genética en las poblaciones y como éstas han ido cambiando a través del tiempo. Además, debe acudir a los conocimientos que otras disciplinas de la ciencia le puedan brindar, para conseguir así la mejor interpretación posible de los cambios evolutivos.



Igual que sucede en cualquier otro área de la ciencia, en la genética forense, poblacional o clínica, el avance tecnológico no se detiene, surgiendo constantemente nuevas técnicas que generan un gran número de resultados en el menor tiempo posible, permitiéndonos hacer nuevos descubrimientos y facilitando todavía más, el progreso y evolución de estos campos.

## 1. El genoma humano

Es en el genoma del *Homo sapiens*, donde se encuentra toda su información hereditaria codificada en el ADN. Está constituido por aproximadamente 3.000 millones de pares de bases de ADN. Se estima que contiene entre 20.000-25.000 genes que codifican para proteínas y que constituyen alrededor del 1,5% de la longitud total del genoma (Venter y cols., 2001). El resto está formado por genes de ARN que no se traducen a ninguna proteína, secuencias reguladoras, intrones, pseudogenes, elementos repetitivos, transposones y una gran cantidad de secuencia de ADN que no se incluye bajo ninguna clasificación y que si tiene alguna función de momento es desconocida.

Deberíamos incluir como parte del genoma humano, además del ADN nuclear, el ADN de la mitocondria.

La mitocondria es un orgánulo citoplasmático encargado de producir energía para la célula en forma de ATP mediante un proceso de fosforilación oxidativa. En lo que nos concierne, el aspecto más interesante de la mitocondria es que contiene su propio ADN que consiste en una molécula circular de doble cadena de alrededor de 16,5 kb de longitud, de la cual se conoce la secuencia completa (Anderson y cols., 1981; Andrews y cols., 1999).

El número de mitocondrias varía en función del tipo de célula, así tendremos células nerviosas o musculares con miles de mitocondrias en su interior y otras con sólo unos pocos cientos. Cada oocito contiene alrededor de unas 100.000 mitocondrias con su ADNmit y cada espermatozoide sobre unas 50-75 mitocondrias, pero en el momento de la formación del cigoto, la mayoría de las evidencias sugieren que, ninguna de las mitocondrias paternas entra a formar parte de dicho cigoto. Por lo tanto llegamos a la conclusión de que el ADNmit tiene una herencia exclusivamente materna y esto unido a

otras características que se mencionan a continuación, lo hace especialmente útil para explorar el pasado de los humanos:

- Su alta tasa de mutación hace que sea más sencillo el estudio de la diversidad en muestras poblacionales.
- La ausencia de recombinación y los efectos de la deriva genética, generan estructuración geográfica.
- El elevado número de copias de ADNmit por célula facilita el análisis de muestras degradadas y de ADN antiguo.
- Su herencia exclusivamente materna y su ausencia de recombinación permite el estudio de linajes maternos.

Por otro lado, en el núcleo celular el ADN está empaquetado en cromosomas. En cada individuo existen dos copias de cada tipo de cromosoma, uno heredado de la madre y otro heredado del padre. En total, 46 cromosomas que forman 23 parejas, de las cuales 22 están constituidas por cromosomas autosómicos y la otra pareja es la formada por los cromosomas sexuales X e Y. Esta tesis se centrará en algunas de las características que presenta el cromosoma Y sobre todo para estudios poblacionales y forenses.

## **2. El cromosoma Y**

### **2.1. Definición**

En los mamíferos y por extensión en los humanos, el sexo cromosómico queda determinado en el momento en que un óvulo es fecundado por un espermatozoide portador de un cromosoma X (mujer-sexo homogamético XX) o por un espermatozoide portador de un cromosoma Y (hombre-sexo heterogamético XY). El cromosoma Y porta el gen SRY que genera el progreso fenotípico a hombre durante la gametogénesis, ya que este gen induce la expresión de una cascada de genes, cuyo resultado final será la diferenciación de la gónada primitiva en testículo (Koopman, 1999). Si el gen SRY no está presente, el individuo se desarrollará como mujer, por un mecanismo que parece

estar predeterminado en las células gonadales (Capel, 2000), demostrando así que el papel principal del cromosoma Y es el de la determinación del sexo.

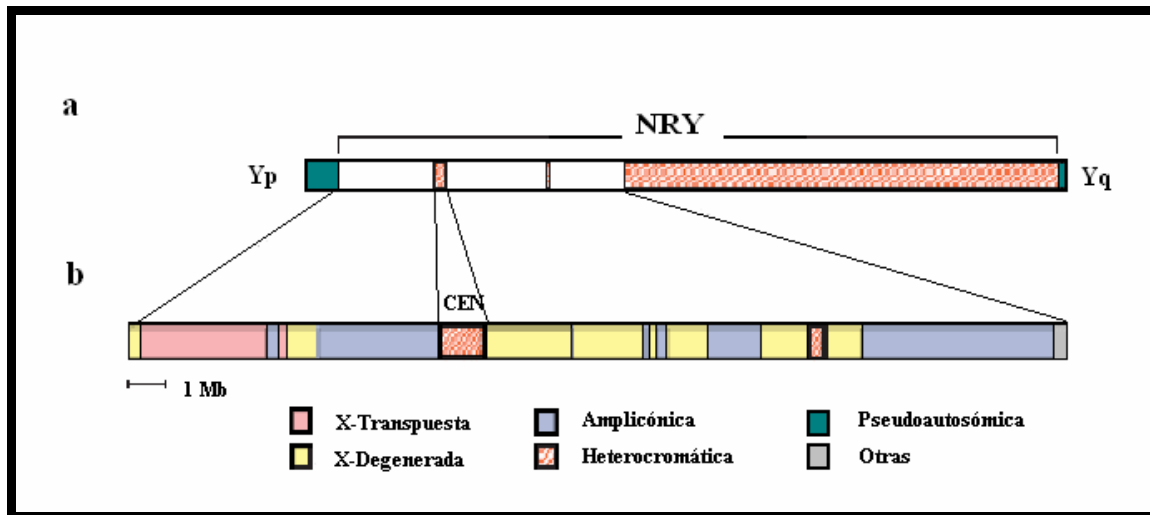
Se cree que los cromosomas sexuales surgieron de un par de autosomas homólogos y la diferenciación entre el cromosoma X e Y comenzaría en el momento que se empieza a reducir la recombinación entre ellos, y esta supresión de la recombinación parece que se produjo en 4 eventos de inversión cromosómica diferentes en la evolución. Estos eventos se originaron en el cromosoma Y, sin embargo, no fue así en el cromosoma X. Cada evento iría suprimiendo la recombinación entre X-Y provocando la diferenciación entre ambos. El primero de los eventos se cree que ocurrió hace aproximadamente 240-320 millones de años, poco después de la divergencia entre el linaje de la aves y el linaje de los mamíferos.

Los eventos segundo (80-130 millones años) y tercero (130-170 millones de años) corresponderían al momento en el que la región pseudoautosómica es ampliada por translocación de material autosómico. El cuarto evento (30-50 millones de años) se situaría durante la evolución de los primates, en el que la diferenciación entre X-Y se encontraría en la etapa más reciente (Lahn y Page, 1999).

## **2.2. Estructura**

El cromosoma Y contiene alrededor de unas 60 Mb y estructuralmente esta constituido por dos brazos: el brazo largo (Yq) y el brazo corto (Yp), ambos separados por el centrómero. Desde un punto de vista más molecular tenemos que mencionar una extensa región, de aproximadamente el 95% de su tamaño, en la que no encontramos recombinación entre X-Y. Esta región tiene distintos nombres: región no recombinante del cromosoma Y (NRY), porción no recombinante del cromosoma Y (NRPY) o región masculina específica (MSY). Sólo en los extremos del cromosoma Y encontramos unas pequeñas regiones pseudoautosómicas en las que sí es frecuente y normal la recombinación entre X-Y (ver figura 1 y 2). La región pseudoautosómica localizada en el extremo del brazo corto se denomina PAR1 y tiene un tamaño de 2.6 Mb. Esta misma presenta una alta frecuencia de recombinación necesaria para la segregación normal de los cromosomas en la meiosis (Ellis y Goodfellow, 1989). La región pseudoautosómica localizada en el brazo largo se denomina PAR2, su tamaño es de 320 Kb y no siempre

experimenta recombinación con el cromosoma X durante la meiosis (Freije y cols, 1992).



**Figura 1.** Región masculino-específica del cromosoma Y. **a**, Representación esquemática del cromosoma Y entero, incluyendo las regiones pseudoautosómicas y heterocromáticas. **b**, Visión aumentada de una porción de 24 Mb del NRY, desde el extremo proximal de la región pseudoautosómica Yp al extremo proximal de la gran región heterocromática de Yq. Se muestran tres clases de secuencias euromáticas, así como secuencias heterocromáticas (Skaletsky y cols., 2003).

Podemos dividir la NRY en dos porciones: una heterocromática de aprox. 40 Mb (forma condensada no activa o transcripcionalmente inactiva) constituida por secuencias altamente repetidas y otra euromática de unas 23 Mb (entramado más delicado, es la región de ADN que va a ser transcrita); a su vez la región euromática podemos subdividirla en tres clases en función del tipo de secuencia:

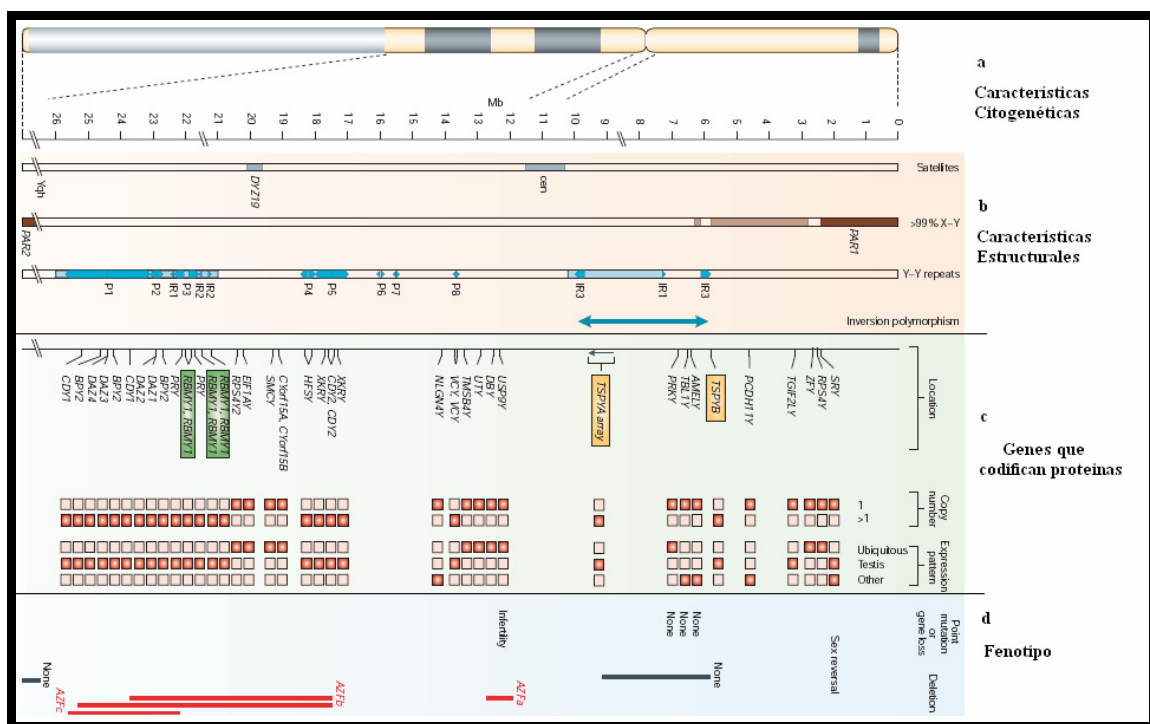
- X-transpuesta: tiene un tamaño aproximado de 3.4 Mb y presenta un 99% de similitud con el cromosoma X. Hay pocos genes funcionales en esta zona y se cree que son genes intercambiados conjuntamente con el cromosoma X hace entre 3 y 4 millones de años, cuando se produjo la divergencia en líneas separadas entre humanos y chimpancés.
- X-degenerada: son secuencias antiguas que corresponden con el momento en el que cromosoma X e Y evolucionaron por primera vez a partir de un cromosoma simple o autosómico. Los genes dentro de estas

secuencias muestran signos de decaimiento constante por mutaciones, dejando muchos de ellos de ser funcionales. Es aquí donde se encuentra el gen SRY, que es un factor de transcripción encargado de determinar la diferenciación de la gónada primitiva en testículo, mediante la activación de otros genes.

- Amplicónica: el tamaño es de aproximadamente 10,2 Mb, posee 8 grandes palíndromos que se repiten y además tiene la mayor densidad de genes de toda la región eucromática del cromosoma Y. Aquí también encontramos nueve familias de genes que codifican para proteínas que se expresan, todas o casi todas, en los testículos. Después de la secuenciación del cromosoma Y en el 2003 se ha visto en la región MSY 156 unidades transcripcionales localizadas en la región eucromática, parece, al menos de momento, que no hay evidencias de estas unidades transcripcionales en la región heterocromática de la región MSY (Skaletsky y cols., 2003).

Según Rozen y cols., 2003 y Skaletsky y cols., 2003, muchos de los genes del cromosoma Y (ver figura 2), y prácticamente todos los de la región amplicónica, se encuentran por pares. Estos pares de genes se pueden intercambiar entre sí, con su compañero correspondiente en un palíndromo idéntico, por un proceso de conversión génica. Parece que este intercambio de copias múltiples del mismo gen dentro de la propia estructura del cromosoma Y es una especie de mecanismo para proteger su integridad genética. Pero no se puede confundir con la recombinación, que no actúa en el cromosoma Y salvo en los extremos mencionados anteriormente. Esta ausencia de recombinación va a permitir la acumulación de mutaciones deletéreas de modo irreversible mediante el proceso denominado *Muller's Ratchet*, descrito por Hermann Joseph Muller (1964) para explicar el éxito de reproducción sexual sobre la asexual (Crow, 2006), que inicialmente parecería más ventajosa ya que no requiere formación de gametos, ni fertilización y el proceso de reproducción es más rápido y necesita menos energía. No obstante, el genoma de la progenie es exactamente igual al del progenitor salvo las mutaciones que se puedan producir en la progenie, haciendo muy

susceptible a la población a cambios ambientales debido a la ausencia de variabilidad. Esto no ocurre en la reproducción sexual, donde la progenie tiene un genoma distinto al de sus progenitores debido a la recombinación. Aún así, los individuos que se reproducen de forma sexual también experimentan el fenómeno de Muller's Ratchet en regiones del genoma, donde no se produce recombinación, como es el caso del cromosoma Y.



**Figura 2.** Genes del cromosoma Y. **a**, Características citogenéticas del cromosoma y su localización aproximada, las cuales están numeradas desde el telómero Yp. **b**, Características estructurales que incluyen 3 regiones satélite (cen, DYZ 19 y Yqh), segmentos de identificación X-Y (PAR1 y PAR2, coloreados en marrón) y de alta similitud (coloreado en marrón claro) y secuencias repetidas Y-Y (IR = repeticiones invertidas, P = palíndromos). **c**, Localización de 27 genes distintos que codifican proteínas específicas del Y, algunas están presentes en más de una copia. **d**, Fenotipo que está asociado con la inactivación o pérdida de genes, algunas eliminaciones producen un fenotipo no detectable (negro) y representan polimorfismos en la población, mientras que otras resultan en infertilidad (AZFa, AZFb y AZFc) (rojo), aunque la contribución de los genes eliminados es poco clara (Jobling y Tyler-Smith, 2003).

### 2.3. Herencia

Una de las características más interesantes del cromosoma Y es el modo en que va pasando de generación en generación. Únicamente pasa de padre a hijo varón siendo, por tanto, específico del género masculino. La ausencia de recombinación en la práctica totalidad del cromosoma le permite pasar de generación en generación como un bloque haplotípico, esto nos lleva a tener una estructura idéntica que pasa de una generación a la siguiente. La única fuerza que favorece la diversificación haplotípica de este cromosoma sería la mutación, siendo este un proceso exclusivamente intra-alélico. Además, si usamos polimorfismos con una muy baja tasa de mutación, como pueden ser los SNPs, podremos construir una filogenia fácilmente.

Hay que resaltar, debido a esta forma de herencia, que el *tamaño efectivo de población* del cromosoma Y es cuatro veces inferior al de cualquier cromosoma autosómico y tres veces más pequeño que el del cromosoma X. La razón de esto radica en el hecho de que cada descendiente de una pareja puede heredar 4 posibles cromosomas autosómicos (dos del padre y dos de la madre) y 3 posibles cromosomas X (dos de la madre y uno del padre), mientras que para el cromosoma Y sólo hay una posibilidad (el cromosoma Y del padre), en caso de tener un hijo varón (Jobling y cols., 1995).

Este reducido tamaño efectivo de población del cromosoma Y lo hace más susceptible a experimentar los efectos de la deriva genética, que provoca cambios aleatorios en las frecuencias de los haplotipos de una generación a la siguiente, acelerando de este modo la diferenciación de grupos de cromosomas Y entre diferentes poblaciones. Los efectos de la deriva genética convierten a este cromosoma en una interesante herramienta para investigar eventos poblacionales ocurridos en el pasado, aunque en ocasiones, la deriva puede producir cambios muy rápidos en las frecuencias haplotípicas dificultando la interpretación de ciertos aspectos cuantificables de estos eventos pasados, como por ejemplo, proporciones de mezcla entre poblaciones o cambios demográficos ocurridos en el pasado (Jobling y Tyler-Smith, 2003).

Además, los patrones de herencia del cromosoma Y pueden influir en la diversidad genética en función de la conducta migratoria de sus portadores. Es obvio el diferente comportamiento de los sexos en la migración. Una hipótesis intuitiva nos hace

pensar que los hombres han migrado distancias mayores que las mujeres en el mundo moderno; ejemplos de migraciones intercontinentales, además de documentos históricos recientes, muestran que exploradores, comerciantes y soldados han sido casi exclusivamente hombres. Sin embargo, si consideramos el impacto de la migración en la diversidad genética, no sólo debemos de examinar los patrones de migración a grandes distancias, sino también, las migraciones locales a pequeña escala.

Cuando hablamos de migraciones locales, son críticos los dos casos de patrones de residencia marital conocidos:

**Patrilocalidad:** es el fenómeno por el cual la residencia de la unidad familiar se ubica cerca o en la misma localidad que la familia del hombre.

**Matrilocalidad:** este fenómeno es a la inversa, el hombre se mudaría a la localidad de la familia de la mujer.

Se estima que en aproximadamente el 70% de la población mundial se da la patrilocalidad (Burton y cols., 1996). Esto implica que el ADNmit (de herencia exclusivamente materna) es el que se desplaza principalmente entre poblaciones en cada generación, mientras que el cromosoma Y permanecería dentro de cada población. Hay que tener en cuenta que la migración conlleva una reducción de la diferenciación poblacional y según algunos trabajos (Seieldstad y cols., 1997; Kayser y cols., 2001a) parece que el cromosoma Y presenta un mayor grado de diferenciación genética entre poblaciones que el ADNmit, lo que sugiere una mayor tasa de migración femenina en el pasado. Estos trabajos demuestran que estas pequeñas migraciones tienen más peso, desde el punto de vista de los patrones geográficos de la diversidad genética, que los procesos migratorios que implican grandes distancias. Estas características migratorias unido al reducido tamaño efectivo de población del cromosoma Y ayudan a interpretar la baja variabilidad dentro de las poblaciones que encontramos para este cromosoma.

### 3. Polimorfismos del cromosoma Y

Los genomas de las células eucariotas están llenos de secuencias de ADN repetidas. Estas secuencias de ADN repetidas tienen todo tipo de tamaños y se diferencian por la longitud de la unidad que se repite y por el número de veces que se



puede repetir una unidad a continuación de la anterior entre distintos individuos. Por lo tanto, nos podemos encontrar con unidades de repetición de gran tamaño (de varios cientos a varios miles de bases) como son el *ADN satélite*, que se encuentra en zonas próximas a los centrómeros cromosómicos, mientras que por otro lado están las repeticiones de tamaño medio (10-100 bases) denominadas *ADN minisatélites* o *VNTR* y finalmente, tenemos las unidades repetitivas de pequeño tamaño (2-6 bases de longitud) llamadas *ADN microsatélites*, *SSRs (Simple Sequence repeats)* o *STRs (Short Tandem Repeats)*. Además de este tipo de ADN repetitivo en tándem, existe otro tipo de ADN repetitivo disperso donde las unidades repetitivas están repartidas por todo el genoma (Smith y cols., 1987).

- **SINEs (Short Interspersed Nuclear elements):** el tamaño de su unidad de repetición está entre 100 y 500 bases. Los elementos más abundantes en el genoma humano son los Alu, de los que hay alrededor de un millón de copias. Los elementos SINEs constituyen el 13% del genoma humano y se cree que sus elementos repetitivos dispersos se generan por conversión génica intracromosómica.
- **LINEs (Long Interspersed Nuclear elements):** el tamaño de su unidad repetitiva suele ser mayor de 500 bases. Se estima que en el genoma hay unos 900.000 LINEs que constituyen el 23% del genoma humano y la mayoría pertenecen a la familia LINE-1 (L1), de la cual sólo 50 elementos son genes funcionales. Se cree que se generan sus elementos repetitivos por conversión génica intercromosómica (Limita su posible presencia en el cromosoma Y).

Otro tipo de variabilidad es la generada por mutaciones puntuales que implican el cambio de un nucleótido por otro (SNPs) y la inserción-delección de uno o varios nucleótidos (Indels).

Todo este tipo de variabilidad se convierte en polimorfismos en el momento en que uno de los 2 posibles alelos (en algunas ocasiones puede haber más de dos posibilidades alélicas) se encuentre en una frecuencia del 1% o más en la población.

### **3.1. Polimorfismos binarios (SNPs)**

#### **3.1.1. Definición**

Es un tipo de polimorfismo simple basado en el cambio de una base nucleotídica por otra. Estos marcadores, algunas veces son referidos como polimorfismos de evento único (UEPs) porque tienen una tasa de mutación muy baja, aprox.  $10^{-8}$  mutaciones/generación ( $10^{-3}$  aprox. en STRs) (Nachman y cols., 2000; Thomson y cols., 2000; Kayser y cols., 2000). Además de los SNPs, otro tipo importante de marcadores binarios son los Indels, que consisten en inserciones/delecciones de un nucleótido o varios. Básicamente los procesos que generan este tipo de polimorfismo son: errores de incorporación de nucleótidos durante la replicación y, mutagénesis causada por la modificación química de las bases o por daños físicos producidos por radiación ionizante o ultravioleta. Se estima que aparecen con una frecuencia de 1 cada 500-100 pares bases en el genoma humano y la mayoría se encuentran en regiones no codificantes, que no tienen ningún efecto conocido sobre el fenotipo del individuo. Estos SNPs son muy útiles como marcadores en genética poblacional y en estudios evolutivos, pero también pueden servir como marcadores genéticos para identificar enfermedades genéticas: mediante estudios de ligamiento en familias, observando el desequilibrio de ligamiento en poblaciones aisladas, estudios de pérdida de heterocigosidad en tumores y por análisis de asociación entre pacientes y controles.

A pesar de que los SNPs, individualmente, son menos informativos que otros marcadores genéticos usados frecuentemente, presentan la ventaja de la mayor abundancia y su gran potencial para la automatización.

#### **3.1.2. Búsqueda de SNPs**

Teniendo en cuenta la gran cantidad de SNPs que han sido descritos en la actualidad, lo mejor sería comenzar la búsqueda del SNP o SNPs a estudiar en alguna de las bases de datos existentes como el NCBI (National Center for Biotechnology Information) dbSNP <http://www.ncbi.nlm.nih.gov/projects/SNP/index.html>. Los SNPs

de esta base de datos fueron descubiertos en parte por la secuenciación de librerías de clones BAC (cromosomas artificiales de bacterias) y por el SNP Consortium. Otro ejemplo de base de datos de SNPs es ALFRED database <http://alfred.med.yale.edu/alfred/> con un fin más antropológico (Osler y cols., 2002). El problema de estas bases de datos es que no encontramos la misma información para todos los SNPs y en algunos casos ni tan siquiera se puede confirmar la existencia de ciertos SNPs.

Para la búsqueda de nuevos SNPs todavía no descritos recurrimos a DHPLC (denaturing high-performance liquid chromatography) (Underhill y cols., 1997), a SSCP (Single-strand conformation polymorphism) (Zerjal y cols., 1997) y, actualmente, se recurre a la secuenciación automática de alto rendimiento del ADN basada en el método de Sanger cuyo principio clave es el uso de didesoxinucleótidos trifosfato (ddNTPs).

Mediante DHPLC (Xiao y cols., 2001), después de amplificar el fragmento, en el que se encuentra el SNP a estudiar, este se desnaturaliza y renaturaliza lentamente, permitiendo así la formación de cadenas homoduplex (unión correcta de las dos cadenas de ADN) y cadenas heteroduplex (unión incorrecta entre las cadenas -una de las cadenas correspondería a un homoduplex con el alelo salvaje y la otra, a una de las cadenas del homoduplex con el alelo mutado-). Permitiendo observar las mutaciones por retención diferencial de homoduplex y heteroduplex de DNA en un soporte cromatográfico de fase reversa bajo desnaturalización parcial.

En SSCP el proceso inicial es parecido a DHPLC, pero en lugar de un cromatógrafo usamos un gel de Poliacrilamida desnaturalizante en el que vamos a cargar los heteroduplex y homoduplex renaturalizados. Observaremos distintos patrones de migración en la electroforesis debido a la conformación que adopten nuestros fragmentos en el gel.

### **3.1.3. Métodos de genotipado de SNPs**

La rapidez a la que se descubren nuevos SNPs y la necesidad de estudiarlos en un gran número de muestras lleva asociada la búsqueda de nuevas tecnologías de alto rendimiento para el genotipado.

Para llevar a cabo el genotipado es interesante diferenciar el método de detección del principio de reacción.

El *principio de reacción* se basa en la propiedad bioquímica del método de genotipado y básicamente podemos hablar de 6 principios diferentes (ver figura 3):

#### ***Enzimas de restricción (RFLPs)***

Las endonucleasas de restricción se usan para producir cortes alelo específicos en un punto del ADN, en algunos casos, un SNP puede alterar la secuencia que es reconocida por un enzima de restricción, por tanto, tendremos moléculas de ADN con su secuencia inalterada que serán cortadas, mientras que las moléculas de ADN que presenten la variante del SNP no serán cortadas y viceversa.

#### ***Hibridación con sondas oligonucleotídicas alelo específicas (ASOs)***

Diseñamos dos sondas, cada una contiene en el medio de la secuencia uno de los nucleótidos complementarios a la variante alélica del SNP. Sólo en el caso de que se produzca la unión correcta de la sonda con el nucleótido correspondiente a la variante alélica, generará un híbrido estable.

#### ***Extensión de primer alelo específica***

Se basa en el diseño de dos primers que se unirán a la secuencia inmediatamente adyacente al SNP; cada uno de los primers tiene en su extremo 3' uno de los nucleótidos complementarios a la variante alélica. Sólo cuando se produzca una unión correcta en el extremo 3' del primer con el SNP, se llevará a cabo la extensión por la polimerasa que cataliza la reacción.

#### ***Minisequenciación***

Ha sido la técnica escogida en esta tesis para llevar a cabo el estudio de SNPs (Syvänen y cols., 1993 y Pastinen y cols., 1997). Es un proceso similar a la extensión de primer alelo específico pero con pequeños matices. La principal diferencia es que las sondas diseñadas no llevan la posible variante alélica en el extremo 3', por lo que el extremo 3' de la sonda acaba justo en el nucleótido situado antes del SNP. Ahora bien,

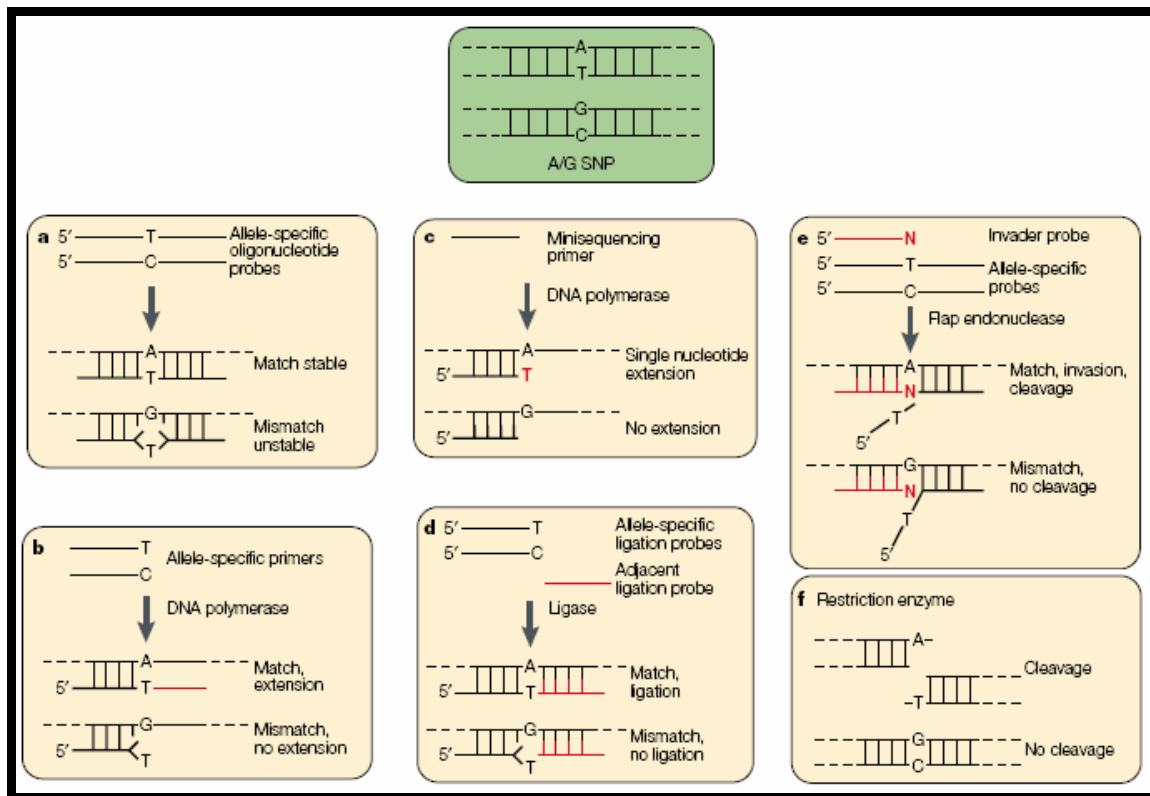
cuando se lleva a cabo la extensión por la polimerasa, lo que tenemos son unos didesoxinucleótidos marcados que se unirán en función de la variante alélica presente. El término didesoxi indica que la ribosa que forma parte del nucleótido carece de -OH en los carbonos 2' y 3'. Sin el 3'-OH un didesoxinucleótido no puede intervenir en la continuación de la polimerización ya que queda impedida la unión con el siguiente didesoxinucleótido o desoxinucleótido. Así, tendremos sondas de un tamaño determinado y con distinta fluorescencia en función del didesoxinucleótido añadido.

### ***“Oligonucleotide Ligation” (OLA)***

Necesitamos crear un par de sondas alelo específicas que presentarán en uno de sus extremos (en el 3' o 5' dependiendo como se diseñen) un nucleótido u otro, en función de la variante alélica. También se diseña una sonda que se unirá a la secuencia justo después del SNP, que mediante una ligasa se unirá a las sondas alelo específicas sólo cuando se produzca la unión correcta entre el nucleótido del extremo de la sonda y la variante alélica.

### ***“Invasive cleavage”***

Diseñamos dos sondas alelo específicas en las cuales a partir del extremo 5' del SNP la secuencia de la sonda no coincide con la secuencia blanco. También se diseña una “sonda invasora” que sí es complementaria a la secuencia 5' del SNP. Cuando la sonda alelo específica se une con la variante correcta del SNP, entran en acción unas enzimas denominadas FLAP endonucleasas que cortan la sonda alelo específica por la zona donde está la variante alélica en 5', y se une a continuación de la sonda alelo específica la sonda invasiva.



**Figura 3.** Los distintos paneles ilustran la detección del alelo A de una transición A-G (Syvänen, 2003). El alelo G sería detectado análogamente en una reacción paralela. **a**, Hibridación con oligonucleótido alelo-específico (ASOs). **b**, Extensión de primer alelo específica. **c**, minisequenciación. **d**, Unión de oligonucleótido. **e**, Corte invasivo. **f**, Corte enzimático (RFLPs).

Para poder analizar el resultado de estos principios de reacción recurrimos a una serie de *métodos de detección* que se basan en aprovechar las propiedades físico-químicas del producto que queremos genotipar, entre estos métodos cabe destacar:

### Métodos de detección por fluorescencia

Se basa en la propiedad de ciertas moléculas que, al ser irradiadas con energía electromagnética de longitud de onda adecuada, emiten radiación de longitud de onda característica permitiendo su cuantificación. Las moléculas de un colorante fluorescente absorben luz en una longitud de onda y convierten la luz absorbida de alta energía (longitud de onda más corta) en luz de menos energía (longitud de onda más larga). Cada fluorocromo tiene un espectro de emisión y excitación característico; si se utilizan 4 con el mismo espectro de excitación pero distinto espectro de emisión, se pueden medir 4 características al mismo tiempo (fluorescencia de 4 colores, una para cada base nucleotídica, por ejemplo). Este método de detección se suele usar para multiplex de

minisequenciación, primer extensión, en arrays de alta densidad de oligonucleotidos y en 'tag' microarrays.

**Método de detección por espectrometría de masas MALDI-TOF** (*Matrix Assisted Laser Desorption Ionization-Time Of Flight*)

Es un sistema de identificación de macromoléculas en base a su masa molecular. En el estudio de los SNPs se usa para detectar productos de “*primer extensión*” (Paracchini y cols., 2002). Estos productos de extensión se inmovilizan sobre un chip y se introducen en un espectrómetro de masas, donde son volatilizados mediante un láser (MALDI), el producto es expelido al tubo de vuelo donde es sometido a un campo eléctrico que lo ioniza y acelera lanzándolo contra el detector que registra el impacto.

Los pesos moleculares son calculados en función de la relación que existe entre la masa molecular y el tiempo que le lleva a la muestra atravesar el tubo de vuelo (TOF) que se encuentra en el interior del espectrómetro de masas.

**Métodos de detección por quimioluminiscencia**

En este método, mediante una reacción química previa, generamos un producto que se encuentra en un estado electrónico excitado, y cuando pierde su exceso energético emite luminiscencia. Un ejemplo de este método es la *Pyrosequencing*<sup>®</sup> de la compañía Biotage AB.

**Método de detección por fluorescencia polarizada**

Consiste en la excitación de una molécula fluorescente mediante luz polarizada plana y la medida de la tasa de despolarización de la fluorescencia. Esta tasa es proporcional a la tasa de caída de la fluorescencia de la molécula. Cuanto más pequeña sea la molécula más rápido cae su fluorescencia, así, moléculas de diferentes tamaños pueden ser diferenciadas.

**Método de transferencia de energía por resonancia de fluorescencia (FRET)**

La transferencia de energía por resonancia de fluorescencia (FRET), también llamada transferencia lineal de energía (en inglés, *Fluorescence Resonance Energy*

*Transfer*) es una interacción que ocurre solo a muy corta distancia ( $<100\text{\AA}$ ) entre dos estados de excitación electrónica de dos moléculas fluorescentes, en las que la longitud de onda de emisión de una de ellas coincide con la de excitación de la otra. Un ejemplo de utilización de este método de detección es TaqMan™ assay (Applied Biosystems).

### 3.1.4. Haplogrupos

El primer marcador bialélico encontrado en el cromosoma Y fue una inserción Alu (DYS287) abreviado YAP (Polimorfismos Alu del cromosoma Y) (Hammer, 1994). En los tres años siguientes no más de una docena de SNPs de la NRY fueron descritos, pero actualmente, debido a modernas técnicas de secuenciación e inicialmente gracias al sistema DHPLC (Underhill y cols., 1997; Shen y cols., 2000; Underhill y cols., 2000; Hammer y cols., 2001) se han descritos varios cientos de SNPs.

La asociación de los polimorfismos binarios con la región no recombinante del cromosoma Y (NRY) nos permite observar el legado genético paterno que se ha preservado de nuestra especie y que ha persistido hasta el presente, permitiendo hacer inferencias de la evolución humana, afinidad poblacional e historia demográfica.

Los terminos haplogrupo y haplotipo, en algunas ocasiones, aparecen en la literatura con definiciones solapadas. Una buena definición es la aportada por Peter de Knijff (2000) en la cual el término haplogrupo se refiere a los linajes de la NRY definidos por polimorfismos binarios, mientras el término haplotipo se reserva para todos los sublinajes de haplogrupos que están definidos por variaciones en los STRs de la NRY.

El constante crecimiento en el número de SNPs y el sistema de nomenclatura casi particular de cada autor para referirse a los haplogrupos del cromosoma Y hacía muy compleja la comparación e intercambio de datos entre laboratorios. Esto llevó en el 2002 a la creación, por parte del YCC (Y Chromosome Consortium) <http://ycc.biosci.arizona.edu/>, de un sistema de nomenclatura para los haplogrupos binarios del árbol del cromosoma Y humano (ver figura 4 y 5). Además de desarrollar una nomenclatura estándar, en este artículo también intentaron mostrar las equivalencias de los haplogrupos de todas las publicaciones anteriores con los haplogrupos correspondientes a la nueva nomenclatura, aunque en muchos casos fue imposible



establecer una correlación debido a la gran inconsistencia que se encontró entre estas primeras nomenclaturas (Su y cols., 1999; Jobling y Tyler-Smith, 2000; Semino y cols., 2000; Underhill y cols., 2000; Capelli y cols., 2001; Hammer y cols., 2001; Karafet y cols., 2001) y la establecida por el YCC.

Hasta esa fecha el número de SNPs del cromosoma Y que habían sido descritos ascendía a 245 y definían un total de 153 haplogrupos.

De los 245 SNPs, 8 eran mutaciones recurrentes, se denominan así porque se producen sobre la variante derivada del SNP y el nucleótido resultante de esa mutación coincide con el que tenía el SNP en su estado ancestral. Estas mutaciones recurrentes se tienen en cuenta como polimorfismos diferentes y fueron observadas en SRY<sub>10831</sub>, 12f2, MSY2, M116, M64, M108, P37 y P41.

La posición de la raíz del árbol filogenético fue determinada mediante la comparación de regiones homólogas de la NRY de especies estrechamente relacionadas (e.g., chimpancés, gorilas y orangutanes) que fueron secuenciadas para determinar el estado ancestral en lugares polimórficos humanos (Hammer y cols., 2001; Underhill y cols., 2001). En el árbol se pueden observar una serie de clados monofiléticos principales, a los que se les asigna una letra mayúscula, que comenzarían con la letra A (sería el primer haplogrupo en el árbol empezando desde arriba después de la raíz en la figura 4) y continuaría por orden alfabético hasta la letra R. En total son 19 letras si incluimos la letra Y, que corresponde al haplogrupo que incluye a todos los haplogrupos desde el A hasta el R. Actualmente el árbol filogenético se ha ampliado enormemente encontrándonos dos clados más que serían el S y el T (Karafet y cols., 2008). Ver figura 5.

Realmente se propusieron 2 nomenclaturas siendo una de ellas opcional (Hammer y Zegura, 2002; YCC, 2002). La primera sería *jerárquica* y se basa en aspectos seleccionados de la teoría para facilitar que los clados monofiléticos puedan ser nombrados sin ambigüedades a todos los niveles. La nomenclatura alternativa nombra el haplogrupo por el nombre de la *mutación* que define al linaje.

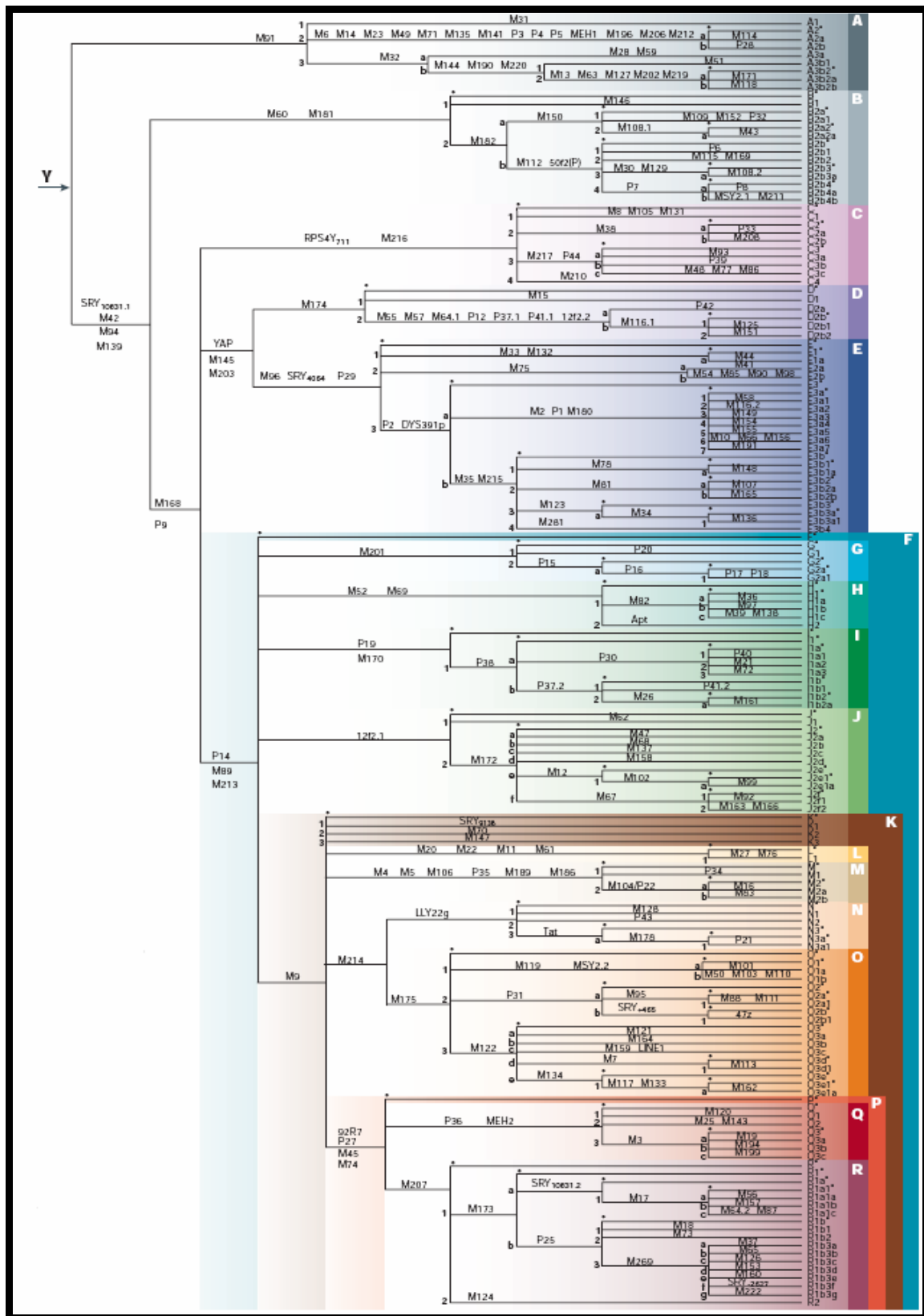


Figura 4. Árbol filogenético de los haplogrupos del cromosoma Y (Jobling y Tyler-Smith, 2003).

En la nomenclatura jerárquica las letras mayúsculas (A-T) se usan para identificar el clado principal y, en los subsecuentes subclados, estas letras mayúsculas constituirían el primer símbolo de la nomenclatura seguidos de un sistema alfanumérico alternativo, hasta llegar a las ramas terminales del árbol.

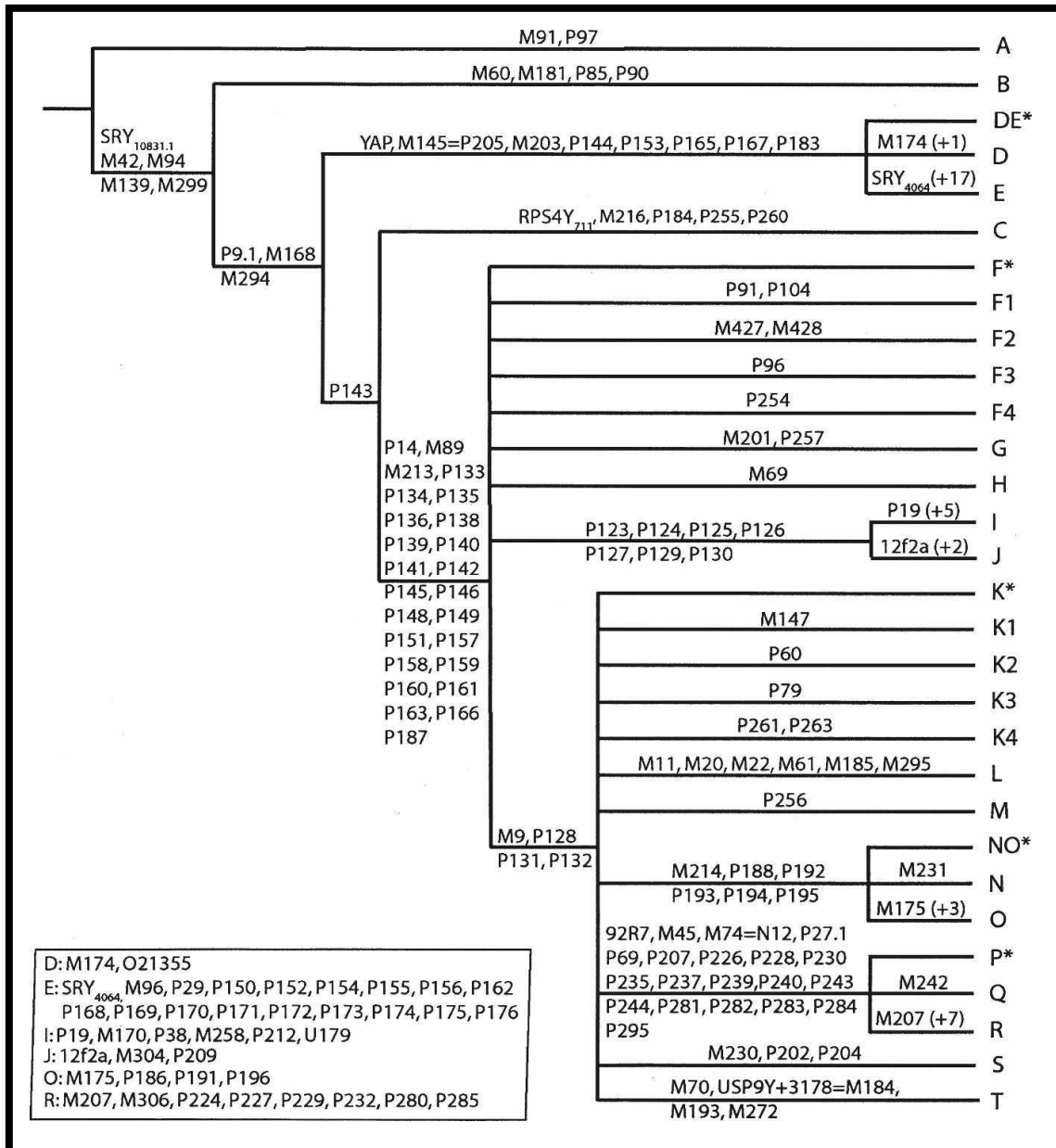


Figura 5. Árbol filogenético con nuevos haplogrupos del cromosoma Y (Karafet y cols., 2008).

En la nomenclatura por mutación primero se pondría la letra mayúscula correspondiente al clado principal, seguida del nombre de la mutación terminal que define al haplogrupo dado y, entre la letra mayúscula y el nombre de la mutación siempre hay que poner un guión (-). Este sistema tiene la propiedad de ser más robusto a cambios en la topología del árbol.

Es importante diferenciar el término *paragrupo* del término haplogrupo. El paragrupo representa a los cromosomas que pertenecen a un clado pero no a sus subclados, es decir serían los cromosomas que comparten el estado derivado de un marcador que define un linaje basal y el estado ancestral de los distintos marcadores descritos hasta el momento que se encuentren situados dentro de ese linaje basal (y que definirían sub-linajes). En la nomenclatura usamos el símbolo asterisco (\*) para referirnos a los paragrupos. Un ejemplo sería el paragrupo Q\* que pertenece al clado Q, sin embargo no pertenece a los haplogrupos Q1, Q2, Q3 (ver figura 4).

Hoy en día, hay aproximadamente 600 SNPs del cromosoma Y descritos y añadidos al árbol filogenético, la mayoría de ellos están representados en la tabla 1. Este aumento en el número de SNPs ha llevado a reajustes a la hora de nombrar ciertos haplogrupos (Karafet y cols., 2008); además, debido al gran número de ramas generadas por los nuevos Y-SNPs, el tamaño del árbol se ha incrementado considerablemente. Como respuesta a todos estos problemas ha surgido una nueva página web interactiva llamada *snpreferencedatabase* <http://www.snp-y.org/>, cuyo principal objetivo es apoyar la validación y aplicación de los Y-SNPs, al mismo tiempo que constituye una colección de marcadores que permitirán la identificación de los principales clusters poblacionales distribuidos por todo el mundo.

SNP	mutación	haplogrupo	SNP	mutación	haplogrupo	SNP	mutación	haplogrupo
M2=SY81	A->G	E3b1a	M215	A->G	E1b1b	P127	C->T	IJ
M3	C->T	Q1a3a	M216	C->T	C	P128	C->T	K-R
M4	A->G	M1	M217	A->C	C3	P129	A->G	IJ
M5=P73	C->T	M1	M218	C->T	B2a1	P130	A->T	IJ
M6	T->C	A2	M219	T->C	A3b2	P131	C->T	K-R
M7	C->G	O3a3b	M220	A->G	A3b	P132	G->T	K-R
M8	G->T	C1	M221	G->A	J2b	P133	G->A	F-R
M9	C->G	K-R	M223	C->T	I2b	P134	C->G	F-R
M10	T->C	E1b1a6	M224	T->C	E1b1b1a1a	P135	C->T	F-R
M11	A->G	L	M226	C->T	S1d	P136	T->G	F-R
M12	G->T	J2b	M227	C->G	I1b	P138	T->C	F-R
M13	G->C	A3b2	M230	A->T	S	P139	G->A	F-R
M14	T->C	A2	M231	G->A	N	P140	G->C	F-R
M15	inserción 9 pb	D1	M236	G->C	B1	P141	G->A	F-R
M16	C->A	M1b1a	M241	G->A	J2b2	P142	G->A	F-R
M17	4G->3G	R1a1	M242	C->T	Q	P143	G->A	CF
M18	inserción 2 pb	R1a1a	M253	C->T	II	P144	G->A	DE
M19	T->A	Q1a3a1	M254	inserción 18pb	S1	P145	A->G	F-R
M20	A->G	L	M258	T->C	I	P146	C->T	F-R
M21	A->T	I1a	M267	T->G	J1	P147	T->A	E1
M22	A->G	L	M269	C->T	R1b1b2	P148	C->T	F-R
M23	A->G	A2	M272	A->G	T	P149	G->A	F-R
M25	G->C	Q1a2	M274	C->T	L2b	P150	C->A	E
M26	G->A	I2a2	M280	G->A	J2b2b	P151	T->C	F-R
M27	C->G	L1	M281	G->A	E1b1b1d	P152	G->C	E
M28	T->G	A3a	M284	ACAA del	I2b1	P153	T->G	D-E
M30	G->A	B2b3	M285	G->C	G1	P154	G->T	E
M31	G->C	A1a	M286	G->A	G2a2	P155	G->A	E
M32	T->C	A3	M287	A->T	G2b	P156	A->G	E
M33	A->C	E1a	M288	C->A	B1	P157	T->C	F-R
M34	G->T	E1b1b1c1	M289	G->A	J2a6	P158	C->T	F-R
M35	G->C	E1b1b1	M290	C->T	E1b1b1c1b	P159	C->A	F-R
M36	T->G	H1a1	M294	C->T	CR	P160	A->C	F-R
M37	C->T	R1b1b2a	M295	T->C	L	P161	G->A	F-R
M38	T->G	C2	M296	C->T	M1	P162	T->C	E
M39	C del	H1a3	M299	T->G	BR	P163	A->T	F-R
M41=P210	G->T	E2a	M300	G->A	O3a5	P164	T->C	O3a3b2
M42	A->T	BR	M304	A->C	J	P165	G->A	DE
M43	A->G	B2a2a	M305	C->T	A3b2	P166	C->T	F-R
M44	G->C	E1a1	M306	C->A	R	P167	G->T	DE
M45	G->A	P	M307	G->A recurrent	I1	P168	G->C	E
M47	G->A	J2a1	M314	A->C	J2b	P169	C->T	E
M48	A->G	C3c	M317	GA del	L2	P170	G->A	E
M49	T->C	A2	M318	T->C	J2a7	P171	G->T	E
M50	T->C	O1a2	M319	T->A	J2a8	P172	C->T	E
M51	G->A	A3b1	M320	T->G	T1	P173	A->G	E
M52	A->C	H1	M321	C->T	J2b2c	P174	G->A	E
M54	G->A	E2b	M322	C->A	J2a1	P175	G->A	E
M55	T->C	D2	M323	C->T	Q1a6	P176	T->G	E
M56	A->T	R1a1a	M324	G->C	O3a	P177	C->T	E1b

SNP	mutación	haplogrupo	SNP	mutación	haplogrupo	SNP	mutación	haplogrupo
M57	+1bp	D2	M327	T->C	J2a2a1	P179	A->C	E1b1
M58	G->A	E1b1a1	M329	G->C	E1b1c	P180	G->A	E1b1
M59	A->C	A3a	M333	G ins	O3a6	P181	C->G	E1b1
M60	+1bp	B	M335	T->A	R1b1c	P182	G->A	E1b1a
M61	C->T	L	M339	T->G	J2a9	P183	C->T	DE
M62	T->C	J1a	M340	G->C	J2a10	P184	T->C	C
M63	G->A	A3b2	M342	C->T	G1	P186	C->A	O
M64.1	A->G recurrent	D2	M343	C->A	R1b	P187	G->T	F-R
M64.2	A->G recurrent	R1a1c	M346	C->G	Q1a3	P188	G->A	NO
M65	A->T	R1b1b2b	M347	A->G	C4	P189	G->A	E1b1a
M66	A->C	E1b1a6	M349	G->T	L2a	P190	A->G	D2
M67	A->T	J2a2	M353	G->A	M2	P191	A->G	O
M68	A->G	J2a3	M356	C->G	C5	P192	A->G	NO
M69	T->C	H	M357	C->A	L3	P193	T->G	NO
M70	A->C	T	M365	A->G	J1b	P194	C->G	NO
M71	C->T	A2	M367	A->G	J1e1	P195	A->G	NO
M72	A->G	I1b1	M368	A->C	J1e1	P196	C->A	O
M73	delección 2pb	R1b1b1	M369	G->C	J1e2	P197	G->T	O3a
M74=N12	G->A	P	M370	C->G	H1b	P19	T->C	O3
M75	G->A	E2	M377	A->G	G2c	P199	A->G	O3a
M76	T->G	L1	M378	A->G	Q1b	P20	T->G	O3a
M77	C->T	C3c	M379	GT del	I2b2	P201=IMS-JST021354	T->C	O3a3
M78	C->T	E1b1b1a	M387	T->G	M2	P202	T->A	S
M81	C->T	E1b1b1b	M390	A ins	J1c	P203	G->A recurrent	O1a1
M82	-2bp	H1a	M407	A->G	C3d	P204	G->T	S
M83	C->T	M1b1b	M410	A->G	J2a	P207	G->A	P
M84	A del	E1b1b1c1a	M419	AAAAG del	J2a11	P209	T->C	J
M85	C->A	E2b1	M427	C->T	F2	P211	T->A	E1b1a
M86	T->G	C3c	M428	G->A	F2	P212	T->A	I
M87	T->C	R1a1c	M450	G->A	I1	P214=M436	G->C	I2b
M88	A->G	O2a1	YAP	Alu- ->Alu+	DE	P215=M438	A->G	I2
M89	C->T	F-R	P1	C->T	E1b1a	P216	C->G	I2b
M90	C->G	E2b	P2	C->T	E1b1	P217	C->T	I2b
M91	9T->8 T	A	P3	G->A	A2	P218	T->G	I2b
M92	T->C	J2a2a	P4	C->T	A2	P219	T->G	I2b
M93	C->T	C3a	P5	C->T	A2	P220	G->T	I2b
M94	C->A	BR	P6	G->C	B2b1	P221	C->A	I2b
M95	C->T	O2a	P7	T->C	B2b4	P223	C->G	I2b
M96	G->C	E	P8	G->A	B2b4a	P224	C->T	R
M97	T->G	H1a2	P9.1	C->A	CR	P225	G->T	R1
M98	G->C	E2b	P9.2	A->C	E1b1a7a1	P226	C->T	P
M99	delección 1pb	J2b2a	P12	A->G	D2a1a1	P227	G->C	R
M101	C->T	O1a1a	P14	C->T	F-R	P228	C->T	P
M102	G->C	J2b	P15	C->T	G2a	P229	G->C	R
M103	C->T	O1a2	P16	A->T	G2a1	P230	G->A	P
M105	C->T	C1	P18	C->T	G2a1a	P231	A->G	R1
m106	A->G	M1	P19	T->G	I	P232	G->A	R
M107	A->G	E1b1b1b1	P20	C del	G1a	P233	T->G	R1
M108.1	T->C recurrent	B2a2	P21	C->A	N1c1a	P234	T->C	R1
M108.2	T->C recurrent	B2b3a	P22=M104	G/A->A	M1b1	P235	A->G	P
M109	C->T	B2a1a	P25	C->A	R1b1	P236	C->G	R1

SNP	mutación	haplogrupo	SNP	mutación	haplogrupo	SNP	mutación	haplogrupo
M110	T->C	O1a2	P27.1	G/G->G/A	P	P237	A->G	P
M111	TT del	O2a1	P27.2=DYS257	G/G->G/A	O3a1	P238	G->A	R1
M112	G->A	B2b	P28	C->T	A2b	P239	G->C	P
M113	A->G	O3a3b1	P29	A->C	E	P240	T->C	P
M114	T->C	A2a	P30	G->A	I1	P242	G->A	R1
M115	C->T	B2b2	P31	T->C	O2	P243	A->G	P
M116.1	A->T	D2a	P32	-T	B2a1a	P244	G->A	P
M116.2	A->C	E1b1a2	P33	T->C	C2a1	P247	T->A	A2
M117	delección 4pb	O3a3c1	P34	G->A	M1a	P248	G->T	A2
M118	A->T	A3b2b	P35	T->C	M1	P249	G->A	R2
M119	A->C	O1a	P36.1	G->A	A2	P254	C->T	F4
M120	T->C	O1a1	P36.2	G->A	Q1	P255	G->A	C
M121	delección 5pb	O3a1	P37.1	T->C	D2	P256	G->A	M
M122	T->C	O3	P37.2	T->C	I2a	P257	G->A	G
M123	G->A	E1b1b1c	P38	A->C	I	P258	A ins	E2b1a2
M124	C->T	R2	P39	G->A	C3b	P259	T->G	I1d
M125	T->C	D2a1	P40	C->T	I1	P260	A->C	C
M126	delección 4pb	R1b1b2h1	P41.1	T->C	D2	P261	G->A	K4
M127	C->T	A3b2	P41.2=M359	T->C	I2a1	P262	G->C	A2c
M128	-2bp	N1a	P42	G->A	D2a1a	P263	A->C	K4
M129	G->A	B2b3	P43	G->A	N1b	P266	A->T	H2b
M131	delección 9pb	C1	P44	G->A	C3	P267	C->G	R2
M132	G->T	E1a	P45	G->A	E2b1a1	P268	T->A	E1b1a9
M133	T del	O3a3c1	P46	C->T	E1b1a	P269	A->G	E1b1a9
M134	-1bp	O3a3c	P47	C->T	D3a	P277	A->G	E1b1a8a
M135	+1bp	A2	P48	T ins	Q1a4	P278	G-A	E1b1a8a
M136	C->T	E1b1b1c1a	P49	A->T	O2b	P279	G-A	J2a13
M137	T->C	J2a4	P50	G del	B2a1a	P280	A->G	R
M138	C->T	H1a3	P51	T->C	M1a1	P281	A->G	P
M139	5G->4G	BR	P53.1	T->C	C3e	P282	A->G	P
M141	T->A	A2	P53.2	T->C	D2a1b1	P283	C->T	P
M143	G->T	Q1a2	P54	A->G	C2a2	P284	C->G	P
134 M144	T->C	A3b	P55	C->A	C6	P285	C->A	R
M145=P205	G->A	DE	P56	A->G	J1d	P286	C->T	R1
M146	A->C	B1a	P57	T->G	S1a	P287	G->T	G2
M147	T ins	K1	P58	T->C	J1e	P289	C->G	A3b
M148	A->G	E1b1b1a3a	P59	A->G	E1b1a8a2	P291	C->G	A3b1
M149	G->A	E1b1a3	P60	T->C	K2	P292	G ins	Q1a3a3
M150	C->T	B2a	P61	G->A	S1b	P293	T->G	E1b1a
M151	G->A	D2a2	P62	C ins	C3f	P294	G->A	R1
M152	C->T	B2a1a	P63	C ins	N1b1	P295	G->C	P
M153	T->A	R1b1b2c	P65	C->A	E1b1b1a2b	P297	G->C	R1b1b
M154	T->C	E1b1a4	P66	G->A	R1b1b2f	12f2.b	present->ausent	D2
M155	G->A	E1b1a5	P67	T->C	N1c1b	12f2a	present->ausent	J
M156	A->G	E1b1a6	P68	G->A	E2	47z	G->C	O2b1
M157	A->C	R1a1b	P69	T->C	P	50f2(P)	G->C	B2b
M158	G->A	J2a5	P70	G->A	B2b4a	92R7	G->A	P
M159	A->C	O3a3a	P71	T->C	A3b1a	Apt	G->A	H2
M160	A->C	R1b1b2b2	P72	G->A	E1b1b1f	DYS391p	C->G	E1b1
M161	C->A	I2a2a	P75	G->A	E1b2	IMS-JST002611	C->T	O3a4
152 M162	C->C/T	O3a3c1a	P76	G->C	G1b	IMS-JST021355	A->G	D

SNP	mutación	haplogrupo	SNP	mutación	haplogrupo	SNP	mutación	haplogrupo
M163	A->C	J2a2b	P77	G->A	T2	IMS-JST022454	T->G	O2b
M164	T->C	O3a2	P78	C->T	I2b3	IMS-JST022457	C->G	D2a1b
M165	A->G	E1b1b1b2	P79	T->C	K3	LLY22g	C/C->C/A	N1
M166	G->A	J2a2b	P80	G->C	H2a	MEH1	C->G	A2
M168	C->T	CR	P81	C->T	J2a12	MEH2	G->T	Q1a
M169	T->C	B2b2	P82	G->T	A1a	MSY2.1	4->3	B2b4b
M170	A->C	I	P83	C->T	S1c	MSY2.2	4->3	O1
M171	G->C	A3b2a	P84	C->T	J2b2d	N1	C->G	D1a
M172	T->G	J2	P85	T->C	B	N2	T->C	D1a1
M173=P241	A->C	R1	P87	A->C	M1b	N4	A->T	O3a3b1a
M174	T->C	D	P8	G->T	Q1a5	N5	T->A	O3a3b1b
M175	-5bp	O	P90	C->T	B	N14	C->A	Q1a1
M178	T->C	N1c1	P91	C->T	F1	PK1	C->A	A2
M179	C->T	D2	P92	C->T	C5a	PK2	T->C	C3
M180=P88	T->C	E1b1a	P93	C->T	O3a	PK3	T->C	L3a
M181	T->C	B	P94	G->A	M1a2	PK4	A->T	O2a1a
M182	C->T	B2	P95	G->T	I2b4	PK5	C->T	R1a1e
M183	A->C	E1b1b1b2	P96	C->A	F3	RPS4Y711=M130	C->T	C
M184=USP9Y+3178T	G->A	T	P97	T->G	A	SRY10831.1	A->G	BR
M185	C->T	L	P98	C->T	R1a1d	SRY10831.2	G->A	R1a
M186	delección 1pb	M1	P99	C->T	D3	SRY2627=M167	C->T	R1b1b2d
M188	C->T	O3a3b1	P100	G->A	A3b1	SRY4064=M40=P20 E	G->A	E
M189	G->T	M1	P101	G->A	O3a3c2	SRY465	C->T	O2b
M190	A->G	A3b	P102	A->C	A3b1a	SRY9138=M177	C->T	M2a
M191=P86	T->G	E1b1a7	P103	G->C	O3a4a	Tat	T->C	N1c
M192	C->T	B2b	P104	G->A	F1	U106=M405	C->T	R1b1b2a
M193	inserción 4 pb	T	P105	G->A	N1c	U152	C->T	R1b1b2h
M194	T->C	Q1a3a2	P106	G->A	Q1a3a3	U174=P252	G->A	E1b1a7a
M195	A->G	E1b1a6	P107	G->A	R1b1b2a2	U175	G->A	E1b1a8
M196	C->G	A2	P108	G->A	A1	U179	G->A	I
M197	T->C	H1a1	P109	G->A	I1c	U181	C->T	E1b1a8a1a
M198	C->T	R1a1	P110	G->A	E1a2	U186	A->G	E1b1a7
M199	G ins	Q1a3a3	P111	C->T	B2a2a	U198=M467	G->A	R1b1b2a1
M200	G->A	E2b1a	P112	G->A	B2c	U209	C->T	E1b1a8a
M201	G->T	G	P113	T->G	E1b1a7a3a	U247=P253	C->T	E1b1a7
M202	T->G	A3b2	P114	G->A	A1b	U250=P222	G->C	I2b
M203	G->C	DE	P115	C->T	E1b1a7a2	U290	T->A	E1b1a8a1
M204	T->G	R1a1c	P116	A->G	E1b1a7a3	USP9Y+3636=M222	G->A	R1b1b2e
M205	T->A	J2b1	P117	G->T	M3	V6	G->C	E1b1b1e
M206	T->G	A2	P118	C del	M3	V12	A->G	E1b1b1a1
M207	A->G	R	P119	G->A	N1c1c	V13	G->A	E1b1b1a2
M208	C->T	C2a	P120	G->T	D2a3	V19	T->C	E1b1b1a3b
M209	A->G	O3a3b1	P121	G->A	C1a	V22	T->C	E1b1b1a3
M210	A->T	C4a	P122	C->A	C1	V27	A->T	E1b1b1a2a
M211	C->T	B2b4b	P123	T->C	IJ	V32	G->C	E1b1b1a1b
M212	C->A	A2	P124	A->C	IJ	V36	T->C	E1b1b1a2
M213=P137	T->C	F-R	P125=M429	T->A	IJ	V65	G->T	E1b1b1a4
M214	T->C	NO	P126	C->G	IJ			

Tabla 1. Colección de los 599 Y-SNPs descritos en la actualidad con sus variantes mutacionales y sus haplogrupos correspondiente (Karafet y cols., 2008).



### **3.1.5. Aplicaciones poblacionales**

En diversos estudios se ha observado una fuerte correlación entre la variación del cromosoma Y y la geografía. Esta propiedad ha sido utilizada para conocer patrones de migración, subestructuración poblacional y mezcla entre diversas poblaciones. En la actualidad, se dispone de una buena representación de los linajes del cromosoma Y alrededor del mundo, aunque es posible que importantes linajes permanezcan sin descubrir por la necesidad de un muestreo más profundo en muchas áreas del planeta.

A continuación se muestra la distribución global de los haplogrupos del cromosoma Y (ver figura 6 y 7), comenzando desde el clado principal “A” en orden alfabético hasta el último haplogrupo principal “T”:

El haplogrupo A, que en el árbol filogenético (ver figura 4 y 5) correspondería con la rama más antigua, aparece únicamente en África; lo mismo sucede con el haplogrupo B que presenta las frecuencias más altas entre grupos cazadores-recolectores de Etiopía y Sudán (Semino y cols., 2002; Cruciani y cols., 2006).

En el haplogrupo C se ha encontrado una mayor dispersión que abarca el Centro, Sur y Este de Asia (Karafet y cols., 2001; Ke y cols., 2001; Underhill y cols. 2001; Wells y cols., 2001; Zerjal y cols., 2002; Kayser y cols., 2003). Dentro de este haplogrupo, el linaje C1 aparece exclusivamente en Japón, el C2 en Nueva Guinea, Melanesia y Polinesia., mientras que el C3 se cree que se ha originado en el Sureste o Centro de Asia y desde aquí se ha expandido hacia el Norte de Asia, América y Europa Central. Por otro lado, el haplogrupo C4 parece estar restringido a poblaciones de aborígenes Australianos donde este linaje es dominante (Hudjashov y cols., 2007). El C5 tiene una presencia significativa en la India.

El clado D presenta las frecuencias más altas en el Tíbet (50%) y Japón (35%), pero también aparece en otras regiones de Asia Central y Sudeste asiático (Karafet y cols., 2001; Cox y cols., 2006).

El haplogrupo E es uno de los que presenta mayor ramificación, con una gran cantidad de subhaplogrupos descritos (Cruciani y cols., 2002, 2004, 2006). Empezando por los linajes E1a y E2, que fueron descritos en el Noroeste de África y seguidos por el E1b1, que presenta una amplia distribución geográfica con dos de sus sublinajes bien

definidos: E1b1 presente en todo África y en Afro-americanos, y E1b1b1 presente en el Oeste de Europa, Norte de África y Oriente Próximo.

Dentro del haplogrupo F los clados menores F\*, F1 y F2 aparecen en población de la India (Kivisild y cols., 2003).

El haplogrupo G aparece con más frecuencia en la región del Cáucaso, sin embargo, también presenta frecuencias significativas en áreas del Mediterráneo y Oriente Medio (Cinnioglu y cols., 2004; Nasidze y cols., 2004).

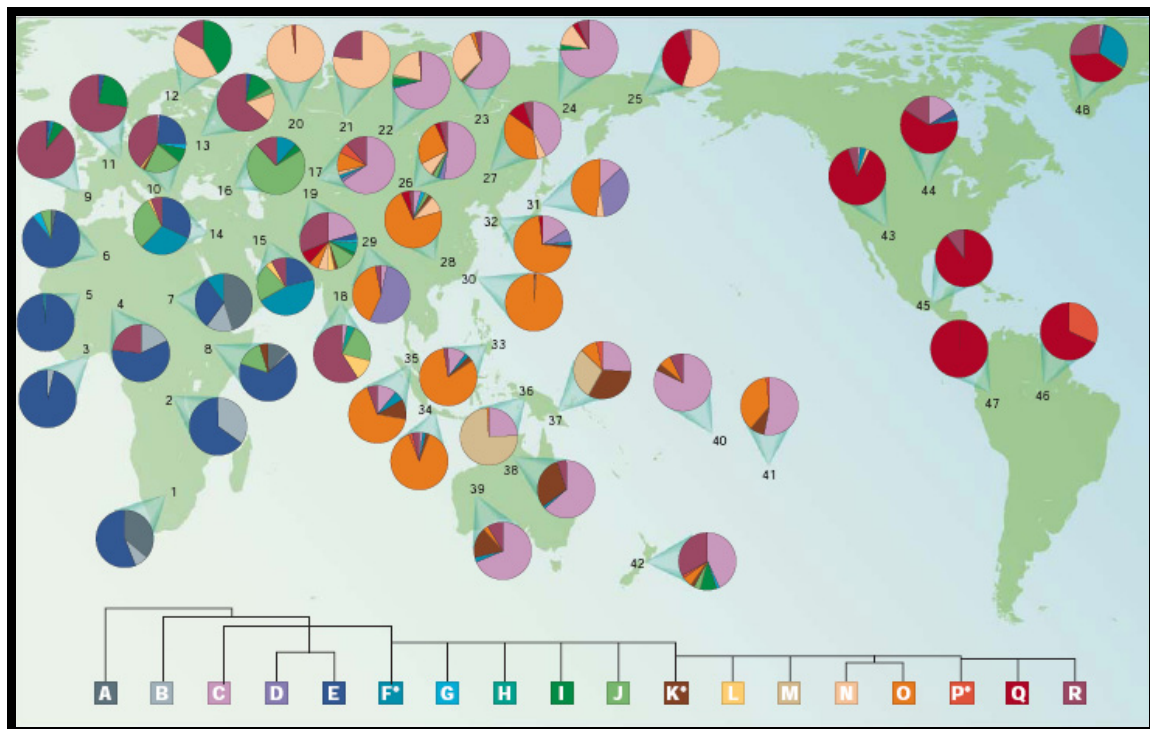
El Haplogrupo H también aparece en la India, como sucedía con el “F”, pero no ha sido estudiado en profundidad (Kivisild y cols., 2003; Sengupta y cols., 2006).

El haplogrupo I es un haplogrupo claramente europeo. Es de los más frecuentes entre las poblaciones del Noroeste de Europa (Rootsi y cols., 2004).

La característica del haplogrupo J, en la que casi hay un acuerdo unánime, es el hecho de la dispersión que experimentó este haplogrupo cuando se produjo la migración hacia el Oeste de individuos de Oriente Próximo, portando así este haplogrupo al Norte de África, Europa, Asia Central, Pakistán y la India (Quintana-Murci y cols., 2001; Sengupta y cols., 2006). Aunque otros como Di Giacomo y cols. (2004) consideran la dispersión de este haplogrupo como una expansión del mundo Griego.

El haplogrupo K es el haplogrupo ancestral de los grupos principales que van desde “L” a “R” (lo mismo sucedía con el haplogrupo F, que sería el ancestral de los clados que van desde “G” a “R”). Además, dentro del clado K encontramos los clados menores K\*, K1, K2, K3 y K4 que están repartidos por todo el mundo en bajas frecuencias (Kayser y cols., 2003).

Otro linaje principal es el que corresponde al haplogrupo L, que se encuentra sobre todo en la India y Pakistán, así como en Oriente Medio y, muy ocasionalmente, en poblaciones Europeas, concretamente en países Mediterráneos (Kivisild y cols., 2003; Sengupta y cols., 2006).



**Figura 6.** Distribución global de haplogrupos del cromosoma Y (Jobling y Tyler-Smith, 2003). Cada círculo representa una población y cada sector coloreado la frecuencia relativa de los 18 haplogrupos principales del cromosoma Y.

El haplogrupo M presenta sus frecuencias más altas en Melanesia, estando su presencia limitada al área geográfica de las lenguas Papuanas (Capelli y cols., 2001; Hurler y cols., 2002; Cox y cols., 2006).

El haplogrupo N actualmente presenta una amplia distribución y debió surgir en el interior de Asia y Sur de Siberia (Rootsi y cols., 2007). Los subclados más frecuentes son el N1c, que probablemente apareció en la región que corresponde a lo que hoy se conoce como China, expandiéndose desde aquí a Siberia y al Este de Europa. El otro subclado sería el N1b, presente en altas frecuencias en poblaciones Urálicas, sobretodo en el grupo poblacional Finno-Ugric.

El linaje O representa aproximadamente el 60% de los cromosomas del Este de Asia; de sus sublinajes el O3 tiene la frecuencia más alta y está ausente fuera del Este de Asia, mientras que los haplogrupos O1 y O2 aparecen en Malasia, Vietnam, Indonesia, Sur de China, Japón y Corea (Shi y cols., 2005; Cox y cols., 2006; Hammer y cols., 2006).

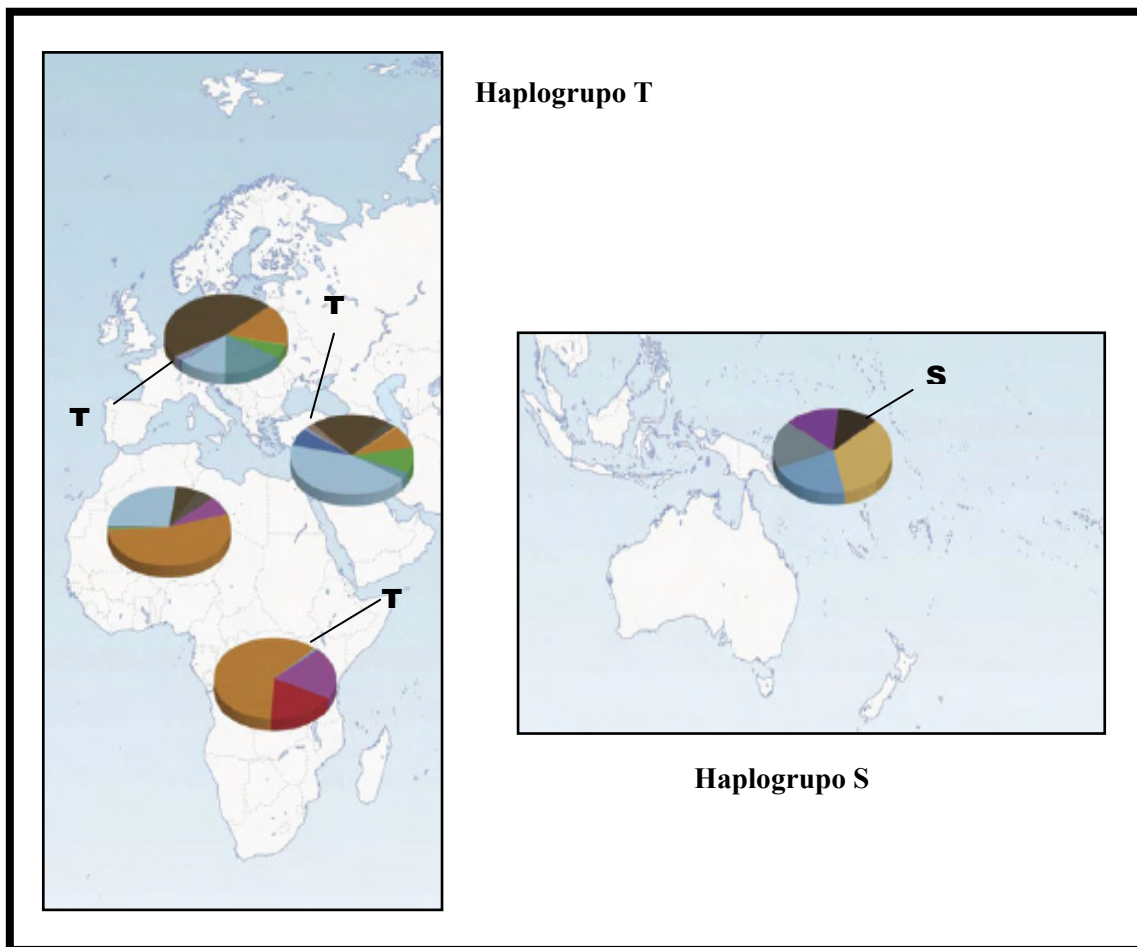
Un linaje detectado en bajas frecuencias en el Cáucaso y la India es el haplogrupo P (Underhill y cols., 2000; Nasidze y cols., 2004).

El haplogurpo Q se encuentra repartido por Asia, América, Europa y Oriente Próximo, mientras que su subclado Q1a3a está asociado de modo casi exclusivo con la población Nativa Americana (Bortolini y cols., 2003; Zegura y cols., 2004).

Hasta la reciente publicación de Karafet y cols., 2008 el último clado principal se correspondía con el amplísimo haplogrupo R. La mayoría de los individuos de este haplogrupo pertenecen al subclado R1, que está representado principalmente por dos linajes: R1a y R1b (Wells y cols., 2001; Cinnioglu y cols., 2004; Brión y cols., 2005; Capelli y cols., 2005). Se cree que el haplogrupo R1b1 corresponde a los descendientes de los primeros humanos modernos que entraron en Europa. En la actualidad es el haplogrupo más frecuente en el Oeste de este continente, también aparece en el Norte de África y en bajas frecuencias en Irán y Korea; curiosamente se encuentra en América y Australia pero parece que es debido a las recientes migraciones europeas. Por otro lado, los haplogrupos R1a y R1a1 se hallan en elevadas frecuencias en el Centro y Oeste de Asia, la India y en población Eslava del Este de Europa; la distribución de estos haplogrupos se cree que está relacionada con la expansión de la cultura Kurgan. El subclado R2 presenta su frecuencia más elevada en el Sur de Asia, encontrándose en bajas frecuencias en el Cáucaso y Asia Central.

En la publicación de Karafet y cols., 2008, mencionada con anterioridad, se describen dos nuevos haplogrupos principales el “S” y el “T”.

El clado S, anteriormente llamado K5 (Scheinfeldt y col., 2006), se encuentra principalmente en Oceanía e Indonesia. Mientras que el clado T, anteriormente se le denominaba como K2, es observado en bajas frecuencias en Oriente Medio, África y Europa.



**Figura 7.** Distribución de los nuevos haplogrupos “S” y “T” del cromosoma Y (Karafet y cols., 2008).

Los métodos para datar el orden filogenético de estos haplogrupos de la región no recombinante del cromosoma Y emplean la diversidad de Y-STRs dentro de cada linaje. Los SNPs son eventos evolutivos raros o únicos, que generan linajes paternos en los cuales podemos estudiar la variabilidad de STRs que se va acumulando con el tiempo. De este modo, cuanto más antiguo sea un linaje, mayor será la variabilidad de STRs hallada.

Además de estimar las edades de las mutaciones que definen los clados principales en el árbol de los haplogrupos de la NRY, también se puede calcular de un modo similar la edad del ancestro común más reciente (TMRCA, time to the most recent common ancestral). Concretamente en el cromosoma Y se han realizado diversas estimas de TMRCA, como el trabajo de Pritchard y cols. (1999), que usando 8 microsatelites estimó un TMRCA de 46.000 años (incertidumbre de 120.000 años), mientras que Thomson y cols. (2000) basándose en la variación de los SNPs de una

secuencia de 64Kb de ADN del cromosoma Y, obtuvo una edad de 59.000 años. Todos estos valores de TMRCA son más bajos de lo esperado sobre todo si los comparamos a los valores obtenidos con ADNmit o con locus autosómicos, esto es debido en parte al reducido tamaño efectivo de población y a la neutralidad evolutiva del cromosoma Y.

Como ya se ha mencionado con anterioridad, los haplogrupos A y B están restringidos a África, se encuentran distribuidos a través de este continente en bajas frecuencias. Después de múltiples estudios se ha llegado a la conclusión de que el haplogrupo A forma parte de un pool genético de cromosomas Y proto-africanos (Cruciani 2004). Hoy en día las frecuencias más elevadas de este linaje se encuentran en etíopes y sobre todo en khoisánidos. El haplogrupo que le sigue en antigüedad es el B, que aparece en población Bantú del Sur de África, Pigmeos del centro de África y en algunos linajes de Sudán y Mali, además de estar presente pero en frecuencias más bajas que para el haplogrupo A, en etíopes y khoisánidos. Datos antropológicos parecen sugerir que, durante el período interglacial ocurrido hace aproximadamente entre 130.000-90.000 años, las poblaciones de humanos recientes se expandieron a través de África, alcanzando el Norte y Sur del Sahara y también el Levante Africano, pero parece que el período comprendido entre 90.000-50.000 es el que está asociado con la expansión de los linajes A y B. A pesar de que los registros fósiles y arqueológicos son demasiado pobres para reforzar esta hipótesis que, sin embargo se ve favorecida por la información paleoclimática, la cual indica que hace unos 70.000 años el clima glacial llevó a una fragmentación del continente africano en ambientes muy diferentes y aislados entre el Noroeste, Noreste y Sur. Este período comprendido entre 90.000-50.000 años antes del presente, podría ser el responsable de parte de la diversidad encontrada fuera de África, como resultado del proceso de diversificación que se produjo dentro del continente africano (Underhill y cols. 2001).

El primer movimiento migratorio hacia fuera del continente debió producirse hace en torno a 50.000-40.000 años (Goebel y cols., 2007) y coincide con la entrada en Eurasia de los primeros humanos modernos, pero previamente a este evento, surge dentro de África la mutación en el Y-SNP M168 (haplogrupo CF). Se cree que los individuos que entraron en Eurasia portaban esta mutación y comenzaron a reemplazar los linajes del cromosoma Y de Neandertales que había hasta ese momento en esta región. La dispersión de este linaje parece que ocurrió desde “el Cuerno de África”

hacia la India y a través del Corredor Levantino situado en Oriente Medio (Cavalli-Sforza y cols., 1994).

En la actualidad, la población subsahariana se caracteriza por la presencia de los haplogrupos principales A, B y E, y de ellos, el más frecuente es el haplogrupo E, que también está presente en una frecuencia alta en Oriente Medio y Sur de Europa, con presencia ocasional en Asia Central, Pakistán y América, mostrando las frecuencias más altas para el subclado E1b1a en el África subsahariana y para el subclado E1b1b1 en el Norte y Este de África, en la cuenca del Mediterráneo y Europa (Cruciani y cols., 2004; Cruciani y cols., 2007). La amplia distribución del linaje E1b1a posiblemente esté relacionada con la “expansión de la agricultura Bantú” hace aprox. 3.000 años que, se supone, fue acompañada de un reemplazamiento de otros cromosomas Y. Por otro lado, en el caso del linaje E1b1b1, lo más probable es que haya surgido en la región del “Cuerno de Africa” hace 24.000-27.000 años, expandiéndose desde aquí hacia el Norte y Noroeste de Africa alcanzando el Mediterráneo. A finales del Pleistoceno comenzaría a dispersarse hacia el Oeste de Asia, introduciéndose más tarde, durante la expansión agrícola del Neolítico, desde Oriente Medio hacia el Sur de Europa (Luis y cols., 2004; Semino y cols., 2004; Cruciani y cols., 2007).

Otra mutación es la ocurrida en el Y-SNP M213; se cree que se originó en África a partir de las poblaciones que ya tenían la mutación en M168, produciéndose su dispersión desde el Este de África, cruzando el Corredor Levantino, hacia Eurasia hace 45.000-30.000 años. Este hecho se ve reforzado por los registros paleantropológicos que muestran una expansión de los humanos modernos desde Levante, coincidiendo con las culturas del Paleolítico Superior y con la extinción de los Neandertales de Oriente Medio y Europa. Las poblaciones portadoras de esta mutación se diversificarían más tarde en los distintos clados y subclados que contiene el haplogrupo parental F.

Otro linaje con una amplia dispersión es el haplogrupo C, que se encuentra principalmente en Asia, Austro-Melanesia, Micronesia, Polinesia y Norte de América. Este haplogrupo surgiría de las poblaciones que salieron de África en dirección Sur de Asia y que portaban la mutación M168 (hace 50.000-40.000 años). Dentro de este haplogrupo el sublinaje C3 es uno de los más representativos ya que se encuentra diseminado por Turquía, Centro y Este de Asia y, además es la única variedad del haplogrupo C encontrada en Norteamérica.

Un haplogrupo exclusivamente asiático es el D, que únicamente se ha encontrado en Japón y el Tibet. Este linaje debió encontrarse en bajas frecuencias, lo que facilitó su extinción en la mayoría de las poblaciones asiáticas, exceptuando Japón y el Tibet, donde la deriva genética habría incrementado su frecuencia (Underhill y cols., 2001). Estos dos haplogrupos, C y D, habrían sido sustituidos en parte por las subsecuentes expansiones de individuos con haplogrupos más recientes.

Los individuos portadores de la mutación M213 también se habrían expandido hacia Asia Central (no sólo hacia el Cáucaso) y, con el tiempo, individuos de estas poblaciones habrían adquirido la mutación del Y-SNP M9. Los portadores de esta mutación se tuvieron que expandir ampliamente ya que han dado lugar a varios linajes con mutaciones independientes, como son los haplogrupos O, N y P, que se fueron distribuyendo por la mayoría de Asia desplazando a otros haplogrupos anteriores, como el C. Un ejemplo es el haplogrupo N, que se expandiría hacia el Oeste desde el interior de Asia/Sur de Siberia hace unos 12.000-14.000 para llegar al Este de Europa (Rootsi y cols., 2007). Otro caso es el del haplogrupo M, que llega a Nueva Guinea y desplaza en parte también al haplogrupo C.

Dentro del haplogrupo parental P, el haplogrupo R surgió en el Norte de Asia en torno al año 30.000 de nuestra era y dentro de este el R1b, que se empezó a dispersar hacia el Oeste alcanzando Europa, la región del Cáucaso, Oriente Medio, Centro de Asia y Norte de la India y Pakistán. Otro linaje dentro del haplogrupo P es el Q, que se expandió ampliamente a través de la estepa del Norte de Asia, así como por el centro de Asia, Norte de India y Pakistán. Dentro de este haplogrupo Q, el sublinaje Q1a3a se caracteriza porque probablemente se expandió más tardíamente desde Siberia a las Américas.

Distintos estudios realizados sobre los SNPs de la región no recombinante del cromosoma Y, revelan indicios de una expansión ocurrida hace 25.000-20.000 años desde Levante hacia la cuenca del Mediterráneo y que está relacionada con la presencia del haplogrupo I en Europa. A estos eventos de expansión les siguió un período de contracción poblacional asociado con la última gran glaciación ocurrida hace 18.000 años y que también se manifiesta en los registros antropológicos. Situación que cambia radicalmente en el momento de mejora climática que conlleva a una expansión del haplogrupo R1a y R1a1 en Europa desde un posible refugio glacial en la región de



Ucrania hace aprox. 10.000-15.000 años, así como la expansión Neolítica del haplogrupo E3b desde Oriente Medio hacia Europa, acompañado de los haplogrupos G y J2.

Otro de los grandes movimientos poblacionales es el relacionado con el haplogrupo O, que tiene elevadas frecuencias en el Este de Asia y también está presente en el Centro de Asia y en la Polinesia. Esta expansión parece que se originó en el Norte de China y su distribución podría estar vinculada con la agricultura del arroz y del mijo (Su y cols., 2000) hace unos 7.000 años y que correspondería aproximadamente con el comienzo de la colonización de la Polinesia, inhabitada hasta ese momento. Existen diversas hipótesis para explicar cuál puede ser la población original que colonizó estas islas (Su y cols., 2000; Hurles y cols., 2002).

Posiblemente uno de los modelos de migración más interesantes, a la vez que recientes, corresponde con la entrada de los primeros humanos en el continente americano durante el Pleistoceno. Los dos principales haplogrupos fundadores encontrados parecen ser el C y el Q. Siendo el linaje Q el que presenta una mayor frecuencia a lo largo de todo el continente, mientras que la frecuencia del haplogrupo C es mucho más reducida. Uno de los principales problemas que nos encontramos es el hecho de explicar como se produjo la entrada de estos cromosomas Y, teniendo en cuenta que el continente americano se encuentra aislado del resto de la gran masa terrestre. Para resolver esto, además de en datos genéticos nos apoyamos en datos paleoclimáticos, los cuales indican que durante la glaciación *Würm o Wisconsin*, iniciada 100.000 años antes del presente y finalizada hace aproximadamente 12.000 años, se produjeron una serie de fenómenos de eustacia glacial que implicaron un descenso en el nivel de los océanos, apareciendo en algunos casos conexiones terrestres entre varios puntos del planeta. Uno de esos puntos fue *Beringia*, una gran masa de tierra que conectó Siberia y Alaska hace unos 40.000 años y que se mantuvo durante 4.000 años para luego volver a hundirse. Pero este puente parece que volvió a surgir hace unos 25.000 años, dejando un corredor abierto por el que pudieron pasar tribus primitivas asiáticas hasta hace unos 11.000 años, momento en el cual se hundiría definitivamente con el final de la última gran glaciación (Elias y cols., 1996; Pitulko y cols., 2004). Estos datos concuerdan con los obtenidos por Seielstad y cols. (2003), que estimaron una edad aproximada de 20.000-15.000 años para el haplogrupo fundador Q.

Por otro lado Zegura y cols. (2004) calcularon la entrada de los primeros pobladores de América en unos 17.000 años, basándose en estudios realizados sobre Y-SNP e Y-STRs. Hay otras dataciones, no sólo con estudios de cromosoma Y sino también con estudios realizados sobre ADN mitocondrial, con resultados muy diferentes. Pero hay que tener en cuenta que los individuos que pudieron pasar inicialmente a través de este corredor se encontrarían con un Canadá totalmente cubierto de hielo, imposibilitando la entrada más allá de Beringia; concretamente se encontrarían con dos gigantescas placas de hielo: la placa Laurentina y la placa Cordillerana. Según distintos estudios, estas dos masas de hielo comenzaron a separarse hace unos 14.000-11.000 años, formando un estrecho corredor (corredor de McKenzie) de unos 25 km, permitiendo así el avance de las tribus asiáticas hacia el interior de América, aunque no hay evidencias que lo prueben. Una hipótesis diferente se basa en una ruta marítima alternativa, en la cual los individuos saldrían desde el Sur de Beringia bordeando la costa Oeste de Norteamérica.

A pesar de que hay cierta discrepancia en el modo de entrada, sí que encontramos un consenso casi total en cuanto a la entrada en América a través de Beringia. Hay otras hipótesis difíciles de demostrar, como las de Mendes Correâ (1925) o Paul Rivet (1943), que se basan en características culturales y lingüísticas para proponer una posible entrada de poblaciones australianas por América del Sur después de cruzar a través de la Antártida, pero de momento no se han encontrado datos genéticos que apoyen este hecho.

Además de la datación de la primera entrada, se plantean otras dudas, como el número de oleadas migratorias que han podido ocurrir.

Durante mucho tiempo se aceptó el modelo de tres migraciones propuesto por Greenberg y cols. (1986), que se basaba en datos de evidencias lingüísticas, de morfología dental y de marcadores genéticos clásicos. En su trabajo proponía una primera migración que ocurriría hace unos 11.000 años y que daría lugar a las poblaciones amerindias; la siguiente sería hace unos 9.000 años y constituiría la población Na-Dene y la tercera se produciría hace unos 4.000 años y sus descendientes serían los Esquimo-Aleutinos. Realmente esta teoría era bastante pobre desde un punto de vista genético, aunque se mantuvo durante casi 20 años. Aún así, los nombres de los tres principales grupos poblacionales descritos por Greenberg se siguen manteniendo, sobre todo en el campo de la lingüística.

En los primeros estudios migratorios, llevados a cabo sobre los polimorfismos del cromosoma Y, se propone un modelo de migración única (Santos y cols., 1996; Underhill y cols., 1996; Bianchi y cols., 1997; Karafet y cols., 1997; Lell y cols., 1997), estudios más tardíos apoyaron un modelo de 2 migraciones (Karafet y cols., 1999; Lell y cols., 2002), aunque trabajos más recientes parece que vuelven a estar a favor del modelo de un origen único (Tarazona-Santos y cols., 2002; Zegura y cols., 2004).

Zegura, en su trabajo del 2004, concluye que los linajes C y Q proceden de una única migración de individuos procedentes de la región de las montañas Altay. Sin embargo Lell, en su trabajo del 2002, propone un modelo de dos migraciones en el cual: la primera entrada correspondería a los individuos portadores del haplogrupo Q, procedente del Centro Sur de Siberia y la segunda oleada incluiría al haplogrupo C y al haplogrupo R procedentes de poblaciones de la zona baja del río Amur en la región de Okhotsk. En contra de esta teoría también se encuentra Tarazona-Santos (2002), que duda del origen Siberiano del haplogrupo R, atribuyéndolo más bien a un proceso de mezcla de las poblaciones amerindias con las europeas a partir de la colonización, después del descubrimiento de América.

Por el momento, y a pesar de que cada autor presenta datos con los que pretende demostrar el número de posibles entradas en América usando los polimorfismos de la NRY, es muy difícil llegar a una conclusión definitiva.

### **3.1.6. Aplicaciones forenses**

Los Y-SNPs son utilizados principalmente en la antropología molecular para estudios evolutivos. Sin embargo, aunque presentan un poder discriminativo inferior al de los STRs, poseen otras ventajas de gran interés en la rutina forense. Una de ellas es la posibilidad de ser analizados en pequeños amplicones, que es una ventaja a la hora de trabajar sobre muestras degradadas; otra de sus propiedades es la posibilidad de automatización y análisis con plataformas de alto rendimiento, pero posiblemente la ventaja más salientable sea la distribución geográfica observada en los haplogrupos obtenidos a partir de los Y-SNPs. Esta característica puede servir de herramienta para predecir el posible origen geográfico del linaje paterno de cualquier vestigio biológico masculino encontrado en un escenario forense. En pericia criminal la interpretación

estadística en caso de exclusión, es directa, pero en caso de coincidencia de perfiles, la marcada distribución geográfica hace necesarias las correcciones estadísticas (Jobling, 2001a). Además los Y-SNPs se muestran como marcadores prometedores para usar en complicados casos de mezcla masculina y femenina, a pesar de que los STRs del cromosoma Y sean los marcadores de elección en estos casos, no se puede negar la utilidad de la información complementaria proporcionada por los Y-SNP (Sanchez y cols., 2004).

Esto nos lleva a concluir que, sería realmente interesante el desarrollo de multiplexes que incluyan el mayor número posible de Y-SNPs que caracterizen los clados principales del árbol filogenético del cromosoma Y, ya que permitirían la discriminación geográfica en una simple reacción (Sanchez y cols., 2003, Brión y cols., 2004; Brión y cols., 2005).

Como se mostró con anterioridad, el número de métodos de genotipado de SNPs es muy variado y continuamente surgen nuevos métodos y técnicas, por lo tanto es muy difícil decir cuál es mejor para un uso forense, más bien es el usuario el que debe decantarse por una técnica u otra en función de sus requerimientos, que en genética forense suelen ser multiplexes de gran capacidad y alta precisión (Sobrinó y cols. 2005).

## **3.2. Microsatélites o STRs**

### **3.2.1. Definición**

Los microsatélites como ya se ha comentado, son bloques de 2-6 bases que se repiten en tándem. El número de repeticiones de un STR puede ser muy variable entre individuos (los más interesantes como marcadores genéticos tienen entre 10 y 30 repeticiones), por lo que esta propiedad lo convierte en una herramienta muy interesante para la identificación humana.

En la actualidad hay miles de microsatélites polimórficos identificados en el ADN humano. Se estima que aparecen en el genoma con una media de uno cada 10000 nucleótidos (Edwards y cols., 1991). Están presentes en todos los cromosomas, pero en esta tesis nos hemos centrado únicamente en los STRs del cromosoma Y. Estos

microsatélites de la región no recombinante del cromosoma Y tienen un papel importante en la genética forense, donde han llegado a ser los marcadores de elección, particularmente en los casos de agresión sexual o en pruebas de paternidad, sobre todo cuando el presunto padre no está disponible (Jobling y cols., 1997; Kayser y cols., 1997); también son muy útiles en investigaciones genealógicas (Jobling y cols., 2001b) y en estudios evolutivos como marcadores de linajes masculinos (Kayser y cols., 2001b; Stumpf y Goldstein, 2001; Jobling y Tyler-Smith, 2003).

Básicamente los Y-STRs son usados con 2 objetivos (Kayser y cols., 2004): el primero es distinguir linajes, para ello el número de marcadores y su variabilidad determinará el grado de discriminación, el segundo objetivo sería suministrar información sobre la relación entre linajes.

En estudios evolutivos se han llegado a utilizar hasta 16 Y-STRs, apareciendo en algunas poblaciones gran número de individuos que compartían el mismo haplotipo para los 16 Y-STRs (Zerjal y cols., 2003; Hedman y cols., 2004). Incluso en casos excepcionales padre e hijo difieren en su haplotipo de Y-STRs debido a que se ha producido una mutación (Kayser y cols., 2000). Esa posibilidad debe de ser considerada en los cálculos realizados en forense y en las pruebas de paternidad (Kayser y Sajantila, 2001; Rolf y cols., 2001). Para tener resultados fiables es necesario usar un gran número de microsatélites y conocer sus frecuencias haplotípicas, así como conocer lo mejor posible las características mutacionales de cada locus (Stumpf y Goldstein, 2001).

### **3.2.2. Tasas de mutación**

El conocimiento del mecanismo mutacional que genera el alto nivel de variabilidad de estos marcadores nos ayudará a conocer mejor sus aplicaciones.

La gran similitud en los tamaños de los alelos en cada locus inspiró la hipótesis del mecanismo o modelo “*stepwise mutation*” (Ohta y Kimura, 1973), donde podría estar implicado el slippage replicativo (Tautz y Renz, 1984; Levinson y Gutman, 1987; Stephan, 1989; Weber, 1990), ya que estudios de microsatélites sobre enfermedades genéticas (Mahtani y cols., 1993; Petrukhin y cols., 1993; Weber y cols., 1993) muestran que más del 90% de las mutaciones provienen del aumento o disminución de una sola unidad repetitiva con respecto al alelo inicial (Cooper y cols., 1996; Calafell y

cols., 1998), esto sobretodo ocurre en unidades repetitivas de 2 a 4 pb (Jin y cols., 1996). Parece que este modelo no tiene en cuenta todos los tipos de variación encontrados (Shriver y cols., 1993; Valdes y cols., 1993; Deka y cols., 1995), esto llevó a proponer un nuevo modelo con 2 *fases mutacionales* diferentes (DiRienzo y cols., 1994): por un lado tendríamos las mutaciones con la variación de una única unidad repetitiva, que sería lo mas frecuente, y por otro lado tendríamos los cambios mutacionales en los que hay varias unidades repetitivas implicadas, que suponen variaciones de longitud a mayor escala y son menos frecuentes.

Como se acaba de mencionar en el texto, el slippage de cadena durante la replicación es sugerido como el mecanismo más probable implicado en la generación de mutaciones en microsatélites (Levinson y Gutman, 1987; Schlötterer y Tautz, 1992; Strand y cols., 1993; Heyer y cols., 1997; Kayser y cols., 1997; Brinkmann y cols., 1998). El slippage es un fenómeno intracromátida que tiene lugar durante el proceso de replicación cuando se produce un desplazamiento de la ADN polimerasa sobre la región repetida de la hebra molde de ADN y el reensamblamiento en una posición más avanzada. A consecuencia de este deslizamiento, la enzima ADN polimerasa incorpora las bases complementarias, provocando un emparejamiento erróneo entre ambas hebras de ADN. Dependiendo de que dicho error se produzca en la hebra parental o en la de nueva síntesis, el resultado es la delección o inserción respectivamente de nuevas unidades repetitivas (Sia y cols., 1997).

Una buena estima de la tasa de mutación de Y-STRs es un requisito fundamental para hacer una buena datación de los linajes del cromosoma Y definidos por los SNPs (Zerjal, 1997; Bianchi y cols., 1998), así como para la interpretación de datos en pruebas de paternidad y casos forenses (Jobling y cols., 1997; Kayser y cols., 1997).

Para determinar las tasas de mutación hay básicamente dos modos de enfocar los estudios (Holtkemper y cols., 2001):

- Estudio de patrones alélicos en pedigríes bien conocidos (Heyer y cols., 1997)
- Estudio de ADN de células espermáticas (Leeflang y cols., 1995; Monckton y cols., 1995).

De momento el número de estudios de tasa de mutación de los STRs del cromosoma Y es escaso y se centran siempre sobre los mismo STRs. Uno de los estudios más

amplio es el realizado por Gusmao y cols. (2005a) donde utilizan 16 Y-STRs y obtienen una estima de la tasa de mutación de  $1,998 \times 10^{-3}$ , que se mueve más o menos dentro de los valores obtenidos también por Kayser y cols. (2000) y Dupuy y cols. (2004) que son de  $2,8 \times 10^{-3}$  y  $2,31 \times 10^{-3}$  respectivamente. Otro estudio con resultados prácticamente iguales fueron los de Heyer y cols. (1997), que estudió la tasa de mutación en 7 Y-STRs de pedigríes bien conocidos y obtuvo una tasa de mutación de  $2 \times 10^{-3}$ , esto sugirió en su momento que las frecuencias de mutación de los Y-STRs eran comparables a las de los STRs autosómicos después de ver los resultados de Weber y Wong (1993); ambos habían usado en sus estudios STRs con repeticiones tetranucleotídicas. Por otro lado tenemos las estimas de Foster y cols. (2000), que fueron de los primeros en reconocer la importancia del significado evolutivo de la tasa de mutación para realizar una datación adecuada de los eventos poblacionales. Utilizó 6 Y-STRs y 5 SNPs para su estudio basado en network filogenéticos. Su estima de la tasa de mutación fue de  $2,6 \times 10^{-4}$  mutaciones/20 años, casi 7 veces más bajas que las de Heyer y cols. (1997).

Resultados parecidos a los de Foster son los obtenidos por Zhivotousky y cols. (2004), que estimó la tasa de mutación media por repetición por STR en  $6,9 \times 10^{-4}/25$  años, usando datos de la variación de los Y-STRs dentro de haplogrupos de cromosoma Y definidos por SNPs en poblaciones con historias bien documentadas. Estos datos dificultan mucho la tarea a la hora de seleccionar una tasa de mutación adecuada para nuestros estudios, ya que como propone Zhivotousky y cols. (2005) es posible que exista una discrepancia entre la tasa de mutación desde el punto de vista genealógico y evolutivo.

### 3.2.3. Métodos de detección

Hace unos años, la detección de los STRs se realizaba mediante sistemas *electroforéticos manuales*, basados en la separación por tamaño de los productos de PCR en geles de poliacrilamida o agarosa, que luego eran teñidos con bromuro de etidio o nitrato de plata para visualizar estos productos y caracterizar los alelos.

Actualmente los *sistemas electroforéticos automáticos* son los que se han impuesto. En estos sistemas las muestras pasan por delante de un haz de luz UV procedente de un láser. La fluorescencia emitida por cada fragmento de ADN, es

recogida por una cámara CCD (Charge-Coupled Device), encargándose luego el software del equipo de interpretar la fluorescencia emitida. Dentro de estos sistemas electroforéticos automáticos podemos hablar de dos tipos:

- *Sistemas electroforéticos automáticos en gel* basado en la polimerización de un gel de acrilamida/bisacrilamida entre dos cristales montados sobre un soporte que posteriormente se acopla en el secuenciador automático para proceder a cargar las muestras dentro de los pocillos marcados en el gel y llevar a cabo la secuenciación o el análisis de los fragmentos.
- *Sistemas electroforéticos automáticos capilares* proporcionan una clara alternativa al sistema basado en geles. En este sistema el gel ha sido sustituido por un polímero que se inyecta de forma automática en unos capilares, esto ha de ocurrir antes de que la muestra sea introducida también de forma automática en el capilar donde se lleva a cabo la electroforesis.

Para que el láser pueda detectar los productos, estos deben de estar marcados con fluorescencia. Hay tres modos básicamente para conseguir esto:

- Incorporar la fluorescencia en el amplicón a través de un primer oligonucleotídico marcado en el extremo 5' (es el más usado).
- Incorporar dNTPs marcados con fluorescencia en la PCR.
- Intercalar entre las bases de ADN del producto amplificado moléculas fluorescentes (tiene muchas limitaciones).

Cuando usamos este tipo de sistemas electroforéticos automáticos debemos de tener cuidado a la hora de interpretar los resultados, ya que algunas veces se producen artefactos que pueden llevarnos a error. Para evitarlos es bueno establecer rigurosos criterios de lectura de los resultados.

Los artefactos más frecuentes son:

- Stutters o picos tartamudos, se originan durante la PCR por el deslizamiento o “slippage” de la enzima ADN polimerasa, generando así picos con una unidad de repetición menos que el verdadero alelo. Estos picos n-1 suelen ser de menor tamaño que el del alelo real, aunque esto depende del locus analizado. En algunas situaciones puede llegar a ser un problema a la hora de diferenciar si un



pico pequeño es un alelo real producto de una mezcla con un contribuidor menor o un stutter producto de un alelo adyacente.

- Pull-up son picos artefactuales resultado de la entrada de un color en el canal espectral de otro, esto ocurre cuando el pico está fuera de escala debido generalmente a sobreamplificación, la dilución de la muestra reduce este efecto o incluso lo llega a eliminar.
- Artefactos inespecíficos, originados por hibridaciones inespecíficas de los primers durante el proceso de amplificación.

Recientemente han aparecido en el mercado una serie de kits para la amplificación de STRs del cromosoma Y, que evitan o disminuyen la aparición de gran parte de los artefactos debido a su sensibilidad, especificidad y estabilidad. Los Y-STRs que contienen estos kits los convierten en una herramienta realmente útil en la genética forense y también en el campo de la genética poblacional:

**AmpFISTR® Y filer™ PCR Amplification kit** (Applied Biosystems): DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 y Y GATA H4.

**PowerPlex® Y System** (Promega): DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439.

**Y-PLEX™ 5, 6, 12** (Reliagene Technologies. New Orleans, LA): **Y-PLEX™ 5:** DYS389I/II, DYS392, DYS438, DYS439. **Y-PLEX™ 6:** DYS19, DYS389II, DYS390, DYS391, DYS393, DYS385a/b. **Y-PLEX™ 12:** DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439 y Amelogenina.

**Mentype® Argus Y-MH<sup>QS</sup>** (Biotype AG. Dresden, Alemania): DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393.

**GenRES® DYSplex-1 y DYSplex-2** (Serac. Bad Homburg, Alemania): **DYSplex-1:** DYS390, DYS391, DYS385a/b, DYS389I/II, Amelogenina. **DYSplex-2:** DYS19, DYS389I/II, DYS392, DYS393.

**GeneKin® Y-STR systems I, II, III, IV** (GKT Inc., Korea): basado en separación electroforética en gel y tinción con plata. **I:** DYS392, DYS19, DYS388. **II:** DYS391, DYS391X, DYS390, DYS393. **III:** DYS389II, DYS389I, DXYS156Y, DXYS156X. **IV:** DYS385a/b, DYS156Y, DXYS156X.

Se han hecho diversos estudios para comparar el rendimiento de los diferentes kits (Johns y cols., 2006; Mayntz-press y cols., 2007), los datos reflejan que los kits de Applied Biosystems y Promega son de los mejores hasta el momento, además de ser los que contienen un mayor número de Y-STRs.

#### **3.2.4. Aplicaciones poblacionales**

Los Y-STRs nos permiten obtener haplotipos que caracterizan en detalle a cada linaje masculino.

Debido a la mayor sensibilidad que presentan los STRs del cromosoma Y a la deriva genética y al efecto fundador los convierte en una muy buena herramienta en la detección de diferencias genéticas entre poblaciones si los comparamos con los STRs autosómicos (Jorde y cols., 2000).

Según Roewer y cols. (2005) estos marcadores son de gran utilidad para estudios de estructuración poblacional local y en estudios de la historia demográfica reciente por sus características de hipervariabilidad que le permitieron observar en su estudio, de Y-STRs en hombres europeos, niveles más altos de variabilidad interindividual así como la identificación de estratos masculinos previamente no encontrados en otros trabajos donde se utilizaron Y-SNPs en lugar de Y-STRs (Rosser y cols., 2000; Semino y cols., 2000).

Ya se había propuesto con anterioridad que los Y-STRs eran muy útiles para comparar poblaciones estrechamente relacionadas en las que posiblemente usando STRs autosómicos no se podrían considerar por separadas (de Knijff P y cols., 1997). Sin embargo, su uso en un contexto evolutivo debería de ser combinado con polimorfismos más estables del cromosoma Y como podrían ser los SNPs (Jobling y cols., 1997), porque la alta tasa mutación de estos STRs, aunque le proporciona unas interesantes características también se convierte en un inconveniente provocando en algunos casos la presencia de haplotipos idénticos por estado que no lo son por descendencia (de Knijff y cols., 2000), homogeneizando así haplotipos de descendientes diferentes. Esto podría ser solventado en parte haciendo un amplio muestreo de la población de estudio para asegurar que todas la líneas paternas comunes de la población a estudiar estén

representadas por clases haplotípicas que contengan perfiles separados por una o unas pocas mutaciones Single-step (Roewer y cols., 2005).

### **3.2.5. Aplicaciones forenses**

En los últimos años, los STRs han sido los marcadores elegidos en genética forense por su alto poder de discriminación y la facilidad para sus análisis. Dentro de la gran cantidad de STRs, los específicos del cromosoma Y son una importante herramienta para la mayoría de los laboratorios que trabajan en genética forense aunque no permitan una individualización con el grado que lo harían los STRs autosómicos, debido a que un haplotipo de Y-STRs es un linaje no recombinante que puede ser compartido por muchos individuos (Gusmao y cols., 2006).

Los Y-STRs se suelen usar en casos de paternidad cuando tenemos un hijo varón pero no disponemos del padre, pero sí de algún pariente masculino de la línea paterna como por ejemplo: un hermano del padre, el abuelo paterno del hijo varón, etc (Kayser y cols., 1997; Roewer y cols., 2000; Gill y cols., 2001). Además son especialmente útiles en casos de agresión sexual ya que los Y-STRs nos permitirán identificar el componente masculino a partir de fluidos corporales donde exista mezcla de células masculinas y femeninas, así como determinar el número de varones implicados en casos de agresiones sexuales múltiples. Otra posible aplicación incluye la detección a partir de células epiteliales en eyaculados de individuos vasectomizados (Kayser, 1997).

El constante descubrimiento de Y-STRs estaba comenzando a suponer un problema desde el punto de vista de la nomenclatura, debido a que no todos los laboratorios nombraban de igual modo al mismo STR y a sus variantes alélicas (White y cols., 1999; Ayub y cols., 2000; Gonzalez-Neira y cols., 2001; Gusmao y cols., 2002). Esto llevó a que la comisión de ADN de la ISFG (Internacional Society of Forensic Genetics) crease unas normas para establecer el uso de una nomenclatura común, algo crucial para permitir la comunicación y el intercambio de datos entre los distintos laboratorios forenses (Gómez y Carracedo, 2000; Gill y cols., 2001; Gusmao y cols., 2006).

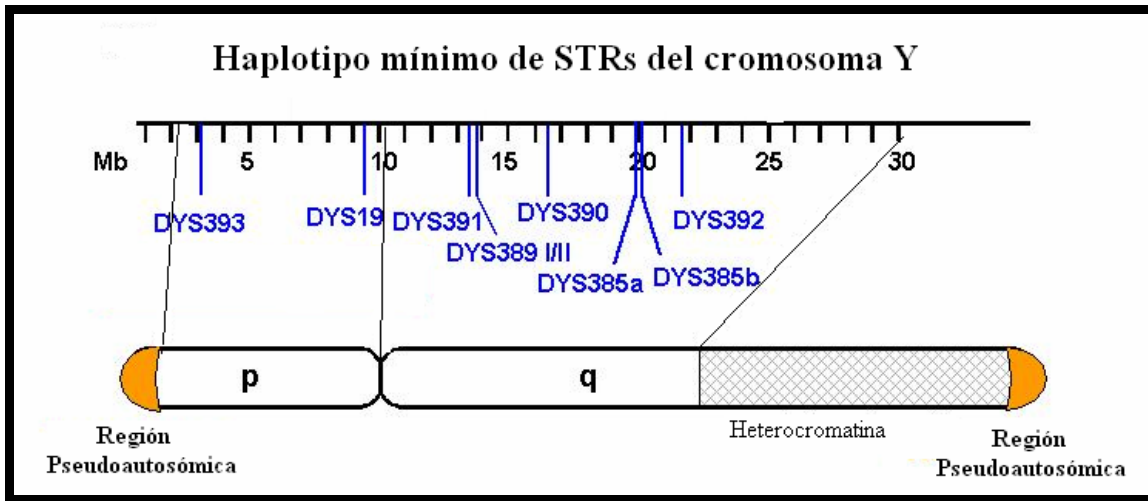
La comisión de ADN de la ISFG publica regularmente nuevas recomendaciones relacionadas con la aplicación de los Y-STRs en la identificación humana además de

otras muchas recomendaciones para otros polimorfismos y modos de procedimiento en el campo de la genética forense (Gill y cols., 2006; Prinz y cols., 2007).

Cuando tenemos coincidencia del haplotipo de Y-STR encontrado, en por ejemplo, una mancha de sangre sobre la víctima y el sospechoso, nos encontramos con la necesidad de conocer la frecuencia de dicho haplotipo en la población para poder calcular la probabilidad de match. Esto implica la necesidad de crear grandes bases de datos con haplotipos completos de Y-STRs de los individuos de la población, para así poder realizar estimas realistas de las frecuencias haplotípicas (Roewer y cols., 2001; Lessig y cols., 2003; Hammer y cols., 2005). Tres de las principales características que debería reunir una base de datos con propósitos forenses serían (Roewer y cols., 2001):

- Disponer de polimorfismos capaces de discriminar entre la mayoría de linajes no relacionados en una población dada.
- Datos representativos de la estructura étnico-geográfica de la población de interés.
- Aspirar a conseguir un tamaño de la base de datos que permita estimar con precisión la frecuencia de los haplotipos raros.

Hoy en día una de las bases de datos más amplia es la YHRD, que inicialmente contenía los 8 Y-STR que constituyen el haplotipo mínimo (ver figura 8) usado en genética forense: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b (Pascali y cols., 1998). Esta base de datos es el fruto de un proyecto internacional surgido en 1994 que buscaba optimizar las posibilidades que ofrecían los Y-STRs al campo forense. Está disponible de forma gratuita en la red <http://www.yhrd.org> y recientemente se le han añadido 8 nuevos Y-STRs: DYS 438, DYS 439, DYS 437, DYS 445, DYS 456, DYS 558, DYS 635, Y GATA H4. El tamaño de la YHRD aumenta día a día y esto ha permitido observar a través de los haplotipos de Y-STRs que en la población europea hay subestructuración ya que se han observado grupos poblacionales que no presentan diferencias significativas dentro del propio grupo pero sí entre grupos (Roewer y cols., 2005). Estos efectos deben ser considerados cuando se lleven a cabo las estimas de las frecuencias haplotípicas (Gusmao y cols., 2005b).



**Figura 8.** Haplotipo mínimo de STRs del cromosoma Y (NIST “short tandem repeat DNA Internet database”).

Desde que fue descubierto el primer STR del cromosoma Y (Roewer y cols., 1992) hasta la actualidad, se han descubierto aproximadamente unos 220 nuevos Y-STRs, pero de todos ellos, los que tienen o pueden tener un mayor interés forense (ver tabla 2) han sido compilados por The National Institute of Standards and Technology (NIST) desde 1997 en una base de datos denominada STRBase, disponible en <http://www.cstl.nist.gov/biotech/strbase/> (Ruitberg y cols., 2001), en la que se puede encontrar gran cantidad de información relacionada con estos STR.

MARCADOR	RANGO DE ALELOS	MOTIVO REPETITIVO	GENBANK
DYS19	10-19	TAGA	AC017019
DYS385 a/b	7-28	GAAA	AC022486
DYS389 I	9-17	(TCTG) (TCTA)	AC004617
DYS389II	24-34	(TCTG) (TCTA)	AC004617
DYS390	17-28	(TCTA) (TCTG)	AC011289
DYS391	6-14	TCTA	AC011302
DYS392	6-17	TAT	AC011745
DYS393	9-17	AGAT	AC006152
YCAII a/b	11-25	CA	AC015978
DYS388	10-18	ATT	AC004810
DYS425	10-14	TGT	AC095380
DYS426	10-12	GTT	AC007034
DYS434	9-12	TAAT (CTAT)	AC002992
DYS435	9-13	TGGA	AC002992
DYS436	9-15	GTT	AC005820
DYS437	13-17	TCTA	AC002992
DYS438	6-14	TTTTC	AC002531
DYS439	9-14	AGAT	AC002992
DYS441	12-18	CCTT	AC004474
DYS442	10-14	TATC	AC004810
DYS443	12-17	TTCC	AC007274
DYS444	11-15	TAGA	AC007043
DYS445	10-13	TTTA	AC009233

MARCADOR	RANGO DE ALELOS	MOTIVO REPETITIVO	GENBANK
DYS446	10-18	TCTCT	AC006152
DYS447	22-29	TAAWA	AC005820
DYS448	20-26	AGAGAT	AC025227
DYS449	26-36	TTTC	AC051663
DYS450	8-11	TTTTA	AC051663
DYS452	27-33	YATAC	AC010137
DYS453	9-13	AAAT	AC006157
DYS454	10-12	AAAT	AC025731
DYS455	8-12	AAAT	AC012068
DYS456	13-18	AGAT	AC010106
DYS458	13-20	GAAA	AC010902
DYS459 a/b	7-10	TAAA	AC010682
DYS460 (A7.1)	7-12	ATAG	AC009235
DYS461 (A7.2)	8-14	(TAGA) CAGA	AC009235
DYS462	8-14	TATG	AC007244
DYS463	18-27	AARGG	AC007275
DYS464 a/b/c/d	11-20	CCTT	X17354
DYS481	20-30	CTT	AC016991
DYS485	10-18	TTA	AC009233
DYS490	12-16	TTA	AC01905
DYS495	12-18	AAT	AC004474
DYS497	13-16	TAT	AC017020
DYS504	11-19	TCCT	AC006157
DYS505	9-15	TCCT	AC012078
DYS508	8-15	TATC	AC006462
DYS510	14-20	TAGA	AC010972
DYS513	10-15	TATC	AC007007
DYS520	18-26	ATAS	AC007275
DYS522	8-17	GATA	AC007247
DYS525	7-13	TAGA	AC010104
DYS531	11-13	AAAT	AC016991
DYS532	9-17	CTTT	AC016991
DYS533	9-14	ATCT	AC053516
DYS534	10-20	CTTT	AC053516
DYS540	10-13	TTAT	AC010135
DYS542	14-18	ATAG	AC007043
DYS544	11-14	GATA	AC010135
DYS549	10-14	AGAT	AC010133
DYS552	23-27	TCTA	AC011302
DYS556	8-12	AATA	AC011745
DYS557	11-22	TTTC	AC007876
DYS561	13-17	GATA (GACA)	AC010726
DYS565	9-14	TAAA	AC010726
DYS570	12-23	TTTC	AC012068
DYS572	8-12	AAAT	AC012078
DYS573	8-11	TTTTA	AC009489
DYS575	8-12	AAAT	AC007247
DYS576	13-21	AAAG	AC010104
DYS587	17-22	ATACA	AC017020
DYS593	15-18	AAAAT	AC010135
DYS594	8-11	TAAAA	AC010137
DYS632	9-10	CATT	AC006371
DYS635 (C4)	17-27	TSTA	AC004772
DYS641	9-11	TAAA	AC018677
DYS643	7-15	CTTTT	AC007007
Y-GATA-H4	8-13 (25-30)	TAGA	AC011751
Y-GATA-C4	20-25	TSTA	G42673
Y-GATA-A10	13-18	TAGA	AC011751

**Tabla 2.** Descripción de los principales STRs de la región no recombinante del cromosoma Y (NIST “short tandem repeat DNA Internet database”; Dai y cols., 2004; Butler y cols., 2006; Gusmao y cols., 2006; Kayser y cols., 2004).

Recientemente se ha podido observar la utilidad que presentan los denominados *mini-STRs*. Su particularidad radica en el inferior tamaño de los amplicones generados en la reacción de PCR multiplex, con respecto al mayor tamaño que usualmente presentan los amplicones de los STRs. Esta posibilidad de amplificar fragmentos más pequeños nos va a permitir examinar y analizar ADN degradado de un modo más eficiente (Butler y cols., 2003). Asamura y cols. (2007) diseñaron el primer multiplex para mini Y-STRs que incluye un total de 8 loci de Y-STR (DYS522, DYS508, DYS632, DYS556, DYS570, DYS576, DYS504 y DYS540).

## **II. Justificación y Objetivos**





En los últimos años las investigaciones sobre los marcadores polimórficos del cromosoma Y han experimentado un gran crecimiento. Se puede hablar básicamente de 2 razones principales por las que es interesante el estudio del cromosoma Y: una de ellas es la especificidad masculina, útil sobre todo para pruebas forenses en las que encontramos mezcla de ADN y, la otra razón, estaría relacionada con la capacidad que presenta este cromosoma para permitirnos rastrear linajes paternos. Dentro de los marcadores del cromosoma Y cabe destacar el papel que juegan tanto los Y-STRs como los Y-SNPs en la genética forense, ya que son aplicados en un gran número de pruebas de identificación humana, incluyendo análisis forenses de evidencias de agresiones sexuales, investigaciones de personas desaparecidas, además de ayudar a complementar pruebas de paternidad deficientes e investigaciones genealógicas. Para todo este tipo de casos los Y-SNPs complementan el papel de los Y-STRs, que son realmente los marcadores de elección por sus características polimórficas, mientras que los Y-SNPs son de los marcadores preferidos por la genética poblacional para intentar arrojar más luz sobre la historia de la humanidad y sus patrones de migración.

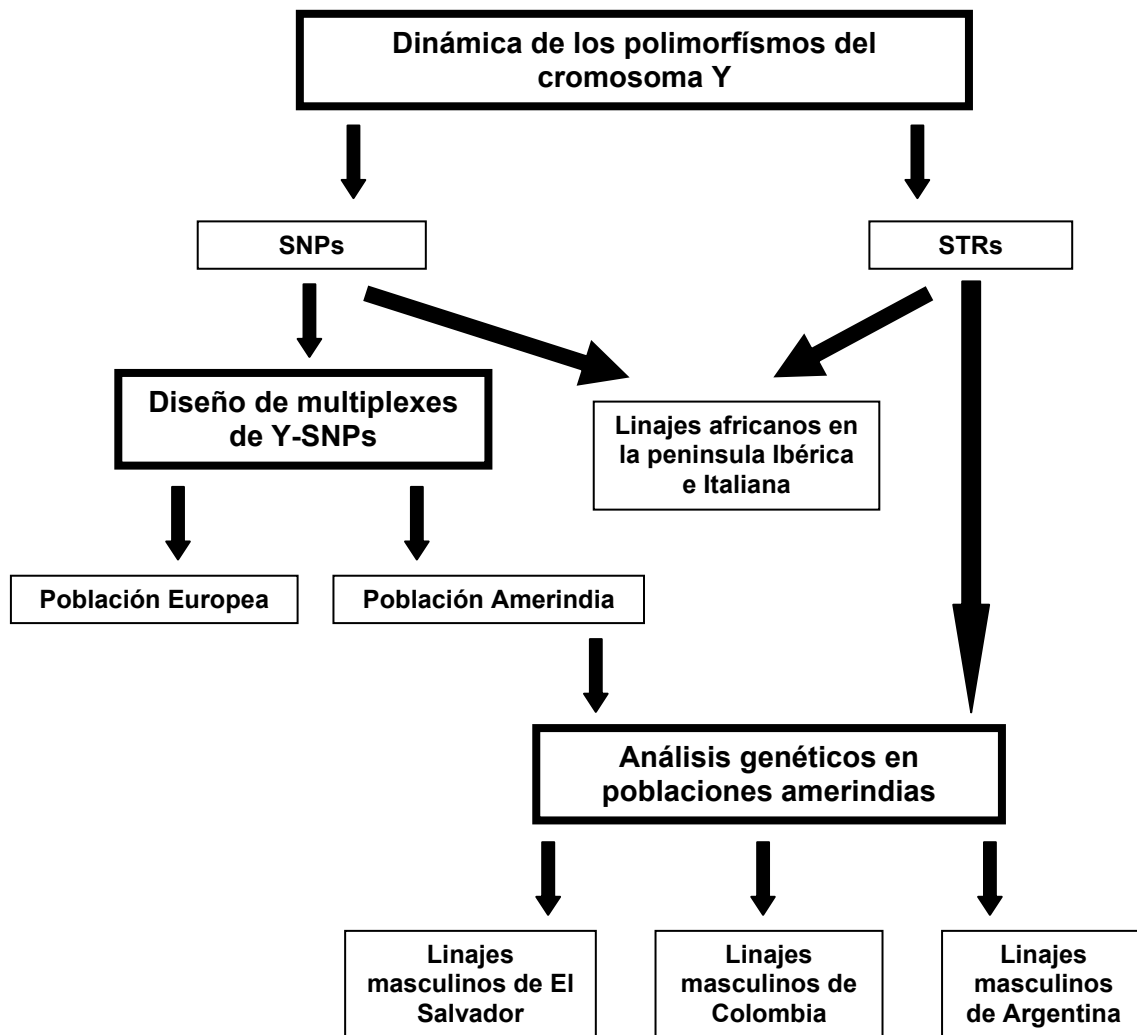
En este trabajo, gracias a las posibilidades que ofrece nuestro laboratorio a la hora de llevar a cabo líneas de investigación tanto en el campo de la genética forense como en el de la genética poblacional, nos planteamos profundizar en el estudio de los polimorfismos del cromosoma Y (SNPs y STRs), tratando de entender la dinámica evolutiva del cromosoma Y actual, como un marcador genético, y de los cromosomas Y actuales, como marcadores de linajes. En este sentido, los objetivos planteados podríamos agruparlos en tres grandes categorías:

- Conocimiento de la dinámica evolutiva de los marcadores del cromosoma Y, centrándonos en el estudio de dos polimorfismos binarios: P25 y 92R7 y en el del microsatélite DYS19.
- Desarrollo de grandes multiplexes que nos permitan incluir en una sola reacción de amplificación el mayor número posible de marcadores.
- Estudio de la variabilidad de los polimorfismos del cromosoma Y para interpretar las historias evolutivas en poblaciones bien diferenciadas del continente americano y del continente europeo.

De forma sintética los objetivos planteados fueron:

1. Estudio de la dinámica evolutiva de los marcadores del cromosoma Y:
  - 1.1. Demostrar que los polimorfismos binarios P25 y 92R7 son variantes de secuencia parálogas originadas a partir de duplicación segmental.
  - 1.2. Uso de diferentes tecnologías para la detección de los alelos de P25 y 92R7.
  - 1.3. Estudio de la correlación cronológica entre los cromosomas que portan la duplicación en el microsatélite DYS19 del cromosoma Y.
2. Desarrollo de multiplexes:
  - 2.1. Desarrollo de una estrategia jerárquica para analizar 30 marcadores binarios del cromosoma Y en 4 multiplexes para detectar los haplogrupos más frecuentes en poblaciones Europeas.
  - 2.2. Desarrollo de un multiplex con 29 Y-SNPs para la identificación de los principales haplogrupos del cromosoma Y.
  - 2.3. Desarrollo de un multiplex de 5 SNPs de cromosoma Y para caracterizar los linajes amerindios.
3. Estudio de la variabilidad de los polimorfismos del cromosoma Y:
  - 3.1. Estudio de la variabilidad de los Y-SNPs en población gallega con la estrategia jerárquica desarrollada en el punto 2.1.
  - 3.2. Estudio de los haplogrupos principales de cromosoma Y en más de 1000 individuos de género masculino no emparentados, distribuidos alrededor del mundo, utilizando el multiplex de 29 SNPs desarrollado en el punto 2.2.
  - 3.3. Estudio de Y-SNPs e Y-STRs en 4 grupos étnicos y en población general de El Salvador.
  - 3.4. Análisis de la distribución de los haplogrupos del cromosoma Y en afro-colombianos y caucásico-mestizos de la región Sur-Occidental de Colombia.
  - 3.5. Estudio de Haplogrupos de SNPs y haplotipos de STRs del cromosoma Y en 3 poblaciones indígenas de Argentina, utilizando el multiplex de linajes amerindios desarrollado en el punto 2.3. de los objetivos.
  - 3.6. Estudio del legado genético masculino del Noroeste africano en dos poblaciones del Sur de Europa

A continuación se presenta un esquema en el que mostramos en grandes bloques los distintos estudios realizados y la conexión que existe entre ellos.





# **III. Material y Métodos**



En este apartado se han descrito más en profundidad los distintos aspectos de material y métodos que no han sido mostrados en los artículos, obviando otros que sí se han analizado en detalle en cada uno de los artículos.

## **1. Muestras**

Las muestras fueron recogidas bajo la supervisión de personal habilitado a tal efecto, en las mejores condiciones higiénico-sanitarias posibles y con la autorización de las personas implicadas en el estudio. En todos los casos las muestras son de sangre procedente de diferentes poblaciones, tanto indígena como población general, de las cuales se seleccionaron los varones sin relación de parentesco. La muestra de sangre es almacenada en forma de mancha sobre un tejido especial o sobre FTA. En los casos que tenemos excedente de sangre se guarda en un congelador de -20 °C para su conservación.

## **2. Extracción de ADN**

### **2.1. Extracción a partir de sangre total**

#### **2.1.3. Protocolo Fenol-Cloroformo-Alcohol Isoamílico**

Basado en el protocolo seguido por Diez y cols., (1992)

- Partimos de alícuotas de sangre congelada a -20°C que se descongela en un baño a 37°C.
- Sometemos la muestra descongelada a un pequeño vortex y luego centrifugamos 2’.
- Extraemos el sobrenadante con una pipeta, pero dejando un poco de sangre sobre el pellet (aprox. 100µl).



- Resuspendemos el pellet de leucocitos en 800µl de tampón de lisis (Tris HCl 50mM, NaCl 150mM, EDTA 100mM, a pH 8).
- Separamos el pellet resuspendido en 2 tubos eppendorf (450µl en el tubo eppendorf original y los otros 450µl en uno nuevo).
- Añadimos 25µl de SDS al 20% en el tubo con el pellet y resuspendemos por inversión.
- Añadimos 6µl de proteinasa K de 20mg/ml e incubamos en el baño 1h. a 60°C.
- Damos un spin para arrastrar al fondo todo el fluido que haya podido quedar por las paredes o en la tapa, al realizar la incubación.
- Añadimos 80µl de perclorato sódico 5M.
- Añadimos 1 vol. (500µl) de fenol:cloroformo:alcohol isoamílico en una relación 25:24:1. Mezclar bien por inversión.
- Centrifugar 10' a 12500-14000 r.p.m.
- Recoger el sobrenadante con una pipeta y pasarlo a otro tubo.
- Añadimos 1 vol. (500µl) de fenol:cloroformo:alcohol isoamílico en una relación 25:24:1 al tubo nuevo con el sobrenadante. Mezclar por inversión
- Centrifugar 10' a 12500-14000 r.p.m.
- Recoger el sobrenadante y pasarlo a un tubo nuevo como hicimos con anterioridad.
- Añadimos 1 vol. de Cloroformo:alcohol isoamílico (500µl), mezclamos por inversión y centrifugamos 10' a 12500-14000 r.p.m.
- Recoger el sobrenadante como antes y pasarlo a un nuevo tubo.
- Añadir 2 vol. (1ml) de etanol absoluto frío (-20°C). Mezclar por inversión.
- Centrifugar de 2' o 3' a 12500-14000 r.p.m. y retirar el etanol por decantación.
- Añadimos 500µl de etanol frío (4°C) al 80%.
- Centrifugar 5' a 12500-14000 r.p.m. y retirar el etanol por decantación o con una pipeta.
- Dejar secar a temperatura ambiente durante 10'.
- Resuspender en agua milliQ estéril.
- Dejar en el baño a 55°C entre 9-12 horas.
- Spin y resuspendemos con una pipeta; podemos mezclar el producto del otro tubo con este ya que es el mismo ADN. Guardar los ADNs a -20°C.

### 2.1.2. Protocolo resina quelante Chelex<sup>®</sup> 100 Resin

El Chelex es una resina de intercambio iónico que captura los iones metálicos multivalentes que catalizarían la digestión del DNA. Nos basamos en el protocolo de Singer-Sam y cols., (1989):

- Partimos de un tubo eppendorf de 1,5ml con 1ml de agua miliQ estéril.
- Se le añaden 4µl de sangre aprox. e incubamos a temperatura ambiente durante 30'. Agitar de vez en cuando.
- Centrifugar 3' a 10000-15000 r.p.m.
- Retirar el sobrenadante con una pipeta hasta dejar 50µl.
- Añadir Chelex al 5% hasta un volumen final de 200µl.
- Incubar en un baño a 56°C durante 30' y luego dar un vortex de 10''.
- A continuación poner el tubo con la muestra en agua hirviendo durante 8'.
- Dar un vortex de 10'' y centrifugar 3' a 10000-15000 r.p.m.
- Recoger el sobrenadante con nuestro ADN y pasarlo a otro tubo eppendorf estéril, el chelex quedaría en el tubo original.

### 2.1.3. Protocolo QIAamp<sup>®</sup> DNA blood mini kit de QIAGEN

Se usan los componentes suministrados con el kit (Proteinase K, Buffer AL, AW1, AW2, AE y columnas QIAamp), es especialmente interesante porque nos permite realizar la extracción en muy poco tiempo.

- Cargamos en un tubo eppendorf de 1,5ml un volumen de 20µl de proteasa QIAGEN.
- Añadimos 200µl de muestra y 200µl de tampón AL. Mezclar por vortex durante 15''.
- Incubar a 56°C durante 10' y luego dar un breve centrifugado para eliminar las gotas que pudieran quedar adheridas a la tapa.
- Añadir 200µl de etanol (96-100%) a la muestra, mezclar mediante vortex durante 15''.
- Cargamos esta mezcla con nuestra muestra en una columna QIAamp que introduciremos en un tubo colector de 2ml.

- Centrifugamos a 8000 r.p.m. durante 1', descartamos este tubo colector que contiene el filtrado y pasamos la columna a un tubo colector nuevo.
- Añadimos 500µl de tampón AW1 a la columna, centrifugamos a 8000 r.p.m. durante 1'. Descartamos el tubo con el filtrado.
- Ponemos la columna en un nuevo tubo colector y añadimos 500µl de tampón AW2, centrifugar a 14000 r.p.m. durante 3'.
- Situamos la columna en un tubo eppendorf de 1,5ml, cargamos 200µl de tampón AE sobre la columna. Dejamos a temperatura ambiente 1'.
- Centrifugamos a 8000 r.p.m. durante 1', y nos queda la muestra extraída en el tubo eppendorf de 1,5ml. Cuantificamos y almacenamos a -20°C.

## **2.2. Extracción a partir de mancha de sangre**

### **2.2.1. Protocolo Fenol-Cloroformo-Alcohol Isoamílico**

- Cortamos aprox. 1cm<sup>2</sup> de mancha de sangre en trozos pequeños y resuspendemos el tejido en 500µl de tampón de lisis (Tris/HCl 1M pH 7.4, NaCl 5 M, EDTA 0.5 M pH 8.0 y agua bidestilada).
- Añadimos 50µl de SDS al 10% y 5µl de proteinasa K (20mg/ml).
- Incubar a 56°C (toda la noche) con agitación suave.
- Añadimos 20µl de NaCl 5M y 575µl de fenol:cloroformo:alcohol isoamílico en una relación 25:24:1 y mezclar por inversión.
- Centrifugar 3' a 12000 r.p.m., pasar el sobrenadante a otro tubo.
- Añadir 575µl de cloroformo:alcohol isoamílico al nuevo tubo con el sobrenadante y centrifugar 3' a 12000 r.p.m.
- Añadir 1ml de etanol frío al 96% y poner el tubo con la muestra a -80°C durante 15'.
- Centrifugar a 12000 r.p.m. durante 15' y decantar.
- Dejar que el precipitado se seque a temperatura ambiente.
- Resuspender el precipitado en 100µl de TE o agua milliQ y poner en baño a 56°C entre 2 y 16 horas.

- Ahora ya estarían listas para cuantificar y almacenar.

### 2.2.2. Protocolo resina quelante Chelex<sup>®</sup> 100 Resin

El protocolo es igual que para sangre total, pero en este caso cortamos aprox. 1 cm<sup>2</sup> de mancha de sangre, en lugar de coger un volumen.

## 3. Cuantificación

### 3.1. Cuantificación fluorimétrica de ADN

Se usó un *Hoefer DyNA Quant 200 Fluorometer<sup>®</sup>* de Pharmacia Amershan GE diseñado específicamente para la cuantificación de concentraciones bajas de ADN usando el marcador Hoechst 33258<sup>®</sup> (también llamado Bisbenzimidida), que experimenta cambios en sus características fluorescentes en función de la cantidad de ADN a la que se una, siendo así la medida de la fluorescencia un indicador directo de la concentración de ADN.

Para un correcto calibrado del instrumento necesitamos un estándar (solución control) que le proporcione un punto de referencia.

Solución control (10 mL):

H33258: 1,0 µL  
TNE (10X): 1,0 mL  
H<sub>2</sub>O: 9,0 mL

Cuantificación:

1. Encender el equipo y esperar 15 minutos para que se estabilice.
2. Cargar 1 ml de solución control en el aparato, pulsar “Zero”.

3. Calibrar el equipo con 1 ml de solución control y 1  $\mu$ l de estándar (100 ng/ml), pulsar “Zero”.
4. Repetir el paso número 2.
5. Para medir las muestras cargamos 1 ml de solución control y 1  $\mu$ l de muestra, pulsar “Zero”.
6. Repetir los pasos 2 y 5 para el resto de las muestras.

### 3.2. Cuantificación espectrofotométrica del ADN

Usamos el *NanoDrop*<sup>®</sup> *ND-1000 spectrophotometer* que permite la realización de una cuantificación muy precisa de ácidos nucleicos partiendo tan sólo de 1 $\mu$ l de muestra. Una de sus principales características es su remarcada reproducibilidad en los resultados.

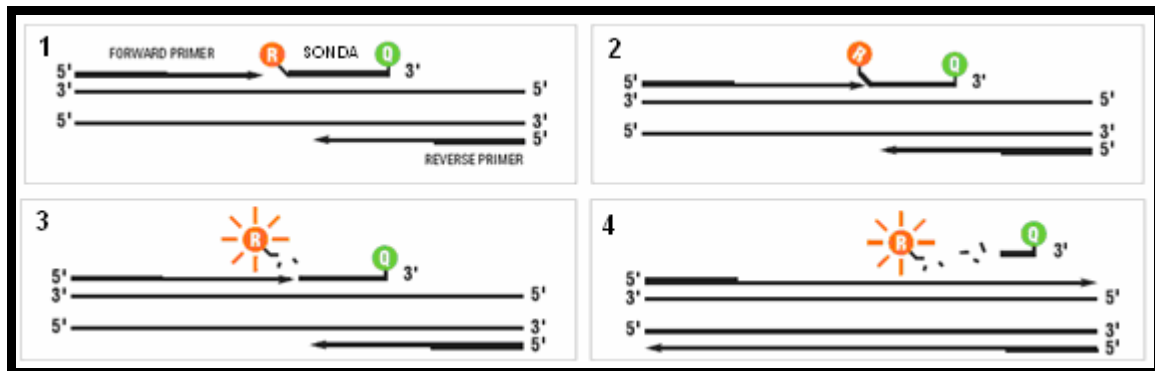
Este aparato combina la tecnología de la fibra óptica junto con las propiedades inherentes a la tensión superficial que presentan las muestras líquidas para desarrollar un sistema de retención de la muestra capaz de medir la absorbancia sin la necesidad de las tradicionales cubetas o capilares. La muestra líquida quedaría retenida entre 2 superficies ópticas debido a su tensión superficial durante el tiempo de medida, que dura aproximadamente 10seg., pudiendo recuperar la muestra posteriormente.

Este espectrofotómetro utiliza un rango de longitud de onda comprendido entre 220 y 750nm. El límite de detección se mueve entre 2 y 3700ng/ $\mu$ l para el ADN de doble cadena, lo que reduce la necesidad de hacer diluciones de la muestra a cuantificar.

### 3.3. Quantifiler<sup>™</sup> Human DNA Quantification Kit y Quantifiler<sup>™</sup> Y Human Male DNA Quantification Kit

El funcionamiento de este sistema de cuantificación del DNA se basa en el uso combinado de una enzima 5' nucleasa (Real-Time PCR) y la tecnología TaqMan<sup>®</sup>. Partimos de una sonda TaqMan<sup>®</sup> que en el extremo 3' lleva un *quencher* y en el 5' lleva un *reporter* fluorescente (ver figura 9). Mientras la sonda está intacta el *reporter* no emite fluorescencia porque se lo impide el *quencher*. Durante cada ciclo de extensión, la

DNA polimerasa corta el reporter marcado de la sonda, quedando liberado del quencher y comenzando así a emitir su fluorescencia. La PCR en tiempo real nos permite monitorizar la acumulación del producto de PCR marcado con fluorescencia ciclo a ciclo (Heid y cols., 1996; Green y cols., 2005).



**Figura 9.** Pasos de la reacción Real-time PCR. (R) reporter, (Q) quencher.

La amplificación es robusta, específica e identifica muestras con concentración de ADN humano entre los rangos de 0,023 µg/L y >50 µg/L.

Una de las principales ventajas es la amplia variedad de tipos de muestras de las que podemos partir para la identificación humana incluyendo: saliva, sangre, semen y muestras de tejido. Además, el protocolo de cuantificación es compatible con varios métodos de extracción de ADN comúnmente usados incluyendo Fenol-cloroformo, Chelex, etc.

El quantifiler Y amplifica una diana específica presente en el cromosoma Y humano suministrando de este modo una información adicional cuando analizamos muestras que podrían contener mezcla de ADN masculino y femenino.

Los componentes del kit son idénticos para quantifiler y quantifiler Y, sólo varía el mix de primers que para quantifiler son unas sondas que amplifican el gen de la reverso transcriptasa de la telomerasa humana, mientras que para el quantifiler Y las sondas están diseñadas para amplificar el gen SRY.

## 4. Amplificación

Las condiciones de PCR suelen variar para cada multiplex (Shuber y cols., 1995; Henegariu y cols., 1997) o para cada singleplex. En cada uno de los artículos incluidos en esta tesis se describen perfectamente las condiciones en las que fue realizada cada amplificación. Es interesante resaltar la eficacia de un nuevo producto llamado QIAGEN<sup>®</sup> Multiplex PCR Kit, que está especialmente diseñado para minimizar la necesidad de optimizar una PCR multiplex. El kit contiene una mezcla de: ADN Taq Polimerasa HotStar, Cl<sub>2</sub>Mg, dNTPs y un buffer, todo en concentraciones preoptimizadas.

## 5. Purificación

En algunas ocasiones no obtenemos producto amplificado aunque partamos de muestra cuantificada. Esto puede ser debido a la presencia de algún inhibidor en el producto extraído. Uno de los modos de eliminar muchos de estos posibles inhibidores es usar un mecanismo de filtrado por centrifugación como el CENTRICON<sup>®</sup> Centrifugal Filter Devices de MILLIPORE. Consiste en una columna de ultrafiltrado con una membrana hidrofílica de celulosa de baja absorción en la que quedarían retenidos los inhibidores después de una centrifugación a 1500 r.p.m.

### 5.1 Protocolo CENTRICON<sup>®</sup> Millipore

- Añadir 500 µl de TE en un CENTRICON<sup>®</sup>.
- Añadir la muestra a purificar y 1 ml más de TE.
- Centrifugar a 4000 r.p.m. durante 30´.
- Añadir 2 ml TE y 30´ a 4000 r.p.m. (sino baja lo suficiente centrifugar 20´ más).
- Invertir la columna del CENTRICON<sup>®</sup> colocando el tubo recolector del ADN.
- Desechar el vial de filtrado con los restos de filtrado.
- Centrifugar (con el CENTRICON<sup>®</sup> invertido) a 4000 r.p.m. durante 5´.

- Obtención de la muestra en el vial de retención. Se puede pasar el ADN a un tubo eppendorf o guardar en el propio vial, para lo cual dispone de tapa.

## **6. Detección de marcadores de cromosoma Y**

### **6.1. Detección de STRs**

Para el estudio de los STRs se han usado dos kits comerciales de amplificación. Inicialmente usamos el kit Powerplex<sup>®</sup> Y de Promega que incluye 11 STRs: DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438 y DYS439. Posteriormente se sustituyó este kit por otro de la casa comercial Applied Biosystems denominado AmpFISTR<sup>®</sup> Y Filer<sup>™</sup>, contiene 16 STRs compartiendo 11 con el Kit de Promega: DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS456, DYS458, DYS635, GATA H4, DYS448.

Para detectar estos STRs, una vez amplificados, se ha utilizado un secuenciador automático de ADN, ABI PRISM<sup>®</sup> 3100 Genetic Analyzer de Applied Biosystems que posteriormente fue adaptado para convertirse en un ABI PRISM<sup>®</sup> 3130xl Genetic Analyzer. Los softwares para la lectura de los resultados fueron GeneScan<sup>®</sup> Análisis Version 3.7 de Applied Biosystems en el secuenciador ABI PRISM<sup>®</sup> 3100 y GeneMapper<sup>®</sup> ID Software v3.2 también de Applied Biosystems en el secuenciador más evolucionado ABI PRISM<sup>®</sup> 3130xl, que con la ayuda de los ladder alélicos que vienen en cada uno de los kits nos permite asignar la variante alélica para cada uno de los sistemas en cada muestra.

### **6.2. Detección de SNPs**

En la detección de los SNPs se usaron los mismos secuenciadores automáticos y los mismos softwares que en el caso de los STRs.



Una vez realizada la amplificación de los SNPs por PCR tenemos que llevar a cabo una reacción de minisequenciación (Pastinen y cols., 1997), en la que se produce la extensión de un sólo nucleótido en la sonda utilizada, mediante el kit comercial SNaPshot<sup>®</sup> Multiplex System de Applied Biosystems.

Antes de iniciar la reacción de minisequenciación necesitamos purificar el producto amplificado, para ello usamos el compuesto ExoSAP-IT<sup>®</sup> de USB<sup>®</sup> que contiene 2 enzimas hidrolíticas:

- Exonucleasa I: enzima encargada de degradar primers residuales monocatenarios y cualquier producto de ADN monocatenario extraño producido en la reacción de PCR.
- Shrimp Alkaline Phosphatase: hidroliza los restos de dNTPs no utilizados durante la amplificación.

Esta mezcla de enzimas se carga directamente sobre el producto amplificado y se deja incubando a 37°C durante 15', luego se procede a la inactivación del enzima a 85°C durante otros 15'.

A continuación, se realiza la reacción de minisequenciación y, antes de cargar el producto de esta reacción en el secuenciador automático, necesitamos realizar una nueva purificación, ahora con SAP<sup>®</sup> de USB<sup>®</sup>, que contiene solamente la enzima Shrimp Alkaline Phosphatase encargada de eliminar por defosforilación los ddNTPs no incorporados durante la reacción de minisequenciación. Necesita una incubación a 37°C durante 30'- 60' y otro de inactivación a 85°C durante 15'.

## 7. Análisis estadísticos

Los cálculos estadísticos fueron llevados a cabo con el programa ARLEQUIN ver 2.000 (Schneider y cols., 2000). ARLEQUIN proporciona una gran cantidad de métodos y pruebas estadísticas que nos permiten extraer información de las características genéticas y demográficas de nuestras muestras poblacionales.

Los análisis de Arlequin pueden ser realizados sobre datos englobados principalmente en 2 categorías: categoría intra-poblacional y categoría inter-poblacional. En la primera categoría la información es extraída independientemente de

cada población, mientras que en la segunda categoría las muestras son comparadas entre ellas.

Los análisis que nosotros realizamos con nuestras muestras para aprovechar la información que nos podían proporcionar Y-STRs y Y-SNPs fueron:

Análisis intra-poblacionales:

- Índices de diversidad estándar.
- Estimaciones de las frecuencias haplotípicas.

Análisis inter-poblacionales:

- Haplotipos compartidos entre poblaciones.
- Análisis AMOVA (análisis de la varianza molecular) para evaluar la cantidad de estructuración genética que presenta la población (Excoffier y cols., 1992).
- Distancias genéticas basadas en el estadístico  $F_{ST}$  para tiempos de divergencia cortos y análisis de la diversidad en microsatélites con el estadístico  $R_{st}$  (Slatkin, 1995).
- Test Exacto de diferenciación poblacional basado en frecuencias haplotípicas, que nos permite observar la distribución no aleatoria de los haplotipos de las muestras dentro de la población.

Además de usar el programa Arlequín se recurrió a otros para realizar los siguientes análisis:

- *Análisis de Componentes Principales* (ACP) es una técnica estadística de síntesis de la información o reducción de la dimensión (número de variables), es decir, ante un conjunto de datos con muchas variables, el objetivo sería reducirlas en número perdiendo la menor cantidad posible de información. Estos nuevos componentes principales son una combinación lineal de las variables originales y además también son independientes entre sí. Para el Análisis de Componentes Principales nosotros hemos recurrido al software SPSS versión 15.0.
- *Escalonamiento Multidimensional* (MDS) generalmente se usa para generar representaciones gráficas en dos o tres dimensiones de las distancias

interpoblacionales o intermoleculares usando el dato crudo de las frecuencias alélicas o las distancias genéticas. Nosotros hemos llevado a cabo los análisis MDS de las distancias  $F_{ST}$  y  $R_{ST}$  usando el software SPSS version 15.0.

- *Análisis Networks* son representaciones gráficas para visualizar las relaciones evolutivas entre entidades biológicas. Un simple network contiene varios árboles filogenéticos. De este modo, si dos árboles son parecidos y se apoyan en sus datos no tendríamos que elegir entre uno u otro, ya que se pueden sumarizar en un network. Los networks representan más la información contenida en los datos de lo que podría mostrar cualquier árbol filogenético. Pueden ser contruidos a partir de distancias o datos de caracteres, basándonos en estas características nos encontramos con un número de métodos alternativos de construcción:
  - Método de split decomposition (basado en distancias genéticas).
  - Método minimum spanning networks (basado en datos de caracteres, como pueden ser los haplotipos de STRs).
  - Método de median networks (basado en datos de caracteres).

Los network filogenéticos han sido calculados con el programa Network 4.1.1.2 (<http://www.fluxus-engineering.com/>).

# **IV. Resultados**



## **Bloque I. Dinámica evolutiva de los STRs y SNPs del cromosoma Y**

A lo largo del cromosoma Y es frecuente encontrar regiones con loci duplicados que van a producir un fenómeno de doble señal después de amplificar estos loci en PCR.

Estos dobles haploides tienen mucho interés desde un punto de vista poblacional, evolutivo y también genético forense. Este interés implica la necesidad de un profundo conocimiento de la historia evolutiva de las duplicaciones del cromosoma Y.

El fenómeno de la duplicación afecta tanto a STRs como a SNPs, que han sido los polimorfismos del cromosoma Y elegidos para los distintos estudios realizados en esta tesis.

Se ha sugerido recientemente que aproximadamente el 40% del cromosoma Y está constituido por duplicaciones inter e intracromosómicas, con un 90-98% de similitud entre las secuencias duplicadas (Bailey y cols., 2002). También se ha estimado que alrededor del 10% de los SNPs reportados en las bases de datos son variantes de secuencias parálogas (PSVs) originadas de duplicación segmental, por tanto, en la posible aplicación forense de los SNPs, es muy importante elegir los que tengan posiciones únicas en los cromosomas para que no nos induzcan a error a la hora de interpretar los resultados.

Esto nos ha llevado a profundizar en el estudio de dos Y-SNPs, que además son de los más usados en estudios poblacionales: los PSVs 92R7 y P25. En este trabajo, uno de los objetivos era recurrir a diferentes metodologías para confirmar las duplicaciones y su localización.

El segundo estudio de duplicación se centra en el DYS19 que fue el primer STR descubierto del cromosoma Y (Roewer y cols., 1992). Actualmente se dispone de gran cantidad de datos en la literatura científica. Nosotros hemos hecho una recopilación de éstos para comprobar, mediante análisis network y genotipado de SNPs, si las duplicaciones del DYS19 se produjeron más de una vez durante la evolución humana.



**Artículo 1: “Duplicaciones de los loci P25 y 92R7 específicos del cromosoma Y y sus implicaciones forenses”**

Sanchez JJ, Brión M, Parson W, **Blanco-Verea AJ**, Borsting C, Lareu M, Niederstätter H, Oberacher H, Morling N, Carracedo A.  
Forensic Science International **140** (2004): 241-250

Resumen:

En el presente estudio, demostramos que 2 polimorfismos binarios del cromosoma Y, P25 y 92R7, son variantes de secuencias parálogas (PSVs) originadas por duplicación segmental; por lo menos una de las variantes de secuencia en cada grupo de loci es polimórfica. Se usaron diferentes metodologías para detectar los alelos de cada SNP y las PSV de los loci. Todos los resultados obtenidos con las distintas técnicas apoyan la misma conclusión. La distribución de los alelos de los marcadores binarios se analizó en más de 600 hombres que pertenecen a 7 haplogrupos diferentes. Para P25, el alelo ancestral C se encontró en varias muestras de diferentes haplogrupos, mientras que el alelo derivado A siempre estuvo presente con una variante adicional C. El haplogrupo P está definido por el alelo derivado A del SNP 92R7. Sin embargo, el alelo ancestral G siempre va asociado con una variante A debido a la duplicación.







ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Forensic Science International 140 (2004) 241–250

**Forensic  
Science  
International**
[www.elsevier.com/locate/forsciint](http://www.elsevier.com/locate/forsciint)

## Duplications of the Y-chromosome specific loci P25 and 92R7 and forensic implications

Juan J. Sanchez<sup>a,\*</sup>, Maria Brión<sup>b,1</sup>, Walther Parson<sup>c,1</sup>, Alejandro J. Blanco-Verea<sup>b</sup>,  
Claus Børsting<sup>a</sup>, Maviki Lareu<sup>b</sup>, Harald Niederstätter<sup>c</sup>, Herbert Oberacher<sup>c</sup>,  
Niels Morling<sup>a</sup>, Angel Carracedo<sup>b</sup>

<sup>a</sup>Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen,  
11 Frederik V's Vej, DK-2100 Copenhagen, Denmark

<sup>b</sup>Institute of Legal Medicine, University of Santiago de Compostela, E-15705 Santiago de Compostela, Spain

<sup>c</sup>Institute of Legal Medicine, University of Innsbruck, A-6020 Innsbruck, Austria

Received 30 July 2003; received in revised form 11 October 2003; accepted 20 November 2003

### Abstract

In the present study, we demonstrate that two commonly used Y-chromosome single nucleotide polymorphisms (SNPs), P25 and 92R7, are paralogous sequence variants (PSVs) originating from segmental duplications and that at least one of the sequence variants in each group of loci is polymorphic. Several methodologies were used in order to detect the SNP alleles and the PSVs of the loci. All results obtained with the various typing techniques supported the conclusion. The allele distributions of the binary markers were analysed in more than 600 males with seven different haplogroups. For P25, the ancestral allele C was found in several samples from different haplogroups. The derived allele A was always present with an additional C variant. Haplogroup P was defined by the derived allele A at the 92R7 locus. However, the ancestral allele G was always associated with an A variant due to the duplication.

© 2004 Elsevier Ireland Ltd. All rights reserved.

**Keywords:** Y-chromosome; Biallelic polymorphisms; SNP; PSV; P25; 92R7

### 1. Introduction

Single nucleotide polymorphisms (SNPs) are promising markers for forensic analysis due to some special characteristics of these simple polymorphisms, most notably, the possibility to be analysed in short amplicons, making these markers very appropriated for degraded samples. In addition, SNPs exhibit low mutation rates and give the possibility of automation and high-throughput analysis. Y-chromosome SNPs are promising markers for use in complicated male–female mixture cases [1] and although Y-chromosome STRs are more informative and therefore are the markers of first

choice in forensic cases, the complementary information provided by Y-chromosome SNPs can be very useful.

In the last few years, there has been a significant increase in the use of Y-chromosome specific markers, especially binary polymorphisms (mostly SNPs) in population and evolutionary studies. The use of these polymorphisms to trace paternal lineages is particularly useful for understanding the male roles in migration and admixture [2–5].

Binary markers are common in the non-recombining region of the Y-chromosome. Haplogroups constructed by using binary polymorphisms have proved to be very useful in the investigation of evolutionary events [4–7] and they have a potential interest in forensic analysis since, although more prone to population stratification, this characteristic make them good markers for predicting ethnicity.

Since different Y-chromosome SNPs can define the same haplogroup, it is important to select the most appropriate ones for forensic purposes. Genome and SNP databases are

\* Corresponding author. Tel.: +45-35-32-61-10;

fax: +45-35-32-61-20.

E-mail address: [juan.sanchez@forensic.ku.dk](mailto:juan.sanchez@forensic.ku.dk) (J.J. Sanchez).

<sup>1</sup>The first three authors were the main contributors and contributed equally to the work.

essential for solving this task. It was recently suggested that as much as 40% of the Y-chromosome are inter- or intra-chromosomal duplications with 90–98% sequence identity [8]. The number of duplicated regions with more than 98% identity is difficult to estimate because four out of five such alignments have been shown to be false positives [9,10]. On the other hand, sequences with more than 98% identity may have been considered to be allelic copies and not independent loci during the assembly of genome databases, and consequently, some sequences have not been included in the databases [11]. Single nucleotide differences between nucleotides at the same position in duplicated regions are often erroneously reported as SNPs. It was recently estimated that as many as 10% of the reported SNPs in the databases are paralogous sequence variants (PSVs) originating from segmental duplications [12]. For forensic purposes, we need to choose SNPs mapped to unique positions on the chromosomes.

Here, we report that two of the most widely used Y-chromosome SNPs in population studies, P25 previously described in association with the DYS194 locus [13] and 92R7 described by Hurles et al. [3] are PSVs and that at least one of the sequence variants in each group of loci is polymorphic. Several methodologies were implemented in order to confirm the duplications and to locate them. The variability of both loci and their paralogous fragments were analysed in various haplogroups to determine the ancestral state of the mutations.

2. Material and methods

2.1. Samples

Samples from an EDNAP exercise were collected from laboratory staff under informed consent and previous accep-

tance of the local ethical committee. Additional samples to analyse the duplication in depth were collected from local ethically approved databases (unrelated Galicians, Danes and other population groups). DNA was extracted using conventional phenol–chloroform extraction or using the QIAmp DNA Blood Mini Kit (Qiagen).

2.2. PCR amplification of 92R7 and P25

PCR amplification was performed according to the primers and conditions described in Tables 1 and 2.

2.3. PCR amplification of autosomal STR loci

PCR amplification of autosomal STR loci was performed on 1 ng chromosomal DNA using the AmpFISTR Profiler Plus kit (Applied Biosystems) as recommended by the manufacturer.

2.4. RFLP analysis

Digestions with *Cac8I* (New England BioLabs) for P25 and *HindIII* (Amersham Biosciences) for 92R7 were performed as recommended by the manufacturers. The digested PCR products (Table 3) were separated by electrophoresis with an ALF express II sequencer and fragment sizes were determined either automatically using the AlleleLinks 1.00 software or in an 11% polyacrylamide gel in 1× TBE buffer using a 10 bp DNA ladder (Invitrogen) as marker.

2.5. Minisequencing analysis

PCR amplified fragments were either cleaned by MinE-lute PCR purification columns from Qiagen or mixing of 15 µl of the PCR product with 5 U SAP (shrimp alkaline

Table 1  
Primer sequences

	Primer sequences (5'–3')	
	P25 (AC016698)	92R7 (AC006986)
<b>Santiago de Compostela</b>		
Amplification	Forward (121 bp) GGACCATCACCTGGGTAAAGT Reverse AGTGCTTGTCCAAGGCAGTA	Forward (55 bp) TGCATGAACACAAAAGACGTA Reverse GCATTGTAAATATGACCAGC
Minisequencing	Forward 7C + TCTGCCTGAAACCTGCCTG	Reverse 4C + GCATGAACACAAAAGACGTAGAAG
<b>Copenhagen</b>		
Amplification	Forward (109 bp) TGGACCATCACCTGGGTAAAGT Reverse GGCAGTATAAGGTTGTCACATCACAT	Forward (89 bp) TTAAATCCCTCCTATTTGTGCTAACC Reverse AATGCATGAACACAAAAGACGTAGA
Minisequencing	Forward TGCCTGAAACCTGCCTG Reverse CGAGATACGAGCACAAATTCTATTT	Reverse CATGAACACAAAAGACGTAGAAG
<b>Innsbruck</b>		
Amplification	None	Forward (72 bp) CAAATTAGCATTGTAAATATGACCAGCAAAGACA Reverse TCAAGAAAAATGCATGAACACAAAAAGACGTAGAAG

Table 2  
PCR amplification conditions

	Santiago de Compostela	Copenhagen	Innsbruck
Buffer	1×	1×	1×
DNTPs	200 μM	200 μM	200 μM
MgCl <sub>2</sub>	1.5 mM	1.5 mM	1.5 mM
Primers	0.25 μM each one	0.2 μM each one	0.25 μM each one
Taq DNA polymerase	0.5 U	–	–
AmpliTaq gold polymerase	–	0.6 U	2.5 U
DyNAzyme EXT DNA polymerase	–	1.0 U	–
DNA	5 ng	1 ng	10 ng
Cycling conditions	30 cycles: 94° 1', 58° 30', 72° 30''	95° 10', 31 cycles: 95° 30', 60° 30'', 72° 30''	95° 10', 40 cycles: 94° 30'', 60° 30'', 72° 30''

Table 3

The recognition sites of the restriction enzymes *Cac8I* (P25) and *HindIII* (92R7) are in capital letters. The arrow indicates the site of cleavage

P25

Allele C: *Cac8I* digested fragment (44 + 77 bp): ggaccatcacctgggtaaagtgaattatctgcctgaaacctGCC↓TGCAaatagaattgtctcgtatctcggccatc-  
atgtatgtgatgtgacaaccttatactgccttgacaagcact

Allele A: Undigested fragment (121 bp): ggaccatcacctgggtaaagtgaattatctgcctgaaacctGCC TGAaaatagaattgtctcgtatctcggccat-  
catgtatgt gatgtgacaaccttatactgccttgacaagcact

92R7

Allele G: *HindIII* digested fragment (28 + 27 bp): gcattgttaaataatgaccagcaaaagacA↓AGCTTctacgtcttttgtttcatgca

Allele A: Undigested fragment (55 bp): gcattgttaaataatgaccagcaaaagacAAACTTctacgtcttttgtttcatgca

phosphatase, Amersham Pharmacia Biotech) and 2 U Exo I (Applied Biosystems) at 37° for 1 h.

The minisequencing reaction was performed with the SNaPshot multiplex kit (Applied Biosystems) according to the manufacturer's recommendations with the primers described in Table 1 in a concentration of 0.2 μM. The following programme was used: 25 cycles of 96 °C for 10 s, 50 °C for 5 s, 60 °C for 30 s. To remove the unincorporated ddNTPs, the product was cleaned with 1 U SAP.

Fragments were investigated on an ABI 3100 genetic analyser and the results were analysed using the Genescan 3.7 and Genotyper 2.0 software (Applied Biosystems).

## 2.6. Hybridisation on NanoChip microarray

The physical properties of the NanoChip microarray have been described elsewhere [14,15]. A total of 7–10 μl biotinylated PCR product or 1–6 nM biotinylated control oligo (Table 4) was diluted in 50 mM histidine in a total volume of

Table 4

Oligonucleotides for hybridisation on NanoChip microarrays

P25	
Reporters	5'CY3-ACCTGCCTGC 5'CY5-AACCTGCCTGA
Stabilizer	AAATAGAATTGTGCTCGTATCTCGGTCCAT
Controls	5'Biotin-ATGATGGACCGAGATACGAGCACAAATTCTATTTTCAGGCAGGTTTCA 5'Biotin-ATGATGGACCGAGATACGAGCACAAATTCTATTTTCAGGCAGGTTTCA
92R7	
Reporters	5'CY3-GACGTAGAAGC 5'CY5-AGACGTAGAAGT
Stabilizer	TTGTCTTTGCTGGTTCATATTTTACAATGCTAATTTGGTTA
Controls	5'Biotin-GTTAAATATGACCAGCAAAGACAAGCTTCTACGTCTTTTGTGTTTCATG 5'Biotin-GTTAAATATGACCAGCAAAGACAAGCTTCTACGTCTTTTGTGTTTCATG

60  $\mu$ l. The DNA was electronically addressed to individual pads on the NanoChip using the Nanogen Molecular Biology Workstation (Nanogen). Loading time for samples and control oligos was 3 and 2 min, respectively. After loading, the NanoChip was treated with 0.1 M NaOH for 3 min to remove the un-labelled strand of the PCR product. The NanoChip was washed three times in 150  $\mu$ l high salt buffer (500 mM NaCl, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.0). The last wash was left in the NanoChip for 5 min to ensure that the permeation layer above the electrodes was saturated with high salt buffer. One micromolar of each fluorescently labelled reporter oligo and 250 nM stabiliser oligo (Table 4), diluted in 150  $\mu$ l high salt buffer, was added to the NanoChip, and hybridisation was performed at room temperature for 5 min. After the hybridisation, the NanoChip was washed three times in 150  $\mu$ l high salt buffer. The NanoChip was loaded in the Nanogen Molecular Biology Workstation and the NanoChip was washed with 150  $\mu$ l low salt buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7.0). The fluorescent signal was scanned at different temperatures using the laser and the detector on the Nanogen Molecular Biology Workstation [14,15]. The NanoChip was washed twice with low salt buffer after the NanoChip was warmed up to a new temperature. The optimum temperatures for P25 and 92R7 were 37–39 and 38–40 °C, respectively. The data were collected and analysed with the nLAB 1.09.09c software (Nanogen).

### 2.7. Ion-pair reversed phase chromatography–electrospray ionisation mass spectrometry (ICEMS)

Molecular masses of the 92R7 amplification products were determined using ICEMS as described previously [16]. The measured masses are given in Table 5.

Acetonitrile (HPLC gradient-grade) was obtained from Merck (Darmstadt, Germany). Stock solutions (0.50 M) of butyldimethylammonium bicarbonate (BDMAB) was prepared by passing carbon dioxide gas (AGA, Vienna, Austria) through a 0.50 M aqueous solution of butyldimethylamine (Fluka, Buchs, Switzerland) at 5 °C until pH 8.4 was

Table 5  
Measured molecular masses of the 92R7 amplification products of the EDNAP samples #3 and #4

	Measured molecular mass	Assignment
EDNAP sample #3	22,095	A <sub>for</sub>
	22,112	G <sub>for</sub>
	22,247	G <sub>rev</sub>
	22,260	A <sub>rev</sub>
EDNAP sample #4	22,096	A <sub>for</sub>
	22,258	A <sub>rev</sub>

The strand assignment was deduced from the theoretical masses. #3 shows 2 peaks (A/G alleles), whereas #4 displayed only the A allele.

reached. For preparation of all solutions, HPLC-grade water (Merck) was used.

The HPLC system consisted of a low-pressure gradient micro pump (model Rheos 2000, Flux Instruments, Base, Switzerland) controlled by a personal computer, a vacuum degasser (Knauer, Berlin, Germany), a column thermostat made from 3.3 mm o.d. copper tubing which was heated by means of a circulating water bath (model K 20 KP, Lauda, Lauda-Königshofen, Germany), and a microinjector (model C4-1004, Valco Instruments Co. Inc., Houston, TX, USA) with 500 nl internal sample loop. ESI-MS was performed on a Bruker Daltonics Esquire 3000+ quadrupole ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ion source. The 60 mm  $\times$  0.2 mm i.d. monolithic capillary column was prepared according to the published protocol [17] and connected directly to the spray capillary (fused silica, 90  $\mu$ m o.d., 20  $\mu$ m i.d., Polymicro Technologies) by means of a microtight union (Upchurch Scientific, Oak Harbor, WA, USA). A syringe pump equipped with a 250  $\mu$ l glass syringe (Unimetrics, Shorewood, IL, USA) was used for the post-column addition of a 3.0  $\mu$ l/min flow of acetonitrile by means of a tee-piece. For analysis with pneumatically assisted ESI, an electrospray voltage of 3.0 kV, a dry gas pressure of 30 p.s.i., and a nebulizer gas flow of 4 l/min were employed. The temperature of the dry gas was set to 300 °C. Total ion chromatograms and mass spectra were recorded on a personal computer with the Bruker Daltonics esquire software Version 5.0 (Bruker Daltonics).

Mass calibration and coarse tuning were performed in the positive ion mode by direct infusion of a solution of Electrospray Calibrant Solution (Agilent Technologies, Palo Alto, CA, USA). Fine tuning for ESI-MS of oligodeoxynucleotides in the negative ion mode was performed according to [18].

### 2.8. Sequencing analysis

PCR amplified fragments were purified with Microspin S-300 HR columns (Amersham Pharmacia Biotech). A dideoxy cycling sequencing reaction was carried out using dRhodamine Terminator Cycle sequencing kit and BigDye Terminator Cycle sequencing kit (Applied Biosystems). The products were purified using an MgCl<sub>2</sub>/ethanol based protocol and ran in a 6% denaturing gel on an ABI 377 sequencer, or an ABI 3100 genetic analyser (Applied Biosystems). The results were analysed using the Data Collection Software 377-18.

### 2.9. SNP haplogroups

Samples used to confirm the duplication were previously typed [19] for eight binary markers (M9, SRY-2627 [M167], 12f2, YAP [M1], SRY-8299 [M40], sY81 [M2], Tat [M46], LLY22g) and classified in seven haplogroups [20,21]: P\*(xR1b3f,R1a), BR\*(xDE,JK), R1a, J, E\*(xE3a), R1b3f and K\*(xN,P) (Fig. 1).

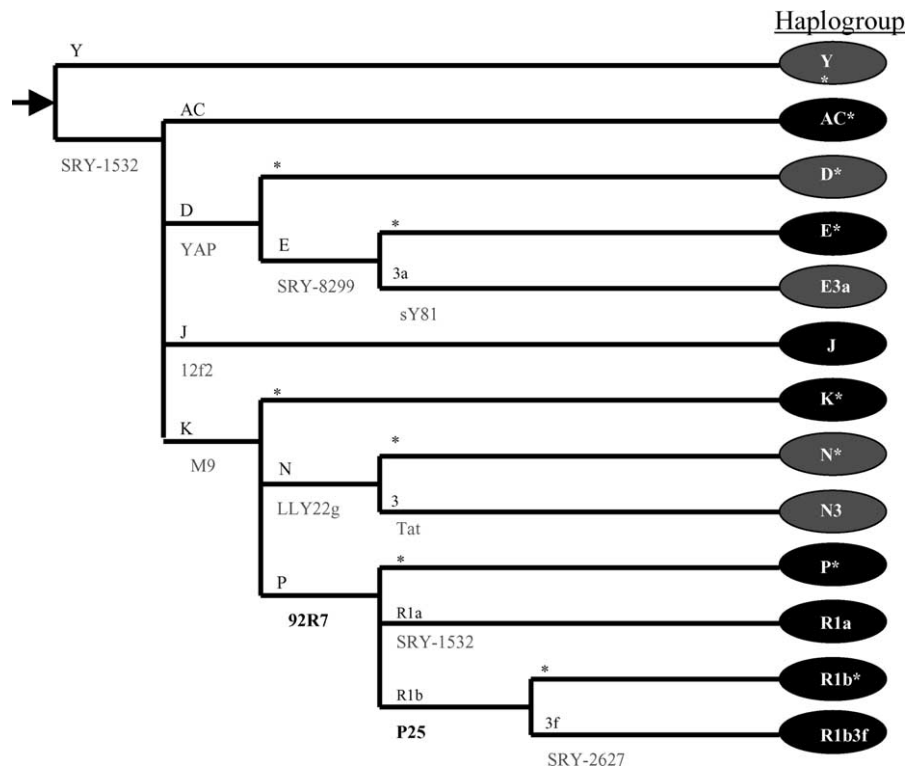


Fig. 1. Y-chromosome haplogroup tree including the binary polymorphisms analysed.

### 3. Results and discussion

#### 3.1. Two signals in P25 and 92R7

Independently, the groups from Santiago de Compostela and Copenhagen observed two signals in P25 and 92R7, and the group from Innsbruck also observed two signals in 92R7. Simultaneously, laboratories participating in an exercise of the European DNA Profiling Group (EDNAP) on Y-chromosome SNPs (manuscript in preparation) reported double signals or non-consensus results for the two markers. No signs of double signals were observed in following SNPs included in the EDNAP study: SRY<sub>10831</sub>, SRY<sub>4064</sub> [M40], M35, M89, M213, M9, M17, M18, M153, SRY<sub>2627</sub> [M167] and Tat [M46]. Therefore, additional research was performed on P25 and 92R7 by three of the EDNAP laboratories (Copenhagen, Santiago de Compostela and Innsbruck).

#### 3.2. The two signals in P25 and 92R7 are not due to contamination

DNA samples from 10 individuals that had shown two SNP signals in P25 and/or 92R7 were STR typed with the AmpF/STR Profiler Plus kit. Each DNA sample showed only one DNA STR profile demonstrating that the DNA preparations were not contaminated.

#### 3.3. The two signals in P25 and 92R7 are Y-chromosome specific

No amplification of female DNA was observed indicating that the results are due to PSVs on the Y-chromosome.

#### 3.4. The P25 SNP PSVs

The P25 SNP was reported as a C–A transversion in the q11.223 band of the Y-chromosome [13]. By using the published sequence, we designed two sets of PCR primers that amplified 109 (Copenhagen) and 121 bp (Santiago de Compostela) fragments containing the P25 SNP from chromosomal DNA from males. The 109 and 121 bp fragments were used as targets in minisequencing reactions. Sequencing primers were hybridised to the targets immediately upstream of the position of the SNP, and the primers were extended by one base using *Taq* DNA polymerase and fluorescently labelled dideoxynucleotides.

Fig. 2A and B show typical results from the minisequencing reaction analysed by capillary electrophoresis. The minisequencing primers were extended with ddATP and ddCTP in numerous samples. In the remaining samples, the minisequencing primers were extended with only ddCTP. To date, we have analysed more than 600 males of different ethnical origins (Caucasian, Inuit and African) and we have not found any sample with only the A allele.

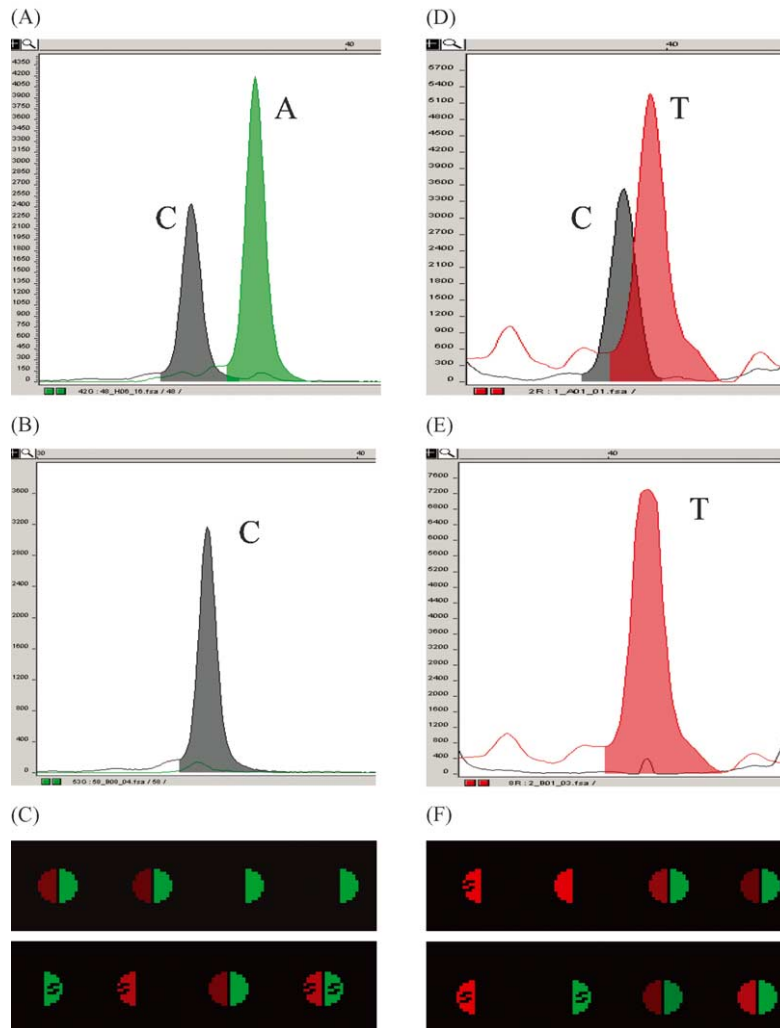


Fig. 2. Example of minisequencing with SNaPshot and hybridisation on the NanoChip microarray, (A) and (B) minisequencing reaction performed on P25 PCR products from two different samples; (D) and (E) minisequencing reaction performed on 92R7 PCR products from two different samples; (C) and (F) the top panel shows the results of the hybridisation performed on P25 (C) and 92R7 (F) PCR products from four different samples. The lower panel shows the hybridisation performed on control oligos. Each control oligo was addressed to the NanoChip alone (the first two pads on the lower panel) and together with the oligo control holding the other allele (the last two pads on the lower panel). The letter 's' inside the coloured circle indicates that the fluorescent signal is saturated.

In order to confirm the results, we performed the same PCR reaction with one of the primers labelled with biotin in the 5' end. A total of 0.2 µl of the PCR product was used in a minisequencing reaction and 7 µl of the PCR product was electronically addressed to a NanoChip microarray. The permeation layer above the electrodes on the NanoChip contained streptavidin and the biotin-labelled PCR product was bound to the array. The unlabelled strand was stripped of with 0.1 M NaOH and the labelled oligos were hybridised to the biotin-labelled strand. The oligo containing the C allele was labelled with CY3 (green) and the oligo containing the A allele was labelled with CY5 (red). Fig. 2C shows the results of hybridisation of four different samples. The first

two samples were scored as A and C in the minisequencing reaction and the last two samples were scored as C. In the hybridisation experiment, we observed a green and a red signal for the first two samples, and we observed a green signal for the last two samples, thus, confirming the results from the minisequencing analysis. We analysed more than 100 samples in parallel by minisequencing and by hybridisation on the NanoChip and we obtained the same results with both techniques for all samples.

For P25, *Cac8I* restriction enzyme digestion of the PCR product of allele C resulted in the generation of two smaller fragments. However, in individuals with both the C and the A alleles, the large fragment remained intact indicating that



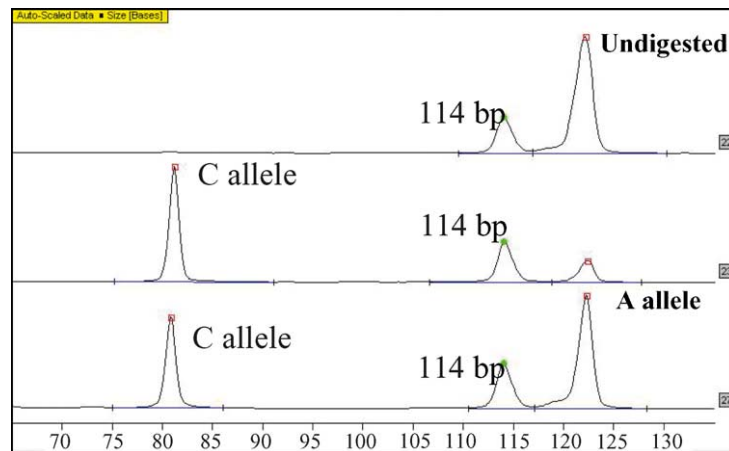


Fig. 3. Y-chromosome SNP P25 PCR/RFLP results. Electrophoreses on ALF Express II (Amersham Biosystems) sequencer of *Cac8I* digested and undigested PCR amplified P25 fragments and 114 bp control DNA fragments. The numbers on the abscissa indicate the fragment sizes. Upper part: undigested 121 bp P25 fragments; middle part: the larger *Cac8I* digested P25 fragments (77 bp) from a man of type C; and lower part: the undigested 121 bp P25 fragment and the larger *Cac8I* digested P25 fragment (77 bp) from a man of type C + A. In case of *Cac8I* digestion of P25 fragments, the smaller 44 bp fragment was also present but after digestion only the fragment with the labelled primer was detected.

the restriction enzyme site was lost in allele A. (Fig. 3). The samples where a part of the P25 product remained uncut always belonged to the haplogroup P\*(xR1b3f,R1a) as would be expected according to the maximum parsimony tree described by the Y-chromosome Consortium [20,21].

We sequenced the 121 and the 109 bp PCR product from more than 40 different samples. Fig. 4A and B show the forward and reverse sequences of where the C and the A alleles are clearly visible, whereas the sample in Fig. 4C and D has only the C allele. Therefore, we conclude that at least

two different regions on the Y-chromosome are amplified by the PCR strategy outlined above. The regions contain P25 and at least one of the regions is polymorphic.

All sequenced samples had a T insertion 14 bp downstream to the polymorphic position of P25 compared to the sequence reported by [13].

We designed new PCR primers that amplified different fragments containing P25. The fragments covered almost 9 kbp of the Y-chromosome surrounding P25. We analysed the fragments of six samples by minisequencing and by

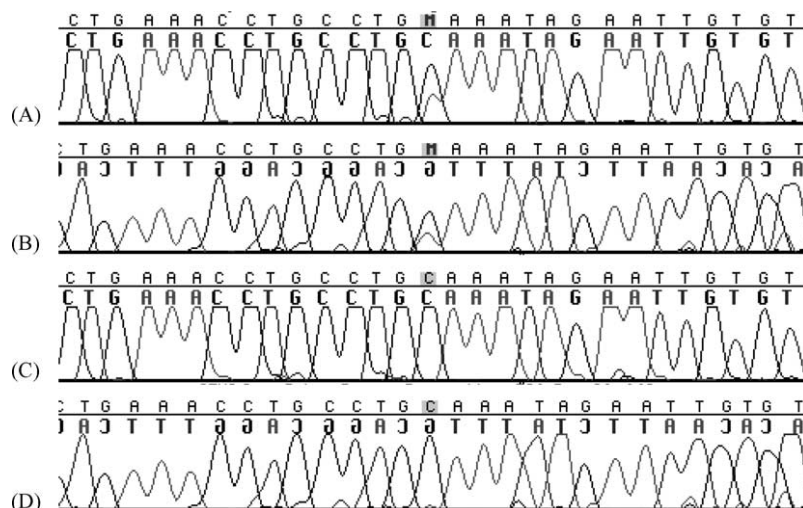


Fig. 4. Sequencing results of P25. The upper two sequences demonstrate forward (A) and reverse (B) sequences of P25 PCR fragment(s) that were SNP typed to be P25 C + A. At the P25 SNP position both a C and an A reaction are observed. The lower two sequences demonstrate forward (C) and reverse (D) sequences of P25 PCR fragments that were SNP typed to be P25 allele C. The nucleotide on the P25 SNP position is C.



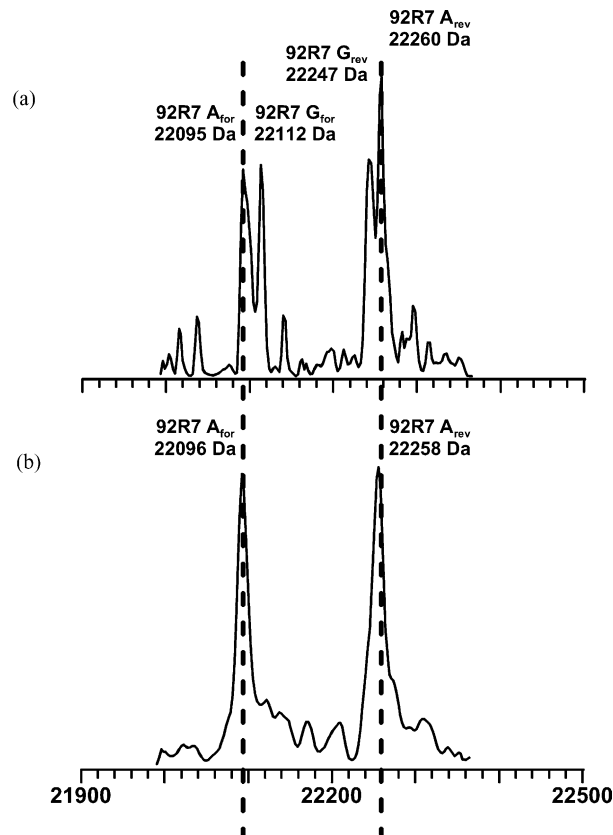


Fig. 5. Deconvoluted mass spectrum of the 92R7 locus. (a) Two signals (A and G) and (b) one signal (A).

*Cac8I* digestion, and we detected the same allele(s) as were obtained with the large 109 bp fragment (data not shown). This shows that the segmental duplication of the region containing P25 is larger than 9 kbp.

### 3.5. The 92R7 SNP duplication

The 92R7 SNP was reported as an A–G transversion in the p11 band of the Y-chromosome [22].

We amplified a 55 bp fragment (Santiago de Compostela) and an 89 bp fragment (Copenhagen) containing the 92R7 SNP from male samples. The PCR products were analysed by minisequencing as described for P25. The 89 bp fragment was also analysed by hybridisation on the NanoChip microarray as describe for P25 (Fig. 2). In all male samples where we detected the G allele, we also detected the A allele. The remaining samples all had the A allele. To date, we have analysed more than 600 males of different ethnic origin and we have not found any sample containing only the G allele.

Similar results were obtained using ion-pair reversed phase chromatography–electrospray ionisation mass spectrometry (Fig. 5).

The G allele contains a site for the endonuclease *HindIII* whereas the A allele can not be digested with this enzyme. If

allele G is present, digestion of the 55 bp PCR product generated two fragments of 28 and 27 bp (Table 3). In the presence of the G → A mutation, there is no restriction site for this enzyme and, after digestion with *HindIII*, the fragment size does not change. It has previously been reported that the G allele is only partially digested with *HindIII*, suggesting that PCR products containing the G allele also contains the A allele and suggesting that there are more than one copy of the 92R7 locus on the Y-chromosome [3,23].

Several samples with two signals in 92R7 and belonging to different Y-chromosome haplogroups were sequenced. Samples that had been SNP typed as 92R7 A (i.e. carrying the derived allele A) showed clear sequence results. However, samples that had been SNP typed as 92R7 A + G (i.e. carrying the ancestral allele G at the SNP position) showed heteroplasmic sequence patterns.

### 3.6. Allele distributions

The distribution of the polymorphisms in the context of previously defined haplogroups (according to the nomenclature of YCC [20]) was studied. The results obtained correlated with the haplogroup tree described by the Y-chromosome consortium [20,21]. All samples carried

the P25 variant C and some samples also carried P25 allele A. All samples carried the 92R7 variant allele A and a number of haplogroups carried also the 92R7 allele G.

### 3.7. BLAST search

In order to investigate if P25 and 92R7 had previously been reported to be paralogous sequences, the described sequences were compared to the sequences in the NCBI LocusLink and UniGene databases using the BLAST programme.

The search performed with the P25 sequence showed the presence of at least four loci with a homology of 99% or higher. One sequence was the DYS194 located on the clone BAY 3–8 that corresponds to the ancestral sequence of the mutation described by Hammer et al. [13]. Two loci with 99% homology were detected in different clones of the library RP11 created from the same person with the ancestral base in the polymorphic position and with the T insertion 14 bp downstream to the mutation. A third clone with high homology to P25 was detected in a different clone of the RP11 library with the T insertion and the mutation (A allele) in the P25 SNP position.

In the case of 92R7, the mutated sequence was found with 100% homology in two different clones of the RP11 library (RP11-155J5 and RP11-492C2). In both clones, another fragment with 78% homology was detected. This strongly indicates duplication events. In another clone of the RP11 library, a sequence with a 91% homology was found.

When similar sequences are detected in different clones of a library, it cannot be excluded that the same sequence was reported twice. However, when similar sequences are reported on the same clone twice, or the sequence with a mutation is reported in a different clone of the same library, there are good reasons to assume that the results are due to duplications as suggested by detailed analysis of the Y-chromosome map [24].

## 4. Final remarks

The two Y-chromosome loci P25 and 92R7 are highly polymorphic in European populations and they have been widely used in population studies [3,13]. They represent two important branching points on the haplogroup tree constructed from biallelic Y-chromosome markers [21]. A high degree of discrimination could therefore be obtained by typing P25 and 92R7, which makes the two loci potentially very interesting for genetic testing.

In this paper, we show that P25 and 92R7 previously thought to be simple SNP loci with binary polymorphisms are not unique loci and that each marker exhibits at least one paralogous sequence variant.

In this work, methods based on RFLPs, minisequencing, mass spectrometry, electronically activated chips and sequencing analysis were used in order to evaluate the

presence of duplicated fragments. Among the technologies used for typing these SNPs, direct sequencing or minisequencing methods as well as the use of ion-paired reversed phase chromatography electrospray ionisation mass spectrometry and electronically activated chips are in our opinion all convenient and appropriate methods, since they allow unambiguous typing of both sequences. However, RFLP methods are more prone to errors if the possible duplications have not previously been characterised.

The observations clearly illustrate the necessity of selecting unique loci for genetic testing. During the ongoing search for new polymorphic loci, it is important that we use the database information available, but it is equally important that we do not have blind faith in the data we obtain from the databases. Several recent papers have discussed the problems of handling paralogous sequence variants in genome databases [9,10,12,20,21] and it is clear that a large number of duplicated loci have been falsely assigned. Any SNP that we wish to analyse therefore needs to be notoriously tested. We especially wish to stress the importance of testing new SNPs with more than one technique. Denaturing-HPLC is a widely used technique for SNP typing. However, polymorphic PSVs like P25 and 92R7 will be scored as SNPs by D-HPLC, because D-HPLC detects variations from “wild type DNA” and not the PCR product itself. Other techniques have different limitations. Minisequencing or hybridisation techniques depend on the sequences immediately upstream or downstream of the SNP. Any mutations in these regions may impair SNP scoring. Confirmation of the results by a different SNP typing technique is therefore essential during the search for new polymorphic markers.

## Acknowledgements

We thank Ms. Rikke W. Hansen and Ms. Annette H. Birk for technical assistance. The work was supported by grants to Juan Sanchez from Ellen and Aage Andersen's Foundation, the Manuel Morales Foundation, and the “Xunta de Galicia” (PGIDT01PXI20806PR) and from the “Ministerio de Ciencia y Tecnología” (DGCYT.P4.-BIO2000-0145-P4-02).

## References

- [1] M. Lareu, J. Puente, B. Sobrino, B. Quintáns, M. Brión, A. Carracedo, The use of the LightCycler for the detection of Y chromosome SNPs, *Forensic Sci. Int.* 118 (2001) 163–168.
- [2] P. De Knijff, M. Kayser, A. Cagliá, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccini, M. Prinz, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, L. Rower, Chromosome Y microsatellites:

- population genetics and evolutionary aspects, *Int. J. Legal Med.* 110 (1997) 134–149.
- [3] M.E. Hurles, R. Veita, E. Arroyo, M. Armenteros, J. Bertranpetit, A. Pérez-Lezaun, E. Bosch, M. Shlumukova, A. Cambon-Thomsen, K. McElreavey, A. López de Munain, A. Röhl, I.J. Wilson, L. Singh, A. Pandya, C. Tyler-Smith, M.A. Jobling, Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism, *Am. J. Hum. Genet.* 65 (1999) 1437–1448.
- [4] Z.H. Rosser, T. Zerjal, M.E. Hurles, M. Adojaan, D. Alavantic, A. Amorim, W. Amos, M. Armenteros, E. Arroyo, G. Barbujani, G. Beckman, L. Beckman, J. Bertranpetit, E. Bosch, D.G. Bradley, G. Brede, G. Cooper, H.B. Corte-Real, P. de Knijff, R. Decorte, Y.E. Dubrova, O. Evgrafov, A. Gilissen, S. Glisic, M. Golge, E.W. Hill, A. Jeziorowska, L. Kalaydjieva, M. Kayser, T. Kivisild, S.A. Kravchenko, A. Krumina, V. Kucinskis, J. Lavinha, L.A. Livshits, P. Malaspina, S. Maria, K. McElreavey, T.A. Meitinger, A.V. Mikelsaar, R.J. Mitchell, K. Nafa, J. Nicholson, S. Norby, A. Pandya, J. Parik, P.C. Patsalis, L. Pereira, B. Peterlin, G. Pielberg, M.J. Prata, C. Previdere, L. Roewer, S. Rootsi, D.C. Rubinsztein, J. Saillard, F.R. Santos, G. Stefanescu, B.C. Sykes, A. Tolun, R. Villems, C. Tyler-Smith, M.A. Jobling, Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language, *Am. J. Hum. Genet.* 67 (2000) 1526–1543.
- [5] P.A. Underhill, G. Passarino, A.A. Lin, P. Shen, M. Mirazón Lahr, R.A. Foley, P.J. Oefner, L.L. Cavalli-Sforza, The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations, *Ann. Hum. Genet.* 65 (2001) 43–62.
- [6] O. Semino, G. Passarino, A. Brega, M. Fellous, A.S. Santachiara-Benerecetti, A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers, *Am. J. Hum. Genet.* 59 (1996) 964–968.
- [7] M.E. Hurles, E. Maund, J. Nicholson, E. Bosch, C. Renfrew, B.C. Sykes, M.A. Jobling, Native American Y chromosomes in Polynesia: the genetic impact of the Polynesian slave trade, *Am. J. Hum. Genet.* 72 (2003) 1282–1287.
- [8] J.A. Bailey, Z. Gu, R.A. Clark, K. Reinert, R.V. Samonte, S. Schwartz, M.D. Adams, E.W. Myers, P.W. Li, E.E. Eichler, Recent segmental duplications in the human genome, *Science* 297 (2002) 1003–1007.
- [9] International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.
- [10] J.A. Bailey, A.M. Yavor, H.F. Massa, B.J. Trask, E.E. Eichler, Segmental duplications: organization and impact within the current human genome project assembly, *Genome Res.* 11 (2001) 1005–1017.
- [11] J.A. Bailey, E.E. Eichler, Are 100000 “SNPs” useless?, *Science* (2002) 1509a.
- [12] X. Estivill, J. Cheung, M.A. Pujana, K. Nakabayashi, S.W. Scherer, L.C. Tsui, Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome, *Hum. Mol. Genet.* 11 (2002) 1987–1995.
- [13] M.F. Hammer, A.J. Redd, E.T. Wood, M.R. Bonner, H. Jarjanaz, T. Karafet, S. Santachiara-Benerecetti, A. Oppenheim, M.A. Jobling, T. Jenkins, H. Ostrer, B. Bonn -Tamir, Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes, *Proc. Natl. Acad. Sci. U.S.A.* 97 (2000) 6769–6774.
- [14] R.G. Sosnowski, E. Tu, W.F. Butler, J.P. O’Connell, M.J. Heller, Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control, *Proc. Natl. Acad. Sci.* 94 (1997) 1119–1123.
- [15] P.N. Gilles, D.J. Wu, C.B. Foster, P.J. Dillon, S.J. Chanock, Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips, *Nat. Biotech.* 17 (1999) 365–370.
- [16] B. Berger, G. H lzl, H. Oberacher, H. Niederst tter, C. Huber, W. Parson, SNP genotyping by on-line liquid chromatography-mass spectrometry in forensic science of the Y-chromosomal locus M9, *J. Chromatogr. B.* 782 (2002) 89–97.
- [17] A. Premstaller, H. Oberacher, C.G. Huber, High-performance liquid chromatography-electrospray ionisation mass spectrometry of single- and double stranded nucleic acids using monolithic capillary columns, *Anal. Chem.* 72 (2000) 4386–4393.
- [18] H. Oberacher, W. Walcher, G.C. Huber, Effect of instrument tuning on detectability of biopolymers in electrospray ionisation mass spectrometry, *J. Mass. Spectrom.* 38 (2003) 108–116.
- [19] M. Brion, A. Salas, A. Gonz lez-Neira, M.V. Lareu, A. Carracedo, Insights over the Iberian population origin through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs and the MSY1 minisatellite, *Am. J. Phys. Anthropol.* 122 (2003) 147–161.
- [20] The Y Chromosome Consortium, A Nomenclature system for the tree of human Y-chromosomal binary haplogroups, *Genome Res.* 12 (2002) 339–348.
- [21] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, *Nat. Genet. Rev.* 4 (2003) 598–612.
- [22] N. Mathias, M. Bay s, C. Tyler-Smith, Highly informative compound haplotypes for the human Y chromosome, *Hum. Mol. Genet.* 3 (1994) 115–123.
- [23] P. Forster, A. R hl, P. L nnemann, C. Brinkmann, T. Zerjal, C. Tyler-Smith, B. Brinkmann, A short tandem repeat-based phylogeny for the human Y chromosome, *Am. J. Hum. Genet.* 67 (2000) 182–196.
- [24] C.C. Tilford, T. Kuroda-Kawaguchi, H. Skaletsky, S. Rozen, L.G. Brown, M. Rosenberg, J.D. McPherson, K. Wylie, M. Sekhon, T.A. Kucaba, R.H. Waterston, D.C. Page, A physical map of the human Y chromosome, *Nature* 409 (2001) 943–945.

**Artículo 2: “Evidencias filogenéticas para múltiples eventos de duplicación independiente en el locus DYS19”**

Capelli C, Brisighelli F, Scarnicci F, **Blanco-Verea A**, Brión M, Pascali VL.  
Forensic Science Internacional: Genetics (2007) 1: 287-290.

Resumen:

Los eventos de duplicación en los loci STRs del cromosoma Y han sido descritos repetidas veces en poblaciones humanas. El DYS19 es probablemente el ejemplo mejor conocido y su estado duplicado aparece en individuos de todos los continentes. A pesar de la gran cantidad de datos disponibles, la relación evolutiva entre los cromosomas que portan la duplicación en el DYS19 no ha sido muy investigada. En este estudio se buscó la correlación cronológica entre dichos cromosomas mediante el análisis de nuevas duplicaciones identificadas en el DYS19 del cromosoma Y, por genotipado de SNPs y análisis network usando microsatélites. Los SNPs y los análisis network muestran que las duplicaciones del DYS19 del cromosoma Y están asociadas con diferentes linajes del cromosoma Y. Estos resultados indican que las duplicaciones del DYS19 ocurrieron más de una vez durante la evolución humana.



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Forensic Science International: Genetics 1 (2007) 287–290

[www.elsevier.com/locate/bsfig](http://www.elsevier.com/locate/bsfig)

Short communication

## Phylogenetic evidence for multiple independent duplication events at the DYS19 locus

Cristian Capelli<sup>a,\*</sup>, Francesca Brisighelli<sup>b,c</sup>, Francesca Scarnicci<sup>b</sup>,  
Alejandro Blanco-Verea<sup>c</sup>, Maria Brion<sup>c</sup>, Vince L. Pascali<sup>b</sup>

<sup>a</sup> Department of Zoology, University of Oxford, Tinbergen Building, South Parks Road, OX1 3PS Oxford, UK

<sup>b</sup> Istituto di Medicina Legale e delle Assicurazioni, Università Cattolica del S. Cuore, Rome, Italy

<sup>c</sup> Grupo de Medicina Xenómica, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, Galicia, Spain

Received 2 February 2007; received in revised form 10 May 2007; accepted 1 June 2007

### Abstract

Duplication events at Y chromosome STR loci have been repeatedly described in human populations. DYS19 is probably the best known example and it exhibits duplicate state in individuals from all continents. Despite the large amount of available data, evolutionary relationship between DYS19 duplication-bearing chromosomes has not been so far investigated. We address the genealogical correlation among such chromosomes by analysing newly identified DYS19 duplicated Y chromosomes by SNP genotyping and microsatellite-based network analysis. SNP and network analysis show that DYS19 duplicated Y chromosomes associate with different Y chromosome lineages. These results indicate that DYS19 duplication occurred more than once during human evolution.

© 2007 Elsevier Ireland Ltd. All rights reserved.

**Keywords:** DYS19; Duplication; Y chromosome microsatellites; Network analysis; Y chromosome genealogy

### 1. Introduction

Y chromosome STRs have today a steady place in several fields of molecular analysis. In population genetics and in forensics, a repertory of these Y-linked markers is routinely used to help identify sexual assaulters, analyze genealogical relationships and reconstructing ancient population migrations. The community of Y chromosome analysis users has been consequently – and since the beginning – made familiar with the typical haploid pattern of these markers, almost invariably reproducing the PCR analysis scheme of ‘one locus-one band/peak’.

On the other hand, locus duplications placed along some highly redundant Y chromosome portions [1] can exist and create a PCR-doublets phenomenon. For example, DYS389 and DYS385 STR loci are duplicated in both humans and chimpanzees [2]. Recently, other locus duplication examples have been described at Y chromosome loci ([3], and reference therein).

This phenomenon is interesting from a population and evolutionary point of view. When duplication of a given STR locus is seen across the generality (or the vast majority) of human populations, the obvious question arises as to whether it occurred once or by independent reiteration of the same event. Noticing haploid doublets has implications for the forensic genetics practitioner too [3].

In view of this series of considerations, a better understanding of the evolutionary history of Y chromosome duplication is essential. We here try to tackle this issue by focusing on the well know example of DYS19 locus duplication, whose abundance of observational data may supply the ideal subject for phylogenetic analysis.

### 2. Materials and methods

A collection of newly identified DYS19 duplicated chromosomes were available to us, framed within a high-resolution male haplotypic set of data ([4,15]; Table 1). All duplicated chromosomes have been confirmed by amplification using the PowerPlex Y system ([5]; see Table 1). Haplogroups assignment was done by scoring P15, P16 and M286 markers using the following primers: P15-F-ccaatgcttgaggtctgaat,

\* Corresponding author at: Tel.: +44 1865 271261; fax: +44 1865 310447.  
E-mail address: [cristian.capelli@zoo.ox.ac.uk](mailto:cristian.capelli@zoo.ox.ac.uk) (C. Capelli).

Table 1  
Y chromosome haplotypes bearing *DYS19* duplications found in Italy, all within the G2\*(xG2a, b) lineage

n	<i>DYS19</i>	<i>DYS389I</i>	<i>DYS389II</i>	<i>DYS390</i>	<i>DYS391</i>	<i>DYS392</i>	<i>DYS393</i>	<i>DYS385</i>	<i>DYS437</i>	<i>DYS438</i>	<i>DYS439</i>
2	13–14	14	16	22	10	11	14	13-13	16	10	12
1	13–14	15	16	22	10	11	14	13-13	17	10	12
1+ 1 <sup>a</sup>	14–16	14	16	22	11	11	14	13-14	16	10	12
1 <sup>b</sup>	15–16	14	18	22	10	11	13	13-15	16	10	12
1 <sup>a</sup>	15–16	14	18	22	10	11	13	12-15	16	10	12
1 <sup>a</sup>	15–17	12	17	22	11	11	13	14-14	16	10	11
1 <sup>a</sup>	15–17	14	16	22	11	11	13	14-16	16	10	12
1	15–17	13	19	24	10	11	15	14-15	15	10	9

Number of repeats at *DYS389II* locus was considered after subtracting *DYS389I* repeats.

<sup>a</sup> Haplotype described in Capelli et al. [15].

<sup>b</sup> Haplotype described in Capelli et al. [4].

P15-R-agagcctcaatccatcatcg; P16-F-agcacacagtggagcaatgg, P16-R-ccggcaacagatcagaaaact; M286-F-cgttgccctgtgttccat, M286-R-ggattcgagcatcagcta. The three markers identify G2, G2a and G2b haplogroups, respectively. Phylogenetic network was calculated by Network 4.112 (fluxus-engineering.com [7]) applying the Median Joining algorithms sequentially. A total of 133 *DYS19* duplicated Y chromosome were collected from available literature (88 from the [www.yhrd.com](http://www.yhrd.com)—release 20; 35 from Zerjal et al. [8]; 10 from Capelli et al. [4,15]). The duplicated haplotypes described by Nasidze et al. [9] were already included in the YHRD. Of these, 96 genotyped for the extended Y-STR haplotype loci plus *DYS438* and *DYS439*, were available and were included in the network analysis. *DYS19* and *DYS385* were excluded due to difficulties in assigning each of the duplicated allele to the corresponding locus, and *DYS392* resulted monomorphic for allele 11. The following loci were used for network calculation: *DYS389I*, *DYS389II* (subtracting *DYS389I* repeats), *DYS390*, *DYS391*, *DYS393*, *DYS438*, *DYS439*. One haplotype from the Y-STR repository (YHRD) showing a tri-allelic *DYS19* pattern was included in the analysis. A weighting scheme was applied giving to each locus a score as follows:  $[V_A \times 10]/V_L$ , considering  $V_A$  as the average variance of the repeat score estimated across loci and  $V_L$  the variance at the repeat score at each specific locus.

### 3. Results and discussion

It is well known that *DYS19* locus PCR doublets exist worldwide in humans (see [www.yhrd.org](http://www.yhrd.org)). In the context of a SNP genotyping study, Zerjal et al. [8] showed that Asian individuals bearing this locus duplication belong to C3c haplogroup (Y chromosome genealogy [10]) (identified by marker M48, Chris Tyler-Smith, personal communication). More recently, Nasidze et al. [9] described *DYS19* duplications in populations from the Caucasus and they classified them within haplogroup C3c. In the course of a large Italian Y chromosome screening, we have now identified 10 additional *DYS19* duplication-bearing chromosomes—with the relevant frequency of occurrence amounting to 1% ([4,15, unpublished data]). By additional typing work, we can report that all these chromosomes classify in the haplogroup G2\*(xG2a,b)—a branch of its own in the Y chromosome genealogical hierarchy,

identified by P15, P16 and M286 markers. Our finding clearly implies that duplication at *DYS19* had to occur at least twice, driven by as many *independent* genetic events.

Starting from this line of evidence, we collected all *DYS19* duplication-bearing haplotypes available in the literature and used the inherent specific archive of data to generate a microsatellite network. The resulting structure (Fig. 1) plots in two clusters. The first cluster mostly accommodates chromosomes of European origin, including the Italian chromosomes that bear the P15 mutation. The second cluster is made mostly of Russian and Kazakh individuals carrying the M48 mutation. The most common *DYS19* doublets were 15–16 in group 1 and 16–17 in group 2. The European cluster appears to be more heterogeneous than the second one, as suggested by the number of multiple mutational events connecting the various haplotypes and a larger microsatellite variance (0.45 versus 0.12). This could reflect partial knowledge (by ineffective sampling) of the whole set of G2 *DYS19* duplicated chromosomes—and potentially also additional, independent, still uncharacterised duplication events.

In the light of the current existence of the two lineages, the single-origin hypothesis implies that the original duplication event had to occur very early in the timescale of human evolution [10]. However, the currently available world-wide set of genotyped samples from different Y chromosome genealogy branches shows a limited number of *DYS19* duplication-bearing chromosomes, suggesting that, in the case of single origin, multiple reversion events should have occurred. The multiple-origin hypothesis, requesting so far a minimum of only two independent duplication events, appears, as a more parsimonious alternative, the most likely.

We note that the *DYS19* locus lies within the chromosomal region delimited by the IR3/IR3 repeats that have experienced several molecular rearrangements [11]. While reporting on the structural variation of a larger overlapping area, Jobling et al. [12] recently noticed a number of independent deletion events to be generated by non-allelic homologous recombination. According to these authors, a parallel event of reciprocal duplication should have occur at the same time—as it is suggested by the existence of two chromosomes showing the duplicated state at the *DYS458* locus mapping within the deleted/potentially duplicated region. The same mechanism could account of the *DYS19* duplication, by a larger

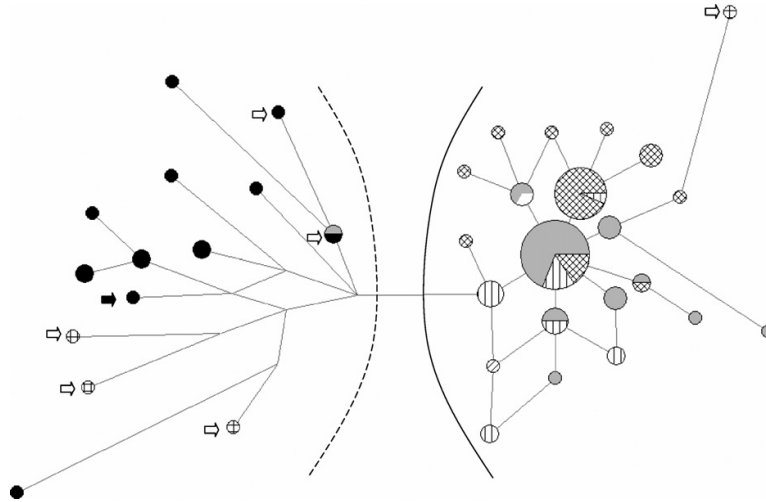


Fig. 1. Phylogenetic network. Total number of individuals in the network is 96. The size of the pies is proportional to the number of individuals represented, being one the minimum number of chromosome per indicated node. The different individuals' origins are coded within the pies as follows: grey, Russia; diagonal lines, Kazakhstan; vertical lines, Asia excluding Russia and Kazakhstan; white, Africa; black, Europe; crossed lines, African American, South American and Hispanics. G2\*(xG2a,G2b) and C3c haplotypes are defined by the dashed and the solid lines, respectively. Arrows point to nodes containing no SNP defined haplotypes. For nomenclature consistency, three repeats have been added to the DYS389I alleles from Zerjal et al. [8] dataset (Chris Tyler-Smith, personal communication).

recombination event involving the entire IR3/IR3 region. Gene conversion could as well be an alternative mechanism. Several regions homologous to DYS19 have been identified within the Y chromosome [13], among which the IR3 regions bear probably the highest homology (99.75% [1]). Regions as these encompass up to 30% of the Y chromosome euchromatin [1] and have experienced multiple gene conversion events [14]. Therefore, some of them are ideal candidates as counterparts in the process of DYS19 duplication.

The status of neighbouring markers should be of decisive help to identify the exact mechanism of production (duplication-deletion versus gene conversion). Further molecular characterisation of DYS19 duplication-bearing chromosomes will help to clarify this issue.

#### Acknowledgements

Authors would like to thank all DNA donors who made this work possible. CC additionally would like to acknowledge Chris Tyler-Smith and Lutz Roewer for kindly providing data. This study was partially funded by the Italian Ministry of University (PRIN-MIUR 2002, N.2002063871).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2007.06.001.

#### References

- [1] H. Skaletsky, 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, D.C. Page, The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes, *Nature* 423 (2003) 825–837.
- [2] L. Gusmao, A. Gonzalez-Neira, C. Alves, P. Sanchez-Diz, E.M. Dauber, A. Amorim, A. Carracedo, Genetic diversity of Y-specific STRs in chimpanzees (*Pan troglodytes*), *Am. J. Primatol.* 57 (2002) 21–29.
- [3] J.M. Butler, A.E. Decker, M.C. Kline, P.M. Vallone, Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation, *J. Forensic Sci.* 50 (2005) 853–859.
- [4] C. Capelli, B. Arredi, L. Baldassari, I. Boschi, F. Brisighelli, A. Caglia, M. Dobosz, F. Scarmicci, G. Vetrugno, V.L. Pascali, A 9-loci Y chromosome haplotype in three Italian populations, *Forensic Sci. Int.* 159 (2006) 64–70.
- [5] B.E. Krenke, L. Viculis, M.L. Richard, M. Prinz, S.C. Milne, C. Ladd, A.M. Gross, T. Gornall, J.R. Frappier, A.J. Eisenberg, C. Barna, X.G. Aranda, M.S. Adamowicz, B. Budowle, Validation of male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex, *Forensic Sci. Int.* 151 (2005) 111–124.
- [6] H.-J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, *Mol. Biol. Evol.* 16 (1999) 37–48.
- [7] T. Zerjal, 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, C. Tyler-Smith, The genetic legacy of the Mongols, *Am. J. Hum. Genet.* 72 (2003) 717–721.
- [8] I. Nasidze, D. Quinque, I. Dupanloup, R. Cordaux, L. Kokshunova, M. Stoneking, Genetic evidence for the Mongolian ancestry of Kalmyks, *Am. J. Phys. Anthropol.* 128 (2005) 846–854.
- [9] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, *Nat. Rev. Genet.* 4 (2003) 598–612.
- [10] S. Repping, S.K. van Daalen, L.G. Brown, C.M. Korver, J. Lange, J.D. Marszalek, T. Pyntikova, F. van der Veen, H. Skaletsky, D.C. Page, S. Rozen, High mutation rates have driven extensive structural polymorphism among human Y chromosomes, *Nat. Genet.* 8 (2006) 463–467.
- [11] M.A. Jobling, I.C. Lo, D.J. Turner, G.R. Bowden, A.C. Lee, Y. Xue, D. Carvalho-Silva, M.E. Hurles, S.M. Adams, Y.M. Chang, T. Kraaijenbrink, J. Henke, G. Guanti, B. McKeown, R.A. van Oorschot, R.J. Mitchell, P. de



- Knijff, C. Tyler-Smith, E.J. Parkin, Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y, *Hum. Mol. Genet.* 16 (2007) 307–316.
- [13] J.M. Butler, R. Schoske, Duplication of DYS19 flanking regions in other parts of the Y chromosome, *Int. J. Legal Med.* 118 (2004) 178–183.
- [14] S. Rozen, H. Skaletsky, J.D. Marszalek, P.J. Minx, H.S. Cordum, R.H. Waterston, R.K. Wilson, D.C. Page, Abundant gene conversion between arms of palindromes in human and ape Y chromosomes, *Nature* 423 (2003) 873–876.
- [15] C. Capelli, F. Brisighelli, F. Scarnicci, B. Arredi, A. Caglia', G. Vetrugno, V. Onofri, S. Tofanelli, G. Paoli, V.L. Pascali, Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic–Neolithic encounter, *Mol. Phyl. Evol.* 44 (2007) 228–239.



Elista, Russia (Kalmykian)	16,17	14	17	24	9	11	13	10	11	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	24	9	11	13	10	11	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	24	9	11	13	10	11	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	24	9	11	13	10	11	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	24	9	11	13	10	11	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	24	9	11	13	10	11	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Mongolia	16,17	14	17	24	9	11	13	10	11	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Mongolia	16,17	14	17	24	9	11	13	10	11	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	24	9	11	13	10	12	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	24	9	11	13	10	12	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	24	9	11	13	10	12	C3e	<a href="http://www.yhrd.org">www.yhrd.org</a>
Siberia, Russia (Tuvan)	16,17	14	17	24	9	11	13	na	na	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Siberia, Russia (Tuvan)	16,17	14	17	24	9	11	13	na	na	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Taraz, Kazakhstan	16,17	14	17	24	9	11	13	na	na	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Taraz, Kazakhstan	16,17	14	17	24	9	11	13	na	na	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Taraz, Kazakhstan	16,17	14	17	24	9	11	13	na	na	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
England-Wales, UK (Chinese)	16,17	14	17	24	9	11	13	na	na	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
England-Wales, UK (Chinese)	16,17	14	17	24	9	11	13	na	na	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Mongolian	16,17	14	17	24	10	11	13	10	11	C3e	Zerjal et al, 2003
Mongolian	16,17	14	17	24	10	11	13	10	11	C3e	Zerjal et al, 2003
Mongolia	16,17	14	17	24	10	11	13	10	11	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Mongolia	16,17	14	17	24	10	11	13	10	11	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	25	9	11	12	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	25	9	11	13	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	14	17	25	9	11	13	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Tunisia (Andalusian Arabs)	16,17	14	17	25	9	11	13	10	11	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Kyrgyz	16,17	14	17	26	9	11	13	10	11	C3c	Zerjal et al, 2003
Elista, Russia (Kalmykian)	16,17	14	18	24	9	11	13	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Siberia, Russia (Tuvan)	16,17	14	18	24	9	11	13	na	na	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Siberia, Russia (Tuvan)	16,17	14	18	24	9	11	13	na	na	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Kyrgyz	16,17	14	18	24	10	11	13	10	11	C3c	Zerjal et al, 2003
Elista, Russia (Kalmykian)	16,17	14	19	24	9	11	13	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Mongolian	16,17	14	19	24	10	11	13	10	11	C3c	Zerjal et al, 2003
Mongolia	16,17	14	19	24	10	11	13	10	11	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	15	16	23	9	11	13	10	12	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	15	17	24	9	11	13	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	15	17	24	9	11	13	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	16,17	15	17	24	9	11	13	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Mongolian	16,17	15	18	24	9	11	13	10	11	C3c	Zerjal et al, 2003
Mongolia	16,17	15	18	24	9	11	13	10	11	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Mongolian	16,18	14	18	24	9	11	13	10	11	C3c	Zerjal et al, 2003
Mongolia	16,18	14	18	24	9	11	13	10	11	na	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	17,18	14	17	24	9	11	13	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>
Elista, Russia (Kalmykian)	17,18	14	17	24	9	11	13	10	11	C3c	<a href="http://www.yhrd.org">www.yhrd.org</a>

Appendix A: Supplementary data:

Haplotypes from [www.yhrd.org](http://www.yhrd.org) to release 20.

DYS389I alleles from Zerjal et al, 2003 have been changed by adding 3 repeats for nomenclature consistency (Chris Tyler-Smith, personal communication).

Haplogroups information for the Kalmyks is from Nasidze et al, 2005.

Only haplotypes genotyped at all the indicated markers have been included in the Network analysis.

na: data not available.

## **Bloque II. Utilidad del desarrollo de multiplexes en genética poblacional y forense**

En el siguiente bloque de artículos se va a tratar el diseño y utilidad de los multiplexes de SNPs del cromosoma Y. Actualmente los marcadores binarios del cromosoma Y comienzan a jugar un papel de cierta relevancia en el campo de la genética forense, debido a que se muestran especialmente útiles en el estudio de muestras procedentes de agresiones sexuales o en casos de paternidades en los que es interesante suministrar evidencias del linaje paterno. Cuando los SNPs autosómicos se usan en un contexto forense el número de SNPs requerido ha de ser muy superior al de STRs para poder aportar niveles comparables de información. Sin embargo, si nuestro propósito es utilizar SNPs del cromosoma Y con fines forenses, es muy importante realizar una selección adecuada de los mismos ya que poseen una marcada diferenciación genética, la cual genera unos perfiles que concluyen en diferentes haplogrupos dependiendo de la región del mundo estudiada. Esta característica también convierte a estos SNPs en una potentísima herramienta de trabajo en estudios de genética poblacional.

Estas propiedades de los Y-SNPs nos han empujado a crear una serie multiplexes: en un primer artículo se realizaron 4 multiplexes que engloban un total de 30 Y-SNPs, permitiéndonos determinar los haplogrupos más frecuentes presentes en la población europea. Por otro lado, en un segundo artículo se creó un único multiplex que contiene un total de 29 Y-SNPs. Ambos proyectos perseguían el mismo fin; demostrar la utilidad de los multiplexes de polimorfismos binarios del cromosoma Y para inferir el origen geográfico de un linaje masculino con el menor número posible de amplificaciones. Es de esperar que en un futuro la combinación de estos marcadores bialélicos unido a la variabilidad de los Y-STRs nos permitirá identificar regionalmente cualquier linaje masculino implicado en un caso forense.



**Artículo 3: “Análisis jerárquico de 30 SNPs del cromosoma Y en poblaciones europeas”**

Brión M, Sobrino B, **Blanco-Verea A**, Lareu MV, Carracedo A.

Int J Legal Med (2004) **119**: 10-15

Resumen:

Los análisis de haplogrupos del cromosoma Y definidos por polimorfismos binarios, han llegado a ser un estándar para abordar el estudio del origen de las poblaciones humanas recientes y para medir la variabilidad entre ellos. Además, la simplicidad y especificidad poblacional de los polimorfismos binarios permite inferir el origen poblacional de cualquier muestra masculina de interés para propósitos forenses. De todos los polimorfismos binarios descritos en el *Y Chromosome Consortium tree* que pueden ser analizadas por PCR, nosotros hemos seleccionado 30 marcadores. Estos 30 marcadores han sido agrupados en 4 multiplexes, para determinar los haplogrupos más frecuentes en Europa son suficientes un multiplex o dos. De este modo evitamos el tipado innecesario de SNPs para definir el haplogrupo final ahorrando esfuerzo y coste, así sólo necesitamos tipar 9 SNPs en el mejor de los casos y en el peor, no más de 17 SNPs para definir el haplogrupo. El método elegido para discriminar los alelos fue una reacción de “*single base extensión*” usando el kit de SNaPshot multiplex. Un total de 292 muestras de 8 comarcas diferentes de Galicia (Noroeste de España) fueron analizadas con esta estrategia. No se detectaron diferencias significativas entre las distintas comarcas, excepto en la Mariña Lucense, que presenta una frecuencia de haplogrupos diferentes pero no valores  $\Phi_{ST}$  más altos.



## ORIGINAL ARTICLE

M. Brion · B. Sobrino · A. Blanco-Verea · M. V. Lareu  
A. Carracedo**Hierarchical analysis of 30 Y-chromosome SNPs in European populations**Received: 11 December 2003 / Accepted: 2 March 2004 / Published online: 17 April 2004  
© Springer-Verlag 2004

**Abstract** Analysis of Y-chromosome haplogroups defined by binary polymorphisms, has become a standard approach for studying the origin of modern human populations and for measuring the variability between them. Furthermore, the simplicity and population specificity of binary polymorphisms allows inferences to be drawn about the population origin of any male sample of interest for forensic purposes. From the 245 binary polymorphisms that can be analysed by PCR described in the Y Chromosome Consortium tree, we have selected 30 markers. The set of 30 has been grouped into 4 multiplexes in order to determine the most frequent haplogroups in Europe, using only 1 or 2 multiplexes. In this way, we avoid typing unnecessary SNPs to define the final haplogroup saving effort and cost, since we only need to type 9 SNPs in the best case and in the worst case, no more than 17 SNPs to define the haplogroup. The selected method for allele discrimination was a single base extension reaction using the SNaPshot multiplex kit. A total of 292 samples from 8 different districts of Galicia (northwest Spain) were analysed with this strategy. No significant differences were detected among the different districts, except for the population from Mariña Lucense, which showed a distant haplogroup frequency but not higher  $\Phi_{st}$  values.

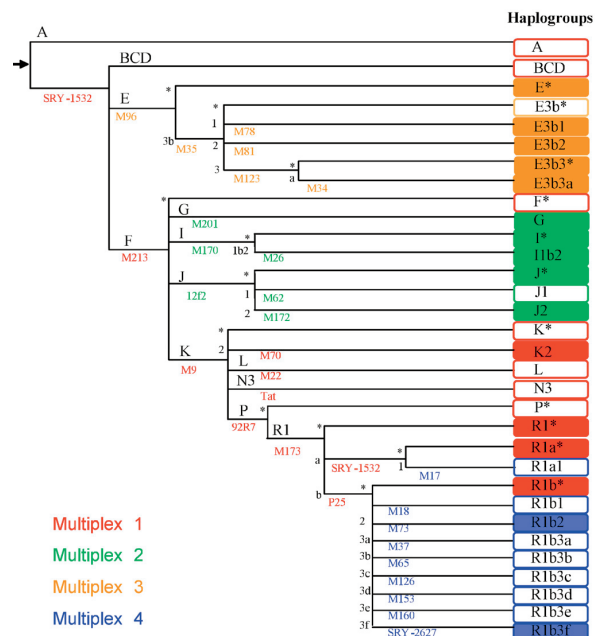
**Keywords** Y chromosome · Single nucleotide polymorphisms · SNPs · Single base extension · SNaPshot reaction · Europe

M. Brion (✉) · B. Sobrino · A. Blanco-Verea · M. V. Lareu  
A. Carracedo  
Institute of Legal Medicine,  
University of Santiago de Compostela, San Francisco s/n,  
15782 Santiago de Compostela, Spain  
Tel.: +34-981582327, Fax: +34-981580336,  
e-mail: brioniml@usc.es

**Introduction**

Single nucleotide polymorphisms (SNPs, including small insertions and deletions) are the markers of choice for many applications, including the location and identification of disease susceptibility genes, pharmacogenomics and pharmacogenetics, studying the origin of modern human populations or measuring the variability between them (Jorde et al. 2001; Zhao et al. 2003).

Markers located on the Y chromosome have specific interest as forensic tools, because most of the chromosome



**Fig. 1** Phylogenetic tree defined with the binary Y-chromosomal polymorphisms analysed. Marker names are indicated below the lines and lineage names are shown above the lines, but the length of each branch has no significance. Colours represent multiplex groups, and the coloured area represents the presence in Galicia



does not undergo recombination. In particular, Y-chromosome SNPs, because of their abundance, simplicity and low mutation rate, are becoming an extensively used marker set. Forensic laboratories are starting to implement Y chromosome SNP analysis, in order to investigate the forensic usefulness of these markers (Sanchez et al. 2003; Borsting et al. 2004).

In the case of autosomal SNPs, the required number of SNPs to give comparable levels of information to STRs (Gill 2001), the most widespread markers used in the forensic field has already been determined. However, in the case of the Y chromosome, the selection of SNPs is complicated by the far more extensive genetic differentiation exhibited. This differentiation across geographical distance (Seielstad et al. 1998) results in markedly different haplogroup profiles, depending on the region of the world studied (Karafet et al. 1999; Rosser et al. 2000).

At the time of writing this manuscript, the Ensembl database lists 36,449 Y-SNPs, however most of them could be paralogous sequence variants (Jobling and Tyler-Smith 2003; Sanchez et al. 2004), highlighted by comparing true Y-chromosomal sequences with similar sequences elsewhere. Currently more than 240 Y-chromosome SNPs are available and well characterised. They define a highly re-

solved tree of binary haplogroups with a unified nomenclature proposed by the Y Chromosome Consortium (YCC 2002; Jobling and Tyler-Smith 2003).

Checking the literature, an extensive search has been performed looking for the allele frequencies of each SNP in European populations (Rosser et al. 2000; Semino et al. 2000; Brion et al. 2003). As a result of this search, a set of 30 SNPs was selected in order to determine 32 of the most frequent haplogroups present in European populations (Fig. 1). The inclusion of a male sample in one of these haplogroups might be useful in the identification of the regional origin within Europe. In combination with Y-chromosome STR variation, the regional identification of any male sample involved in a forensic case could be possible in the future.

A large number of SNP genotyping methods are now available (Chen and Sullivan 2003), and usually the choice of the appropriate method depends on the number of SNPs and the number of individuals that need to be typed. As this study typed 30 SNPs in 292 samples, the genotyping method selected was a multiplex PCR followed by the single base extension reaction using the SNaPshot multiplex kit (Applied Biosystems, Foster City, CA).

**Table 1** SNP primer sequences and PCR concentrations of the SNPs used in this study, grouped by multiplexes

SNP	Primer (5'→3')		Size (bp)	Conc. (µM)	
	Forward	Reverse			
92R7	TGCATGAACACAAAAGACGTA	GCATTGTAAATATGACCAGC	55	0.20	Multiplex1
M70	TCATAGCCCCTACTACTTTGGAC	CTGAGGGCTGGACTATAGGG	81	0.20	
M22	GCTGATAGTCCCTGGTTCCCTA	TGAGCATGCCTACAGCAGAC	106	0.20	
Tat	GACTCTGAGTGTAGACTTGTGA	GAAGGTGCCGTAAAAGTGTGAA	112	0.20	
P25	GGACCATCACCTGGGTAAGT	AGTGCTTGTCCAAGGCAGTA	121	0.20	
SRY1532	TCCTTAGCAACCATTAATCTGG	AAATAGCAAAAAGTACACAAGGC	167	0.20	
M173	GCACAGTACTACTTTAGGTTTGC	GCAGTTTTCCAGATCCTGA	172	0.20	
M213	GGCCATATAAAAACGCAGCA	TGAATGGCAAATTGATTCCA	208	0.30	
M9	GCAGCATATAAAAACCTTCAGG	AAAACCTAAGCTTGTCTCAAGC	340	0.35	
I2f2	CAGTACTGATCAAAAATGCTTACAGAT	GGATCCCTTCTTACACCTTATACA	90	0.25	Multiplex2
M201	TCAAATTGTGACACTGCAATAGTT	CATCCAACACTAAGTACCTATTACGAA	144	0.25	
M26	AGCAGAAGAGACCAAGACAGC	GACGAAATCTGCAGCAAAAA	147	0.25	
M170	TGCAGCTCTTATTAAGTTATGTTTCA	CCAATTAAGTTTCAACATTTAAGACC	158	0.30	
M172	TCCTCATTACCTGCCTCTC	TCCATGTTGGTTTGGAAACAG	187	0.25	
M62	ACTAAAACACCATTAGAAAACAAAGG	CTGAGCAACATAGTGACCCC	309	0.25	
M96	GTGATGTGTAACCTGGAACACAGG	GGACCATATATTTGCCATAGGTT	88	0.25	Multiplex3
M34	CACAGTGTTTTCTCATGTTAATGC	GGGACCCCAATAATCATAA	92	0.25	
M81	TTATAGTTTCAATCCCTCAGTAATTTT	TGTTTTCTTCTGGTTTGTGTGAGTA	176	0.25	
M35	GCATGGTCCCTTTCTATGGAT	GAGAATGAATAGGCATGGGTTT	198	0.25	
M123	CACAGAGCAAGTACTCTCAAAG	TCTTCCCTCAACATAGTTATCTCA	248	0.25	
M78	CTTCAGGCATTAATTTTTTTGGT	ATAGTGTTCCTTACCTTTCCTT	301	0.25	
M65	AAGGCTACCCATTCCTCAAAT	AAGTCTGGCATCTGCAAAAATC	71	0.15	Multiplex4
M126	GTGCTTGAACCCGAGTTTGT	TCGGGAAACACAATTAAGCA	83	0.15	
M73-M160	AAAACAATAGTTCCAAAACCTTCTGA	CCTTGTGATTCTCTGAACG	98	0.5	
M37	ATGGAGCAAGGAACACAGAA	AAGAAAGGAGATTGTTTTCAATTTT	124	0.3	
M167	GAGGCTGGGCCAAGTTAAGG	CTTCTCGGAACCACTACCA	130	0.15	
M17-M18	CTGGTCATAACACTGGAAATC	AGCTGACCACAACTGATGTAGA	171	0.10	
M153	TCTGACTTGGAAAGGGGAAA	TTTTCTCTCATTATTTGTCTTCA	239	0.5	

Instead of typing the whole set of polymorphisms in each sample, our strategy was to group the SNPs in a hierarchical way following the YCC tree. Four multiplex PCR/primer extension reactions were developed, allowing the assignment of a sample to 1 of 32 possible haplogroups using only 1 or 2 multiplexes.

## Material and methods

### DNA samples

A total of 292 male subjects, belonging to 8 different districts of Galicia (northwest Spain) were analysed. Appropriate informed consent was obtained from all individuals. Blood was collected by venous puncture using EDTA as anticoagulant. Genomic DNA was extracted using a phenol-chloroform method.

### Multiplex PCRs

PCR multiplexes were performed in 25 µl final volume, with 1× buffer, 300 µM of dNTPs, 2 mM of MgCl<sub>2</sub>, 2 U of AmpliTaq Gold polymerase (Applied Biosystems) and 10 ng of genomic DNA. The cycling conditions were 95°C for 10 min then 32 cycles of 94°C for 30 s, 59°C for 30 s, 70°C for 30 s, and a final extension at 65°C for 15 min. Despite the fact that primer designs for PCR amplification of these SNPs are available in the literature, most of

them were redesigned using the Primer3 software ([http://www-genome.wi.mit.edu/cgi-bin/primer/primer3\\_www.cgi](http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi)) and checked for possible secondary structures with the Oligonucleotide properties calculator software v. 3.02 (<http://www.basic.nwu.edu/bio-tools/oligocalc.html>). Primer sequences and detailed individual concentrations are shown in Table 1.

Amplicon sizes have been designed to be different enough to allow checking the amplification by electrophoresis. Conventional polyacrylamide electrophoresis (T=9, C=5) with silver stain detection was used for checking the amplification products.

### Multiplex single base extensions

Before single base extension (SBE), 1 µl of the PCR product was cleaned up with 0.5 µl of ExoSAP-IT (Amersham Biosciences) and incubated at 37°C for 15 min followed by 85°C for 15 min to inactivate the enzyme.

Multiplex single base extension reactions were performed in a 5 µl final volume, combining 2 µl of SNaPshot ready reaction mix (Applied Biosystems), 1.5 µl of cleaned PCR product and extension primers. The cycling conditions were 96°C for 10 s, 50°C for 5 s and 60°C for 30 s, for 25 cycles. The same primer design software used to develop PCR primers helped to select the SBE primers. However, in this case each primer had varying lengths of poly (dC) non-homologous tails attached at the 5' end. All SBE primer sequences and concentrations are shown in Table 2.

To remove the unincorporated ddNTPs, the final product was incubated with 1 U of shrimp alkaline phosphatase (Amersham

**Table 2** SBE primer sequences and concentrations of the SNPs used in this study, grouped by multiplexes

SNP		Minisequencing primer (5'→3')	Size (bp)	Conc (µM)	
M22	For	CCGCCATTCCTGGTGGCTCT	20	0.10	Multiplex 1
P25	For	CCCCCCTCTGCCTGAAACCTGCCTG	26	0.15	
92R7	Rev	CCCCGCATGAACACAAAAGACGTAGAAG	28	0.20	
SRY1532	For	CCCCCCTGTATCTGACTTTTTCACACAGT	30	0.20	
M70	Rev	CCCCCCTAGGGATTCTGTTGTGGTAGTCTTAG	34	0.15	
M173	For	CCCCCCCCCTTACAATTCAAGGGCATTAGAAC	34	0.20	
Tat	Rev	CCCCCCCCCCCCCCCCCTCTGAAATATTAATTAACAAC	42	0.20	
M213	Rev	CCCCCCCCCCCCCCCCCTCAGAATTAACAATCTCGTTAC	45	0.25	
M9	For	CCCCCCCCCCCCCCCCCGAAACGGCCTAAGATGGTTGAAT	48	0.20	
M170	Rev	ACACAACCCACACTGAAAAAAA	22	0.45	
M62	Rev	CCCCCCTAATGTTGTTGGCCATGGA	27	0.50	
M172	For	CCCCCCCCCCCCCAACCCATTTTGATGCTT	32	0.10	
M26	Rev	CCCCCCCCCCCCCATAGGCCATTAGTGTCTCTG	37	0.25	
M201	For	CCCCCCCCCCCCCGATCTAATAATCCAGTATCAACTGAGG	42	0.05	
M34	Rev	TTGCAGACACACCACATGTG	20	0.15	Multiplex 3
M81	For	CCCCCTAAATTTTGCTTTTTTGAA	27	0.20	
M78	For	CCCCCCCCCACACTTAACAAAGATACTTCTTTC	34	0.35	
M35	Rev	CCCCCCCCCCCCCCCCAGTCTGTGCCTGTGTC	36	0.03	
M96	For	CCCCCCCCCGTAACCTGGAAAAACAGGTCTCTCATAATA	40	0.05	
M123	Rev	CCCCCCCCCCCCCCCCCCCCCTTCTAGGTATTCAGGCGATG	51	0.35	Multiplex 4
M167	For	CCAAGCCCCACAGGGTGC	19	0.45	
M153	For	AAAGCTCAAAGGGTATGTGAACA	23	0.30	
M17	For	CCAAAATCACTTAAAAAAACCC	23	0.20	
M18	For	CCCCAGTTGTGGTTGCTGGTTGTTA	26	0.15	
M126	For	CCCCGCTTGAACCGAGTTTGTACTTAATA	30	0.05	
M37	For	CCGGAACACAGAAAATAAATCTATGTGTG	30	0.35	
M73	Rev	CCCCCCCCCGATTCTCTGAACGTCTAACCA	33	0.30	
M65	Rev	CCCCCCCCCCCCCCCCCCCCACCCGCGGTAAG	36	0.05	
M160	For	CCCCCCTTACAAGTTTAATACATACAACTTCAATTTTC	40	0.20	

Biosciences) at 37°C for 1 h, and at 85°C for 15 min to inactivate the enzyme.

Electrophoretic detection

The products of the single base extension reactions were run on an ABI 3100 Genetic Analyser (Applied Biosystems). Analysis of electropherograms was performed using the GeneScan 3.7 software (Applied Biosystems), determining the size of the fragments based on GeneScan-120 LIZ size standards.

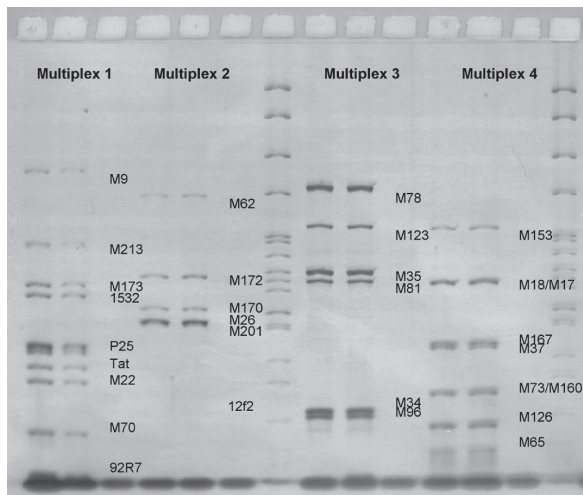
Statistical analysis

Binary marker haplogroup frequencies were calculated and Arlequin 2.0 software (Schneider et al. 2000) was used to test the hypothesis of a random distribution of the individuals between pairs of populations, through an exact test of population differentiation, and to calculate genetic distances, as pair-wise values of  $\Phi_{ST}$ . A multi-dimensional scaling (MDS) analysis was performed using SPSS version 11.5 software package, with the genetic distances.

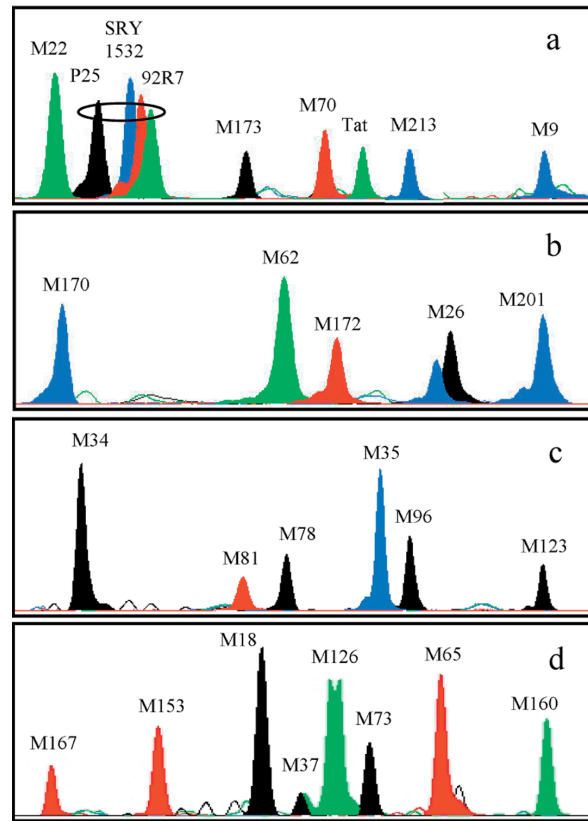
Results

The 30 SNPs were divided into 4 multiplex PCR/SBEs (Table 1), according to their location on the Y Chromosome Consortium tree. Multiplex 1 allows the detection of the more frequent major clades in Europe (Rosser et al. 2000; Semino et al. 2000; Brion et al. 2003), multiplex 2 determines haplogroups G, I and J, multiplex 3 subdivides haplogroup E, and multiplex 4 subdivides haplogroup R1b.

The products of the PCR multiplex reactions were designed to give different fragment sizes, allowing checking of the results in polyacrylamide gels. The sizes of the amplicons are represented in Table 1, and all fragments were



**Fig. 2** Multiplex PCR products separated in a polyacrylamide gel and silver stained. Two samples and a negative control were run for each multiplex, in this order. Lane 7 and 14 show a pBR322 DNA-Msp I digest molecular weight standard (New England Bio-Labs). Multiplex 2 includes the 12f2 deletion, which was typed by presence/absence in the PCR amplification



**Fig. 3a-d** Four SNaPshot multiplexes from different samples. **a** Multiplex 1 from sample assigned to haplogroup R1b, the P25 shows a duplicated pattern, **b** multiplex 2 from sample assigned to Hg I\*(xI1b2), M26 always shows an artefactual blue peak, **c** multiplex 3 from sample assigned to Hg E3b2, **d**, multiplex 4 from sample assigned to Hg R1b3f

unambiguously identified, even those fragments most similar in size, with only 3 bp differences (Fig. 2). None of the 292 samples analysed failed PCR amplification, therefore the multiplexes described seem to be a robust methodology for Y chromosome SNP typing.

In the primer extension reactions most of the samples gave a full profile (Fig. 3), however, some failed to give a detectable peak for certain SNPs or gave a very weak signal. In all these cases the SNPs which failed primer extension, also showed a weak signal in the PCR amplification. This problem was always resolved by repeating the PCR to produce a better amplification signal for the SNPs.

Multiplex 1 is a combination of 9 SNPs (Fig. 3a), however in the SNaPshot results more than 9 different primer extension peaks could be seen. The reason is because two SNPs showing paralogous sequence variants (PSV), P25 and 92R7 (Sanchez et al. 2004) form part of the multiplex. SBE multiplex 2 comprises 5 SNPs (Fig. 3b), however the results always produced 6 primer extension peaks. De-

**Table 3** Haplogroup frequencies and diversity in 8 different districts of Galicia

Haplotype	Galicia	Noroeste	Golfo Artabro	Mariña Lucense	Lugo	Santiago	Ourense	Rias Baixas	Montes Baixo Miño
E*(E3b)	0.0034	0	0	0	0.0164	0	0	0	0
E3b1	0.0205	0	0.0385	0	0.0164	0.0435	0	0	0.0714
E3b2	0.0411	0.0345	0	0	0.0820	0.0435	0.0541	0	0.0714
E3b3*(xE3b3a)	0.0034	0	0	0	0	0	0	0	0.0357
E3b3a	0.0103	0	0	0	0	0.0435	0	0.0323	0
G	0.0308	0.0690	0.0385	0.0588	0.0164	0.0435	0	0	0.0357
I*(xI1b2)	0.0959	0.1379	0.1154	0	0.0984	0.0870	0.0811	0.1290	0.1429
I1b2	0.0171	0.0345	0	0.0294	0	0	0.0270	0.0645	0
J*(xJ1,2)	0.0445	0.0345	0	0.1765	0.0164	0.0217	0.0541	0	0.0714
J2	0.1301	0.1724	0.1154	0.1471	0.1148	0.0870	0.1892	0.1290	0.1071
K2	0.0240	0	0	0.0294	0.0328	0.0435	0.0270	0	0.0357
R1*(xR1ab)	0.0103	0	0	0	0.0328	0.0217	0	0	0
R1a	0.0137	0	0.0385	0	0	0.0217	0.0270	0.0323	0
R1b*(xR1b1,2,3a-3f)	0.5377	0.5172	0.6538	0.5588	0.5738	0.5217	0.5405	0.5161	0.3929
R1b2	0.0034	0	0	0	0	0	0	0.0323	0
R1b3f	0.0137	0	0	0	0	0.0217	0	0.0645	0.0357
n	292	29	26	34	61	46	37	31	28
HgD	0.6806	0.6995	0.5631	0.6488	0.6486	0.7169	0.6757	0.7118	0.8228

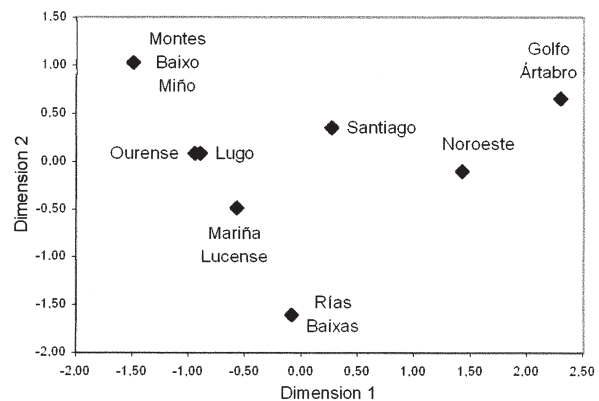
tailed analysis indicated the SNP M26 has an artefactual blue peak a few bases smaller than the true SBE peak. Because M26 is a G to A transition and we are analysing the reverse strand, the artefact does not affect the interpretation of results in this case. Multiplex 3 and 4 comprise 6 and 8 SNPs, respectively (Fig. 3c,d), and none of them exhibited artefactual results.

In order to check the reproducibility of the four PCR/SBE multiplexes and to know the haplogroup composition of the Galician population (northwestern Spain), 292 samples, taken from locations scattered throughout the whole region, were analysed. The strategy adopted was to perform multiplex 1 in all of the samples, and depending on the results obtained, to continue with the appropriate additional multiplex to define the haplogroup more precisely.

All the results were completely consistent with the Y Chromosome Consortium tree (Y Chromosome Consortium 2002; Jobling and Tyler-Smith 2003). The 30 SNPs analysed describe a total of 32 haplogroups, however, for the 292 samples analysed only 16 of these haplogroups were detected (Fig. 1). The frequencies and haplogroup diversity values are represented in Table 3, and it can be seen that the highest diversity value was present in Montes Baixo Miño (0.823).

An exact test of sample differentiation based on haplogroup frequencies was also performed, and only Mariña Lucense was significantly differentiated from Lugo and Rias Baixas ( $P=0.0089\pm 0.0021$  and  $P=0.0065\pm 0.0027$ , respectively).

We calculated genetic distances between populations as pairwise  $\Phi_{ST}$  values. The distance matrix was represented in two-dimensional space using multidimensional scaling (Fig. 4). Golfo Artabro, Montes Baixo Miño and Rias Baixas appeared as the furthest outliers, while Lugo and Ourense appeared as the closest populations.



**Fig. 4** Multidimensional scaling plot of Galician populations, from pairwise  $\Phi_{ST}$  distances based on Y-chromosome haplogroup frequencies. Stress value =0.003

## Discussion

We have developed a strategy for Y-chromosome SNP typing, which allows the quick assignment of a sample to one of 32 haplogroups defined by 30 SNPs, after only 1 or 2 multiplex reactions.

Despite the presence of some ambiguous positions (Weale et al. 2003), the phylogenetic tree of binary Y-chromosomal haplogroups (Y Chromosome Consortium 2002; Jobling and Tyler Smith 2003) has a stable structure at this moment, allowing researchers to assign the haplogroup of any sample, through a hierarchical analysis following the branches of the tree. This avoids unduly extensive SNP typing, saving time, costs and target DNA. The only disadvantage is the possibility of missing occa-



sional recurrent mutations, which are rare, due to the low mutation rate of the SNPs.

From a technical point of view, in the case of PCR multiplexes no problems emerged. The development of the PCR reaction was performed according to the guidelines of Henegariu et al. (1997), trying to establish similar melting temperatures for all the primers and to use the same standard conditions, in order to allow the incorporation of new SNPs or the regrouping of the SNPs used. In the development of the primer extension reactions, crucial factors are the purity of the primers, (HPLC purification is highly recommended) and the cleaning of the PCR product, with the minimal possible residual primers and dNTPs from the previous PCR (Sanchez et al. 2003). In our experience there is a clear correlation between the PCR results and the primer extension results; allele designation was been possible for all the markers in all the samples. However, all the cases of weak PCR amplification resulted in primer extension results that were difficult to interpret.

The results always correlated with the Y Chromosome Consortium tree. Haplogroup composition and frequencies are in agreement with previous publications (Rosser et al. 2000; Semino et al. 2000; Brion et al. 2003) including European studies. Once more, the high degree of population homogeneity present in Europe has been confirmed, since more than half the samples belong to the same paragrroup (157 individuals belong to paragrroup P). In addition, when the diversity was checked among the different districts belonging to Galicia, no significant differences were detected, except for the population from Mariña Lucense, which showed a distant haplogroup frequency but  $\Phi_{st}$  values comparable to the other regions (Fig. 4). It is clear that in a microgeographical study of binary haplogroups in a general population, no significant differentiation is expected. But this general pattern is not always true, and examples of clear differentiation at a local level have been shown in the literature (Brion et al. 2003). This can have an important impact in the forensic evaluation of the Y chromosome evidence and for this reason it is important to check for possible differentiation at a local level.

With a hierarchical strategy adapted for European populations extensive typing of SNPs was avoided, and therefore the time and cost involved in the study was reduced. Primer extension reactions using the SNaPshot multiplex kit, allowed with previous multiplex amplification and with an automatic sequencer, the quick development of a sufficiently large study, without investment in new technologies.

Despite the fact that Y-chromosome SNP analysis is becoming increasingly accessible and the number of polymorphisms detected is being extended even further, the population-specific diversity exhibited by these polymorphisms should be interpreted with great caution. In comparison to SNPs, microsatellites, which are variable in all populations, provide a less biased measure of diversity.

This must always be borne in mind in forensic analysis, when the haplogroup frequency of a random sample is estimated and interpreted according to anthropological data.

**Acknowledgements** The technical assistance of Meli Rodriguez and Raquel Calvo is highly appreciated. This work was supported by the grant from the "Ministerio de Ciencia y Tecnología" (DGCYT.P4.BIO2000-09822).

## References

- Borsting C, Sanchez JJ, Morling N (2004) Multiplex PCR, amplicon size and hybridization efficiency on the NanoChip electronic microarray. *Int J Legal Med* (in press)
- Brion M, Salas A, González-Neira A, Lareu MV, Carracedo A (2003) Insights over the Iberian population origin through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs and the MSY1 minisatellite. *Am J Phys Anthropol* 122:147–161
- Chen X, Sullivan PF (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenom J* 3:77–96
- Gill P (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* 114:204–210
- Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH (1997) Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques* 23:504–511
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–612
- Jorde LB, Watkins WS, Bamshad MJ (2001) Population genomics: a bridge from evolutionary history to genetic medicine. *Hum Mol Genet* 10:2199–2207
- Karafet TM, Zegura SL, Posukh O et al. (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817–831
- Rosser Z, Zerjal T, Hurler MH et al. (2000) Y chromosomal diversity within Europe is clinal and influenced primarily by geography, rather than language. *Am J Hum Genet* 67:1526–1543
- Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, Morling N (2003) Multiplex PCR and minisequencing of SNPs – a model with 35 Y chromosome SNPs. *Forensic Sci Int* 137: 74–84
- Sanchez JJ, Brion M, Parson W et al. (2004) Duplications of the Y-chromosome specific loci P25 and 92R7 and forensic implications. *Forensic Sci Int* 140:241–250
- Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland
- Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278–280
- Semino O, Passarino G, Oefner PJ et al. (2000) The genetic legacy of paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290:1155–1159
- Weale ME, Shah T, Jones AL et al. (2003) Rare deep-rooting Y chromosome lineages in humans: lessons of phylogeography. *Genetics* 165:229–234
- Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348
- Zhao Z, Fu YX, Hewett-Emmett D, Boerwinkle E (2003) Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* 312:207–213

**Artículo 4: “Multiplex de Y-SNPs para el análisis de los haplogrupos principales útil en la predicción del origen geográfico de linajes masculinos”**

Brión M, Sanchez JJ, Balogh K, Thacker C, **Blanco-Verea A**, Borsting C, Stradmann-Bellinghausen B, Bogus M, Syndercombe-Court, Schneider PM, Carracedo A, Morling N.

Electrophoresis **26** (2006): 4411-4420

Resumen:

El Consorcio Europeo “análisis de alto rendimiento de polimorfismos binarios para identificación forense de personas – SNP-ID” ha llevado a cabo una selección de Y-SNPs para deducir el origen geográfico de una muestra desconocida. De los más de 200 SNPs compilados en el árbol filogenético publicado por el Y-Chromosome Consortium en el momento de esta publicación, y considerando los estudios poblacionales publicados previamente, se ha seleccionado un paquete de 29 SNPs para la identificación de los principales haplogrupos poblacionales. Desarrollando así un “kit de tipado de los haplogrupos principales del cromosoma Y” que permite la amplificación en una única reacción de multiplex de los 29 SNPs. El genotipado de los alelos se realizó mediante una reacción “single base extensión” (minisequenciación) detectada por CE (electroforesis capilar). Para la validación del multiplex se usaron un total de 1126 hombres no relacionados distribuidos entre 12 poblaciones repartidas por todo el mundo. Este kit presenta ventajas a la hora de observar la distribución geográfica específica de los haplogrupos del cromosoma Y y demuestra la utilidad de los polimorfismos binarios para inferir el origen de un linaje masculino.



María Brión<sup>1</sup>  
 Juan J. Sanchez<sup>2</sup>  
 Kinga Balogh<sup>3</sup>  
 Catherine Thacker<sup>4</sup>  
 Alejandro Blanco-Verea<sup>1</sup>  
 Claus Børsting<sup>2</sup>  
 Beate Stradmann-Bellinghausen<sup>3</sup>  
 Magdalena Bogus<sup>3</sup>  
 Denise Syndercombe-Court<sup>4</sup>  
 Peter M. Schneider<sup>3</sup>  
 Angel Carracedo<sup>1</sup>  
 Niels Morling<sup>2</sup>

<sup>1</sup>Institute of Legal Medicine, National Genotyping Center (CEGEN), University of Santiago de Compostela, Spain

<sup>2</sup>Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen, Denmark

<sup>3</sup>Institute of Legal Medicine, University of Mainz, Germany

<sup>4</sup>Barts and the London, Queen Mary's School of Medicine and Dentistry, London, UK

## Introduction of an single nucleotide polymorphism-based “Major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages

The European Consortium “High-throughput analysis of single nucleotide polymorphisms for the forensic identification of persons – SNPforID”, has performed a selection of candidate Y-chromosome single nucleotide polymorphisms (SNPs) for making inferences on the geographic origin of an unknown sample. From more than 200 SNPs compiled in the phylogenetic tree published by the Y-Chromosome Consortium, and looking at the population studies previously published, a package of 29 SNPs has been selected for the identification of major population haplogroups. A “Major Y-chromosome haplogroup typing kit” has been developed, which allows the multiplex amplification of all 29 SNPs in a single reaction. Allele genotyping was performed with a single base extension reaction (minisequencing) detected by CE. The validation of the multiplex was performed in a total of 1126 unrelated males distributed among 12 worldwide populations. The approach takes advantage of the specific geographic distribution of the Y-chromosome haplogroups and demonstrates the utility of binary polymorphisms to infer the origin of a male lineage.

**Keywords:** Capillary electrophoresis / Geographic origin / Multiplex / Single base extension / Y-SNPs  
DOI 10.1002/elps.200500293

### 1 Introduction

The utility of Y-chromosome-specific markers has been already well established in many different areas. Their genetic diversity can be used to provide information on male-specific patterns of migration in the past, and the origin and diversity of specific populations [1–4]. In addition, in the forensic field they are shown to be useful for screening samples from sexual assault cases as well as to provide evidence on the male lineage in deficiency paternity cases [5–7].

Because of the lack of recombination in this genomic region, the polymorphic markers described cannot be considered as independent and therefore haplotypes have to be constructed. Desirable characteristics of markers are a high level of polymorphism in the popula-

tion, the lowest possible incidence of recurrent mutations, and a maximum of simplicity, allowing their detection in multiplex reactions and in small fragments of DNA. Polymorphisms such as microsatellites, historically used in the forensic field, show high levels of polymorphism but they are prone to recurrent mutational events that may generate haplotypes identical by state [8]. In contrast, single nucleotide polymorphisms (SNPs) which may include small insertions and deletions are amenable to classic phylogenetic analysis; however, they are not very polymorphic. This disadvantage is compensated by the simplicity of the markers, allowing the analysis of a high number of SNPs in a single multiplex reaction, using small quantities of DNA [9, 10].

In the case of autosomal SNPs used in a forensic context, the required number of SNPs to give comparable levels of information to STRs has already been determined [11]. However, as the Y-chromosome exhibits extensive genetic differentiation, the selection of SNPs is far more complicated. This differentiation across geographical distance [12] results in markedly different haplogroup profiles distribution, depending on the region of the world studied [1, 13].

In 2002, the Y-Chromosome Consortium (YCC) [14] has compiled 245 binary polymorphisms in a parsimonious tree defining 153 haplogroups. Subsequent publications

**Correspondence:** Dr. María Brión, Institute of Legal Medicine, National Genotyping Center (CEGEN), University of Santiago de Compostela, San Francisco s/n. E-15782, Santiago de Compostela, Spain

**E-mail:** brioniml@usc.es

**Fax:** +34-98-158-0336

**Abbreviations:** AMOVA, molecular analysis of the variance; SBE, single base extension; SNP, single nucleotide polymorphism; YCC, Y-Chromosome Consortium



[15–17] have increased the number of SNPs, but all of them are classified into the 18 major clades (A–R) defined by the YCC. It has been described that most of the haplogroups belonging to each of these major clades tend to be confined to specific continents [15], so this geographical distribution can be used to make inferences on the possible geographic origin of any sample of interest. In a forensic scenario this may serve as a tool to predict the origin of the paternal lineage of any stain contributor.

A large number of SNP genotyping methods are now available [18, 19], and usually the choice of the appropriate method depends on the number of SNPs and the number of individuals that need to be typed. The work presented here was planned to be used for laboratories involved in forensic DNA analysis, so the selected genotyping method was a multiplex PCR followed by single base extension (SBE) reaction using the SNaPshot multiplex kit. The main reason for this selection was the possibility of genotyping SNPs without investment in new technologies, simply using a CE instrument, which is available in most of these laboratories for multiplex STR typing by fragment analysis.

The main objectives of the present study were, first, the selection of candidate Y-SNPs to provide clues on the possible geographic origin of a sample. The second objective was the development of multiplex strategies, allowing the analysis of as many SNPs as possible in a single reaction. Finally, the last goal was the validation of the strategy in a sample of more than one thousand unrelated males distributed all around the world, to check the accuracy and reproducibility of the multiplex, as well as the potential to differentiate between the major population groups.

## 2 Materials and methods

### 2.1 Samples and DNA purification

Nine human DNA samples from eight males and one female and one dog male were distributed among the participating labs, in order to compare the results of the new multiplex assay. The dog sample was used to check the possibility of amplification because of homologous regions between the human and dog Y chromosome.

A total of 1 126 unrelated males from Denmark (150), Greenland (90), Turkey (51), Thailand (84), China (55), Japan (40), Germany (150), Galicia (130), Mozambique (130), Somalia (105), Argentina (81 racially mixed individuals), and Colombia (60 ethnically interbred individuals) were typed. The samples were distributed equally among three of the participating labs (Copenhagen,

Mainz, and Santiago de Compostela) to be typed for the 29 SNPs. Blood on FTA cards, DNA purified by phenol/chloroform extraction or the QIAamp DNA blood mini kit (Qiagen) was used. DNA concentration was determined by real-time PCR using the Quantifiler™ Human DNA Quantification Kit and the ABI 7300 instrument (Applied Biosystems, Foster City, CA, USA) or using the NanoDrop® ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies, Rockland, DE). The protocols were approved by the local ethical committees (Copenhagen: KF-01-037/03, Santiago USC-FM 030104).

### 2.2 Selection of Y-chromosome markers

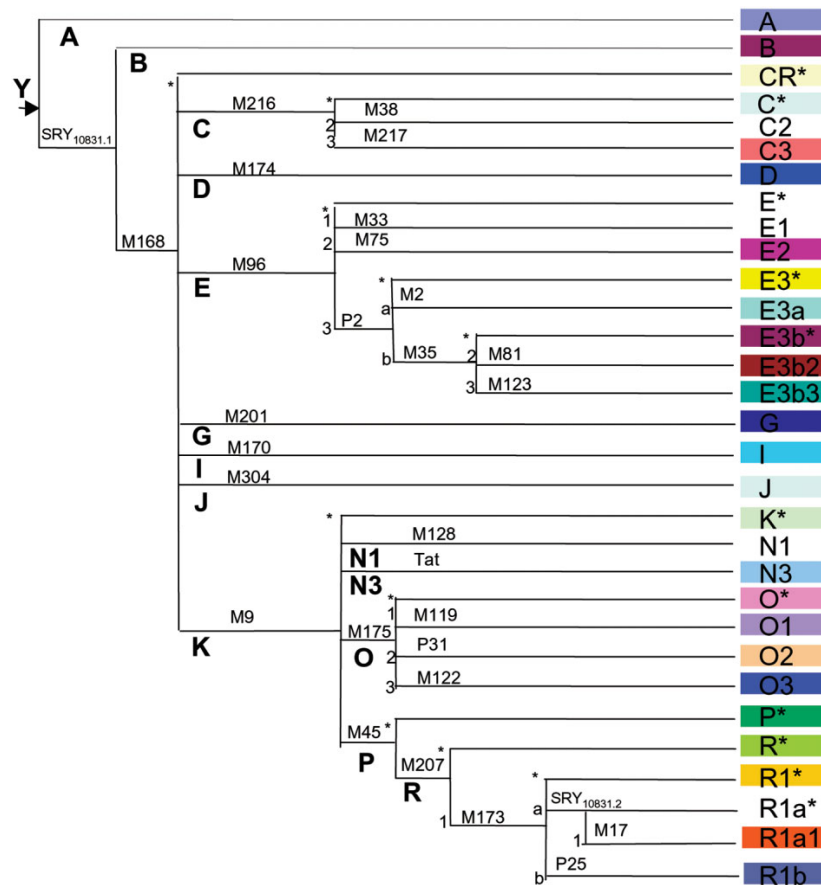
From population studies previously published [1, 20–22] a package of 29 SNPs was selected for the identification of major population haplogroups present all around the world. The 29 SNPs and the haplogroup tree defined with them are represented in Fig. 1. Twelve SNPs define major clades and the others subdivide some of these clades.

### 2.3 Primer design

PCR primers (Table 1) and SBE primers (Table 2) were designed according to the guidelines outlined in Sanchez *et al.* [9]. HPLC-purified oligonucleotides were purchased from DNA Technology A/S (Aarhus, Denmark). The quality of the oligonucleotides was investigated by MALDI-TOF analysis.

### 2.4 PCR conditions and purification of PCR products

The Major Y Haplogroup multiplex PCR was amplified in a 25 µL reaction volume containing 1–10 ng DNA, 1 × PCR buffer, 8 mM MgCl<sub>2</sub>, 700 µM of each dNTP, 0.26–7.15 µM of each primer, and 2 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems) using the following cycling program: denaturation at 94°C for 5 min followed by 35 cycles at 95°C for 30 s, 60°C for 30 s, and 65°C for 30 s, followed by 7 min at 65°C. All PCR amplifications were performed in GeneAmp PCR system 9600 or 9700 (Applied Biosystems) or Eppendorf Mastercycler gradient instruments. PCR products were purified on a MinElute™ PCR purification spin column (Qiagen) and the DNA eluted in 20 µL of Milli-Q water. Alternatively, the PCR products were purified by adding 1 µL ExoSAP-IT™ (Amersham Biosciences), or by adding 0.75 U SAP (Amersham Biosciences) and 0.225 U Exol (USB Corporation) to 2.5 µL PCR product. The mix was incubated at 37°C for 15 min or 37°C for 1 h, respectively. The SAP and Exol enzymes were inactivated by incubation at 75°C for 15 min.



**Figure 1.** Phylogenetic tree defined with the binary Y-chromosomal polymorphisms analyzed. Marker names are indicated above the lines and lineage names are shown below the lines, but the length of each branch has no significance. Only the haplogroups with a colored square appear in the population samples analyzed.

## 2.5 SBE reaction and purification of the SBE products

The SBE reactions were performed in a total volume of 8  $\mu$ L with 0.3  $\mu$ L purified PCR product, 4  $\mu$ L SNaPshot™ reaction mix (Applied Biosystems), and 1  $\mu$ L SBE primer mix (Table 2). The SBE primer mix was stored in 160 mM ammonium sulfate (Sigma-Aldrich). The SBE reaction was performed using the following program: 30 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 30 s. After the SBE reaction, 1 U of SAP (Amersham Biosciences) was added and the mix was incubated at 37°C for 45 min. SAP was inactivated by incubation at 75°C for 15 min.

## 2.6 Detection and analysis of the SBE products

Two microliters SAP-purified SBE products were mixed with 20  $\mu$ L Hi-Di formamide and analyzed by CE using an AB 310 or AB 3100 genetic analyzer with 36-cm capillary arrays (Applied Biosystems), POP-4 polymer (Applied Biosys-

tems), and 22 s injections at 2000 V. Analysis was done using GeneScan Analysis 3.7 with a minimum peak height of 100 relative fluorescence units (RFUs) for all dyes. Automated allele calls were made using Genotyper 3.7 macros.

## 2.7 Sequencing analysis

The PCR-amplified fragments corresponding to the locus M174 were purified on a MinElute PCR purification spin column (Qiagen) following the manufacturer. The DNA was eluted in 10  $\mu$ L of Milli-Q water. Both strands of the PCR product from four males and one female were sequenced by DNA Technology A/S (Arhus, Denmark).

## 2.8 Statistical analysis

Binary marker haplogroup frequencies were calculated in the 12 populations analyzed. Arlequin 2.0 software [23] was used to calculate genetic distances, as pairwise

**Table 1.** Y-chromosome markers and primer sequences for the multiplex PCR amplification

Locus	GenBank or DbSNPs accession	Mutation	PCR primers (5' → 3')		μM	Amplicon size, bp
			Forward primer	Reverse primer		
M2	rs3893	A > G	acggaaggagtctctaaaattcagg	aaaatacagctcccctttatctct	1.63	128
M9 <sup>b)</sup>	rs3900	C > G	aggaccctgaaatacagaactg	aaatattcaacatttcacaaaggaa	6.50	186
M17 <sup>b)</sup>	rs3908	4G > 3G	cctggctataacactggaaatc	agctgaccacaaactgatgtaga	1.95	170
M33 <sup>b)</sup>	AC009977	A > C	tgaccgtcataggctgagaca	ttgaagcccccagagagac	3.58	160
M35 <sup>b)</sup>	AC009977	G > C	agggcatggtccctttctat	tccatgcagactttcgaggt	3.90	96
M38	AC009977	T > G	agatctgttgctactgtacccta	gagctggcacatctgtcataata	1.30	177
M45 <sup>b)</sup>	Rs2032631	G > A	gagagaggatatcaaaaattggcagt	tgacagtgccaccaagggtc	0.39	138
Tat <sup>b)</sup>	AC002531	T > C	tatatggactctgagtgtagactgtga	ggtgccgtaaaagtgtgaaataatc	0.52	115
M75	AC010889 <sup>c)</sup>	G > A	cttgcaaaagccaaacagatttc	ttgaacagaggcattttggaat	5.20	120
M81 <sup>b)</sup>	rs2032640	C > T	catctcttaacaaaa gaggtaaatttgtcc	cattgtgtacatggcctataat attcagt	2.60	179
M96 <sup>b)</sup>	AC010889 <sup>c)</sup>	G > C	tgccctctcacagagcactt	ccaccactttgttgctttg	7.15	143
M119	AC010889 <sup>c)</sup>	A > C	gatgtagaaaaatgttatgggtattcc	gaaatgccaaggtaaatgactcac	0.78	188
M122 <sup>a)</sup>	AC010889 <sup>c)</sup>	T > C	gttgccctttgaaatgaataaatc	gaatttgattctaggtattcagg	0.65	175
M123 <sup>a)</sup>	AC010889 <sup>c)</sup>	G > A	gttgccctttgaaatgaataaatc	gaatttgattctaggtattcagg	0.65	175
M128	AC010889 <sup>c)</sup>	-2 bp	gcaaaagagaactgcctctta taaaatc	gcagaattatctacctcttcaactg	2.60	139
M168	Rs2032595	C > T	gtggagtatgtgtggaggtga	cctctatcagaccatggaatctca	2.60	145
M170 <sup>b)</sup>	rs2032597	A > C	cagctcttattaagttatgtttca tattctgtg	gtcctcattttacagtgagacacaaac	1.63	119
M173 <sup>b)</sup>	rs2032624	A > C	ttttctacaattcaaggcatttag	ctgaaacaaaactggtctatca	0.98	81
M174	rs2032602	T > C	gcttctctgaataccttctggagt	tcttgaaggaaaagtgtgc	0.46	94
M175 <sup>b)</sup>	rs2032678	-5 bp	gatttaactctctgaatcaggcacat	tttactgatacctttgttctgttcattc	0.65	79
M201	rs2032636	G > T	tcagatctaataatccagtatc aactgag	gttcaaatccatattccagca	0.65	93
M207	AC006376	A > G	ggggcaaatgtaagtcaagc	tcactcaacctctgttgga	0.33	83
M216	AC010877	C > T	ccaatggaaatattataaccaca	tgacactgctagtattatgatacctgttg	1.63	160
M217	rs2032668	A > C	tctgtttcgagatcattctaactactg	ctgctgtggctttcatcaaaata	1.95	145
M304	AC007678	A > C	caaagaaaagcaggagagttgtaa	aaactgtctataccaaaatattcaccag	0.65	90
P2	AC010137 <sup>c)</sup>	C > T	agccatctttcttaagagaggt	gcttgagttctcttaggttttga	0.26	104
P25 <sup>b)</sup>	rs150173	C > CA	tgaccatcacctgggtaagt	ggcagataagggtgtcacatcacat	0.26	109
P31	AC002992	T > C	aacaggtaggtgtattttgttaca	gacccgatatcgtgccatt	5.85	122
SRY <sub>10831</sub> <sup>b)</sup>	rs2534636 <sup>c)</sup>	A > G	tcatccagtccttagcaaccatta	ccacataggtgacacctgaaaatg	1.63	150

a) Markers on the same DNA fragment.

b) PCR primers were published previously [24].

c) These accession numbers contain opposite sequences compared to the information reported in YCC 2002.

values of  $F_{ST}$ , and to perform a molecular analysis of the variance (AMOVA) analysis. A multidimensional scaling (MDS) analysis was performed on the genetic distances using SPSS version 11.5 software package.

### 3 Results

#### 3.1 Y-chromosome SNP selection

From a forensic point of view, and because the Y-chromosome cannot be used to individualize samples, our interest was focused on selecting SNPs to make an inference of the geographic origin of any male

sample. For this purpose, from more than 250 SNPs compiled in the phylogenetic tree published by the YCC, a package of 29 SNPs has been selected for the identification of major population haplogroups.

In the first step SNPs were selected according with their position in the parsimony tree and according to the haplogroup determined by them and their frequencies among populations. In the second step, some of the SNPs previously selected were changed due to problems in multiplex amplification or because new SNPs likely to be more useful were discovered during the interval between the two steps.

**Table 2.** SBE primer sequences for the detection of 29 Y-chromosome markers

Locus	Poly (dNTP) (5' → 3')	Sequence of 5'-tail (5' → 3')	Target-specific sequence (5' → 3')	<sup>a)</sup>	μM	Primer size, nt
M174	None	None	taccttctggagtgcc	F	0.06	17
M45	None	caa	ctcagaaggagcttttgc	R	0.14	22
M2	None	gacaa	ctttatcctccacagatctca	R	0.20	26
Tat	None	None	gctctgaaatataaattaacaac	R	0.08	26
M170	None	T	ctattttacttaaaaaatcattgtc	F	0.05	30
M217	None	acaa	ttatgtattttccttctgaagagt	R	0.07	30
P25	None	tcgtgaaagtctgacaa	tgccctgaaacctgcctg	F	0.41	34
M201	None	aagtctgacaa	taataatccagatcaactgagg	F	0.02	34
M38	None	gtgaaagtctgac	cagtactattatggaaaaccaact	F	0.10	38
M304	None	gaaagtctgacaa	tgttcaattgaaagtaactgtga	F	0.14	38
M123	None	acgtcgtgaaagtctgacaa	atttctaggtattcaggcgatg	R	0.68	42
M207	None	cgtgaaagtctgacaa	caaatgtaagtcaagcaagaaatta	F	0.24	42
M35	None	gccacgtcgtgaaagtctgacaa	cccaattttccttgggacactg	F	0.54	46
M128	None	cgtgaaagtctgacaa	cctcttcaaacgtaaatgaaaataactg	F	0.11	46
P31	None	gccacgtcgtgaaagtctgacaa	tacataaataaggttttttgggtg	F	0.08	50
M216	None	ggtgccacgtcgtgaaagtctgacaa	gctagtattgtatacctgtgaat	R	0.08	50
M17	None	aactaggtgccacgtcgtgaaagtctgatct	ccaaaattcacttaaaaaaaccc	R	0.07	54
M175	None	ctaaactaggtgccacgtcgtgaaagtctgacaa	cacatgccttctcacttctc	F	0.24	54
M173	None	actaaactaggtgccacgtcgtgaaagtctgacaa	tacaattcaagggcatttagaac	F	0.07	58
M119	None	gactaaactaggtgccacgtcgtgaaagtctgacaa	tattccaattcagcatacaggc	R	0.20	58
M96	None	aactgactaaactaggtgccacgtcgtgaaagtctgacaa	ggaaaacaggtctctcataata	F	0.36	62
M122	DC + 5(dTdC) + dT	ctaaactaggtgccacgtcgtgaaagtctgacaa	tcagattttcccctgagagc	R	0.10	66
M75	4(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	aaagacaattatcaaacacatcc	F	0.02	70
SRY <sub>10831</sub>	5(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	ttgtatctgacttttccacacagt	F	0.06	74
M168	DC + 6 (dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	ctattgttttaattctcagctagc	R	0.03	78
M81	7(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	cttggtttgtgtgagtatactctatgac	R	0.08	82
M9	9(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	catgtctaaattaagaaaaataaagag	R	0.88	86
P2	16(dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	gtgcccttaggaggagaa	R	0.10	90
M33	DC + 12 (dTdC)	aactgactaaactaggtgccacgtcgtgaaagtctgac	cagttacaaaagtataatgtctgagat	R	0.27	94

a) Detection orientation has been probed relative to the YCC information reported in YCC 2002.

The final SNPs selected are represented in Fig. 1, where the haplogroups defined by them and their phylogenetic relationship are also shown.

### 3.2 Y-chromosome SNP genotyping

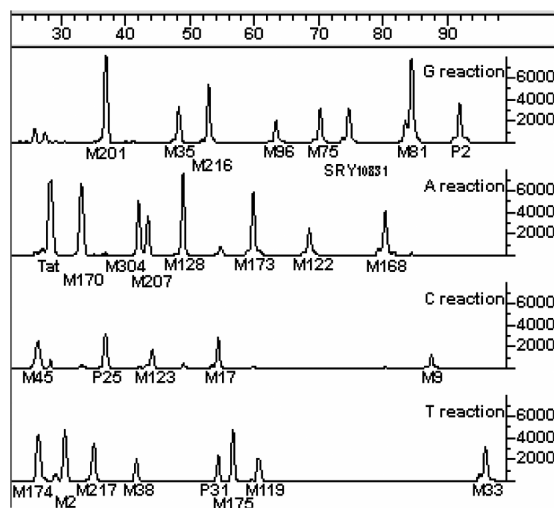
Once the SNPs were selected, a multiplex assay was developed using the guidelines for PCR multiplex design previously described in Sanchez *et al.* [9, 24], allowing the

simultaneous amplification of all 29 loci in a single reaction. The PCR multiplex design was performed by considering (i) the quality of flanking sequence of the candidate SNP, (ii) the predicted interactions between primers, and (iii) homologies to other amplicons in the multiplex and to other sequences in the human genome. It was also important to ensure that the PCR fragments were as short as possible and of uniform size, in order to obtain efficient amplification of all loci. The final result was a stable PCR multiplex reaction with 28 amplicons ranging in size from

79 to 188 bp (the two SNP loci M122 and M123 are located on the same amplicon). The SBE primers were also designed according to the guidelines described in Sanchez *et al.* [9, 24]. The target specific sequences of the SBE primers ranged from 17 to 29 nucleotides and the noncomplementary tails from 0 to 72 nucleotides. Tails were added in order to separate the SBE products efficiently in the subsequent electrophoresis. The SBE products in the final multiplex design ranged from 18 to 95 nucleotides in size. All PCR and SBE primers were first tested in singleplex reactions and later in multiplex reactions. Balanced peak heights in the final SBE reaction were obtained by altering the primer concentrations used in the PCR and SBE multiplexes. The final PCR and SBE primer concentrations are given in Tables 1 and 2, respectively.

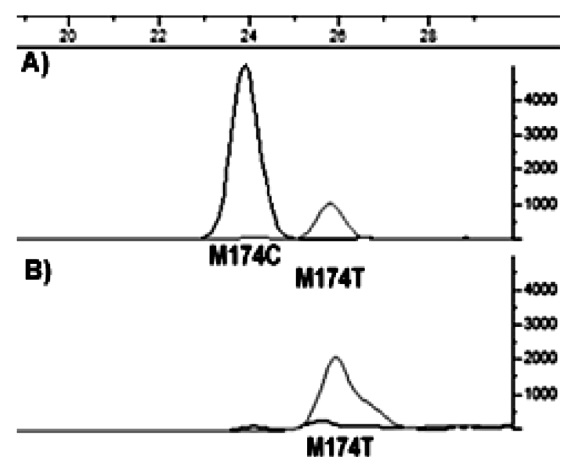
In Fig. 2, a representative electropherogram of the 29-plex is shown. For the short SBE primers a lack of correlation was observed between the observed and expected sizes. This may be due to the sequence composition of the SBE primers and the relative contribution of the respective fluorochrome of the labeled ddNTP. The difference between the observed and the expected size was not observed for the longer SBE primers.

The quality of the electropherograms is completely dependent on the dosage and quality on the initial DNA template. The optimum was found with 1–10 ng; however, good results and complete genotyping were achieved with only 0.25 ng. Fresh samples and old samples frozen for more than 3 years were analyzed and complete profiles were obtained.



**Figure 2.** Electropherogram with 29 Y-chromosome SNP profile from a male donor belonging to the Hg K\*(xN1, N3, OP).

One of the 29 selected SNPs, P25, is the result of a duplication event [25]. When the P25 marker has its ancestral allele, the duplication is not detected, but with the derived allele, the duplication is observed. Similarly, two M174 alleles were detected in 12 Japanese samples and one Chinese sample (example shown in Fig. 3) containing the derived allele of M174, whereas all other samples contained the ancestral allele alone. The sequence of the M174 locus was BLASTed against the database and we realized that this SNP is located in the ubiquitin-specific protease 9 gene (USP9Y), which has a homolog on the X-chromosome (USP9X). Looking at the great homology between the X- and Y-chromosome sequences, with only four-base difference, and since we had observed a peak comparable in size to the SBE products of the ancestral M174 allele in our female control samples, we suspected that the homologous region on the X-chromosome may be amplified in the PCR reaction. We amplified M174 in singleplex from four of the Japanese males and one Danish female, and sequenced the PCR products. A fragment homologous to the M174 locus was amplified from the female sample and the fragment contained the ancestral M174 allele. In addition, at least two fragments were amplified from the Japanese male samples. One fragment contained the derived allele and one fragment contained the ancestral allele. We made mixtures of male and female DNA with a dominant amount of female DNA (0.5, 5, 10, 20, and 46 ng) and a fixed amount of male DNA (0.25 ng). Two different male samples and a female sample were used. A complete 29-plex profile was obtained for all mixtures; however, the signal from the ancestral M174 allele increased proportionally to the amount of female DNA in the mixture (data not shown). Therefore, we conclude that the X-chromosome USP9X



**Figure 3.** GeneScan analysis of SBE reactions performed on M174 PCR products from a male donor belonging to the Y-chromosome haplogroup D (A) and a female donor (B).

locus was also amplified in the multiplex PCR. An alternative M174 primer design is being carried out to avoid possible preferential amplifications in case of mixtures.

### 3.3 Intercomparison exercise

An intercomparison exercise was performed with ten reference samples analyzed by all the participating labs. Haplogroups determined for eight male samples were concordant in all the labs. No results were obtained with the dog DNA, whereas a single extension product similar in size to the ancestral allele of the Y-chromosome M174 marker was detected for the female control sample.

### 3.4 Worldwide population study

A population study was performed to obtain allele and haplogroup frequencies for the major population groups. Among the 31 haplogroups defined by the 29 SNPs analyzed, 26 of them were found in a total of 1126 individuals distributed among four continents, America, Asia, Africa, and Europe. Haplogroup frequencies and diversities are represented in Table 3. The most frequent haplogroups are E3b\*(xE3b2, E3b3) in Somalia and E3a in Mozambique, which represent more than 70% of the samples of these populations, R1b in Galicia and Colombia, with more than 50% of the samples in each population, and O3 in China also representing more than 50% of the samples.

**Table 3.** Haplogroup frequencies

Haplogroup	Colombia (n = 60)	Argentina (n = 81)	Greenland (n = 90)	Thailand (n = 84)	China (n = 55)	Japan (n = 40)	Mozambique (n = 130)	Somalia (n = 105)	Germany (n = 150)	Denmark (n = 150)	Galicia (n = 130)	Turkey (n = 51)
A							0.008					
B							0.177	0.010				
CR*(xCE,G,I,K)				0.036		0.025		0.029			0.008	
C*(xC2,C3)				0.036		0.025						
C2												
C3			0.011	0.036	0.055	0.075						
D					0.018	0.300						
E*(xE1,E2,E3)												
E1												
E2							0.069					
E3*(xE3a,E3b)									0.020		0.008	
E3a	0.083						0.715	0.010	0.020			
E3b*(xE3b2,E3b3)	0.067		0.011					0.743	0.067	0.020	0.046	0.020
E3b2	0.033						0.008	0.029			0.046	
E3b3	0.033	0.012							0.013	0.007	0.008	0.059
G	0.017	0.074						0.010	0.020	0.007	0.015	0.098
I	0.033	0.062	0.222						0.193	0.373	0.131	0.078
J	0.067	0.099	0.022	0.012				0.019	0.067	0.027	0.146	0.333
K*(xN1,N3,OP)	0.017	0.012		0.048	0.055	0.025		0.143	0.007		0.015	0.039
N1												
N3		0.012							0.013			
O*(xO1,O2,O3)				0.024								
O1				0.107	0.164	0.025						
O2				0.250	0.164	0.350						
O3				0.429	0.545	0.175						
P*(xR)	0.050	0.346	0.533						0.020	0.013		0.020
R*(xR1)	0.050											0.020
R1*(xR1a,R1b)										0.020	0.008	
R1a*(xR1a1)												
R1a1	0.017		0.056	0.024				0.010	0.133	0.173	0.008	0.137
R1b	0.533	0.383	0.144				0.023		0.427	0.360	0.562	0.196
HgD <sup>a)</sup>	0.702	0.723	0.649	0.744	0.654	0.768	0.455	0.429	0.757	0.704	0.646	0.826

a) Haplogroup diversity.



Figure 4 represents the observed haplogroup frequencies in the 12 populations in relation to the world map. At first glance the differences among populations are obvious; however, an ANOVA analysis was performed to look at this in greater depth, grouping the samples by continents. The results showed a high percentage of variation among continents, 33.7% ( $p < 0.001$ ) and a low percentage of variation among populations within continents, 6.8% ( $p < 0.001$ ). Most of the variance was attributable to differences within populations, 59.5% ( $p < 0.001$ ). These differences among groups of populations are also evident when a Principal component analysis was performed with the haplogroup frequencies (data not shown) and also when the first distances were multidimensionally scaled (Fig. 5). Both Asian and African populations are clearly grouped in a cluster in both analyses. As the two South American populations analyzed here have a high degree of historical coancestry with Europeans they tend to cluster together with European populations.

#### 4 Discussion

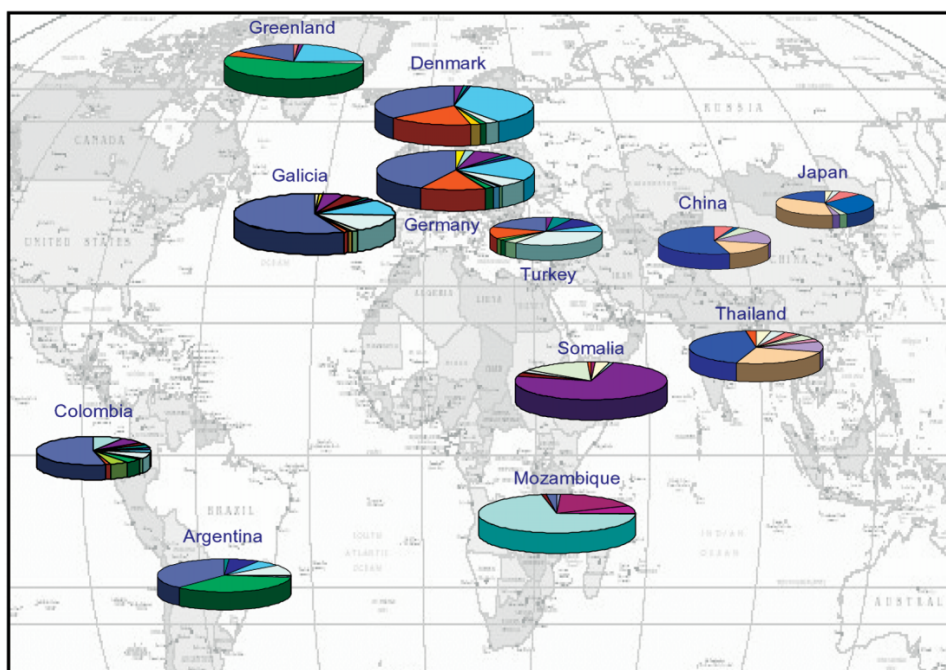
The objective of the SNPforID consortium (<http://www.snpforid.org/>) is the selection of SNPs suitable for forensic typing which should form the core set of markers

for future forensic high-throughput typing technologies. For the Y-chromosome, the interest was focused on making inferences on the geographic origin of an unknown sample. To evaluate this really interesting feature of the Y-SNPs, we selected the 29 SNPs used in this study as a first step of future designs more specific for each major haplogroup.

#### 4.1 Technological remarks

Once the SNPs had been selected, the second objective of the consortium was to find an appropriate technology for typing the SNPs in any interested laboratory. Nowadays, a large number of SNP genotyping methods are available; however, some of them require high investments and new developments, not always accessible in most of the routine laboratories. The selected method in this case was an SBE reaction, using the SNaPshot multiplex kit of Applied Biosystems, with the only requirement being access to an automatic sequencer.

From a technical point of view, we can conclude that the Major Y-chromosome haplogroup typing kit is a reliable technology, since the results for the control samples sent to all the labs were concordant; is a robust system, since



**Figure 4.** Haplogroup frequencies. Diagrams for each population are located over their position in the world map. Size of diagrams is proportional to the sample size, and haplogroup colors are the same as those in Fig. 1.

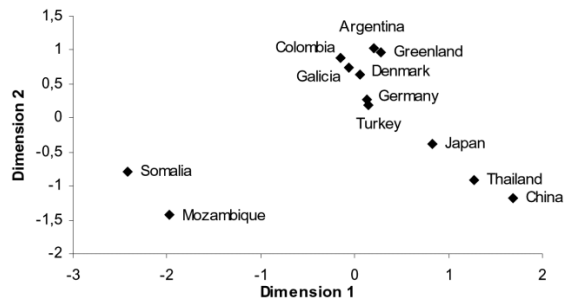


Figure 5. MDS of the  $F_{ST}$  distances. Stress value 0.07.

good results were obtained from diverse samples with different quantities and qualities; and as it was described in many studies [4, 25, 26] the SBE method used is an accurate technique, without the necessity of investment in new technologies. Our strategy of multiplex PCR and primer extension reaction can, however, easily be adapted to other platforms, such as MALDI-TOF-MS [27].

#### 4.2 Population validation

Among the 31 haplogroups defined using the 29 binary markers, 26 were found in the populations analyzed. Setting aside the South American samples (Colombians and Argentinians) and the Greenlanders, because they are considered as mixed populations, six haplogroups appeared only in Asians, three haplogroups only in Africans, and seven haplogroups only in Europeans. We have named these haplogroups “private haplogroups.”

Looking more deeply at the South-east Asian data, their private haplogroups were C\*, D, O\*, and its immediate sub-clades, O1, O2, and O3. As haplogroup C3 is also typical of these Asian populations [28], we decided to consider it as a private haplogroup, in spite of it being detected in one individual from Greenland because of the admixed origin of this population. These Asian haplogroups include 88.2% of the Thai samples, 94.6% of the Chinese samples, and 95% of the Japanese samples; so if we consider these populations to be representative of the South-east Asians as a whole, we could say that approximately 93% of people with a south-east Asian patrilineage could be detected anywhere with our Major Y-chromosome haplogroup typing kit.

Among the African samples, only three private haplogroups were observed, A, B, and E3; in addition, haplogroup E3a has also been described as a haplogroup with a sub-saharan origin [29], so we have considered it as private for Africans. The three German individuals as well as the single individual from Colombia most probably

have a sub-saharan origin. The private haplogroups in this case include 96.9% of the samples from Mozambique, but only the 2% of the Somali samples; so it is clear that these haplogroups are not representative of all African populations, but the sub-saharan ones. Most of the samples from Somalia (74%) belong to haplogroup E3b, which was described to have a sub-saharan origin, expanding into North Africa and the Near East, and then into southern Europe [30, 31]. The Somali samples were included in a group of 201 Somali immigrants to Denmark in whom the frequency of E3b1 was 77.6% [31].

Europeans were the most abundant group of samples in this study, and presented seven private haplogroups, E3\*(xE3a, E3b, E3b3, I, N3, P\*(xR), R\*(xR1), and R1\*(R1a, R1b). However, haplogroup R1b described as an European haplogroup [1] also appeared in three samples from Mozambique, and can probably be explained by European admixture. If we set aside these three samples and consider R1b as a private haplogroup, they represent 68.6% of individuals in the German population sample, 77.3% in Denmark, 71.7% in Galicia, but only 37.3% of individuals in Turkey. The highest frequency of the seven haplogroups, 89.9%, was found in Greenland. This population despite the fact of being historically influenced by European migration [32], shows a high frequency of haplogroup P\*(xR) because the Native American haplogroup Q was not differentiated with the selected set of markers. On average, setting aside the Greenlanders, we could say that 62.1% of the people with a European origin could be detected anywhere with our Major Y-chromosome haplogroup typing kit. When other haplotypes are considered which do not have a private distribution, then a clear prediction is not possible. Nevertheless, one might consider making a relative prediction based on a likelihood ratio. In this case, two alternative hypotheses need to be formulated which can be calculated on the basis of the relative proportions among the different populations exhibiting a given haplogroup.

#### 4.3 Concluding remarks

The possibility to make inferences on the geographical origin of an unknown sample contributor has an enormous value in the investigative phase of a criminal procedure. However, it cannot be used as scientific evidence in the trial, where the individual DNA profile of the accused person is directly compared with the respective profiles of the samples collected at the scene of the crime.

Looking at the results obtained with the population samples selected, the Major Y-chromosome haplogroup typing kit seems to have utility in making inferences about the possible geographic origin of any sample of interest. With



this first panel of SNPs it can be possible to discriminate only between the major human groups present all around the world. To provide discrimination between more closely related populations, a new selection of markers will be required. Depending on the composition of the populations, different multiplexes will be necessary in combination to assign a sample to a specific haplogroup to give a more accurate prediction on the geographic origin of a sample. It is evident, however, that a Y haplotype-based prediction will always be limited to the elucidation of the patrilineage and thus could be quite misleading regarding the phenotype of a given individual. It cannot be excluded that a person exhibiting a European phenotype may carry a sub-saharan or Asian Y-chromosome, which has been passed on over many generations. To overcome this pitfall, a more complex approach is required which has to include autosomal SNPs with population-specific distributions [33], as well as the assignment to a mitochondrial DNA haplogroup [34] to include information as well about the maternal lineage. When all these three independent sources of genetic variation are considered equally by combining them into a single likelihood calculation, an accurate prediction of the geographic origin of an unknown sample contributor may be possible.

Furthermore, it has to be kept in mind that if we want to take advantage of the specific properties of Y-chromosome inheritance, we should always check that the markers we are analyzing do not have homologous loci on the X-chromosome, as it is the case for the M174 polymorphism analyzed here and in previous studies [21, 35].

*We want to thank Amparo Acosta from the University of Cauca (Colombia) for providing the Colombian samples. The technical assistance of Annemette Holbo Birk, Meli Rodríguez, and Raquel Calvo is highly appreciated. This work was partially supported by the European Community (GROWTH programme, SNPforID, contract G6RD-CT-2002-00844), the "Ministerio de Ciencia y Tecnología" (BMC2003-09822) and Ellen and Aage Andersen's Foundation.*

Received April 15, 2005

Revised July 24, 2005

Accepted August 2, 2005

## 5 References

- [1] Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., *et al.*, *Am. J. Hum. Genet.* 2000, 67, 1526–1543.
- [2] Arredi, B., Poloni, E. S., Paracchini, S., Zerjal, T., *et al.*, *Am. J. Hum. Genet.* 2004, 75, 338–345.
- [3] Semino, O., Magri, C., Benuzzi, G., Lin, A. A., *et al.*, *Am. J. Hum. Genet.* 2004, 74, 1023–1034.
- [4] Brion, M., Quintans, B., Zarrabeitia, M., Gonzalez-Neira, A., *et al.*, *Gene* 2004, 329, 17–25.
- [5] Gusmao, L., Brion, M., Gonzalez-Neira, A., Lareu, M., Carracedo, A., *Leg. Med.* 1999, 1, 55–60.
- [6] Jobling, M. A., Pandya, A., Tyler-Smith, C., *Int. J. Legal Med.* 1997, 110, 118–124.
- [7] Jobling, M. A., *Forensic Sci. Int.* 2001, 118, 158–162.
- [8] Pereira, L., Prata, M. J., Amorim, A., *Forensic Sci. Int.* 2002, 130, 147–155.
- [9] Sanchez, J., Borsting, C., Morling, N., *Methods Mol. Biol.* 2004, 297, 209–228.
- [10] Brion, M., *Methods Mol. Biol.* 2004, 297, 229–242.
- [11] Gill, P., *Int. J. Legal Med.* 2001, 114, 204–210.
- [12] Seielstad, M. T., Minch, E., Cavalli-Sforza, L. L., *Nat. Genet.* 1998, 20, 278–280.
- [13] Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., *et al.*, *Am. J. Hum. Genet.* 1999, 64(3), 817–831.
- [14] Y-Chromosome Consortium, *Genome Res.* 2002, 12, 339–348.
- [15] Jobling, M. A., Tyler-Smith, C., *Nat. Rev. Genet.* 2003, 4, 598–612.
- [16] Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., *et al.*, *Am. J. Hum. Genet.* 2004, 75, 128–137.
- [17] Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., *et al.*, *Hum. Genet.* 2004, 114, 127–148.
- [18] Chen, X., Sullivan, P. F., *Pharmacogenomics J.* 2003, 3, 77–96.
- [19] Sobrino, B., Carracedo, A., *Methods Mol. Biol.* 2004, 297, 107–126.
- [20] Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., *et al.*, *Nat. Genet.* 2000, 26, 358–361.
- [21] Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., *et al.*, *Ann. Hum. Genet.* 2001, 65, 43–62.
- [22] Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., *et al.*, *Science* 2000, 290, 1155–1159.
- [23] Schneider, S., Roessli, D., Excoffier, L., *Genetics and Biometry Laboratory*, University of Geneva, Switzerland 2000.
- [24] Sanchez, J. J., Borsting, C., Hallenberg, C., Buchard, A., *et al.*, *Forensic Sci. Int.* 2003, 137, 74–84.
- [25] Sanchez, J. J., Brion, M., Parson, W., Blanco-Verea, A. J., *et al.*, *Forensic Sci. Int.* 2004, 140, 241–250.
- [26] Quintans, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., *et al.*, *Forensic Sci. Int.* 2004, 140, 251–257.
- [27] Mengel-Jørgensen, J., Sanchez, J. J., Borsting, C., Kirpekar, F., Morling, N., *Anal. Chem.* 2004, 76, 6039–6945.
- [28] Zerjal, T., Wells, R. S., Yuldashova, N., Ruzibakiev, R., Tyler-Smith, C., *Am. J. Hum. Genet.* 2002, 71, 466–482.
- [29] Luis, J. R., Rowold, D. J., Regueiro, M., Caeiro, B., *et al.*, *Am. J. Hum. Genet.* 2004, 74, 532–544.
- [30] Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., *et al.*, *Am. J. Hum. Genet.* 2004, 74, 1014–1022.
- [31] Sanchez, J. J., Hallenberg, C., Borsting, C., Hernandez, A., Morling, N., *Eur. J. Hum. Genet.* 2005, 13, 856–866.
- [32] Bosch, E., Calafell, F., Rosser, Z. H., Norby, S., *et al.*, *Hum. Genet.* 2003, 112, 353–363.
- [33] Phillips, C., Lareu, M., Salas, A., Fondevila, M., *et al.*, *Progr. in Forensic Genetics 10, Int. Congress Series* 2004, 1261, 233–235.
- [34] Quintans, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M. V., Carracedo, A., *Forensic Sci. Int.* 2004, 140, 251–357.
- [35] Karafet, T., Xu, L., Du, R., Wang, W., *et al.*, *Am. J. Hum. Genet.* 2001, 69, 615–628.

### **Bloque III. Estudios poblacionales basados en la variabilidad de los polimorfismos del cromosoma Y**

El poblamiento del Nuevo Mundo siempre ha atraído la atención de los investigadores de la genética poblacional, sobre todo desde dos puntos de vista: ¿quiénes fueron los primeros humanos en entrar en este continente y por dónde? y ¿cuál fue el número de grandes migraciones que se produjeron hacia América? Greenberg y cols. (1986) fueron los primeros en proponer un modelo de migración basándose en datos genéticos, además de datos lingüísticos y de morfología dental, este modelo proponía tres grandes movimientos migratorios a través del Estrecho de Bering. Actualmente estos datos están muy superados, sobre todo desde el punto de vista de la genética poblacional, donde los estudios realizados sobre los SNPs del cromosoma Y han encontrado un linaje predominante en nativos americanos definido por el Y-SNP “M3” (Underhill y cols., 1996) que designa al haplogrupo Q1a3a (Karafet y cols., 2008). Este haplogrupo se ha encontrado también en algunas poblaciones de Siberia, próximas al Estrecho de Bering, que podrían haber adquirido este haplogrupo por migración inversa. Otro haplogrupo característico de la población indígena americana, pero localizado casi exclusivamente en la región de Norteamérica, es el haplogrupo C. Frecuente también en muchas zonas del Este de Asia, aunque estudios con microsatélites sobre este haplogrupo C encontrado en las Américas parece indicar una procedencia más concretamente de la región del Lago Baikal (Lell y cols., 2002 y Karafet y cols., 1999).

Actualmente la distribución de haplogrupos en el continente americano es muy diversa debido, en gran parte, al proceso de mezcla que se ha originado a partir de la colonización europea y el tráfico de esclavos africanos llevado a cabo en determinadas zonas del continente.

Nosotros hemos realizado tres estudios poblacionales usando sobretodo Y-SNPs e Y-STRs para aumentar el número de datos disponibles sobre la distribución de las frecuencias de los haplogrupos y la diversidad haplotípica en el Nuevo Mundo.

En el primer trabajo se estudiaron STRs y SNPs del cromosoma Y en población general y en individuos de unas regiones muy concretas de El Salvador. En el segundo

artículo únicamente se usó una batería de Y-SNPs para conocer la distribución de haplogrupos en dos grupos poblacionales muy diferentes de Colombia. El tercer estudio se llevo a cabo en tres poblaciones indígenas de Argentina sobre la que aplicamos una amplia batería de Y-SNPs e Y-STRs.

En un último trabajo nos centramos en 2 grupos poblacionales europeos, en los que usando Y-SNPs e Y-STRs se intenta buscar la reciente contribución haplotípica del Noroeste africano en la península Ibérica y Sicilia basándonos en datos históricos.

## **Artículo 5: “El legado genético masculino de El Salvador”**

Lovo-Gómez J, **Blanco-Verea A**, Lareu MV, Brión M, Carracedo A.

Forensic Science International **171** (2007): 198-203

### Resumen:

En este artículo para llevar a cabo el estudio de las frecuencias alélicas y los análisis de haplogrupos y haplotipos hemos utilizado 16 marcadores binarios y 8 STRs del cromosoma Y (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393) que se estudiaron en 93 muestras de hombres no relacionados del área metropolitana de El Salvador y 67 individuos de género masculino de grupos étnicos diferentes: Conchagua, San Alejo, Panchimalco, Izalco y Nueva Concepción. Se evaluaron los niveles de mezcla entre áreas metropolitanas y rurales y se analizó la subestructuración poblacional.

Se encontraron un total de 13 haplogrupos y 136 haplotipos diferentes. El haplogrupo más frecuente en la población general metropolitana fue el R1b de origen posiblemente europeo, mientras que en la población indígena, si la consideramos como una unidad, el haplogrupo más frecuente fue el Q1a3a\* característico de la población amerindia.



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

Forensic Science International 171 (2007) 198–203

[www.elsevier.com/locate/forsciint](http://www.elsevier.com/locate/forsciint)

Rapid communication

## The genetic male legacy from El Salvador

J. Lovo-Gómez<sup>a,b</sup>, A. Blanco-Verea<sup>b</sup>, M.V. Lareu<sup>b</sup>, M. Brión<sup>b,\*</sup>, A. Carracedo<sup>b</sup>

<sup>a</sup> *Laboratorio de Genética Forense, Instituto de Medicina Legal Dr. Roberto Masferrer, Corte Suprema de Justicia, San Salvador, El Salvador*

<sup>b</sup> *Grupo de Medicina Xenómica, Institute of Legal Medicine, Faculty of Medicine, University of Santiago de Compostela, C/San Francisco s/n, 15782 Santiago de Compostela, Spain*

Received 8 May 2006; received in revised form 5 July 2006; accepted 6 July 2006

Available online 17 August 2006

### Abstract

Allele frequencies and haplotype and haplogroup analysis have been performed for 16 Y-chromosome binary markers and 8 Y-chromosome STRs (DYS19, DYS385I and II, DYS389I and II, DYS390, DYS391, DYS392, DYS393). Data was obtained from a general sample of 93 unrelated individuals living in metropolitan areas from El Salvador, and 67 individuals from different historical ethnic groups, Conchagua, San Alejo, Panchimalco, Izalco and finally Nueva Concepción with white people. Levels of admixture among metropolitan and rural areas were evaluated and population substructure measured.

A total of 13 haplogroups and 136 different haplotypes were found. The most frequent haplogroup in the general metropolitan population was the European R1b, while in the Indigenous samples considered as a whole the most frequent was the Amerindian haplogroup Q3.

© 2006 Elsevier Ireland Ltd. All rights reserved.

**Keywords:** Y-chromosome; STRs; SNPs; El Salvador; Metropolitan populations; Indigenous groups

### 1. Introduction

Haplotypes based on genetic markers of the non-recombining part of the Y-chromosome are becoming an important tool for the forensic investigation [1,2].

Because of the extremely high diversity of Y-chromosome haplotypes in most of the populations and because the product role cannot be used to estimate Y-STR haplotype frequencies, extensive databases are needed and therefore, they have been performed. The first free online database created was the Y-haplotype reference database (YHRD, <http://www.yhrd.org>), including the established standards of minimal and extended haplotypes.

In the case of Y-chromosome binary polymorphisms, based on studies of global populations there is extensive knowledge of its high geographic specificity. SNP haplogroups can be used to estimate admixture levels among populations and to evaluate population stratifications [2].

El Salvador is the smallest country of Centro America (21,000 km<sup>2</sup>), but it is the most densely populated of Latin America. It has approximately 6.2 millions inhabitants. Near

90% of the population is mixed, between Europeans and Indigenous, and only 1% are Indigenous people. Metropolitan area from San Salvador represents 1.7 million people, and approximately 42% of the population lives in rural areas [3].

In this study general population from metropolitan areas was compared with Indigenous samples scattered in four different rural areas, Conchagua, San Alejo, Izalco and Panchimalco. In addition some white individuals from the rural area of Nueva Concepción were also included.

Population from Conchagua lived in the islands of Fonseca's Gulf, which is shared nowadays by El Salvador, Honduras and Nicaragua. In 1522 it was discovered by a Spanish expedition. The Spaniards and the English pirate invasions caused the "Lencas" emigration towards firm land, the place in which, now, the natives are settled down in the eastern area of El Salvador.

San Alejo is a very successful land (in its best moment) placed in the eastern area of El Salvador, which was inhabited by "Lencas". It was colonized by Spaniards in the XVI century.

The Izalco population originating from Yaqui or "Pipil" is placed in the western area of El Salvador. It was one of the most important native settlements of the pre-Columbian period whose beginnings date back to the second half of XI century after Christ. It was conquered by Spaniards in the first half of the XVI century.

\* Corresponding author. Tel.: +34 981582327; fax: +34 981580336.

E-mail address: [brioniml@usc.es](mailto:brioniml@usc.es) (M. Brión).

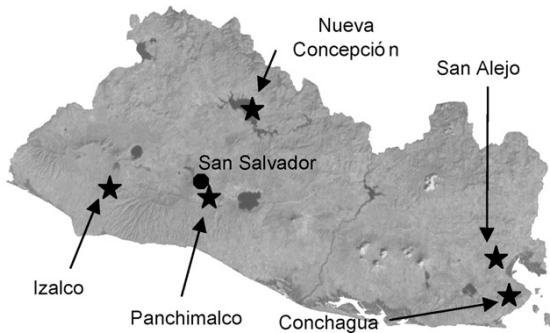


Fig. 1. Map of El Salvador. Stars represent the location of the populations analysed.

Panchimalco is also a population originated from pre-Columbian period. It is placed in the Central area of El Salvador, established by the “Pipil” emigrants as a result of the dispersion of the different tribes “Nahoas” during the XI and XII centuries after Christ. In the first decades of the XVI century it was also colonized by Spaniards.

In this study, we examine the male lineage structure of the population from El Salvador, by means of analysing the highly variable Y-STR haplotypes and the geographically structured Y-chromosome haplogroups. Levels of admixture among metropolitan and rural areas were evaluated and population substructure measured.

2. Materials and methods

2.1. DNA samples

A total of 187 males with different geographic origin from El Salvador have been analysed (Fig. 1). The selected samples include 93 individuals from a general population from El Salvador and 67 individuals from four historical

ethnic groups (Fig. 1): Conchagua ( $n = 23$ ), San Alejo ( $n = 12$ ), Panchimalco ( $n = 11$ ), Izalco ( $n = 12$ ) and finally Nueva Concepción ( $n = 9$ ) with non-Indigenous people living in rural areas. Although the number of samples for some of the population groups is quite reduced, we decided to send this data to publication anyway, because of the high difficulty in getting additional samples.

DNA was extracted from blood stains using a standard phenol–chloroform method. DNA quantification was performed using fluorescence detection with a DyNAQuant 200 (Amersham Biosciences, Uppsala, Sweden).

2.2. Y-chromosome SNP analysis

We typed 16 Y-chromosome binary markers (Fig. 2): SRY-1532, M213, M201, M170, M26, 12f2.1, M62, M172, M9, M70, M22, Tat (or M46), 92R7, M3, M173 and P25. Conditions described by Brión et al. [4] were followed to analyse nine of the SNPs in a multiplex 1, and six markers in a multiplex 2. The remaining polymorphism, M3, was analysed in a single PCR reaction using 200  $\mu$ M of each dNTP, 1.5 mM of  $MgCl_2$ , 0.5 U of taq DNA polymerase, and 0.25  $\mu$ M of each of the following primers F: CTGCCAGGGCTTCAATAG and R: AAGGGCATCTTTTCATTTAGGT. Cycling conditions were 30 cycles of 94°30’, 57°1’, 72°1’. Detection was performed by single base extension, with the SNaPshot multiplex kit (Applied Biosystems). The probe used for the detection reaction was GGGTCACCTCTGGGACTGA.

Haplogroups were named according to the proposals of the Y-Chromosome Consortium [5,2].

2.3. Y-chromosome STR analysis

The eight Y-STRs included in the European minimal haplotype described in the Y-STR haplotype reference database (<http://www.ystr.org>) were analysed in the whole sample. In addition the samples belonging to the Indigenous ethnic groups were also analysed for the DYS437, DYS438 and DYS439.

In the case of the general population from El Salvador PCR reactions and detection were performed as previously described in Lovo et al. [6]. STR analysis in the Indigenous samples was performed using the commercial kit Powerplex<sup>®</sup> Y System (Promega Corporation), following the recommendations of the manufacturer. Products were run on an ABI 3100 Genetic Analyser (Applied Biosystems) and analysed using the GeneScan 3.7 or the Genotyper 3.7 software (Applied Biosystems).

Allele designation for the STRs was performed according to Gill et al. [7] using allelic ladders available with the kit.

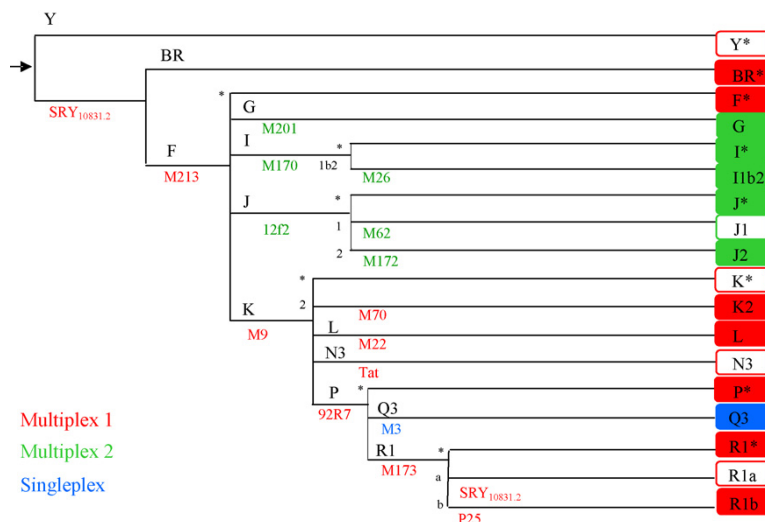


Fig. 2. Maximum-parsimony tree of 18 Y-chromosome haplogroups. Major clades are labelled with large capital letters above each branch. Mutations names are given below the branches. The length of each branch is not proportional to the age of the mutation. Different colours represent different amplification and detection reaction. Haplogroups found in our study appear with coloured squares. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

Table 1  
Haplogroup frequencies

	BR*(xF)	F*(xG,IK)	G	I*(xI1b2)	I1b2	J*(xJ1,2)	J2	K2	L	P*(xQ3,R1)	Q3	R1*(xR1b)	R1b
Metropolitan population	0.043		0.011	0.011	0.032	0.022	0.022	0.032		0.075	0.312	0.011	0.430
Rural population	0.075	0.030			0.015	0.015			0.030	0.104	0.493		0.239

Table 2  
Y-chromosome diversity values

	Metropolitan population	Rural population
SNPs		
<i>n</i>	93	67
<i>n</i> haplogroups	11	8
Hg/Ht diversity	0.7146 ± 0.0331	0.6920 ± 0.0454
STRs		
<i>n</i>	93	64
<i>n</i> haplogroups	83	54
Hg/Ht diversity	0.9953 ± 0.0031	0.9945 ± 0.0039

2.4. Statistical analysis

Haplogroup and haplotypes frequencies and diversities were calculated by use of the Arlequin 2.0 software [8]. The same program was also used to test the hypothesis of a random distribution of the individuals between pairs of populations, through an exact test of population differentiation.

3. Results and discussion

Analysis of the genetic male legacy from El Salvador was performed at two different levels, on one hand through the analysis of Y-chromosome SNPs, which show ancient relationships, an on the other hand through the analysis of the highly variable Y-chromosome STRs.

Among the 67 samples belonging to the different ethnic rural groups, 3 of them failed to amplify microsatellites but non-SNPs. Considering both, SNPs and STRs, a total of 61 different

compound haplotypes were observed, which can be subdivided in 55 different haplotypes based in STRs and 8 different haplogroups based in SNPs.

Although the number of samples from each population group is not enough, in order to increase the statistical significance samples were grouped in two main groups, metropolitan population and rural population.

The haplogroup tree defined with the 16 SNPs analysed are represented in Fig. 2. Among the 18 possible haplogroups only 13 appear in our samples, 11 in the metropolitan population and 8 in the rural populations.

Table 1 shows the haplogroup frequencies found in each of the main population groups and as it would be expected the most frequent haplogroup in the metropolitan population was the R1b, representing an important European component in this mixed population. By contrast, the most frequent haplogroup in the rural samples was the Amerindian Q3, as in most of the native South American populations analysed until now [9].

Table 2 shows the haplogroup and haplotype diversities calculated by means of the SNP and STR data, respectively. Diversities values were similar in both groups of populations, however the values obtained in the metropolitan areas were always higher, as it would be expected from an ethnically mixed population. Haplotype diversities were calculated considering the minimal haplotype in both populations.

STR data for the metropolitan population has already been published in Lovo et al. [6], here are presented both, the data from the rural populations in Table 3 and the data from the

Table 3  
Y-chromosome haplotypes in a rural population from El Salvador

H	<i>n</i>	HG	DYS19	DYS385	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	Population <sup>a</sup>
H1	1	Q3	12	17,17	13	30	24	10	15	13	14	9	11	SA
H2	1	Q3	12,13	14,18	13	29	24	10	15	14	14	11	12	Con
H3	1	Q3	13	10,19	14	34	24	10	12	15	14	11	11	Con
H4	2	Q3	13	13,17	13	30	25	10	16	13	14	11	12	NC
H5	1	Q3	13	13,18	13	29	24	11	13	15	14	11	13	Iz
H6	1	Q3	13	14,15	13	33	25	11	14	12	14	11	11	SA
H7	1	Q3	13	14,17	13	29	23	10	14	13	15	11	12	Con
H8	1	Q3	13	14,17	13	29	24	10	15	14	15	11	12	Con
H9	1	P* (xQ3,R1)	13	14,17	14	31	24	10	14	13	14	12	12	Con
H10	1	P* (xQ3,R1)	13	14,17	14	31	24	10	14	12	14	11	12	SA
H11	1	Q3	13	14,17	14	31	24	10	14	13	14	11	11	Iz
H12	2	Q3	13	14,18	13	29	24	10	15	14	14	11	12	Con
H13	1	Q3	13	14,18	13	30	24	10	13	14	14	11	12	Con



Table 3 (Continued)

H	n	HG	DYS19	DYS385	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	Population <sup>a</sup>
H14	1	Q3	13	15,15	13	32	25	9	15	12	14	11	11	Pan
H15	1	Q3	13	15,15	14	32	24	10	14	12	14	11	11	SA
H16	1	P*(xQ3,R1)	13	15,16	13	29	25	10	14	13	14	9	11	Pan
H17	1	Q3	13	15,16	13	32	25	9	15	12	14	11	11	Pan
H18	1	Q3	13	15,16	13	32	22	10	15	12	14	11	11	Iz
H19	1	Q3	13	15,16	14	32	24	10	15	13	13	11	12	Con
H20	1	Q3	13	15,17	12	28	24	9	12	13	14	11	12	Con
H21	2	Q3	13	15,17	12	28	24	9	14	13	14	11	12	Con
H22	1	Q3	13	15,17	13	29	24	10	14	14	15	11	12	Pan
H23	1	R1b	13	15,17	13	29	24	10	14	14	15	11	12	Iz
H24	1	Q3	13	15,19	12	28	24	9	14	13	14	11	12	Pan
H25	1	P*	13	16,16	13	29	25	10	14	13	14	9	11	Pan
		(xQ3,R1)												
H26	1	BR*(xF)	13	16,17	13	30	24	10	11	13	14	10	12	SA
H27	1	R1b	14	10,15	13	29	24	11	13	13	15	12	12	SA
H28	1	R1b	14	11,14	12	28	23	12	13	13	15	12	12	Iz
H29	1	R1b	14	11,14	13	30	23	12	14	13	15	12	12	Con
H30	1	R1b	14	11,14	13	29	24	11	13	13	15	12	12	Con
H31	1	R1b	14	11,14	13	29	24	11	13	13	15	11	12	NC
H32	1	R1b	14	11,14	13	29	24	11	13	13	14	12	11	NC
H33	1	R1b	14	11,14	13	28	24	11	8	13	14	12	11	SA
H34	1	R1b	14	11,14	14	30	24	11	13	13	15	13	12	Con
H35	1	R1b	14	11,14	14	29	23	11	13	13	14	12	11	NC
H36	1	R1b	14	11,14	14	30	23	11	13	13	15	12	12	NC
H37	1	R1b	14	11,14	14	30	24	11	13	13	15	12	12	Pan
H38	1	R1b	14	11,14	14	31	24	10	13	13	14	12	12	Iz
H39	1	R1b	14	11,14	14	30	25	11	13	13	15	12	12	Iz
H40	1	R1b	14	12,14	12	28	25	10	13	13	15	12	12	Iz
H41	1	BR*(xF)	14	13,14	14	30	24	9	13	13	15	10	10	Iz
H42	1	J*(xJ1,2)	14	13,15	13	31	23	10	11	12	14	10	14	Con
H43	1	Q3	14	13,16	13	30	23	10	13	13	14	12	12	Con
H44	2	Q3	14	13,16	13	30	24	10	13	13	14	12	12	Con
H45	1	Q3	14	13,17	13	30	24	10	14	13	14	11	14	Pan
H46	1	Q3	14	14,17	13	30	24	11	16	12	14	10	12	SA
H47	1	Q3	14	14,18	13	29	24	10	15	13	14	11	11	Pan
H48	1	P*	14	15,16	13	29	24	10	14	13	14	9	11	Pan
		(xQ3,R1)												
H49	1	Q3	14	13,2,18	13	31	24	11	14	13	14	11	11	NC
H50	1	L	15	9,17	13	29	22	10	14	12	15	10	12	NC
H51	1	L	15	9,17	13	30	21	10	14	11	16	10	13	SA
H52	1	F*(xG,IK)	15	12,12	12	27	23	10	11	13	15	10	12	Iz
H53	1	BR*(xF)	15	14,16	12	28	22	10	11	13	17	8	12	Con
H54	1	P*(xQ3,R1)	15	14,16	13	30	23	11	15	13	14	11	11	Pan
H55	1	F*	15	15,17	14	32	21	10	11	13	14	11	13	Con
		(xG,IK)												
H56	1	BR*(xF)	15	15,17	14	32	21	10	11	13	14	11	13	Con
H57	1	BR*(xF)	15	16,18	13	31	21	11	11	13	14	11	11	Iz
H58	1	R1b	16	11,14	13	29	24	10	13	13	15	12	12	Con
H59	1	I1b2	17	12,12	13	28	23	10	11	13	15	10	12	NC
H60	1	Q3	–	16,20	12	25	24	10	–	12	14	–	13	SA
		(xQ3,R1)												
	1	P*(xQ3,R1)	–	–	–	–	–	–	–	–	–	–	–	SA
	1	Q3	–	–	–	–	–	–	–	–	–	–	–	SA
	1	Q3	–	–	–	–	–	–	–	–	–	–	–	Iz

<sup>a</sup> SA: San Alejo; Con: Conchagua; NC: Nueva Concepción; Iz: Izalco; Pan: Panchimalco.

metropolitan population in Table 4, including SNP and STR data. Only one haplotype was shared between metropolitan and rural samples, belonging in both populations to the most frequent haplogroup in the metropolitan areas, R1b.

Looking at the haplotypes it must be noted that DYS19 appeared duplicated in one sample from the Indigenous population of Conchagua. Duplications at this microsatellite

are relatively frequent, since in the reference database 74 duplications have been reported until now. Another special feature, was the presence of a \*.2 variant allele at the DYS385 in the population of Nueva Concepcion. This kind of variant alleles are also reported in the YHRD, where 37 samples appeared with \*.2 intermediate alleles, and 12 of them showing 13.2 alleles.

Table 4  
Y-chromosome haplotypes in a metropolitan population from El Salvador

H	n	HG	DYS19	DYS385	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393
H1	1	Q3	13	13,16	13	29	27	10	13	12
H2	1	Q3	13	13,16	13	30	25	10	13	12
H3	1	Q3	13	13,17	12	29	24	10	14	13
H4	1	Q3	13	13,17	13	29	24	10	12	14
H5	1	Q3	13	13,17	13	30	22	10	12	14
H6	1	P*(xQ3,R1)	13	13,18	14	30	23	10	12	14
H7	1	Q3	13	14,16	13	31	24	10	14	12
H8	1	Q3	13	14,16	14	31	23	10	13	13
H9	1	Q3	13	14,16	14	31	24	10	11	13
H10	1	Q3	13	14,16	14	31	24	10	13	13
H11	1	Q3	13	14,17	13	30	23	9	13	13
H12	1	R1b	13	14,17	13	30	25	10	13	13
H13	1	Q3	13	14,17	14	32	23	9	14	13
H14	1	Q3	13	14,17	14	33	24	12	12	12
H15	1	Q3	13	14,18	13	31	24	10	13	13
H16	1	Q3	13	15,15	13	31	25	10	13	12
H17	1	Q3	13	15,15	14	31	23	10	12	13
H18	1	P*(xQ3,R1)	13	15,16	13	29	23	11	14	13
H19	1	P*(xQ3,R1)	13	15,16	13	30	24	10	13	13
H20	1	P*(xQ3,R1)	13	15,16	13	31	25	10	13	13
H21	1	Q3	13	15,17	13	29	23	10	13	15
H22	1	Q3	13	15,17	13	29	24	10	13	14
H23	1	Q3	13	15,17	13	31	24	10	13	13
H24	1	Q3	13	15,18	14	31	24	10	14	13
H25	1	Q3	13	15,19	13	28	24	10	12	14
H26	1	Q3	13	15,20	13	29	24	10	14	12
H27	1	Q3	13	16,16	13	28	23	10	13	13
H28	1	Q3	13	16,16	14	30	23	10	13	13
H29	1	Q3	13	16,17	12	28	24	10	13	13
H30	1	Q3	13	16,17	12	28	24	11	13	13
H31	1	Q3	13	16,17	13	30	22	10	13	13
H32	1	I1b2	13	17,18	12	29	24	10	11	12
H33	1	R1b	14	11,11	13	29	24	10	13	13
H34	1	R1b	14	11,14	12	28	24	10	13	13
H35	1	R1b	14	11,14	12	29	25	11	12	13
H36	1	R1b	14	11,14	13	28	24	10	11	13
H37	1	R1b	14	11,14	13	29	23	10	12	13
H38	1	R1b	14	11,14	13	29	24	8	13	13
H39	3	R1b	14	11,14	13	29	24	9	12	13
H40	1	R1b	14	11,14	13	29	24	9	13	13
H41	5	R1b	14	11,14	13	29	24	10	12	13
H42	4	R1b	14	11,14	13	29	24	11	12	13
H43	1	R1b	14	11,14	13	29	24	11	13	13
H44	1	R1b	14	11,14	13	29	25	10	13	13
H45	1	R1b	14	11,14	14	29	23	10	12	12
H46	1	R1b	14	11,14	14	30	24	10	12	13
H47	1	P*(xQ3,R1)	14	11,14	14	30	24	10	13	13
H48	1	R1b	14	11,14	14	30	24	11	12	13
H49	1	R1b	14	11,14	14	31	24	10	12	13
H50	1	R1b	14	11,15	13	29	24	9	12	13
H51	1	R1b	14	11,15	13	29	24	11	12	13
H52	1	R1b	14	11,15	13	29	25	9	13	13
H53	1	R1b	14	11,15	14	30	25	9	12	13
H54	1	R1b	14	11,15	14	31	24	11	12	13
H55	1	R1b	14	11,16	13	29	24	11	12	13
H56	2	R1b	14	12,14	13	29	24	10	12	13
H57	1	J*(xJ1,2)	14	13,17	13	32	23	9	11	12
H58	1	P*(xQ3,R1)	14	14,15	13	29	24	9	14	13
H59	1	Q3	14	14,16	13	30	23	10	13	13
H60	1	Q3	14	14,18	13	29	24	10	13	13
H61	1	I*(xI1b2)	14	15,16	13	30	23	10	12	14
H62	1	Q3	14	16,16	14	30	25	10	13	13
H63	1	R1b	15	11,14	12	28	24	10	12	13
H64	1	R1b	15	11,14	13	29	23	11	12	13

Table 4 (Continued)

H	n	HG	DYS19	DYS385	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393
H65	1	R1b	15	11,14	14	30	24	9	12	13
H66	1	R1b	15	11,15	14	30	24	11	12	13
H67	1	R1b	15	11,15	15	31	24	9	12	13
H68	1	R1b	15	11,16	14	30	24	11	12	13
H69	1	J2	15	12,12	14	29	23	10	12	13
H70	1	G	15	12,16	12	28	22	10	11	14
H71	1	R1b	15	13,14	12	28	24	9	14	13
H72	1	K2	15	13,16	12	27	24	10	13	13
H73	1	K2	15	14,14	14	29	22	10	12	13
H74	1	BR*(xF)	15	14,15	13	30	22	11	11	13
H75	1	J2	15	14,18	12	29	24	10	11	12
H76	1	K2	15	16,16	13	30	23	10	12	13
H77	1	BR*(xF)	15	16,18	13	31	24	10	11	13
H78	1	I1b2	16	12,12	13	28	22	9	11	13
H79	1	BR*(xF)	16	15,15	12	29	22	11	10	13
H80	1	J*(xJ1,2)	16	16,18	14	31	21	10	11	13
H81	1	I1b2	17	12,12	13	27	23	10	11	13
H82	1	BR*(xF)	17	15,18	13	29	21	10	11	14
H83	1	P*(xQ3,R1)	17	16,16	15	33	21	10	11	14

Finally, a population differentiation test was performed. Both, SNPs and STR data show significant statistical differences when the rural samples are considered as a group to be compared with metropolitan samples ( $p = 0.0291 \pm 0.0066$  and  $0.0000 \pm 0.00$  for SNPs and STRs, respectively).

Although the population sample available in this work was quite reduced, the data presented here can help us to a better understand of the genetic composition of the male lineage in the population of El Salvador. Despite of been a small country densely populated, the country seems to continue having small Indigenous groups of populations more or less isolated, at least in terms of male genetic markers.

#### Acknowledgement

We want to thank Vilma de Aguilar for helping to get the samples.

#### References

- [1] M.A. Jobling, Y-chromosomal SNP haplotype diversity in forensic analysis, *Forensic Sci. Int.* 118 (2001) 158–162.
- [2] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, *Nat. Rev. Genet.* 4 (2003) 598–612.
- [3] El Salvador, Historia de sus pueblos, villas y ciudades. Biblioteca de Historia Salvadoreña, vol. 4, Consejo Nacional para la Cultura y el Arte (CONCULTURA), San Salvador, 2000.
- [4] M. Brión, B. Sobrino, A. Blanco-Verea, M.V. Lareu, A. Carracedo, Hierarchical analysis of 30 Y-chromosome SNPs in European populations, *Int. J. Legal Med.* 119 (2005) 10–15.
- [5] The Y Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups, *Genome Res.* 12 (2002) 339–348.
- [6] J.S. Lovo, M. Fondevila, A. Salas, M. Brión, M.V. Lareu, A. Carracedo, Y-chromosome STR-haplotype typing in El Salvador, *Forensic Sci. Int.* 142 (2004) 45–49.
- [7] P. Gill, C. Brenner, B. Brinkmann, B. Budowle, A. Carracedo, M.A. Jobling, P. de Knijff, M. Kayser, et al., DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, *Forensic Sci. Int.* 124 (2001) 5–10.
- [8] S. Schneider, D. Roessli, L. Excoffier, Arlequin ver. 2.000: A Software for Population Genetics Data Analysis, Genetics and Biometry Laboratory, University of Geneva, Switzerland, 2000.
- [9] S.L. Zegura, T.M. Karafet, L.A. Zhivotovsky, M.F. Hammer, High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosome into the Americas, *Mol. Boil. Evol.* 21 (2004) 164–175.

**Artículo 6: “El componente genético masculino de dos poblaciones del Sur-Oeste colombiano”**

Acosta MA, **Blanco-Verea A**, Lareu MV, Brión M, Carracedo A.

Forensic Science Internacional: Genetics (2008)

(Article in press)

Resumen:

En la historia reciente de Colombia, encontramos dos factores que han contribuido a cambiar la estructura poblacional: la conquista española y el tráfico de esclavos promovido principalmente por Portugal, Inglaterra y España. Como consecuencia la población nativa de Colombia se ha visto reducida y mezclada con los grupos europeos y africanos que han ido llegando.

Para evaluar la ancestralidad masculina de la población del Cauca, hemos examinado la frecuencia de los principales linajes del cromosoma Y genotipando 30 Y-SNPs en dos poblaciones de esta región: 105 individuos Afro-colombianos y 110 individuos Caucásico-mestizos. Entre los 33 posibles haplogrupos definidos con los SNPs analizados, se detectaron un total de 15 haplogrupos, siendo 10 de ellos compartidos por ambas poblaciones. Se usaron muestras de otras poblaciones del Sur de América con el fin de observar el nivel de mezcla y para comparar los antecedentes genéticos con otras poblaciones vecinas.





Contents lists available at ScienceDirect

## Forensic Science International: Genetics

journal homepage: [www.elsevier.com/locate/fsig](http://www.elsevier.com/locate/fsig)

## Announcement of population data

## The genetic male component of two South-Western Colombian populations

María Amparo Acosta<sup>a,b</sup>, Alejandro Blanco-Verea<sup>c</sup>, María Victoria Lareu<sup>b</sup>,  
María Brion<sup>c,\*</sup>, Angel Carracedo<sup>b</sup><sup>a</sup> Pediatrics Department, College of Medicine, University of Cauca, Popayán, Colombia<sup>b</sup> Institute of Legal Medicine, Grupo de Medicina Xenómica-CIBERER, University of Santiago de Compostela, Spain<sup>c</sup> Complejo Hospitalario Universitario de Santiago (CHUS), Grupo de Medicina Xenómica-USC, Spain

## ARTICLE INFO

## Article history:

Received 8 May 2008

Received in revised form 9 June 2008

Accepted 13 June 2008

Available online xxx

## Keywords:

SNPs

Y-chromosome

Population

Afro-Colombian

Caucasian-Mestizo

## ABSTRACT

In the recent history of Colombia, two factors have contributed to change the population structure, the Spanish conquest and the slave trading promoted principally by Portugal, England and Spain. As a consequence the native population of Colombia has been reduced and mixed with the European and African arriving groups.

To assess the male ancestry of the Cauca population, we have examined the frequency of the major Y-chromosome lineages by typing 30 Y-SNPs in two populations from this region: 105 Afro-Colombian individuals and 110 Caucasian-Mestizo individuals. Among the 33 haplogroups defined with the SNPs analysed, 15 haplogroups were detected, 10 of them being shared by both populations. In order to investigate how the level of admixture is, and to compare the genetic background with other neighbour populations, other South American samples previously published were used for comparative analysis.

© 2008 Elsevier Ireland Ltd. All rights reserved.

## 1. Population

The admixture level in Colombia is not homogenous all around the country, and we can use the present genetic lineages to determine the degree of influence of arriving people coming from Europe and Africa, throughout history. We have collected 215 male individuals from the Cauca region located in the south of the country, under informed consent. According to their self-reported ethnicity, individuals were classified as "Caucasian-Mestizos" (110 individuals) and "Afro-Colombians" (195 individuals).

## 2. Extraction

DNA was extracted from blood stains by phenol/chloroform/isoamyl alcohol precipitation.

## 3. PCR

The following biallelic markers were analysed in a single multiplex PCR and a single multiplex base extension reaction as reported by Brion et al. [1]: SRY<sub>10831</sub>, M168, M216, M38, M217, M174, M96, M33, M75, P2, M2, M35, M81, M123, M201, M170,

M304, M9, M128, Tat, M175, M119, P31, M122, M45, M207, M173, M17, P25. The Y-SNP M3 was also analysed using a singleplex PCR and a singleplex base extension reaction, using the SNaPshot multiplex kit (Applied Biosystems) and following manufactures recommendations.

## 4. Typing

Detection of the SBE products was performed on a ABI 3130 genetic analyzer (Applied Biosystems), and POP-7 polymer. Analysis was done using the GeneScan analysis 3.7 and the Genotyper 3.7 software. Haplogroups were assigned and named according to the proposal of the Y Chromosome Consortium [2,3].

## 5. Analysis of data

Haplogroup frequencies were estimated in both populations by gene counting. Arlequin 2.0 software [4] was used to calculate genetic distances, as pairwise values of *F*<sub>ST</sub>, to perform molecular analysis of the variance (AMOVA) and to test the hypothesis of a random distribution of the individuals between pairs of populations, through an exact test of population differentiation. Genetic distances were multidimensional scaled by use of the SPSS version 14.0 software package. We compared the HG frequencies observed in South-West Colombia with published frequencies in other South American populations: two Brazilian Indian tribes (Guarani and

\* Corresponding author. Tel.: +34 98158232; fax: +34 981580336.  
E-mail address: [maria.brion@usc.es](mailto:maria.brion@usc.es) (M. Brion).

Kaingang) living in five different localities [5], one rural populations and one metropolitan population from El Salvador [6], and one metropolitan population from Río de Janeiro [7]. A principal component plot was drawn with the haplogroup frequencies.

## 6. Quality control

Performance of Y-chromosome binary polymorphism analyses was evaluated through collaborative exercise by the European DNA profiling group [1]. Particularly, the multiplex used here was widely validated in a collaborative study of the European Consortium "High-throughput analysis of single nucleotide polymorphisms for the forensic identification of persons—SNPforID" [1].

## 7. Results

### 7.1. Intra-population analysis

Among the 33 haplogroups defined with the 30 SNPs analysed, 15 haplogroups appeared in our Colombian populations, being shared by both populations 10 of them. In [supplementary Fig. 1](#) are represented the frequencies of the haplogroups detected in both populations analysed.

In the Afro-Colombian population 13 different male lineages were shown, being the most frequent hg E3a, which is present in 45% of the individuals. The European hg R1b also reach a high frequency, while none of the remainder lineages reach more than 10% of the individuals. Although all the individuals were self considered as Afro-Colombians descendants, only 1 individual showed the typical Amerindian hg Q3.

In the case of the Caucasian-Mestizo population, 12 different haplogroups were detected and the most frequent was the European hg R1b (49%). Among the others the most frequent lineage was the haplogroup defined by the M304 mutation, Hg J (10.91%), and the second most represented was the Amerindian hg Q3 (8.18%).

### 7.2. Inter-population analysis

The haplogroup frequencies principal component plot show the Caucasian-Mestizo population clustered with both metropolitan samples (Rio de Janeiro and El Salvador), while Afro-Colombian sample appear quite distant ([Supplementary Fig. 2a](#)).

An exact test of sample differentiation indicate that both Colombian populations are statistical significant differentiated from the other South-American population collected from the literature ( $p < 0.001$ ; Markov chain length: 10,000 steps), with the exception of the Mestizo population and the Metropolitan population from Rio de Janeiro ( $p = 0.12715 \pm 0.0077$ ; Markov chain length: 10,000 steps).

AMOVA analysis was also performed, showing a high percentage of variation among populations, 23.42% ( $p < 0.001$ ); however, most of the variance was attributable to differences within populations, 76.78% ( $p < 0.001$ ).

The multidimensional scaling plot performed with the Fst genetic distances ([Supplementary Fig. 2b](#)), again showed the Afro-Colombian population as an outlier, and the metropolitan populations close to the Mestizo population of Colombia.

## 8. Other remarks

Our results indicate a marked difference in both populations analysed; in the case of the Caucasian–Mestizo population a high percentage of European contribution was detected; similar to

the published metropolitan populations from El Salvador and Brazil [5,6]. However, the great majority of the Afro-Colombian individuals showed African influence, represented by lineage E3a.

The Amerindian component, represented by haplogroup Q3, is only represented by 1 individual among the Afro-Colombians, and 9 individuals among the Mestizos. While in the other South American populations considered here, its frequency is rather high, with the exception of the Rio de Janeiro sample.

The lineage C, described by the RPS4Y<sub>711</sub> or the M216 mutation has so far been reported only in North American populations and in two Wayuu individuals from the extreme north of South America [8,9]. In our study the lineage was not observed, supporting the idea of a geographic distribution restricted to North America.

In conclusion, as reported by Carvajal-Carmona et al. [10] for the population of Antioquia, the male ancestry of our Mestizo population from Cauca is predominantly European. The Afro-Colombian individuals showed a 45% South African male ancestry, intermediate figures between those reported by Goncalves et al. [11] for two black populations from big cities of Brazil.

The use of genetic polymorphisms to examine the history of the American people is complicated by the possibility of recent admixture. Moreover, this effect could be higher in the Y-chromosome markers, since there is evidence that admixture in the Americas has preferentially involved immigrant men and native women [12,13]. It will now be of great interest to see whether this fact is confirmed by the analysis of mitochondrial DNA.

## Acknowledgements

We would like to thank Patía-Valley Population in Cauca Department and Army Forces of Colombia, Brigada 29, Batallón de infantería Número 7 "José Hilario López" in Popayán, for their collaboration providing samples. Generous support provided by J Caicedo and Dra. MC Quijano is recognised and appreciated. Thanks are given to M Rodríguez in the Institute of Legal Medicine of Santiago de Compostela for technical support.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.fsigen.2008.06.006](https://doi.org/10.1016/j.fsigen.2008.06.006).

## References

- [1] M. Brión, J.J. Sanchez, K. Valgo, C. Thacker, A. Blanco-Verea, C. Børsting, B. Stradmann-Bellinghausen, M. Bogus, D. Syndercombe-Court, P.M. Schneider, A. Cariacado, N. Morling, Introduction of an single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages, *Electrophoresis* 67 (2005) 4411–4420.
- [2] Y. Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups, *Genome Res.* 12 (2) (2002) 339–348.
- [3] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, *Nat. Rev. Genet.* 4 (8) (2003) 598–612.
- [4] S. Schneider, D. Roessli, L. Excoffier, Arlequin ver 2.000: a software for population genetics data analysis, in: *Genetics and Biometry Laboratory*, University of Geneva, Switzerland, 2000.
- [5] A.R. Marrero, W.A. Silva-Junior, C.M. Bravi, M.H. Hutz, M.L. Petzl-Erler, A. Ruiz-Linares, F.M. Salzano, M.C. Bortolini, Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil, *Am. J. Phys. Anthropol.* 132 (2) (2007) 301–310.
- [6] J. Lovo-Gomez, A. Blanco-Verea, M.V. Lareu, M. Brión, A. Carracedo, The genetic male legacy from El Salvador, *Forensic Sci. Int.* 171 (2007) 198–203.
- [7] D.A. Silva, E. Carvalho, G. Costa, L. Tavares, A. Amorim, L. Gusmao, Y-chromosome genetic variation in Rio de Janeiro population, *Am. J. Hum. Biol.* 18 (6) (2006) 829–837.
- [8] T.M. Karafet, 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, M.F.

- Hammer. Ancestral Asian source(s) of New World Y-chromosomal founder haplotypes. *Am. J. Hum. Genet.* 64 (1999) 817–831.
- [9] J.T. Lell, R.I. Sukernik, Y.B. Starikovskaya, B. Su, L. Jin, T.G. Schurr, P.A. Underhill, D.C. Wallace, The dual origin and Siberian affinities of Native American Y chromosomes. *Am. J. Hum. Genet.* 70 (2002) 192–206.
- [10] L.G. Carvajal-Carmona, R. Ophoff, S. Service, J. Hartiala, J. Molina, P. Leon, J. Ospina, G. Bedoya, N. Freimer, A. Ruiz-Linares. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum. Genet.* 112 (2003) 534–541.
- [11] V.F. Goncalves, F. Prosdocimi, L.S. Santos, J.M. Ortega, S.D. Pena, Sex-biased gene flow in African Americans but not in American Caucasians. *Genet. Mol. Res.* 6 (2) (2007) 156–161.
- [12] L.G. Carvajal-Carmona, I.D. Soto, N. Pineda, D. Ortiz-Barrientos, C. Duque, J. Ospina Duque, M. McCarthy, P. Montoya, V.M. Álvarez, G. Bedoya, A. Ruiz-Linares, Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am. J. Hum. Genet.* 67 (2000) 1287–1295.
- [13] N.R. Mesa, M.C. Mondragon, I.D. Soto, M.V. Parra, C. Duque, D. Ortiz-Barrientos, L.F. Garcia, I.D. Velez, M.L. Bravo, J.G. Munera, G. Bedoya, M.C. Bortolini, A. Ruiz-Linares, Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. *Am. J. Hum. Genet.* 67 (2000) 1277–1286.



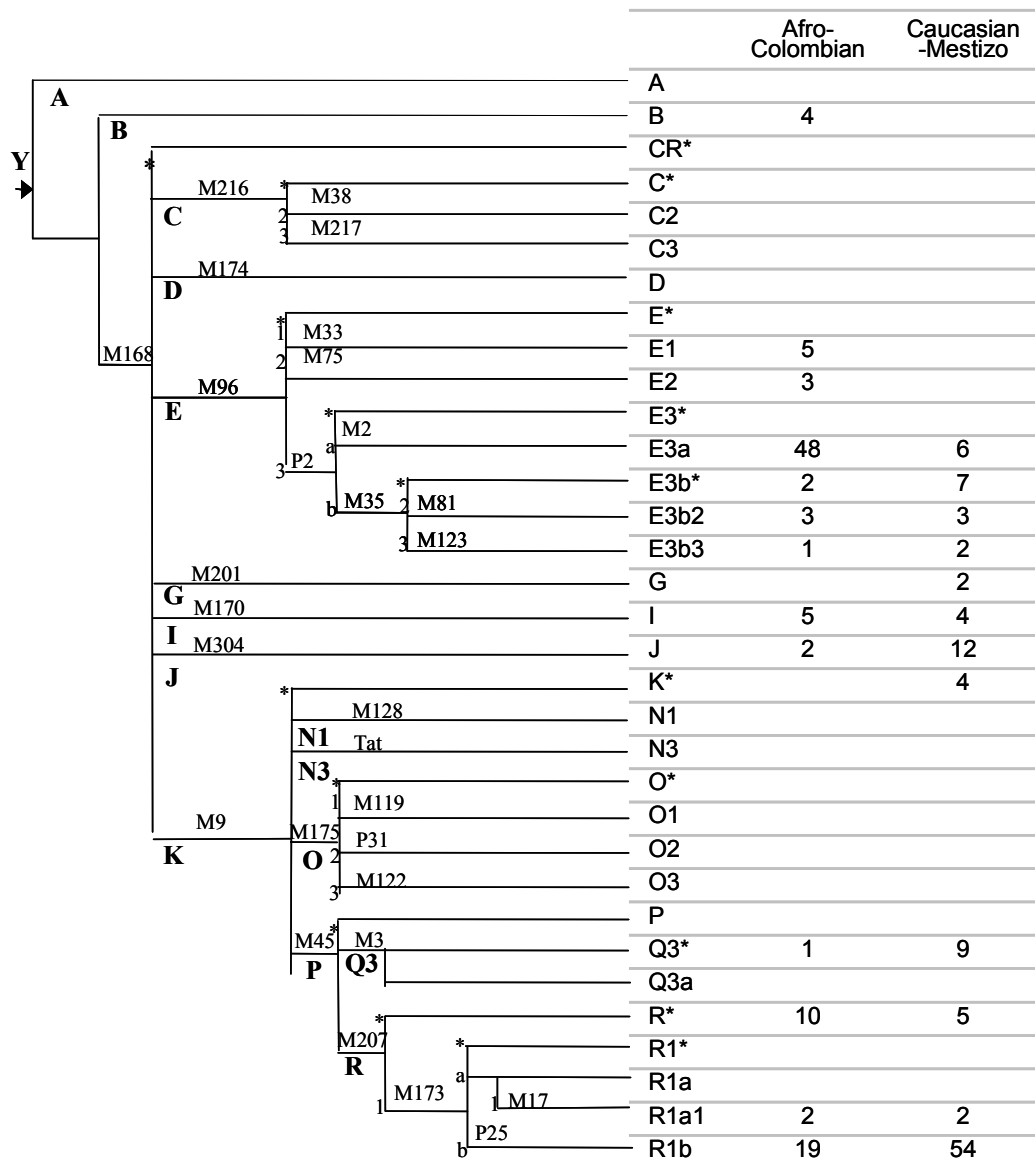


Figure 1. Y-chromosome phylogenetic tree and haplogroup frequencies. Polymorphism names are indicated above the lines and lineage names are shown below the lines. The length of the line has not significance. (Appendix A: Supplementary data).

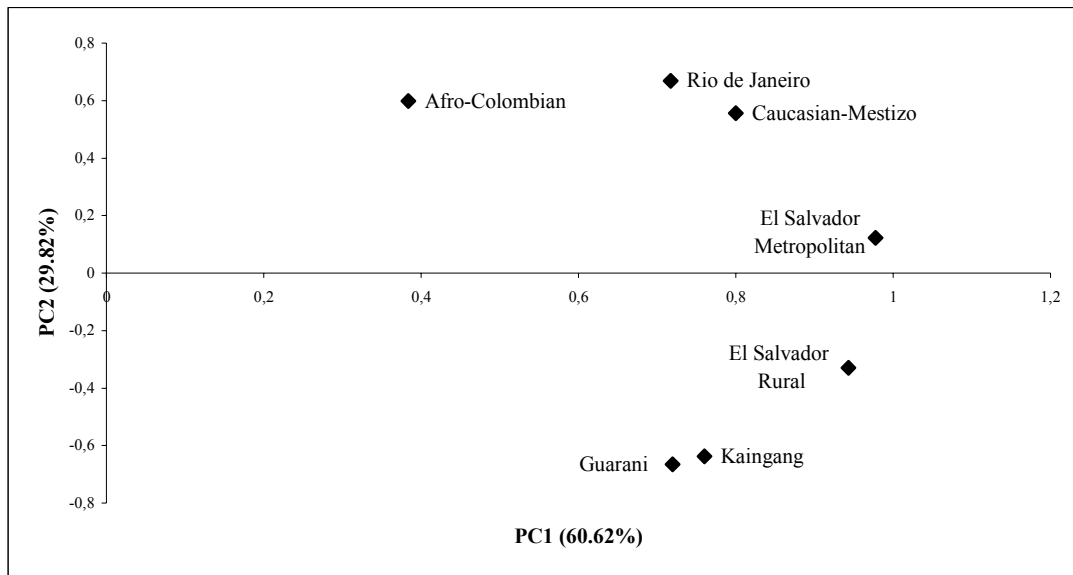


Figure 2a. Principal component analysis of Y-chromosomal haplogroup frequencies. The percentage of variance explained by each component is given on the axes. (Appendix A: Supplementary data).

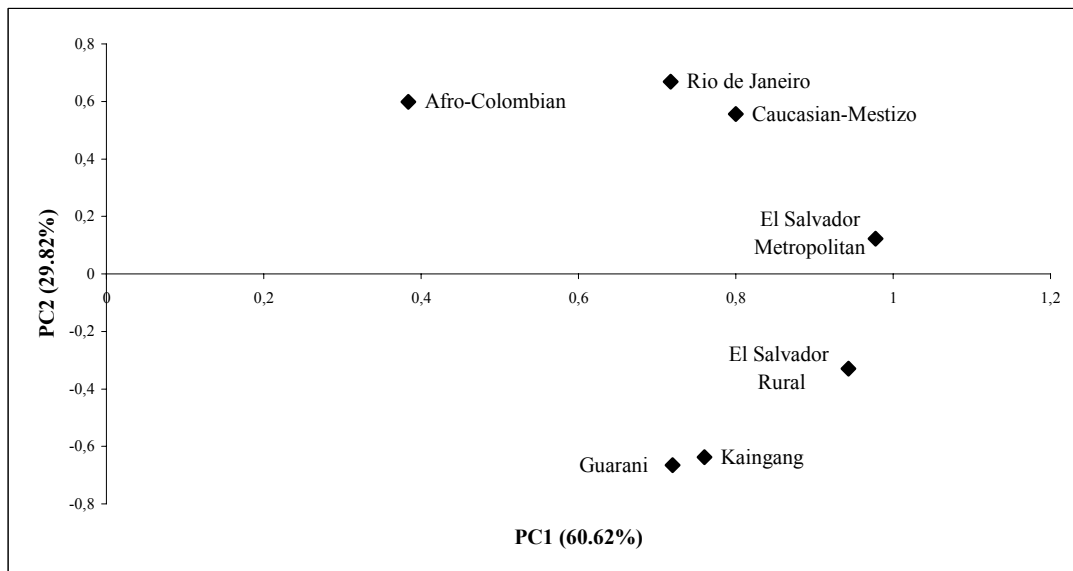


Figure 2b. Multidimensional scaling plot of the Fst distances. Stress value 0.004. (Appendix A: Supplementary data).



**Artículo 7: “Linajes del cromosoma Y en población nativa del Sur de América”**

**Blanco-Verea A**, Jaime JC, Brión M Carracedo A.

(Manuscrito en preparación)

Resumen:

Este trabajo intenta investigar la estructuración poblacional y la variación en poblaciones indígenas amerindias de Argentina. Se recogieron un total de 134 individuos pertenecientes a 3 grupos étnicos (kolla, mapuche, diaguita) localizados en 4 regiones diferentes y se analizaron para 26 Y-SNPs y 11 Y-STRs.

Se analizó la variabilidad interpoblacional, buscando subestructuración poblacional y considerando poblaciones vecinas para análisis genéticos comparativos, además de estimar la contribución del pool genético amerindio y europeo en la población actual.



## **Y chromosome lineages in native South American population**

Blanco-Verea A.<sup>1</sup>, Jaime J.C.<sup>2</sup>, Brión M.<sup>1</sup>, Carracedo A.<sup>2</sup>

<sup>1</sup>Complejo Hospitalario Universitario de Santiago (CHUS), Grupo de Medicina Xenómica-USC

<sup>2</sup>Grupo de Medicina Xenómica-USC, Institute of Legal Medicine. University of Santiago de Compostela. Spain.

### **Abstract.**

The present work tries to investigate the population structure and variation of the Amerindian indigenous populations living in Argentina. A total of 134 individuals from 3 ethnic groups (Kolla, Mapuche and Diaguitas) living in four different regions were collected and analysed for 26 Y-SNPs and 11 Y-STRs.

Intra-population variability was analysed, looking for population substructure and neighbour populations were considered for genetic comparative analysis, in order to estimate the contribution of the Amerindian and the European pool, to the current population.

*Keywords:* Y-chromosome, SNP, STR, population genetics, Argentina, Amerindian

## Introduction

Population genetics historically concerns about the population subdivision and population structure, considering the allele frequency distribution and change under the influence of the four evolutionary forces: natural selection, genetic drift, mutation and gene flow. South America, and especially Argentina, is one of the human world locations where the history, the archaeological knowledge and the genetic information still continue not clear at all. Argentina population harbours many different Amerindian ethnic groups, and many of them with a very low number of individuals ( $\leq 1000$ ). The largest variety of population groups is found in the North, where the most representative ones are the Kollas, Diaguitas, Huarpe, Guaraní, Ava guarani, Wichi, Toba, Mocovi and Guaraní-mbya. While in the centre and south of the country we can find other groups as the Tehuelche, Rankulche, Comechingon, Tonocote, Ona and Mapuche population. (INDEC, Instituto Nacional De Estadística y Censos from Argentina, <http://www.indec.gov.ar> )

During the late 19<sup>th</sup> and the early 20<sup>th</sup> centuries, Argentina received lots of immigrants from different European countries, as Spain and Italy. However, immigrants settled mainly in the city areas, where is concentrated most of the inhabitants, giving rise to a diverse urban population with a highly diluted Amerindian component.

Indigenous population represent only about 1% of the total Argentina population (INDEC 2006), with around 403.000 individuals nowadays distributed all around the country, which shows a western area, with a rugged landscape corresponding to the andinian edge, and a landscape softened towards the eastern area, with plateaus and high plateaus. Among the indigenous groups, Kolla, Diaguita and Mapuche populations, are current representatives following a North to South distribution (Figure 1).

The Kolla group is represented by nearly 71,000 individuals settled in the provinces of Jujuy, Salta and Catamarca. Their languages are Aymara and Quechua.

Mapuche group is the most extensively represented, with almost 114,000 individuals, established along the Neuquen, Rio Negro and Chubut region. Their original language is Mapudungun

A more reduced indigenous group, are the Diaguitas settled in the provinces of Catamarca and Tucuman, which accounts with close to 32,000 individuals. They have a original language too, known with different names “kaka”, “chaka”, “yacampis”, “calchaqui”.

Although a large amount of data on autosomal markers has become available from different regions of South America (Tito et al. 2004, Fernández-Rosado et al. 2002, Martinez-Espin et al. 2003, Durán et al. 2003, Sala et al. 1998, Berardi et al. 2003, Jaime et al. 2004, Toscanini et al. 2007 ), until now, few Y chromosome populational studies have focused on this world region, and especially on Argentina (Sala et al. 1999, Garofalo et al. 2003, Fondevila et al. 2003) . Some of these studies include Y-STRs characterization, but only a few examples have made use of Y chromosome SNPs (Bianchi et al. 1998; Dipierri et al. 1998; Tarazona-Santos et al. 2001; Demarchi et al. 2004; Martinez Marignac et al. 2004). In this work we have joined Y-STRs and Y-SNPs because they are a very useful tool in human evolutionary studies, particularly as concerns male demographic history and migration patterns, furthermore Y-SNPs are especially interesting to find the American Y-chromosome founder haplogroups. There are two very important Y-SNPs which identify Amerindian populations: one of them is the M242 (C>T) (Seielstad et al. 2003), this mutation is located close in time to the entry of the first modern human into the American continent, the other one is M3 (C>T) (Underhill et al. 1996), this polymorphism was described to be occurred in Beringia when the first modern human entered into the American continent (Lell et al. 1997; Bianchi et al. 1998; Ruiz Linares et al. 1999). However, the close phylogenetic position of both mutation in the Y-chromosome haplogroup tree open to debate the possible origin of M3 mutation before the first human settlement in America (Bortolini et al. 2003; Su et al. 1999; Zegura et al. 2004; Lell et al. 2002; Karafet et al 1999)

In order to gain insights into the genetic diversity present in the male lineages of the Argentinian indigeneous populations, in this paper we have analysed the Y chromosome lineages of the Mapuche, Kolla and Diaguita indigenous group. Migration, admixture and antiquity patterns were evaluated, by the analysis of 26 Y chromosome Single Nucleotide Polymorphisms (Y-SNPs) and 11 Y chromosome Short Tandem Repeats (Y-STRs). We have looked for the genetic differences among those three indigenous



populations and moreover, other neighbour populations were considered to value the Amerindian and European genetic contribution.

## **2. Material and Methods**

### *Populations*

We obtained 134 unrelated male blood samples on FTA® Whatman paper, under informed consent. DNA was extracted using a standard phenol chloroform method. Extracted products were quantified on a NanoDrop® ND-1000 Spectrophotometer. Among the 134 samples collected, 34 belong to the Kolla ethnic group, 24 to the Diaguitas ethnic group, and 76 samples were Mapuches. Samples from Mapuche indigenous group were collected from Chubut, a province in the South of the country. Kolla individuals were collected from two locations, Jujuy and Salta provinces, both in Northwest Argentina. Finally, Diaguita samples were collected from Catamarca and La Rioja provinces in the Northwest of Argentina.

### *Y chromosome typing*

In order to carry out our analyses, we used multiplex reactions, both for the Y-SNPs and for the Y-STRs.

The Y-STRs were amplified by means of the PowerPlex Y® System (Promega) multiplex kit, containing 11 STRs (DYS391, DYS389I, DYS439, DYS389II, DYS438, DYS437, DYS19, DYS392, DYS393, DYS390, DYS385). Some samples were further analyzed with AmpFISTR®Yfiler™ multiplex kit (Applied Biosystems). In both cases manufacturer recommendations were followed.

Y-chromosome SNPs were analysed by using the SNaPshot multiplex kit (Applied Biosystems). The 26 Y-SNPs were amplified in the form of three multiplex reactions previously described by Brión et al. 2004, and a novel multiplex reaction containing the SNPs M242, M3, M19, M194 and M199. Design of the amplification primers of the new multiplex was carried out using Primer3 Software. The possible occurrence of

unspecific primer dimer was subsequently explored by means of the autodimer1 program. SNaPShot probes were designed in the same way (Table 1 and 2).

The functioning of the newly created multiplex was optimized using the QUIAGEN® Multiplex PCR kit, using 12.5 µl of PCR Master Mix, 2µM Primer Mix and 5-20ng of genomic DNA, in a final volume of 25µl. Amplification conditions were as follows: 95°C for 10 min, 94°C for 30s for 35 cycles, 58°C for 1min 30s, 72°C for 1min 30s, and a final extension cycle at 72° for 10 min.

Multiplex single base extension reactions were carried out following the Y-SNP amplifications by means of the SNaPshot™ Multiplex kit (Applied Biosystems). Prior to this step, 1µl of PCR product was cleaned with 0.5µl of ExoSAP-IT (Amersham Biosciences) and was incubated at 37°C for 15 min., followed by 85°C for 15 min. to inactivate the enzyme. The 1.5µl cleaned volume was then added to 2µl of the SNaPshot ready reaction mix (Applied Biosystems) and 1.5µl of extension primers mix in 5µl final volume. Amplification conditions were the following: 25 cycles of 96°C for 10 s., 51.5°C for 5 s. and 60°C for 30 s. This was followed by a cleaning step with 0.5 µl of shrimp alkaline phosphatase (Amersham Biosciences) at 37°C for 60 min. and 85°C for 15 min.

The products of both the Y-STR amplifications and the Y-SNP single base extensions, were detected by electrophoresis on a ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems) and analyzed using GeneScan® Analysis Software version 3.7 (Applied Biosystems).

Haplogroup nomenclature was assigned according to The Y Chromosome Consortium (YCC, 2002; Jobling et al. 2003; Karafet et al. 2008).

### *Statistical analysis*

Haplogroup and haplotypes frequencies and diversities were calculated by use of the Arlequin 2.0 software (Schneider et al. 2000). The same programme was also used to test the hypothesis of a random distribution of the individuals between pairs of populations, with an exact test of population differentiation and to perform analysis of molecular variance (AMOVA). Genetic distances were also calculated as pair-wise values of  $F_{ST}$  in the case of the Y-SNP data and as  $R_{st}$  values for the Y-STRs.

Median-joining networks were constructed with the Q1a3a lineages using Network 4.5.0.0 (Bandelt et al., 1999). A weighting scheme was adopted on the basis of the molecular variance of each STR across the whole population. The weights assigned were from 2 to 10, inversely proportional to the STR variance. The resulting network contained high dimensional cubes, which were resolved by using the reduced median algorithm to generate a \*.rmf file, and then applying the median joining network method to this file.

For comparisons, previously published haplogroup data from the following populations were compiled: three Brazilian samples, one metropolitan population from Rio de Janeiro (Silva et al. 2006) and two native samples, Guaraní and Kaingang (Marrero et al. 2007); one metropolitan population and another rural population from El Salvador (Lovo-Gomez et al. 2007); one Afro-Colombian sample and one Caucasian-Mestizo sample from Colombia (Acosta et al. 2008); and finally one European population from Galicia (Brión et al. 2004). Haplogroup frequencies were used for a Principal Component Analysis (PCA) and  $F_{st}$  distances were used for a multidimensional scaling (MDS) analysis. SPSS version 15.0 software package was used for both.

### **3. Results**

#### *Intra-population haplogroup and haplotype distribution*

The combination of the 26 SNP analysed allows the identification of 28 Y-chromosome haplogroups (Fig. 2), and 14 of them were present in our populations, but eight were represented by only one or two individuals. As we can see in the haplogroup frequencies represented in figure 1, six haplogroups accounts for the vast majority of chromosomes, and the two most common haplogroups in the populations analysed were by far haplogroups R1b1 and Q1a3a\*(xQ1a3a1-3), accounting for more than 70% of the total sample.

Considering the three population groups, the highest values of Hg R1b1 were detected in the Mapuches, while in the Kollas and Diaguitas, although high frequencies are also showed, the most frequent haplogroup in this case is the Amerindian haplogroup Q1a3a\*.

In terms of haplogroup diversity the Kolla population showed the highest values (0.8075 +/- 0.0441), Diaguitas showed intermediate values (0.7609 +/- 0.0596) and Mapuches the lowest (0.7042 +/- 0.0372).

A remarkable fact was the presence of haplogroup G in the Kolla population from Salta, reaching values as high as the most frequent haplogroup; however not too much weight should be given to this finding, because of the reduced sample size of the population.

The analysis of eleven Y-chromosomal STRs allowed us to identify 112 different haplotypes in the entire sample (Appendix 1, Electronic supplementary material). The vast majority were observed in single individuals, but some of them were repeated, four haplotypes were present in 3 samples and 14 haplotypes were present in 2 samples. Only one Y-STR haplotype is shared among populations, between Mapuches and Diaguitas, both belonging to haplogroup R1b. In terms of haplotype diversity Kollas and Diaguita population showed similar values (0.9964 +/- 0.0082 and 0.9964 +/- 0.0133 respectively), while Mapuches are the less diverse population (0.9895 +/- 0.0040).

One DYS19 duplicated chromosome was observed in the Kolla population from Salta, but duplicates alleles at this microsatellite have already been included in the YHRD (<http://www.yhrd.org/>).

#### *Inter-population comparative analysis*

Looking at the inter-population variability the percentage of variation observed at the Y-STR level for the four different regions shows that most of the variation is found within populations, but a significant 4.5% of the variation ( $p < 0.001$ . 1023 permutations) could be explained by differences among populations. When the four geographical regions are grouped into three populations, similar results were observed (data not shown).

Population pair-wise differentiation tests with both types of data, did show evidence of differentiation between Kolla sample from Salta and Mapuches in the case of haplogroup data and for Mapuches and the other populations in the case of haplotype data

To provide a context in which to consider the Amerindian and European genetic contribution, haplogroup frequency data were compiled for neighbour populations as Brazil, Colombia and Salvador (Marrero et al. 2007; Silva et al. 2006; Acosta et al. 2008; Lovo-Gomez et al. 2007), and for some European population (Brión et al. 2004). The haplogroup frequencies principal component plot (Fig.3a) shows our Argentinean populations tightly clustered with the populations from El Salvador, and close to the Native Brazilian samples. Populations with a higher European component are closely grouped with the Galician sample, while the Afro-Colombian population appear as an outlying.

The multidimensional scaling plot (MDS) performed with the  $F_{st}$  genetic distances (Fig. 3b) shows a similar picture.

In order to understand more about the molecular basis for the Amerindian haplogroup Q1a3a\* a median-joining network was constructed using the STR data (Fig. 4). Only one haplotype was shared between two populations, and the majority of them were scattered along the network, being the Kolla haplotypes from Jujuy the most distant.

#### **4. Discussion**

We have carried out an extensive analysis of Y-chromosomal diversity in native South-American populations, examining 26 binary polymorphisms and 11 STRs. This has allowed us to study the level of variability and substructure of the actual populations, as well as establishing the genetic contribution of indigenous Amerindian populations or European populations.

The most common haplogroup in the Native-American populations (Bortolini et al, 2003), haplogroup Q1a3a, is also present at high frequencies in our sample, nevertheless in comparison with other south American populations published in the literature the values detected are not as high as we could expect for autochthonous populations (Marrero et al. 2007; Lovo-Gomez et al. 2007; Zegura et al. 2004; Bortolini et al. 2003). We have found that Argentinean populations, as a whole, cluster with other south-American indigenous populations, as would be expected from the general observation that Y-chromosomal DNA variation in many areas of the world is predominantly

structured by geography, rather than language or ethnic affiliation (Zerjal et al. 2001; Rosser et al. 2000).

Most individual populations were not differentiated from one another. Nevertheless, the Kolla population from Salta departed from this general pattern due to the unusual high frequency of haplogroup G, that can be explained by the reduced sample size reported.

Y-STR allele distributions were similar to previous reports of Native Argentinean populations (Toscanini et al. 2008; Altuna et al. 2006; Demarchi et al. 2004).

The Amerindian component, represented by haplogroup Q1a3a, is widely represented in all the populations analysed, but none significant differences were found with other South American samples. In addition, none haplotype substructure was found in the network analysis performed with this haplogroup in the four populations.

The European component, represented by haplogroup R1b was also significant in most of the populations; although in the Kolla sample, the presence of this European haplogroup is reduced.

Our study has shown similar levels of Y-chromosomal differentiation within indigenous Argentinean populations. Pairwise genetic distances ( $R_{st}$ ) calculated among haplotypes (excluding DYS385) did not show significant values. Therefore, some of the most differentiated facts could be explained because of the particular sensitivity of the Y chromosome to genetic drift.

In order to investigate the number of times a particular haplotype presented in our populations appears in other worldwide population, repeated Y-STR haplotypes were searched in the YHRD (<http://www.yhrd.org/>). Only 5 of the 14 repeated haplotypes are present in Europeans and admixed population from Latin American countries, all of them belonging to the widely spread haplogroup R1b; in addition, one haplotype belonging to haplogroup J\* , besides of being present in the above mentioned populations is also present in an Algerian population from North Africa.

The presence of the DYS19 duplication in one Kolla from Salta occurred in a chromosome classified as haplogroup G. Capelli et al (2007) postulated about a possible multiple origin, with only two independent duplication events, one inside haplogroup G2 and another in haplogroup C3c. Our finding lend more weight to the hypothesis, if we consider that our kola duplicated sample could be ancestrally related to the

described Italian duplicated chromosomes belonging to haplogroup G, being Italy one of the great source of immigrants to Argentina during the last century.

#### *Final remarks*

The present study has detected an average clear Amerindian genetic component of around 35%, being lower in Mapuches (30%); a reduced African ancestry (less than 4%) and a predominant European influence, with more than 50% of the haplogroups being typically Europeans.

Our results contrast with those previously reported in similar areas of Argentina. Dipierri et al (1998) have shown a 64% of Amerindian component in a population from the Jujuy region, while the percentage detected in our study, despite of being the highest only reached a 39%.

Previous reports in other Argentinean populations have been shown a patent population substructure. One of the most recent studies (Salas et al, 2008) focus on the general population of the province of Cordoba; but in this case, the European component represent the great majority of the population with a very reduced Amerindian component.

Although all the individuals analysed were assigned to the different ethnic groups by their own accord in a personal interview, the Y chromosome lineages showed a non expected high percentage of European chromosomes for Amerindian populations. Given that gender bias has been previously described in Argentina (Salas et al, 2008), it will now be of great interest to see whether this pattern of differentiation is confined to the Y, or it is shared by mitochondrial DNA and autosomal markers.

#### **Acknowledgments**

We greatly thank all the sample contributors. This work was partially financed by two grants from the Xunta de Galicia, one from the Conselleria de Educación e Ordenación

Univesitaria (2006/117) and another from the Direccion Xeral de I+D (PGIDIT06PXIB208204PR).

## References

- Tito R.Y., López P.W., Iannacone G.C., Lizárraga B. (2004) Genetic diversity of sixteen STRs in the peruvian Mesa Redonda Lima population. *J Forensic Sci* 49(4):1-2.
- Fernández-Rosado F., Martínez-Espin E., Rodríguez T., Entrala C., Alvarez J.C., Lorente J.A., Lorente M., Budowle B. and Villanueva E. (2002) Population data of Ecuador for fifteen STR loci (POWERPLEX<sup>TM</sup> 16). *J Forensic Sci* 48(1):1-3.
- Martínez-Espin E., Fernández-Rosado F., Alvarez J.C., Entrala C., Lorente J.A., Oviedo de Duarte M., Villanueva E., Budowle B. (2003) Paraguayan population data on the fifteen STR loci included in the POWERPLEX<sup>TM</sup> 16 Kit. *J Forensic Sci* 48(1):1-3.
- Durán R., Zarante I., Acevedo M.L., Villegas M.R., Salazar J, Bocanegra BY, Bernal J. (2003) Allelic frequency of six STR loci in five Colombian cities. *J Forensic Sci* 48(4):1-2.
- Sala A., Penacino G., Carnese R., Corach D. (1999) Reference database of hypervariable genetic markers of Argentina: application for molecular anthropology and forensic casework. *Electrophoresis* 20: 1733-1739.
- Berardi G., Toscanini U., Raimondi E. (2003) STR data for PowerPlex 16 System from Buenos Aires population, Argentina, *Forensic Sci Int* 134: 222-224.
- Jaime J.C., Fondevila M., Salas A., Lareu M.V., Carracedo A. (2004) Nine autosomal STRs genotype profiles in a sample from Córdoba (Argentina). *Forensic Sci Int* 139: 81-83.
- Toscanini U., Gusmao L., Berardi G., Amorín A., Carracedo A., Salas A., Raimondi E., (2007) Testing for genetic structure in different urban Argentinian populations. *Forensic Sci Int* 165:35-40.
- Garofalo M.G., Gangitano D.A., Juvenal G.J., Budowle B, Lorente J.A. Padula R.A., (2003) Six Y-chromosome STR Frequencies in a Population from Argentina. *J. Forensic Sci.* Vol. 48 No.2.
- Fondevila M, Jaime JC, Salas A, Lareu MV, Carracedo A. (2003) Y-chromosome STR haplotypes in Córdoba (Argentina). *Forensic Sci Int.* 137(2-3):217-20.



- Bianchi NO, Catanesi C., Baillet G, Martinez-Marignac V.L., Bravi C.M., Vidal-Rioja L.B., Herrera R.J., Lopez-Camelo J.S., (1998) Characterization of Ancestral and Derived Y-Chromosome Haplotypes of New World Native Populations. *Am. J. Hum. Genet.* 63:1862-1871.
- Dipierri J.E., Alfaro E., Matínez-Marignac V.L., Baillet G., Bravi C.M., Cejas S., Bianchi N.O., (1998). Paternal directional mating in two amerindian subpopulations located at different altitudes in Northwestern Argentina. *Human Biology.* 70 (6): 1001-1010.
- Tarazona-Santos E., Carvalho-Silva D. R., Pettener D., De Stefano G. F., Tyler-Smith C., (2001) Genetic Differentiation in South Amerindians Is Related to Environmental and Cultural Diversity: Evidence from the Y Chromosome. *Am. J. Hum. Genet.* 68:1485-1496.
- Demarchi D.A., Mitchell R.J., (2004). Genetic structure and gene flow in Gran Chaco populations of Argentina: evidence from Y-chromosome markers. *Human biology* 76 (3): 413-429.
- Martinez Marignac V.L., Bertoni B., Parra E.J. and Bianchi N.O., (2004). Caracterización de Admixture in an Urban Sample from Buenos Aires, Argentina, Using Uniparentally and Biparentally Inherited Genetic Markers. *Human Biology.* 76 (4): 543-557.
- Seielstad M, Yuldasheva N., Underhill P., Oefner P., (2003) A Novel Y-Chromosome Variant Puts an Upper Limit on the Timing of First Entry into the Americas. *Am. J. Hum. Genet.* 73:700.
- Underhill P.A., Jin L., Zemans R., Oefner P.J. and Cavalli-Sforza L. (1996). A Pre-columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci. USA.* 92:196-200.
- Lell J. T., Brown M.D., Schurr T.G., Sukernik R. I., Starikovsky Y.B. (1997). Y chromosome polymorphisms in Native American and Siberian populations: identification of Native American Y chromosome haplotypes.
- Ruiz-Linares A., Ortiz-Barrientos D., Figueiroa M., Mesa N., (1999) Microsatellite provide evidence for Y chromosome diversity among the founder of the New World. *Proc. Natl. Acad. Sci. USA* Vol. 96: 6312-6317.

- Bortolini M.C., Salzano F.M., Thomas M.G., Stuart S., Nasanen S.P.K. Bau C.H.D., Hutz M.H., Layrisse Z., Petzl-Erler M.L., Tsuneto L.T., Hill K., Hurtado A.M., Castro-de-guerra D., Torres M.M., Groot H., Michalski R., Nymadawa P., Bedoya G., Bradman N., Labuda D., Ruiz-Linares A., (2003) Y-Chromosome Evidence for Differing Ancient Demographic Histories in the Americas. *Am. J. Hum. Genet.* 73: 524-539.
- Su B., Xiao J., Underhill P., Deka R. (1999). Y-chromosome Evidence for a Northward Migration of Modern Humans into Eastern Asia During The Last Ice Age.
- Zegura S.L., Karafet T.M., Zhivotovsky L.A., Hammer M.F. (2004) High-Resolution SNPs and Microsatellite Haplotypes Point to a Single, Recent Entry of Native American Y Chromosome into the Americas. *Mol. Biol. Evol.* 21(1):164-175.
- Lell J.T., Sukernik R. I., Starikovskaya Y.B., Su B., Jin L.,(2002). The Dual Origin and Siberian Affinities of Native American Y Chromosomes. *Am. J. Hum. Genet.* 70:192-206.
- Karafet T.M., Zegura S.L., Posukh, O., Osipova L., Bergen A., (1999) Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes. *Am. J. Hum.* 64:817-831.
- Brión M., Sobrino B., Blanco-Verea A.J., Lareu M.V., Carracedo A., (2004) Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int. J. Legal Med.* 119: 10-15.
- Y-chromosoma Consortium (2002) A nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroup. *Genome Research* 12: 339-348.
- Jobling MA and Tyler-Smith C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4(8): 598-612.
- Karafet T.M., Mendez F.L., Meilerman M.B., Underhill P.A., Zegura S.L., Hammer M., (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 8: 1-9
- Schneider S, Roessli D, Excoffier L. (2000) ARLEQUIN ver. 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory University of Geneva, Switzerland.
- Bandelt HJ, Forster P, Röhl A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16(1): 37-48.

-Silva D.A, Carvalho E., Costa G., Tavares L., Amorim A., Gusmao L, Y-chromosome genetic variation in Rio de Janeiro population, *Am. J. Hum. Biol.* 18 (6) (2006) 829–837.

-Marrero A.R., Silva-Junior W.A., Bravi C.M., Hutz M.H., Petzl-Erler M.L., Ruiz-Linares A., Salzano F.M., Bortolini M.C., Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil, *Am. J. Phys. Anthropol.* 132 (2) (2007) 301–310.

-Lovo-Gomez J., Blanco-Verea A., Lareu M.V., Brión M, Carracedo A., (2007) The genetic male legacy from El Salvador, *Forensic Sci. Int.* 171: 198–203.

-Acosta M.A., Blanco-Verea A., Lareu M.V, Brión M., Carracedo A., (2008) The genetic male component of two South-Western Colombian populations. *Forensic Sci. Int. Genetics.* (In press).

-Zerjal T, Beckman L, Beckman G, Mikelsaar AV, Krumina A, Kucinskas V, Hurles ME, Tyler-Smith C (2001) Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. *Mol Biol Evol* 18: 1077-1087.

-Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorín A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HBSM, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, María S, McElreavey K, Meitinger TA, Mikelsaar A-V, Mitchell RJ, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67: 1526-1543.

-Toscanini U, Gusmao L, Berardi G, Amorim A, Carracedo A, Salas A, Raimondi E.(2008) Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data. *Forensic Sci Int: Genetics* 2: 274-280.

- Altuna ME, Modesti NM, Demarchi DA. (2006) Y-chromosomal evidence for a founder effect in Mbyá-Guaraní Amerindians from Northeast Argentina. *Hum Biol* 78(5): 635-639.
- Capelli C, Brisighelli F, Scarnicci F, Blanco-Verea A, Brión M, Pascali VL. (2007) Phylogenetic evidence for multiple independent duplication events at the DYS19 locus. *Forensic Sci Int: Genetics* 1: 287-290.
- Salas A, Jaime JC, Álvarez-Iglesias V, Carracedo A. (2008) Gender bias in the multiethnic genetic composition of central Argentina. *J Hum Genet* 53: 662-674.

### **Legends to figures**

Figure 1. Map of Argentina showing haplogroups frequencies of each population, geographically localized.

Figure 2. Y-chromosome phylogenetic tree. Polymorphism names are indicated above the lines and lineage names are shown below the lines. The length of the line has not significance.

Figure 3. a. Principal component analysis of Y-chromosomal haplogroup frequencies. The percentage of variance explained by each component is given on the axes. b. Multidimensional scaling plot of the  $F_{st}$  distances. Stress value 0.0295.

Figure 4. Median-joining network of haplogroup Q3. Circles represent haplotypes with an area proportional to frequency, and colors indicate the population of origin (Blue = Kolla Jujuy, Yellow = Mapuche, Cyan = Kolla Salta, Orange = Diaguitas).

SNP	Mutation	Primer (5'-3')		Amplicon Size (bp)	Conc ( $\mu$ M)
		Forward	Reverse		
M242	C > T	atagaaagtttgcaaaaaggga	accttacctagaacaactctgaagc	137	0,4
M3	C > T	ctgccaggcctttcaaatag	aaggcctcttctcatttaggt	93	0,4
M19	T > A	ctggcataaacactggaaatc	agctgaccacaaactgatgtaga	170	0,4
M194	T > C	gcctggatgaggaagtgag	atacagtcgtgcctctcg	127	0,4
M199	> Ins G	cctggttgattctggtctt	tgattcaaggattgtagtctt	197	0,4

Table 1

SNP		Minisequencing primer (5'-3')	Primer Size (bp)	Conc ( $\mu$ M)
M242	For	cccccccccccccccccaaaaagggtgaccaagggtgct	39	0,15
M3	For	ccccccccccccgggtcaactctgggactga	31	0,25
M19	For	caaaactattttgtgaagactgttga	27	0,25
M194	For	ccccccccactttaacatacaggagtggtttt	35	0,15
M199	For	gaaatgttaaatggcttacactg	25	0,3

Table 2

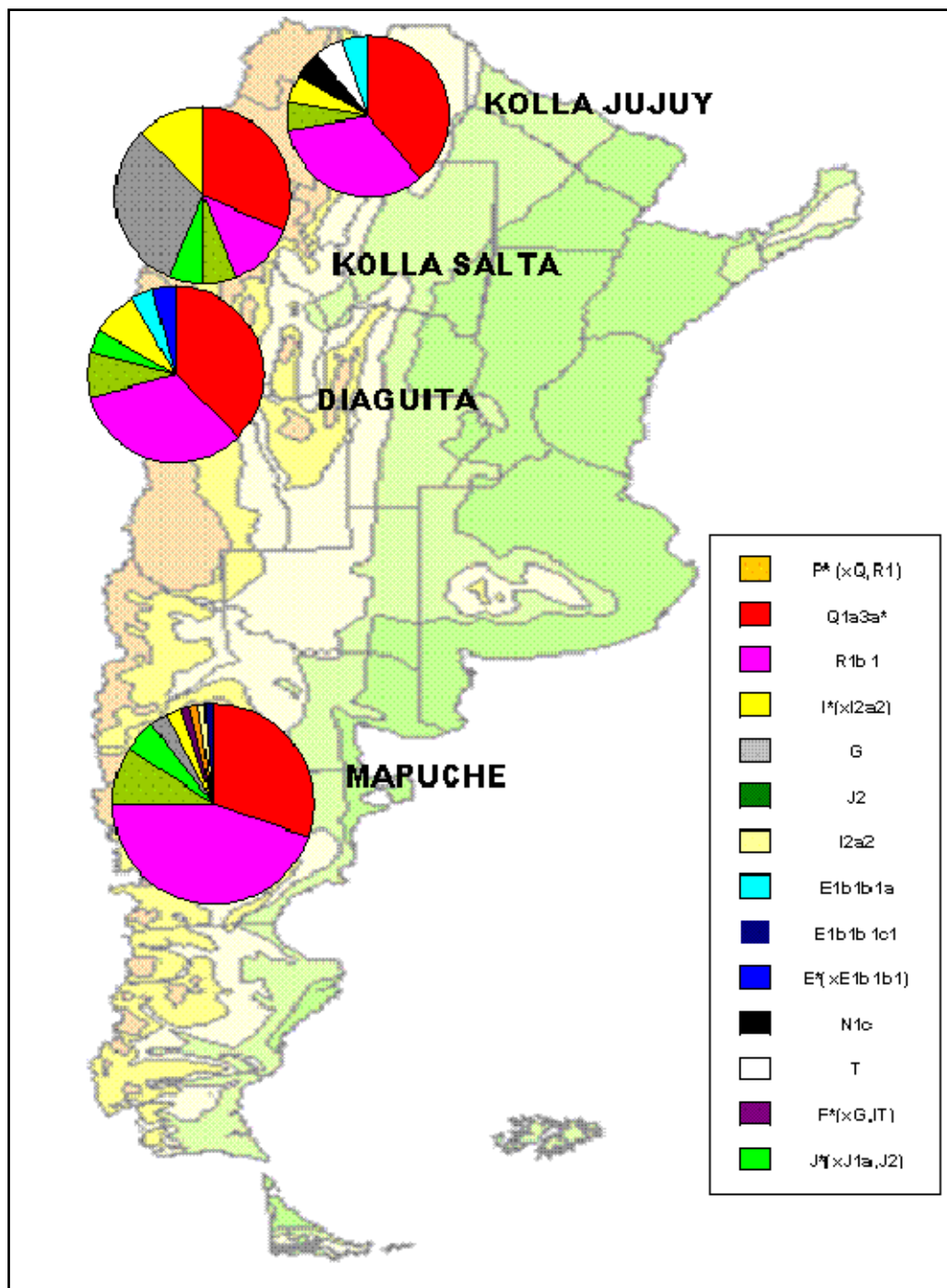


Figure 1

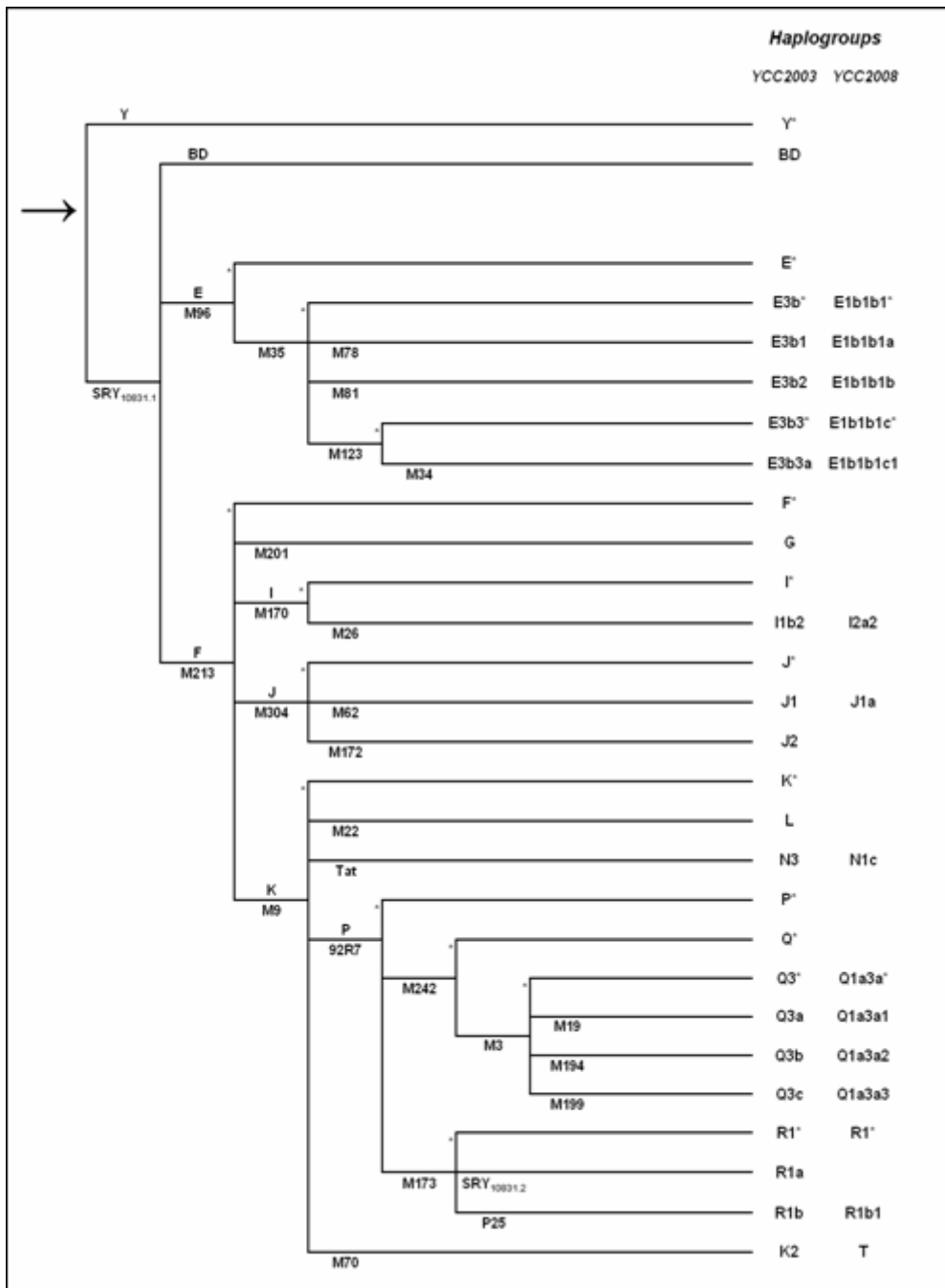


Figure 2



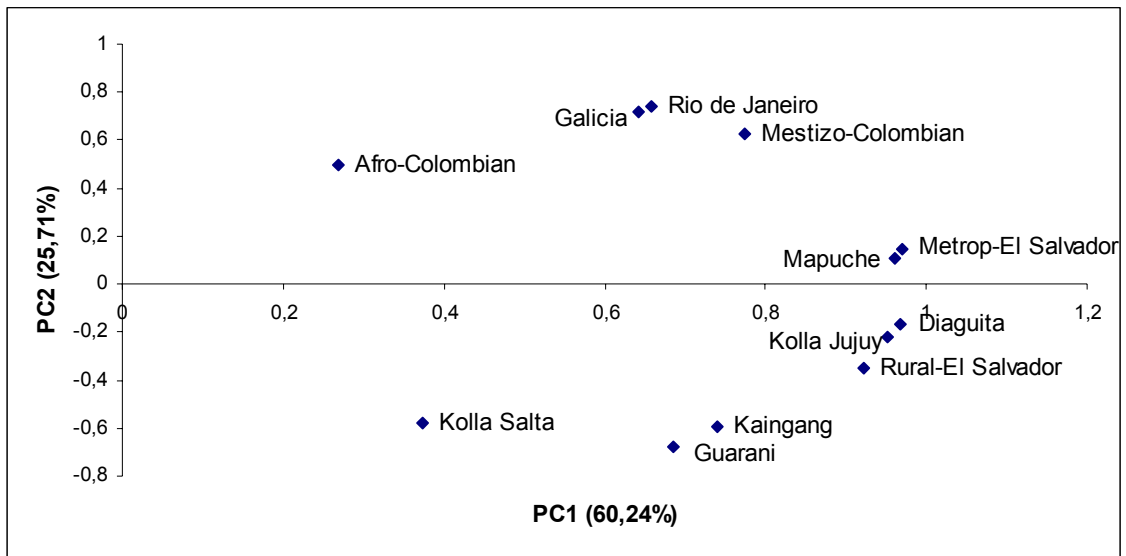


Figure 3a

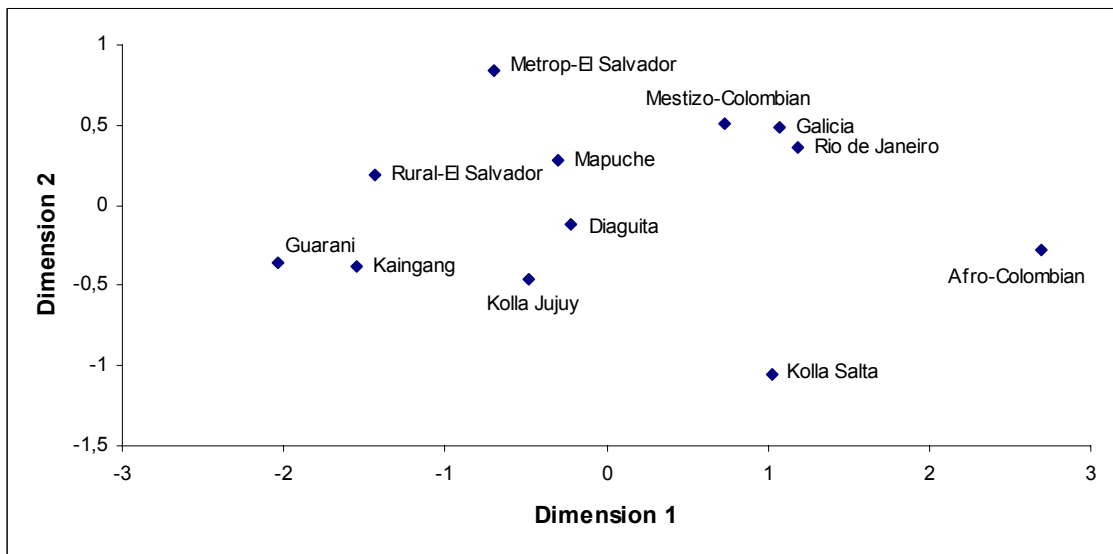


Figure 3b



HT	Sample ID	Population	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385	DYS 437	DYS 438	DYS 439	HG
1	20	Kolla	13	12	29	24	10	15	12	14.15	14	11	13	Q1a3a*
2	25	Mapuche	13	12	30	25	10	14	13	17.17	14	11	13	Q1a3a*
3	14	Kolla	13	13	29	23	10	11	13	14.16	14	10	12	E1b1b1a
4	14	Mapuche	13	13	29	23	10	14	13	14.18	14	11	13	Q1a3a*
5	33	Mapuche	13	13	29	24	10	14	13	14.14	14	11	12	Q1a3a*
5	58	Mapuche	13	13	29	24	10	14	13	14.14	14	11	12	Q1a3a*
6	42	Mapuche	13	13	29	25	10	15	13	14.16	14	11	12	Q1a3a*
6	59	Mapuche	13	13	29	25	10	15	13	14.16	14	11	12	Q1a3a*
7	4	Kolla	13	13	29	26	11	15	13	15.17	14	11	13	Q1a3a*
8	1	Mapuche	13	13	30	22	9	14	13	12.15	14	11	11	Q1a3a*
9	52	Mapuche	13	13	30	23	9	14	13	12.15	14	11	11	Q1a3a*
10	7	Mapuche	13	13	30	23	10	13	13	14.19	14	11	10	Q1a3a*
11	21	Mapuche	13	13	30	23	10	13	13	14.17	14	11	10	Q1a3a*
12	19	Mapuche	13	13	30	24	10	11	14	16.17	14	10	13	E1b1b1c1
13	21	Kolla	13	13	30	24	10	14	13	16.22	14	11	13	Q1a3a*
14	21	Diaguitas	13	13	30	24	10	14	13	14.14	14	11	12	Q1a3a*
15	7	Kolla	13	13	31	24	10	13	12	13.16	14	11	9	Q1a3a*
16	11	Mapuche	13	13	31	23	9	14	13	13.15	14	11	11	Q1a3a*
16	40	Mapuche	13	13	31	23	9	14	13	13.15	14	11	11	Q1a3a*
16	64	Mapuche	13	13	31	23	9	14	13	13.15	14	11	11	Q1a3a*
17	23	Mapuche	13	13	31	23	10	14	13	12.15	14	11	12	Q1a3a*
17	75	Mapuche	13	13	31	23	10	14	13	12.15	14	11	12	Q1a3a*
18	3	Diaguitas	13	13	31	23	10	14	13	15.19	14	11	11	Q1a3a*
19	19	Kolla	13	13	32	24	10	14	13	15.17	14	11	12	Q1a3a*
20	22	Kolla	13	13	33	23	10	14	13	15.18	14	11	12	Q1a3a*
21	39	Mapuche	13	14	29	24	10	14	13	13.16	14	11	13	Q1a3a*
21	61	Mapuche	13	14	29	24	10	14	13	13.16	14	11	13	Q1a3a*
22	10	Diaguitas	13	14	30	24	11	14	13	14.19	14	12	11	Q1a3a*
23	13	Diaguitas	13	14	30	25	10	15	13	14.18	14	10	12	Q1a3a*
23	16	Diaguitas	13	14	30	25	10	15	13	14.18	14	10	12	Q1a3a*
24	70	Mapuche	13	14	30	25	10	15	13	14.14	14	11	12	Q1a3a*
25	22	Diaguitas	13	14	31	23	10	15	14	16.18	14	11	13	Q1a3a*
26	9	Kolla	13	14	31	23	10	16	14	15.18	14	11	14	Q1a3a*
27	30	Kolla	13	14	31	23	10	16	14	15.17	14	11	14	Q1a3a*
28	15	Kolla	13	14	31	23	10	16	14	15.18	14	11	12	Q1a3a*
29	18	Kolla	13	14	31	23	10	16	14	16.19	14	11	13	Q1a3a*
30	18	Diaguitas	13	14	31	23	10	16	14	15.17	14	11	13	Q1a3a*
31	1	Kolla	13	14	31	23	11	14	13	13.19	14	11	12	Q1a3a*
32	9	Diaguitas	13	14	31	24	10	11	13	16.18	14	10	12	E1b1b1a
33	12	Diaguitas	13	14	31	24	10	14	13	14.19	14	11	12	Q1a3a*
34	15	Diaguitas	13	14	31	24	10	16	14	15.19	14	11	13	Q1a3a*
35	25	Kolla	13	14	32	21	10	14	13	14.18	14	11	13	Q1a3a*
36	37	Mapuche	13	14	32	24	10	14	13	14.17	14	10	12	Q1a3a*
37	31	Kolla	14	11	26	24	10	13	13	11.14	15	12	13	R1b1
38	2	Mapuche	14	12	28	22	10	11	14	13.13	16	10	11	I*(xl2a2)
39	26	Mapuche	14	12	28	22	10	11	13	13.14	16	10	12	I*(xl2a2)
40	43	Mapuche	14	12	28	22	10	11	13	12.14	16	10	11	G
41	12	Kolla	14	12	28	23	10	11	12	14.16	15	9	13	J2
42	6	Diaguitas	14	12	28	23	10	11	12	14.16	15	9	12	J2

HT	Sample ID	Population	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385	DYS 437	DYS 438	DYS 439	HG
43	4	Mapuche	14	12	28	24	11	13	13	10.14	15	12	12	R1b1
44	68	Mapuche	14	12	28	25	11	13	13	11.14	16	12	12	R1b1
45	20	Diaguitas	14	12	29	22	10	?	13	?	16	10	11	I*(xI2a2)
46	23	Diaguitas	14	12	29	24	11	13	13	11.14	15	12	12	R1b1
47	72	Mapuche	14	13	28	24	10	13	13	11.14	15	12	11	R1b1
48	24	Mapuche	14	13	28	24	11	13	13	11.14	14	12	12	R1b1
49	65	Mapuche	14	13	28	25	11	13	13	11.14	14	12	12	R1b1
50	8	Diaguitas	14	13	29	22	10	11	12	14.15	15	9	12	J2
51	33	Kolla	14	13	29	23	10	11	12	13.19	14	10	13	J*(xJ1a,J2)
52	17	Kolla	14	13	29	23	10	13	13	11.11	14	12	13	R1b1
53	35	Mapuche	14	13	29	23	11	11	12	13.20	14	10	11	J*(xJ1a,J2)
53	47	Mapuche	14	13	29	23	11	11	12	13.20	14	10	11	J*(xJ1a,J2)
54	54	Mapuche	14	13	29	23	11	11	12	13.17	14	10	11	P*(xQ,R1)
55	11	Kolla	14	13	29	23	11	13	13	11.14	15	12	12	R1b1
55	48	Mapuche	14	13	29	23	11	13	13	11.14	15	12	12	R1b1
56	67	Mapuche	14	13	29	24	10	13	13	11.12	15	12	13	R1b1
57	66	Mapuche	14	13	29	24	10	13	13	11.14	15	12	12	R1b1
58	16	Mapuche	14	13	29	24	10	13	13	12.12	15	12	10	R1b1
59	13	Mapuche	14	13	29	24	10	?	13	14.17	14	11	12	R1b1
60	7	Diaguitas	14	13	29	24	11	13	13	11.14	15	12	12	R1b1
60	53	Mapuche	14	13	29	24	11	13	13	11.14	15	12	12	R1b1
61	62	Mapuche	14	13	29	24	11	13	13	11.14	15	12	13	R1b1
61	63	Mapuche	14	13	29	24	11	13	13	11.14	15	12	13	R1b1
62	69	Mapuche	14	13	29	24	11	13	13	11.14	15	11	11	R1b1
63	30	Mapuche	14	13	29	24	11	13	13	11.15	15	12	11	R1b1
63	51	Mapuche	14	13	29	24	11	13	13	11.15	15	12	11	R1b1
63	57	Mapuche	14	13	29	24	11	13	13	11.15	15	12	11	F*(xG,IT)
64	17	Mapuche	14	13	29	24	11	13	14	11.13	15	12	12	R1b1
65	71	Mapuche	14	13	29	24	11	13	14	11.13	16	12	12	R1b1
66	3	Mapuche	14	13	29	24	11	13	14	11.14	15	12	12	R1b1
66	10	Mapuche	14	13	29	24	11	13	14	11.14	15	12	12	R1b1
67	27	Mapuche	14	13	29	24	11	13	14	11.15	15	12	12	R1b1
68	8	Mapuche	14	13	29	24	12	13	13	11.14	15	12	12	R1b1
69	50	Mapuche	14	13	29	25	10	13	12	11.15	15	12	12	R1b1
70	2	Diaguitas	14	13	29	25	11	13	13	11.14	15	12	12	R1b1
71	6	Kolla	14	13	29	25	11	14	13	11.14	15	12	12	R1b1
72	28	Mapuche	14	13	30	23	10	11	12	13.18	14	10	11	J*(xJ1a,J2)
72	74	Mapuche	14	13	30	23	10	11	12	13.18	14	10	11	J*(xJ1a,J2)
73	38	Mapuche	14	13	30	24	11	13	14	10.14	15	12	13	R1b1
74	6	Mapuche	14	13	30	25	11	14	13	14.15	11	16	13	Q1a3a*
75	18	Mapuche	14	13	31	24	10	13	14	11.14	15	12	12	R1b1
76	15	Mapuche	14	14	30	23	11	13	13	11.14	15	12	12	R1b1
77	29	Mapuche	14	14	30	23	11	13	13	11.14	15	12	11	R1b1
77	46	Mapuche	14	14	30	23	11	13	13	11.14	15	12	11	R1b1
78	9	Mapuche	14	14	30	24	11	13	13	11.14	14	11	11	R1b1
79	1	Diaguitas	14	14	30	24	11	13	13	11.14	15	12	11	R1b1
80	13	Kolla	14	14	30	24	11	13	13	11.15	14	12	11	R1b1
81	27	Kolla	14	14	30	25	11	13	13	11.15	15	12	12	R1b1
82	31	Mapuche	14	14	30	24	11	13	14	10.14	15	12	12	R1b1
82	44	Mapuche	14	14	30	24	11	13	14	10.14	15	12	12	R1b1

HT	Sample ID	Population	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385	DYS 437	DYS 438	DYS 439	HG
82	45	Mapuche	14	14	30	24	11	13	14	10.14	15	12	12	R1b1
83	4	Diaguitas	14	14	31	23	10	11	12	11.18	14	10	12	J*(xJ1a,J2)
84	16	Kolla	14	14	31	24	10	13	13	11.14	14	12	12	R1b1
85	5	Diaguitas	14	14	31	24	10	13	13	11.14	15	12	13	R1b1
86	24	Diaguitas	14	14	31	24	11	13	13	11.14	14	12	12	R1b1
87	26	Kolla	15	12	28	22	10	11	13	14.15	16	10	12	G
88	5	Mapuche	15	12	28	23	10	11	12	14.18	16	9	10	J2
89	24	Kolla	15	12	29	22	10	11	13	14.14	16	11	12	G
90	29	Kolla	15	12	29	22	10	11	14	13.16	16	10	11	G
90	32	Kolla	15	12	29	22	10	11	14	13.16	16	10	11	G
91	28	Kolla	15	12	30	22	10	11	12	12.13	15	10	11	I*(xl2a2)
92	12	Mapuche	15	13	29	22	10	11	14	13.16	16	10	11	G
93	76	Mapuche	15	13	29	23	9	11	12	13.16	14	9	12	J2
94	60	Mapuche	15	13	29	23	10	?	12	?	14	9	11	J2
95	2	Kolla	15	13	29	23	11	13	12	11.14	15	12	12	R1b1
96	32	Mapuche	15	13	29	24	10	13	12	11.15	15	12	13	R1b1
96	49	Mapuche	15	13	29	24	10	13	12	11.15	15	12	13	R1b1
97	9	Kolla	15	13	30	23	11	13	13	15.17	13	9	11	T
98	41	Mapuche	15	13	30	24	10	13	13	14.20	14	11	13	Q1a3a*
99	36	Mapuche	15	13	30	24	10	14	13	14.20	14	11	12	Q1a3a*
100	34	Kolla	15	13	31	23	10	11	12	13.18	14	9	11	J2
101	11	Diaguitas	15	14	30	24	11	13	13	11.14	14	12	12	R1b1
102	3	Kolla	15	14	31	23	10	14	14	11.13	14	10	10	N1c
103	14	Diaguitas	15	14	31	23	11	12	14	11.17	15	12	13	R1b1
104	20	Mapuche	15	14	31	24	10	14	13	16.17	15	11	10	Q1a3a*
105	34	Mapuche	16	12	29	23	9	11	12	14.17	14	9	12	J2
105	55	Mapuche	16	12	29	23	9	11	12	14.17	14	9	12	J2
105	56	Mapuche	16	12	29	23	9	11	12	14.17	14	9	12	J2
106	30	Kolla	16	13	29	23	10	12	14	15.15	16	10	11	I*(xl2a2)
107	17	Diaguitas	16	13	29	23	11	12	13	14.15	15	10	11	I*(xl2a2)
108	22	Mapuche	16	13	29	24	10	11	13	14.17	14	9	12	J2
109	8	Kolla	17	13	28	24	9	11	13	12.12	14	10	13	I*(xl2a2)
110	73	Mapuche	17	14	29	23	10	11	13	12.13	15	10	12	I2a2
111	19	Diaguitas	17	14	31	21	10	11	15	16.19	14	11	9	E*(xE1b1b1)
112	23	Kolla	14-15	14	31	22	10	11	13	14.15	16	10	11	G

Supplementary data: (HT = haplotype, HG = Haplogroup).

**Artículo 8: “Moros y sarracenos en Europa: estima del legado masculino medieval del Norte de África en el Sur de Europa”**

Capelli C, Onofri V, Brisighelli F, Boschi I, Scarnicci F, Masullo M, Ferri G, Tofanelli S, Tagliabracci A, Gusmao L, Amorin A, Gatto F, Brión M, **Blanco-Verea A**, Romano V, Cali F, Pascali V.

European Journal of Human Genetics (manuscrito en revisión).

Resumen:

Para investigar el papel del legado árabe genético masculino en el Sur de Europa durante la época medieval, nos hemos centrado en haplogrupos específicos del Noroeste de África y hemos identificado haplotipos evolutivamente próximos definidos por STRs, en la península Ibérica, Sicilia y la península Italiana. Nuestros resultados señalan una contribución reciente del Noreoeste Africano en la península Ibérica y en Sicilia coincidiendo con los datos históricos. Es sabido que las regiones del Sur de Italia han experimentado largos períodos de presencia árabe mostando un enriquecimiento de linajes del Noroeste Africano. Las implicaciones genómicas y forenses serán discutidas.



**“Moors and Saracens in Europe: Estimating medieval North Africa male legacy into Southern Europe”**

Cristian Capelli<sup>1</sup>, Valerio Onofri<sup>2</sup>, Francesca Brisighelli<sup>3,4</sup>, Iliaria Boschi<sup>4</sup>, Francesca Scarnicci<sup>4</sup>, Mara Masullo<sup>4</sup>, Gianmarco Ferri<sup>5</sup>, Sergio Tofanelli<sup>6</sup>, Adriano Tagliabracci<sup>2</sup>, Leonor Gusmao<sup>7</sup>, Antonio Amorim<sup>7,8</sup>, Francesco Gatto<sup>9</sup>, Maria Brión Martinez<sup>3</sup>, Alejandro Blanco Vereá<sup>3</sup>, Valentino Romano<sup>10</sup>, Francesco Cali<sup>11</sup>, Vincenzo Pascali<sup>4</sup>

<sup>1</sup> Department of Zoology, University of Oxford, UK

<sup>2</sup> Institute of Legal Medicine, Università Politecnica delle Marche, Policlinico Torrette, Ancona, Italy

<sup>3</sup> Medicine Genomic Group, Institute of Legal Medicine, University of Santiago de Compostela, Spain.

<sup>4</sup> Istituto di Medicina Legale, Università Cattolica del S. Cuore, Rome Italy

<sup>5</sup> Department of Diagnostic and Laboratory Service and Legal Medicine, Section of Legal Medicine, University of Modena and Reggio Emilia, Italy

<sup>6</sup> Department of Biology, Anthropology Unit, University of Pisa, Italy

<sup>7</sup> IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto, Portugal

<sup>8</sup> Faculty of Sciences, University of Porto, Portugal.

<sup>9</sup> Biotechnology Unit, Istituto Zooprofilattico Sperimentale Lazio e Toscana, Rome, Italy

<sup>10</sup> Dipartimento di Oncologia Sperimentale e Applicazioni Cliniche Università di Palermo, Italy

<sup>11</sup> Oasi Institute for Research on mental Retardation and Brain Aging (IRCCS), Troina, Italy

**Running Title:** Male Arab legacy in Southern Europe

**Corresponding Author:** Cristian Capelli

Department of Zoology

University of Oxford

South Parks Road

OX1 3PS Oxford, UK

Email: [cristian.capelli@zoo.ox.ac.uk](mailto:cristian.capelli@zoo.ox.ac.uk)

Phone: +44 (0) 1865 271261

Fax: +44 (0)1865 310447

**Abstract**

In order to investigate the male genetic legacy of the Arab rule in Southern Europe during the medieval times, we have focused on specific North West African



haplogroups and identified evolutionary close STR defined haplotypes in Iberia, Sicily and the Italian peninsula. Our results point to higher North West African recent contribution in Iberia and Sicily in agreement with historical data. Southern Italian regions known to have experienced long term Arab presence show an enrichment of North West African types. Forensics and genomic implications are discussed.

**Keywords:** Y chromosome, Arab legacy, Southern Europe

## **Introduction**

After the collapse of the Roman Empire, the Arab dominance across the Mediterranean regions has been one of the most impressive historical events occurring in this region. Arabs appeared on the Southern shores of the Mediterranean in the early 7<sup>th</sup> century and quickly conquered North Africa. They spread their language and religion to the native North West African Berbers populations, the latter representing the bulk of the Muslim army later conquering Southern Europe (Davies, 1988; Hitti 1990). Referred either as Moors (in Iberia) or Saracens (in South Italy and Sicily), their arrival in Europe dates to 711 AD, rapidly subduing most of Spain, Portugal and Sicily (831 AD). Among European kingdoms their presence was seen as constant danger, but only in the 15<sup>th</sup> century the Iberian re-conquest was completed (Norman, 1975). In the 13<sup>th</sup> century Frederick II destroyed the Arab rule in Sicily and in 1221-1226 moved all the Arabs of Sicily to the city of Lucera, North of Apulia (Norman, 1975). Lucera was later destroyed by Charles II (1301) but an Arab community was recorded in Apulia in 1336. Guerrilla warfare was still conducted by Arabs in Sicily after Frederick II action (Norman, 1975). Y chromosome studies attempted to estimate the Arab contribution to Southern Europe focusing almost exclusively on the North African haplogroup E3b1b-M81, only partially taking in consideration the evolutionary relationships among haplotypes (Rosser et al, 2000; Bosch et al, 2001, Brion et al, 2004; Beleza et al, 2007). In order to generate a more comprehensive view of the genetic legacy of the medieval Arab dominance in Europe, we have systematically screened Y chromosome haplotypes within three North West (NW) African specific haplogroups across multiple southern European populations and performed additional genotyping to refine the available genetic data. Our results confirm a general correlation between historical and genetic

data, Iberia and Sicily being the regions with the highest NW African recent male legacy.

## **Material and Methods**

*Identification of recently introgressed NW African haplotypes.* Given the historical indication of a prevalently Berber origin for the Arab groups invading Southern Europe (Norman, 1975; Hitti, 1990), we focus on NW African specific haplogroups as markers of the Arab contribution. The E3b1b (M81 derived), E3b1a- $\beta$  (M78 derived chromosomes showing the rare DYS439 allele 10) and a subset of J1 (M267 derived) haplogroups have been identified by literature screening as NW African specific (Arredi et al, 2004; Cruciani et al, 2004; Semino et al, 2004). By the analysis of 9 Y chromosome STR loci (mutation rate  $2.8 \cdot 10^{-3}$  /  $6.9 \cdot 10^{-4}$  mutations per locus per generation, 25 years generation time; Kayser et al, 2000; Zhivotovsky et al, 2004) an average of 1.25 mutations (0.5-2) are expected per pair of chromosomes who shared a common ancestor 40 generation ago (1,000 years ago). European Y chromosomes within the three haplogroups less than 3 mutational steps away to NW African STR haplotype were considered having a recent common ancestor.

*Samples.* A NW Africa database including more than 400 samples (127 Berbers from Tunisia, Frigi et al, 2006; Cherni et al, 2005; 102 South Tunisians, Ayadi et al, 2006; 109 Moroccan Arab and Berber speakers, Quintana-Murci et al, 2004; 50 Moroccan and 52 Tunisians, unpublished data) genotyped at 9 STRs (DYS19, DYS389I &II, DYS390, DYS391, DYS392, DSY393, and the bi-allelic DYS385) was build-up for STR comparisons. NW African specific haplogroups have been identified by further genotyping samples previously described (Bosch et al, 2001; Brión et al, 2004; Beleza et al, 2006; Onofri et al, 2007; Capelli et al, 2007; Ferri et al, in press) and including two novel Italian samples (Lucera and Veneto, table 1) and a Basque dataset (Garcia et al, 2004; Alonso et al, 2005). E3b1a chromosomes have been scored for the DYS439 locus to identify the E3b1a- $\beta$  cluster and the M267 marker has been investigated in previously identified J\*(xJ2) chromosomes (J1). All the individuals within E3b1b, E3b1a- $\beta$  and J1 were genotyped for 9 STRs (DYS19, DYS389I-II, DYS390, DYS391, DYS392, DYS393 and the bi-allelic DYS385). A Sicilian population was also included

(samples overlapping in Capelli et al, 2006 and Robino et al, 2007). Sicilian genotypes identified E3b\* and J\*(xJ2) lineages and did not include DYS439. Within the E3b\* and J\*(xJ2) haplogroups, 8 and 3 chromosomes respectively were found evolutionary close to NW African types. These samples were made available for further genotyping, to include DYS439, M78, M81 and M267.

## **Results and Discussion**

In order to address the degree of historical NW African contribution, we used a combined SNP-STR approach. The coalescent times for the three NW African specific haplogroups ranges between 5,000 to 24,000 years, spanning a number of historical scenarios each potentially explaining their presence on the Northern Mediterranean shores (Semino et al, 2004; Cruciani et al, 2004). It follows that estimating Arab genetic legacy on the basis of haplogroups' occurrence only would be misleading. To avoid such limitation, we have extended our analysis to STR data whose high mutation rate permit to focus on more recent events. As described in the Material and Methods section, we screened more than 2,300 South European samples to identify those haplotypes evolutionary close to NW African chromosomes (table 1). Frequencies for these types range between 0% and 20 % across Southern Europe, the highest being in Cantabria and comprising a sample from the Pas Valley, previously shown to have an extremely high frequency of the North African haplogroup E3b1b (Cruciani et al, 2004). Iberia and Sicily estimations were among the highest in accordance with the long term Arab rule over the island (Norman, 1975) and resulted not significantly different among them (Chi-square test  $P=0.83$ ) but significantly different from the Italian sample (Chi-square test,  $P<0.01$ ). An inspection of table 1 reveals a non random distribution of Arab types in the Italian peninsula, with at least a two-fold increase over the Italian average estimation in three geographically close samples across Southern Appennine mountains (East Campania, North West Apulia , Lucera). When pulled together, these three Italian samples displayed a local frequency of 5.8%, significantly different from the North and the rest of South Italy ( $P<0.01$ ), but not from Iberia and Sicily ( $P=0.25$  and  $P=0.57$ , respectively). In this areas Arab presence is historically recorded following Frederick the II relocation of Sicilian Arabs (Norman, 1975). Also in Iberia a non random distribution might be present, as suggested by lower estimates in the North East (Basque

region and Catalans), but more samples across the peninsula would be required to properly address this issue. Consistently with historical data (Norman, 1975), no population in Central Europe or the Balkans showed the presence of NW African recently introgressed types (Semino et al, 2004; Cruciani et al, 2004; Berger et al, 2005) beside few chromosomes (2.3%) in Albania and Romania (Bosch et al; 2006).

The increasing use of the highly structured distribution of Y chromosome types to investigate the ethnic/geographic origin of unknown samples (Wetton, Tsang & Khan, 2005) makes the identification in Italy of regions enriched with recently introgressed NW African types of forensic relevance. We found more than 66% of the Italian types here identified as having a recent NW African origin are without a match in a large Italian Y chromosome dataset comprising almost 1200 individuals (Presciuttini et al, 2001). Of these, 35% would instead find perfectly overlapping types in NW African populations, potentially providing misleading advices to investigators. Such results are of interest also in the light of the expanding business of genealogical services offering Y chromosome analysis to identify the ethnic ancestry. Our results clearly confirm that conclusions based on single chromosomes should be taken very cautiously (King et al, 2007). Which are the expected genomic consequences of this historically recent admixture event? Suppose that 40 generations ago there was a 5% male introgression of African DNA into the European gene pool, corresponding to a total contribution of 2.5% of genetic material. Immediately after the admixture event, a fraction of chromosomes within Europe would have African ancestry. Recombination since this event will have substantially reduced the size of fragments of African ancestry within Europe haplotypes, and with these parameters we would expect today to see an approximately exponential distribution (measuring size using genetic distance) of fragment sizes, with a mean value of roughly 2.6cM. Assuming a genome-wide average recombination rate of 1.3cM/Mb (Yu *et al*, 2001), 2.5% of a typical present day southern European genome would consist on average of 2Mb regions of African DNA. We believe that signatures of this event would be correctly identified using modern dense genotype data (Frazer et al, 2007). In general, taking into account population structure is critical in avoiding false positives in case-control association studies (Marchini *et al*, 2004) - so an understanding of similar historical admixture events is likely to aid researchers conducting such studies.

## Acknowledgements

Authors would like to thank Elena Bosch and Walther Parson for kindly providing unpublished data; Giovanni Destro-Bisol for commenting a preliminary version of the manuscript; Dr Trincucci for support in sampling Lucera; Marcello Menegatti, Cristian Sossai and the Associazione Culturale “Borghi dell’Ovest” for the Veneto samples. CC would like to thank Simon Myers for comments and suggestions on the genomic structure implication following recent admixture events.

## References

Ayadi I, Ammar-Keskes L, Rebai A: Haplotypes for 13 Y-chromosomal STR loci in South Tunisian population (Sfax region). *Forensic Sci Int* 2006; **164**: 249-53.

Alonso S, Flores C, Cabrera V, *et al*: The place of the Basques in the European Y-chromosome diversity landscape. *Eur J Hum Genet*. 2005; **13**:1293-302.

Arredi B, Poloni ES, Paracchini S *et al*: A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* 2004; **75**: 338-45.

Beleza S, Gusmão L, Lopes A *et al*: Micro-phylogeographic and demographic history of Portuguese male lineages. *Ann Hum Genet* 2006; **70**: 181-94.

Berger B, Lindinger A, Niederstätter H *et al*: Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. *Int J Legal Med* 2005; **119**: 241-6.

Erratum in: *Int J Legal Med* 2006; **120**: 255.

Brión M, Quintans B, Zarrabeitia M, *et al*: Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis. *Gene* 2004; **329**:17-25

Bosch E, Calafell F, Comas D *et al*: High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet* 2001; **68**: 1019-29.

Bosch E, Calafell F, González-Neira A *et al*: Paternal and maternal lineages in the

Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann Hum Genet* 2006 ; **70**: 459-87.

Capelli C, Redhead N, Romano V *et al*: Population structure in the Mediterranean basin: a Y chromosome perspective. *Ann Hum Genet* 2006; **70**: 207-25.

Capelli C, Brisighelli F, Scarnicci F *et al*: Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter. *Mol Phylogenet Evol* 2007; **44**: 228-39.

Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL *et al*: Y-chromosome diversity characterizes the Gulf of Oman. *Eur J Hum Genet* 2008; **16**: 374-86.

Cherni L, Pereira L, Goios A *et al*: Y-chromosomal STR haplotypes in three ethnic groups and one cosmopolitan population from Tunisia. *Forensic Sci Int* 2005; **152**: 95-9.

Cruciani F, La Fratta R, Santolamazza P *et al*: Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* 2004; **74**: 1014-22.

Davies RHC: A history of Medieval Europe. London, UK, Longmann Group Limited, 1988, pp 83-101.

Ferri G, Alù M, Corradini B *et al*: Slow and fast evolving markers typing in Modena males (North Italy) *Forensic Science International: Genetics in press*

Frazer KA, Ballinger DG, Cox DR *et al*: A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007; **449**:851-61.

Frigi S, Pereira F, Pereira L *et al*: Data for Y-chromosome haplotypes defined by 17 STRs (AmpFLSTR® Yfiler™) in two Tunisian Berber communities. *Forensic Sci Int* 2006; **160**: 80-83

García O, Martín P, Gusmão L, *et al*: A Basque Country autochthonous population

study of 11 Y-chromosome STR loci. *Forensic Sci Int.* 2004; **145**:65-8.

Hitti P: *The Arabs: a short history.* Washington DC, Gateway, 1990.

Kayser M, Roewer L, Hedman M *et al*: Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 2000; **66**: 1580-8.

Keen M: *A History of Medieval Europe.* Essex, UK, The Anchor Press, 1967, pp 23.

King TE, Parkin EJ, Swinfield G *et al*: Africans in Yorkshire? The deepest-rooting clade of the Y phylogeny within an English genealogy. *Eur J Hum Genet* 2007; **15**: 288-93.

Luis JR, Rowold DJ, Regueiro M *et al*: The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet* 2004; **74**: 532-44. Erratum in: *Am J Hum Genet* 2004; **74**: 788.

Marchini J, Cardon LR, Phillips MS *et al*: The effects of human population structure on large genetic association studies. *Nat Genet* 2004; **36**: 512-7.

Nebel A, Landau-Tasseron E, Filon D *et al*: Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa. *Am J Hum Genet* 2002; **70**: 1594-6.

Norman D: *The Arabs and Medieval Europe.* London, UK, Longmann Group Limited, 1975.

Onofri V, Alessandrini F, Turchi C *et al*: Y-chromosome genetic structure in sub-Apennine populations of Central Italy by SNP and STR analysis. *Int J Legal Med* 2007; **121**: 234-7.

Quintana-Murci L, Bigham A, Rouba H *et al*: Y-chromosomal STR haplotypes in Berber and Arabic-speaking populations from Morocco. *Forensic Sci Int* 2004; **140**: 113-5.

Presciuttini S, Caglià A, Alù M *et al*: Y-chromosome haplotypes in Italy: the GEFI collaborative database. *Forensic Sci Int* 2001; **122**: 184-8.

Rosser ZH, Zerjal T, Hurles ME *et al*: Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 2000; **67**: 1526-43.

Robino C, Inturri S, Gino S *et al*: Y-chromosomal STR haplotypes in Sicily. *Forensic Sci Int* 2006; **159**: 235-40.

Semino O, Magri C, Benuzzi G *et al*: Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* 2004; **74**: 1023-34.

Yu A, Zhao C, Fan Y *et al*: Comparison of human genetic and sequence-based physical maps. *Nature* 2001; **409**: 951-953.

Wetton JH, Tsang KW, Khan H: Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs. *Forensic Sci Int* 2005; **152**: 45-53.

Zhivotovsky LA, Underhill PA, Cinnioglu C *et al*: The effective mutation rate at Y-chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 2004; **74**: 50-61.





# **V. Conclusiones**



### **Primera**

Los SNPs del cromosoma Y: P25 y 92R7 son marcadores que presentan al menos una variante de secuencia paráloga originada por duplicación segmental siendo polimórfica una o varias de estas variantes de secuencia. Por lo tanto, no son polimorfismos binarios simples como se creía inicialmente, lo cual puede inducir en algunos casos a error en la interpretación de sus variantes alélicas.

### **Segunda**

El análisis de haplogrupos de cromosoma Y y los análisis networks utilizando Y-STRs, parecen indicar que las duplicaciones presentes en el DYS19 no tienen un origen común, sino que podrían haber ocurrido en tiempos distintos durante la evolución humana por procesos de recombinación entre alelos no homólogos o por conversión génica (intracromátida).

### **Tercera**

Para el estudio de SNPs y particularmente de SNPs del cromosoma Y, la minisequenciación es un método que garantiza un genotipado sin ningún tipo de ambigüedades. Por lo tanto el desarrollo de multiplexes para minisequenciación es especialmente interesante en genética forense y poblacional, ya que nos permite en un tiempo reducido analizar multitud de SNPs en una sola reacción.

### **Cuarta**

En el diseño de multiplexes de Y-SNPs para estudios poblacionales, debido a la marcada subestructuración geográfica, es muy importante adaptar la estrategia de selección de los polimorfismos binarios en función de la población a estudiar, para evitar un tipado de Y-SNPs innecesario.

### **Quinta**

Con el uso de los 4 multiplexes de Y-SNPs desarrollados en este trabajo se ha demostrado una vez más la elevada homogeneidad de la población gallega. A nivel

microgeográfico no se han encontrado diferencias significativas entre las distintas comarcas gallegas estudiadas, salvo en la Mariña Lucense.

### **Sexta**

Los 29 Y-SNPs seleccionados y reunidos en un único multiplex, definen haplogrupos con una marcada estratificación poblacional alrededor del mundo y nos permiten inferir el posible origen geográfico de un linaje. Esta capacidad de discriminación, ha quedado demostrada con el estudio realizado con una muestra de población mundial de 1126 individuos.

### **Séptima**

En el estudio de polimorfismos del cromosoma Y realizado sobre la población de El Salvador se ha encontrado un fuerte componente europeo en población metropolitana que contrasta con la mayor frecuencia del componente nativo americano hallado en poblaciones rurales de El Salvador.

### **Octava**

En el análisis de Y-SNPs en los grupos étnicos Afro-Colombiano y Caucásico-Mestizo del Suroeste de Colombia hemos encontrado que el componente africano y europeo son los más frecuentes respectivamente. El componente Amerindio aparece en frecuencias muy bajas en ambas poblaciones.

### **Novena**

Los linajes del cromosoma Y en población indígena Argentina presentan un componente amerindio y europeo similar al de otras poblaciones nativas del Sur de América. El hábitat urbano de la población mapuche se hace patente en su elevada diversidad y mestizaje europeo.

### **Décima**

El legado árabe genético masculino que se encuentra actualmente en población italiana y de la península Ibérica podría ser explicado por el dominio árabe experimentado por estas 2 poblaciones del Sur de Europa desde el siglo VII al XV.

### **Undécima**

La combinación de Y-SNPs e Y-STRs genera unos patrones de variabilidad genético-geográfica actual que nos permiten encajar los acontecimientos histórico-poblacionales ocurridos en el pasado.

### **Duodécima**

En genética forense los Y-SNPs son de interés en el análisis de muestras degradadas y alcanzan su mayor relevancia en la predicción del origen poblacional de una muestra basándonos en su linaje. Mientras que los Y-STRs son especialmente útiles en casos de mezcla de ADN masculino y femenino y en casos de identificación con ausencia del padre.



## **VI. Bibliografía**





## VI. Bibliografía

- **Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BP, Sanger F, Scherier PH, Smith AJ, Staden R, Young IG.** (1981) *Sequence and organization of the human mitochondrial genome*. Nature **290**: 457-465.
- **Andrews RM, Kubacka I, ChinneryPF, Lightowlers RN, Turnbull DM, Howell N.** (1999) *Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA*. Nature Genet **23**: 147.
- **Asamura H, Sakai H, Ota M, Fukushima H.** (2007) *MiniY-STR quadruplex systems with short amplicon lengths for analysis of degraded DNA samples*. Forensic Science International: Genetics **1**: 56-61.
- **Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Tyler-Smith C.** (2000) *Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information*. Nucl Acids Res **28**: 2-8.
- **Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE.** (2002) *Segmental Duplications: organization and impact within the current Human Genome Project assembly*. Genome Research **11**: 1005-1017.
- **Bianchi NO, Bailliet G, Bravi CM, Carnese RF, Rothhammer F, Martinez-Marignac VL, Pena SD.** (1997) *Origin of Amerindian Y-Chromosome as inferred by the analysis of six polymorphic markers*. Am J Phys Anthropol **102**: 79-89.
- **Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, López-Camelo JS.** (1998) *Characterization of*

*ancestral and derived Y-chromosome haplotypes of new world native populations.* Am J Hum Genet **63**: 1862-1871.

- **Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A.** (2003) *Y-chromosome evidence for differing ancient demographic histories in the Americas.* Am J Hum Genet **73**: 524-539.
- **Brinkmann B, Klitschar M, Neuhuber F, Hühne J, Burkhard R.** (1998) *Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat.* Am J Hum Genet **62**: 1408-1415.
- **Brión M, Sanchez JJ, Balogh K, Thacker C, Blanco-Verea A, Borsting C, Stradmann-Bellinghausen B, Bogus M, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N.** (2005) *Introduction of a single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages.* Electrophoresis **26**: 4411-4420.
- **Brión M, Sobrino B, Blanco-Verea AJ, Lareu MV, Carracedo A.** (2004) *Hierarchical analysis of 30 Y-chromosome SNPs in European populations.* Int J Legal Med **119**(1): 10-15.
- **Burton ML, Moore CC, Whiting JWM, Romney AK.** (1996) *Regions based on social structure.* Current Anthropology **37** (1): 87-123.
- **Butler JM, Decker AE, Vallone PM, Kline MC.** (2006) *Allele frequencies for 27 Y-STRs loci with U.S. Caucasian, African American, and Hispanic samples.* Forensic Sci Int **156**(2-3): 250-260.

- 
- **Butler JM, Shen Y, McCord BR.** (2003) *The development of reduced size STR amplicons as tools for analysis of degraded DNA.* J Forensic Sci **48**(5):1054-1064.
  - **Calafell F, Shuster A, Speed WC, Kidd JD, Kidd KK.** (1998) *Short tandem repeat polymorphism evolution in humans.* Eur J Hum Genet **6**(1): 38-49.
  - **Capel D.** (2000) *The battle of the sexes.* Mechanisms of Development **92**: 89-103.
  - **Capelli C, Redhead N, Romano V, Cali F, Lefranc G, Delague V, Megarbane A, Felice AE, Pascali VL, Neophytou PI, Poulli Z, Novelletto A, Malaspina P, Terrenato L, Berebbi A, Fellous M, Thomas MG, Goldstein DB.** (2005) *Population structure in the Mediterranean basin: a Y chromosome perspective.* Ann Hum Genet **70**: 207-225.
  - **Capelli C, Wilson JF, Richards M, Stumpf MP, Gratrix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM, Goldstein DB.** (2001) *A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania.* Am J Hum Genet **68**: 432-443.
  - **Cavalli-Sforza LL, Menozzi P, Piazza A.** (1994) *The history and geography of human genes.* Charter 2. Princeton, New Jersey: Princeton University Press.
  - **Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA.** (2004) *Excavating Y-chromosome haplotype strata Anatolia.* Hum Genet **114**: 127-148.
  - **Cooper G, Amos W, Hoffman D, Rubinsztein DC.** (1996) *Network analysis of human Y microsatellite haplotypes.* Hum Mol Genet **5**(11): 1759-1766.

- **Cox MP, Lahr MM.** (2006) *Y-chromosome diversity is inversely associated with language affiliation in paired Austronesian-and Papuan-speaking communities from Salomon islands.* Am J Hum Biol **18**: 35-50.
- **Crow JF.** (2006) *H.J. Muller and the “Competition Hoax”.* Genetics **173**: 511-514.
- **Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Cali F, Akar N, Richards M, Torroni A, Novelletto A, Scozzari R.** (2004) *Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and Out of Africa.* Am J Hum Genet **74**: 1014-1022.
- **Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R.** (2006) *Molecular dissection of the Y chromosome haplogroup E-M78(E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers.* Human Mutation **27**: 831-832.
- **Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Beraud Colomb E, Dugoujon JM, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradisnik B, Assun G, Brdicka R, Kozlov AI, Efremov GD, Coppa A, Novelletto A, Scozzari R.** (2007) *Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12.* Mol Biol Evol **24**(6):1300-1311.
- **Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA.** (2002) *A back migration from Asia to Sub-Sahara African is supported by high-resolution analysis of human Y-chromosome haplotypes.* Am J Hum Genet **70**: 1197-1214.

- 
- **Dai HL, Wang XD, Li YB, Wu J, Zhang J, Zang HJ, Dong JG, Hou YP.** (2004) *Characterization and haplotype analysis of 10 novel Y-STR loci in Chinese Han population.* Forensic Sci Int **145**(1): 47-55.
  - **De Knijff P.** (2000) *Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome.* Am J Hum Genet **67**: 1055-1061.
  - **De Knijff P, Kayser M, Cagliá A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, Roewer L.** (1997) *Chromosome Y microsatellites: population genetic and evolutionary aspects.* Int J Legal Med **110**(3): 134-140.
  - **Deka R, Jin L, Shriver MD, Yu ML, DeCruo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R.** (1995) *Population genetics of dinucleotide (dC-dA)<sub>n</sub>(dG-dT)<sub>n</sub> polymorphisms in world populations.* Am J Hum Genet **56**(2): 461-474.
  - **Díez A, Cabrero C, Beamonte D, Alemany J.** (1992) *Análisis del polimorfismo de AND para la determinación de la paternidad y en el seguimiento de los trasplantes de médula ósea.* Sangre **37**: 275-278.
  - **Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, Brdicka R, Barbujani G, Papola F, Ciavarella G, Cucci F, Di Stasi L, Gavrila L, Kerimova MG, Kovatchev D, Kozlov AI, Loutradis A, Mandarino V, Mammi C, Michalodimitrakis EN, Paoli G, Pappa KJ, Pedicini G, Terrenato L, Tofanelli S, Malaspina P, Novelletto A.** (2004) *Y chromosomal*

- haplogroup J as a signature of the post-neolithic colonization of Europe. Hum Genet* **115**: 357-371.
- **Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB.** (1994) *Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA* **91**: 3166-3170.
  - **Dupuy BM, Stenersen M, Egeland T, Olaisen B.** (2004) *Y-chromosomal microsatellite mutation rates : differences in mutation rate between and within loci. Hum. Mutat* **23**: 117-124.
  - **Edwards A, Civitello A, Hammond HA, Caskey CT.** (1991) *DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet* **49**: 746-756.
  - **Elias SA, Shrot SK, Nelson H, Birks HH.** (1996) *Life and ties of the Bering land bridge. Nature* **382**: 60-63.
  - **Ellis N, Goodfellow PN.** (1989) *The mammalian pseudoautosomal region. Trends Genet* **5**(12): 406-410.
  - **Excoffier L, Smouse PE, Quattro JM.** (1992) *Analisis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics* **131**: 479-491.
  - **Foster P, Röhl A, Lünemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B.** (2000) *A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet* **67**: 182-196.
  - **Freije D, Helms C, Watson MS, Donis-Keller H.** (1992) *Identification of a second Pseudoautosomal region near the Xq and Yq telomeres. Science* **258**: 1784-1787.

- 
- **Gill P, Brenner C, Brinkmann B, Budowle E, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr WR, Morling N, Olaisen B, Pascali V, Prinz M, Roewer L, Schneider PM, Sajantila A, Tyler-Smith C.** (2001) *DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-Chromosome STRs*. *Forensic Sci Int* **124**(1): 5-10.
  - **Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS.** (2006). *DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures*. *Forensic Sci Int* **160**(2-3): 90-101.
  - **Goebel T.** (2007) *The missing years for moderns humans*. *Science* **315**: 194-196.
  - **Gómez J, Carracedo A.** (2000) *The 1998-1999 collaborative exercises and proficiency testing program on DNA typing of the Spanish and Portuguese Working Group of the International Society for Forensic Genetics (GEP-ISFG)*. *Forensic Sci Int* **114**(1): 21-30.
  - **Gonzalez-Neira A, Elmoznino M, Lareu MV, Sánchez-Diz P, Gusmao L, Prinz M, Carracedo A.** (2001) *Sequence structure of 12 novel Y chromosome microsatellites and PCR amplification strategies*. *Forensic Sci Int* **122**(1): 19-26.
  - **Green RL, Roinestad IC, Boland C, Hennessy LK.** (2005) *Developmental validation of the Quantifiler™ Real-Time kits for the quantification of human nuclear DNA samples*. *J Forensic Sci* **50** (4): 809-825.
  - **Greenberg JH, Turner CG, Zegura SL** (1986) *The settlement of the Americas: A comparison of the linguistic, dental, and genetic evidence*. *Current Anthropology* **27** (5): 477-497.



- **Gusmao L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM.** (2005b) *DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis.* Int J Leg Med **120**(4): 191-200. Review.
- **Gusmao L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM.** (2006) *DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis.* Forensic Sci Int **157**(2-3): 187-197. Review.
- **Gusmao L, González-Neira A, Alves C, Lareu M, Costa S, Amorin A, Carracedo A.** (2002) *Chimpanzee homologous of human Y specific STRs. A comparative study and a proposal for nomenclature.* Forensic Sci Int **126**(2): 129-136.
- **Gusmao L, Sanchez-Diz P, Calafell L, Martín P, Alonso CA, Álvarez-Fernández F, Alves C, Borjas-Fajardo L, Bozzo WR, Bravo ML, Builes JJ, Capilla J, Carvalho M, Castillo C, Catanesi CI, Corach D, Di Leonardo AM, Espinheira R, Fagundes de Carvalho E, Farfán MJ, Figueiredo HP, Gomes I, Lojo MM, Marino M, Pinheiro MF, Pontes ML, Prieto V, Ramos-Luis E, Riancho JA, Souza-Góes AC, Santapa OA, Sumita DR, Vallejo G, Vidal Rioja L, Vide MC, Vieira da Silva CI, Whitte MR, Zabala W, Zarrabeitia MT, Alonso A, Carracedo A and Amorín A.** (2005a) *Mutation rates at Y chromosome specific microsatellites.* Human Mutation **26**(6): 520-528.
- **Hammer MF.** (1994) *A recent insertion of an Alu element on the Y-chromosome is a useful marker for human population studies.* Mol Biol Evol **11**(5): 749-761.

- 
- **Hammer MF, Chamberlain VF, Kearney VF, Stover D, Zhang G, Karafet T, Walsh B, Redd AJ.** (2005) *Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases.* Forensic Sci Int **164**(1): 45-55.
  - **Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S.** (2006) *Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes.* Hum Genet **51**: 47-58.
  - **Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL.** (2001) *Hierarchical patterns of global human Y-chromosome diversity.* Mol Biol Evol **18**(7): 1189-1203.
  - **Hammer MF, Zegura SL.** (2002) *The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions.* Annu Rev Anthropol **31**: 303-321.
  - **Hedman M, Pimenoff V, Lukka M, Sistonen P, Sajantila A.** (2004) *Analysis of 16 Y STRs loci in the Finnish population reveals a local reduction in the diversity of male lineages.* Forensic Sci Int **142**(1): 37-43.
  - **Heid CA, Stevens J, Livak KJ, Williams PM.** (1996) *Real time quantitative PCR.* Genome Research **6**: 986-994
  - **Henegariu O, Heerema NA, Dlouhy SR, Vance GH y Voght PH.** (1997) *Multiplex PCR: critical parameters and step-by-step protocol.* BioTechniques **23**(3): 504-511.
  - **Henke J, Henke L.** (1999) *Mutation rate in human microsatellites.* Am J Hum Genet **64**: 1473.

- **Heyer E, Puymirat J, Dieltjes P, Bakker E, Knijff P.** (1997) *Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees.* Human Molecular Genetics **6**: 799-803.
- **Holtkemper U, Rolf B, Hohoff C, Foster P, Brinkmann B.** (2001) *Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques.* Human Molecular Genetics **10**: 629-633.
- **Hurles ME, Nicholson J, Bosch E, Renfew C, Sykes BC, Jobling MA.** (2002) *Y chromosomal evidence for the origins of Oceanic-speaking peoples.* Genetics **160**: 289-303.
- **Hudjashov G, Kivisild T, Underhill PA, Endicott P, Sanchez JJ, Lin AA, Shen P, Oefner P, Renfrew C, Villems R, Forster P.** (2007) *Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis.* Proc. Natl. Acad. Sci. **104**: 289-303.
- **Jeffreys AJ, Wilson V, Thein SL.** (1985) *Individual-specific "fingerprints" of human DNA.* Nature **316**: 76-79.
- **Jin L, Macaubas C, Hallmayer J, Kimura A, Mignot E.** (1996) *Mutation rates varies among alleles at a microsatellite locus: Phylogenetic evidence.* Proc Natl Acad Sci USA **93**: 15285-15288.
- **Jobling MA.** (2001b) *In the name of the father: surnames and genetics.* Trends Genet **17**(6): 353-357.
- **Jobling MA.** (2001a) *Y-chromosomal SNP haplotype diversity in forensic analysis.* Forensic Sci Int **118**(2-3): 158-162.
- **Jobling MA, Pandya A, Tyler-Smith C.** (1997) *The Y chromosome in forensic analysis and paternity testing.* Int J Legal Med **110**(3): 118-124.

- 
- **Jobling MA, Tyler-Smith C.** (1995) *Fathers and sons: the Y chromosome and human evolution*. Trends Genet **11**(11): 449-456.
  - **Jobling MA, Tyler-Smith C.** (2000) *New uses for new haplotypes the human Y chromosome, disease and selection*. Trend Genet **16**: 356-362.
  - **Jobling MA, Tyler-Smith C.** (2003) *The human Y chromosome an evolutionary marker comes of age*. Nature Reviews Genetics **4**: 598-612.
  - **Johns LM, Burton RE, Thomson JA.** (2006) *Study to compare three commercial Y-STR testing kits*. International Congress Series **1288**: 192-194.
  - **Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA.** (2000) *The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data*. Am J Hum Genet **66**: 979-988.
  - **Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF.** (2008) *New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree*. Genome Research **18**: 830-838. (Material suplementario está disponible online en [www.genome.org](http://www.genome.org).)
  - **Karafet TM, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF.** (2001) *Paternal population history of East Asia: sources, patterns, and microevolutionary processes*. Am J Hum Genet **69**: 615-628.
  - **Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF.** (1999) *Ancestral Asian source(s) of New World Y-chromosomal founder haplotypes*. Am J Hum Genet **64**: 817-831.

- **Karafet TM, Zegura SL, Vuturo-Brady J, Posukh O, Osipova L, Wiebe V, Romero F, Long JC, Harihara S, Jin F, Dashnyam B, Gerelsaikhan T, Omoto K, Hammer MF.** (1997) *Y chromosome markers and trans-Bering Strait dispersals.* Am J Phy Anthropol **102**: 301-314.
- **Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill PA, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M.** (2003) *Reduced Y-Chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea.* Am J Hum Genet **72**: 281-302.
- **Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill PA, Stoneking M.** (2001a) *Independent histories of human Y chromosomes from Melanesia and Australia.* Am J Hum Genet **68**: 173-190.
- **Kayser M, Cagliá A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinni A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, de Knijff P, Roewer L.** (1997) *Evaluation of Y-chromosomal STRs: a multicenter study.* Int J Legal Med **110**(3): 125-133.
- **Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi Q, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C.** (2004) *A comprehensive survey of human Y-chromosomal microsatellites.* Am J Hum Genet **74**: 1183-1197.
- **Kayser M, Krawczak M, Excoffier L, Deltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Kniff P.** (2001b) *An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations.* Am J Hum Genet **68**: 990-1018.

- **Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger, Krawczac M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, 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.* Am J Hum Genet **66**: 1580-1588.
- **Kayser M, Sajantila A.** (2001) *Mutations at Y-STR loci: implications for paternity testing and forensic analysis.* Forensic Sci Int **118**(2-3): 116-121.
- **Ke Y, Su B, Song X, Lu D, Chen L, Li H, Qi C, Marzuki S, Deka R, Underhill P.** (2001) *African origin of modern humans in East Asia: A tale of 12,000 Y chromosomes.* Science **292**: 1151-1153.
- **Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R.** (2003) *The genetic heritage of the earliest settlers persists both in Indian tribal and Caste populations.* Am J Hum Genet **72**: 313-332.
- **Koopman P.** (1999) *Sry and Sox9: mammalian testis-determining genes.* Cellular and Molecular Life Sciences **55**: 839-856.
- **Lahn BT, Page DC.** (1999) *Four evolutionary strata on the human X chromosome.* Science **286**: 964-967.
- **Lell JT, Brown MD, Schurr TG, Sukernik RI, Starikovskaya YB, Torroni A, Moore LG, Troup GM, Wallace DC.** (1997) *Y chromosome polymorphism in Native American and Siberian populations: identification of Native American Y chromosome haplotypes.* Hum Genet **100**: 536-543.

- **Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC.** (2002) *The dual origin and siberian affinities of Native Americans Y chromosomes.* Am J Hum Genet **70**: 192-206.
- **Lessig R, Willuweit S, Krawczak M, Wu F-C, Pu C-E, Kim W, Henke L, Henke J, Miranda J, Hidding M, Benecke M, Schmitt C, Magno M, Calacal G, Delfin FC, de Ungria MCA, Elias S, Augustin C, Tun Z, Honda K, Kayser M, Gusmao L, Amorín A, Alves C, Hou Y, Keyser C, Ludes B, Klitschar M, Immel UD, Reichenpader B, Zaharova B, Roewer L.** (2003) *Asian online Y-STR Haplotype Reference Database.* Legal Medicine **5**: S160-163.
- **Levinson G, Gutman GA.** (1987) *Slipped-strand mispairing: a major mechanism for DNA sequence evolution.* Mol Bio Evol **4**(3): 203-221.
- **Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioglu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ.** (2004) *The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations.* Am J Hum Genet **74**(3): 532-544.
- **Mahtani MM, Willard HF.** (1993) *A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: implications for mechanisms of mutation at short tandem repeat loci.* Hum Mol Genet **2**(4): 431-437.
- **Mayntz-Press K, Ballantyne MS, Ballantyne J.** (2007) *Performance characteristics of commercial Y-STR multiplex systems.* J Forensic Sci **52**(5): 1025-1034.
- **Mendes Correia AA.** (1925) *O significado genealógico do "Australopithecus" e do crânio de Tabgha e o arco antropofílético indico.* Trabalhos da Sociedade portuguesa de antropologia e etnologia. Pôrto, t. II, fasc. 3.

- **Monckton DG, Wong LJ, Ashizawa T, Caskey CT.** (1995) *Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses.* Hum Mol Genet **4** (1): 1-8.
- **Muller HJ.** (1964) *The relation of recombination to mutational advance.* Mutat Res **106**: 2-9.
- **Nachman MW, Crowell SL** (2000) *Estimate of the mutation rate per nucleotide in humans.* Genetics **156**: 297-304.
- **Nasidze I, Quinque D, Dupanloup I, Rychkov S, Naumova O, Zhukova O, Stoneking M.** (2004) *Genetic evidence concerning the origins of South and North Ossetians.* Ann Hum Genet **68**: 588-599.
- **Ohta T, Kimura M.** (1973) *The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population.* Genet Res **22**: 201-204.
- **Olaisen B, Bar W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr WR, Rand S.** (1998) *DNA recommendations 1997 of the International Society for Forensic Genetics.* Vox Sanguinis **74**: 61-63.
- **Osler MV, Cheung KH, Kidd JR, Pakstis AJ, Miller PL, Kidd KK.** (2002) *ALFRED: an allele frequency database for anthropology.* American Journal of Physical Anthropology **119**: 77-88.
- **Paracchini S, Arredi B, Chlák R, Tyler-Smith C.** (2002) *Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry.* Nucleic Acids Research **30** (6): e27.



- **Pascali VL, Dobosz M, Brinkmann B.** (1999) *Coordinating Y-chromosomal STR research for the courts.* Int J Legal Medicine **112**(1): 1.
- **Pastinen T, Krug A, Metspalu A, Peltonen L, Syvänen AC.** (1997) *Minisequencing: A specific tool for DNA analysis and diagnostics on oligonucleotide arrays.* Genome Research **7**: 606-614.
- **Petrukhin KE, Speer MC, Cayanis E, Bonaldo MF, Tantravahi U, Soares MB, Fischer SG, Warburton D, Gilliam TC, Ott J.** (1993) *A microsatellite genetic linkage map of human chromosome 13.* Genomics **15**(1): 76-85.
- **Pitulko VV, Nikolsky PA, Girya EY, Basilyan AE, Tumskoy VE, Koulakov SN, Astakhov SN, Pavlova EY, Anisimov MA** (2004) *The Yana RHS site: humans in the Arctic before the last glacial maximum.* Science **303**: 52-56.
- **Prinz M, Carracedo A, Mayr WR, Morling N, Parsons TJ, Sajantila A, Scheithauer R, Schmitter, Schneider PM.** (2007) *DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI).* Forensic Sci Int: Genetics **1**: 3-12.
- **Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW.** (1999) *Population growth of human Y chromosomes: a study of Y chromosome microsatellites.* Mol Biol Evol **16**(12): 1791-1798.
- **Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, McElreavey K.** (2001) *Y-chromosome lineages trace diffusion of people and languages in Southwestern Asia.* Am J Hum Genet **6**: 537-542.
- **Rivet P.** (1943) *Les origines de l'homme américain.* Librairie Gallimard, Paris, Francia.

- **Roewer L, Arnemann J, Spurr NK, Grzeschik K-H, Epplen JT.** (1992) *Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts.* Hum Genet **89**: 389-394.
- **Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M.** (2005) *Signature of recent historical events in the European Y-chromosomal STR haplotype distribution.* Hum Genet **116**: 279-291.
- **Roewer L, Kayser M, de Knijff P, Anslinger K, Betz A, Cagliá A, Corach D, Füredi S, Henke L, Hidding M, Kärigel HJ, Lessig R, Nagy M, Pascali V, Parson W, Rolf B, Schmitt C, Szibor R, Teifel-Greding J, Krawczak M.** (2000) *A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males.* Forensic Sci Int **114**(1): 31-43.
- **Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorin A, Anslinger K, Augustin C, Betz A, Bosch E, Cagliá A, Carracedo A, Corach D, Dekairelle A-F, Dobosz T, Dupuy BM, Füredi S, Gehrig C, Gusmao L, Henke J, Henke L, Hidding M, Hohoff C, Hoste B, Jobling MA, Kärigel HJ, de Knijff P, Lessig R, Liebeherr E, Lorente M, Martínez-Jarreta B, Nievas P, Nowak M, Parson W, Pascali VL, Penacino G, Ploski R, Rolf B, Sala A, Schmidt U, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Kayser M.** (2001) *Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes.* Forensic Sci Int **118**(2-3): 106-113.
- **Rolf B, Keil W, Brinkmann B, Roewer L, Fimmers R.** (2001) *Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations.* Int J Legal Med **115**(1): 12-15.

- **Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioglu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Furedi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O. (2004) *Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe*. *Am J Hum Genet* **75**: 128-137.**
- **Rootsi S, Zhivotovsky LA, Baldovic M, Kayser M, Kutuev IA, Khusainova R, Bermisheva MA, Gubina M, Fedorova SA, Ilumäe AM, Khusnutdinova EK, Voevoda MI, Osipova LP, Stoneking M, Lin AA, Ferak V, Parik J, Kivisild T, Underhill PA, Villems R. (2007) *A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe*. *European Journal of Human Genetics* **15**: 204-211.**
- **Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorín A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côté-Real HBSM, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskis V, Lavinha J, Livshits LA, Malaspina P, María S, McElreavey K, Meitinger TA, Mikelsaar A-V, Mitchell RJ, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA. (2000) *Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language*. *Am J Hum Genet* **67**: 1526-1543.**

- **Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC.** (2003) *Abundant gene conversion between arms and palindromes in human and ape Y chromosome.* Nature **423**: 873-876.
- **Ruitberg CM, Reeder DJ, Butler JM.** (2001) *STRBase: a short tandem repeat DNA database for the human identity testing community.* Nucleic Acids Research **29**(1): 320-322.
- **Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, Morling N.** (2003) *Multiplex PCR and minisequencing of SNPs-a model with 35 Y chromosome SNPs.* Forensic Sci Int **137**(1): 74-84.
- **Sanchez JJ, Brión M, Parson W, Blanco-Verea AJ, Borsting C, Lareu M, Niederstätter H, Oberacher H, Morling N, Carracedo A.** (2004) *Duplications of the Y-chromosome specific loci P25 and 92R7 and forensic implications.* Forensic Sci Int **140**(2-3): 241-250.
- **Santos FR, Rodriguez-Delfin L, Pena SDJ, Moore J, Weiss KM.** (1996) *North and South Amerindians may have the same major founder Y chromosome haplotype.* Am J Hum Genet **58**: 1369-1370.
- **Scheinfeldt L, Friedlaender F, Friedlaender J, Latham K, Koki G, Karafet T, Hammer M, Lorenz J.** (2006) *Unexpected NRY chromosome variation in Northern island Melanesia.* Mol. Biol. Evol. **23**(8): 1628-1641.
- **Schlötterer C, Tautz D.** (1992) *Slippage synthesis of simple sequence DNA.* Nucleic Acids Res **20**: 211-215.
- **Schneider S, Roessli D, Excoffier L.** (2000) *ARLEQUIN ver. 2.000: a software for population genetics data analysis.* Genetics and Biometry Laboratory University of Geneva, Switzerland.

- **Seielstad MT, Minch E, Cavalli-Sforza LL.** (1998) *Genetic evidence for a higher female migration rate in humans.* Nature Genetics **20**: 278-280.
- **Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS.** (2003) *A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas.* Am J Hum Genet **73**: 700-705.
- **Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS.** (2004) *Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the Neolithization of Europe and later migratory events in the Mediterranean area.* Am J Hum Genet **74**: 1023-1034.
- **Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA.** (2000) *The genetic legacy of paleolithic Homo sapiens sapiens in extant europeans: A Y chromosome perspective.* Science **290**: 1155-1159.
- **Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA.** (2002) *Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny.* Am J Hum Genet **70**: 265-268.
- **Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA.** (2006) *Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists.* Am J Hum Genet **78**: 202-221.

- 
- **Shi H, Dong YL, Wen B, Xiao CJ, Underhill PA, Shen PD, Chakraborty R, Jin L, Su B.** (2005) *Y-chromosome evidence of Southern origin of the East Asian-specific haplogroup O3-M122*. *Am J Hum Genet* **77**: 408-419.
  - **Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ.** (2000) *Population genetic implications from sequence variation in four Y chromosome genes*. *Proc Natl Acad Sci* **97**: 7354-7349.
  - **Shriver MD, Jin L, Chakraborty R, Boerwinkle E.** (1993) *VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach*. *Genetics* **134**: 983-993.
  - **Shuber AP, Grondin VJ, Klinger KW.** (1995) *A simplified procedure for developing multiplex PCRs*. *Genome Research* **5**: 488-493.
  - **Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD.** (1997) *Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes*. *Molecular and cellular biology*. **17**(5): 2851-2858.
  - **Singer-Sam J, Tanguay RL, Riggs AD.** (1989) *Use of chelex to improve the PCR signal from a small number of cells*. *Amplification: a forum for PCR users* **3**:11.
  - **Skaletsky H, Kurada-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfling T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC.** (2003) *The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes*. *Nature* **423**: 825-837.

- **Slatkin M.** (1995) *A measure of population subdivision based on microsatellite allele frequencies.* Genetics **139**: 457-462.
- **Smith KD, Young KE, Talbot CC, Schmeckpeper BJ.** (1987) *Repeated DNA of the human Y chromosome.* Development **101**: 77-92.
- **Sobrino B, Brión M, Carracedo A.** (2005) *SNPs in forensic genetics: a review on SNP typing methodologies.* Forensic Sci Int **154**(2-3): 181-194.
- **Stephan W.** (1989) *Tandem repetitive noncoding DNA: forms and forces.* Mol Biol Evol **6**(2): 198-212.
- **Strand M, Prolla TA, Liskay RM, Petes TD.** (1993) *Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair.* Nature **365**: 274-276.
- **Stumpf MP, Goldstein DB.** (2001) *Genealogical and evolutionary inference with the human Y chromosome.* Science **291**: 1738-1742.
- **Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deka R.** (2000) *Polynesian origins: Insights from the Y chromosome.* Proc Natl Acad Sci **97**(15): 8225-8228.
- **Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L.** (1999) *Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age.* Am J Hum Genet **65**: 1718-1724.
- **Syvänen AC.** (2003) *Assessing genetic variation: genotyping single nucleotide polymorphisms.* Nat Rev Genet **2**: 930-942.

- **Syvänen AC, Sajantila A, Lukka M.** (1993) *Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing.* Am J Hum Genet **52**: 46-59.
- **Tarazona-Santos E, Santos FR.** (2002) *The peopling to the Americas: a second major migration?.* Am J Hum Genet **70**: 1377-1380.
- **Tautz D, Renz M.** (1984) *Simple sequences are ubiquitous repetitive components of eukaryotic genomes.* Nucleic Acids Res **12**(10): 4127-4138.
- **Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW.** (2000) *Recent common ancestry of human Y chromosome evidence from DNA sequence data.* Proc Natl Acad Sci USA **97**: 7360-7365.
- **Underhill PA, Jin L, Lin AA, Mehdi Q, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ.** (1997) *Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography.* Genome Res **7**: 996-1005.
- **Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL.** (1996) *A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history.* Proc Natl Acad Sci USA **93**: 196-200.
- **Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, PJ Oefner, Cavalli-Sforza LL.** (2001) *The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations.* Ann Hum Genet **65**: 43-62.
- **Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman**



- MW, Cavalli-Sforza LL, Oefner PJ.** (2000) *Y-chromosome sequence variation and the history of human populations*. *Nat Genet* **26**: 358-361.
- **Valdes AM, Slatkin M, Freimer NB** (1993) *Allele frequencies at microsatellite loci: the stepwise mutation model revisited*. *Genetics* **133**(3): 737-749.
  - **Venter JC y cols.** (2001) *The sequence of the human genome*. *Science* **291**: 1304-1351.
  - **Weber JL.** (1990) *Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms*. *Genomics* **7**: 524-530.
  - **Weber J, Wong C.** (1993) *Mutation of human short tandem repeats*. *Hum Mol Genet* **2**(8): 1123-1128.
  - **Weller P, Jeffreys AJ, Wilson V, Blanchetot A.** (1984) *Organization of the human myoglobin gene*. *EMBO Journal* **3**(2): 439-446.
  - **Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodner WF.** (2001) *The Eurasian Heartland: a continental perspective on Y-chromosome diversity*. *Proc Natl Acad Sci* **98**: 10244-10249.
  - **White PS, Tatum OL, Deaven LL, Longmire JL.** (1999) *New, male-specific microsatellite markers from the human Y chromosome*. *Genomics* **57**: 433-437.
  - **Wyman AR, White R.** (1980) *A highly polymorphic locus in human DNA*. *Proc Natl Acad Sci USA* **77**(11): 6754-6758.

- 
- **Xiao W, Oefner PJ.** (2001) *Denaturing high-performance liquid chromatography: a review.* Hum Mutat **17**(6): 439-474.
  - **Xiao W, Stern D, Jain M, Huber CG, Oefner PJ.** (2001) *Multiplex capillary denaturing high-performance liquid chromatography with laser-induced fluorescence detection.* Biotechniques **30**(6): 1332-1338.
  - **YCC.** (2002) *A nomenclature system for the tree of human Y-chromosomal binary haplogroups.* Genome Research **12**: 339-348.
  - **Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF.** (2004) *High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas.* Mol Biol Evol **21**(1): 164-175.
  - **Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C.** (1997) *Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis.* Am J Hum Genet **60**: 1174-1183.
  - **Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C.** (2002) *A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia.* Am J Hum Genet **71**: 466-482.
  - **Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi Q, Tyler-Smith C.** (2003) *The genetic legacy of the Mongols.* Am J Hum Genet **72**:717-721.

- **Zhivotovsky LA, Underhill PA.** (2005) *On the evolutionary mutation rate at Y-chromosome STRs: comments on paper by Di Giacomo et al.* (2004). *Hum Genet* **116**: 529-532.
- **Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L.** (2004) *The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time.* *Am J Hum Genet* **74**: 50-61.

# Glosario



## Glosario

**ADN (o Ácido Desoxiribonucleico):** macromolécula que codifica la información genética.

**Alelo:** una de las dos o más formas alternativas de un gen o secuencia de ADN en una localización específica.

**Amelogenina:** gen que codifica para proteínas que se encuentra en el esmalte dentario. Está presente en el cromosoma X e Y, se utiliza para diferenciación sexual mediante la técnica de PCR.

**Amplicon:** secuencia de ADN amplificada en una reacción específica de PCR.

**Annealing:** estado de una reacción de PCR en el cual el cebador se une a su lugar complementario en la cadena simple que sirve de molde; típicamente se lleva a cabo a temperaturas de 50°- 60°C.

**Base:** la parte informativa de una molécula nucleotídica. Las bases del ADN son adenina (A), guanina (G), timina (T) y citosina (C).

**Beringia:** es una región asiático-americana ubicada dentro del Círculo Polar Ártico, intergrada por Alaska, el extremo Este de Siberia, el estrecho de Bering, los mares de Bering y Chukchi. Además también se incluyen dentro de esta región las Islas Aleutianas y la Península de Kamchatka.

**Bottleneck:** reducción de la diversidad genética que resulta de una dramática reducción en el tamaño de población.

**Centrómero:** constricción primaria de un cromosoma, separando el brazo corto del brazo largo.

**Clado:** rama evolutiva.

**Cladograma:** árbol evolutivo que engloba los ancestros de diferentes taxones.

**Clina:** gradiente de frecuencias génicas de una región a otra.

**Conversión génica:** intercambio no recíproco de información de secuencia entre una molécula de ADN y otra.

**Cromosoma:** estructura que contiene parte del genoma nuclear; las células humanas diploides tienen 46.

**Crossing-over o Crossover:** intercambio de ADN entre cromosomas durante la meiosis.

**Deriva genética:** fluctuación aleatoria de las frecuencias de los alelos en una población finita debido a variaciones al azar en la contribución de cada individuo a la generación siguiente.

**Desnaturalización:** separación de las hebras de ADN, a menudo por calor o por tratamiento con alkali; en PCR es el primer paso de un ciclo, en el cual la temperatura es típicamente de 94°C.

**DHPLC:** método para la detección de mutaciones en el cual los heteroduplex son detectados por la alteración en su tiempo de retención en una cromatografía en columna bajo condiciones desnaturalizantes.

**Diploide:** organismo o célula que tiene dos copias del genoma.

**Efecto fundador:** reducción de la diversidad en una población fundada por un pequeño número de individuos.

**Electroferograma:** es un modo de visualizar el resultado de la electroforesis en gel o en matriz capilar. En el caso de la electroforesis capilar en un secuenciador automático se muestra la intensidad de la fluorescencia como una función del peso molecular; los picos en el análisis corresponden con una determinada longitud de onda (color) que procede de una molécula marcada específicamente y que tiene un tamaño particular.

**Electroforesis:** separación de macromoléculas usando un gradiente de voltaje; se lleva a cabo usando un gel o una matriz capilar.

**Eucromatina:** la parte del genoma que contiene ADN transcripcionalmente activo y que se encuentra en una conformación relativamente extendida

**Eustacia Glacial:** descenso en el nivel del mar generado en parte por la congelación de las aguas continentales en períodos glaciales.

**Extensión:** el paso de un ciclo de la PCR en el cual la Taq polimerasa se extiende por las nuevas cadenas de ADN desde el cebador oligonucleotídico.

**Fenotipo:** características observables de una célula u organismo

**Filogenética:** estudio de la diversidad genética a través de la construcción de árboles evolutivos.

**Filogenia:** clasificación que refleja la historia evolutiva de una especie o grupo, para ello utiliza un árbol como estructura que representa relaciones evolutivas entre un grupo de taxones.

**Gen:** parte de una molécula de ADN que codifica una proteína o una molécula de ARN funcional.

**Genética poblacional:** estudio de la diversidad genética en poblaciones y como esta cambia a través del tiempo.

**Genotipo:** combinación de estados alélicos de un grupo de marcadores polimórficos procedentes de un par de cromosomas (la combinación de dos haplotipos).

**Glaciación Würm o Wisconsin:** fue la última glaciación que conoció la tierra, comenzó hace unos 100.000 años y finalizó hace unos 12.000 años. Con ella se considera terminado el Pleistoceno.

**Haplogrupo:** Usualmente se ha aplicado a un grupo de “haplotipos” de ADN mitocondrial o del cromosoma Y que se definían por marcadores con una tasa de mutación baja, y que tienen más estabilidad filogenética que otros “haplotipos”. Para el cromosoma Y, los haplogrupos son definidos por marcadores binarios como los SNPs, mientras que los “haplotipos” son usualmente definidos por microsatélites.

**Haploide:** organismo o célula que tiene una única copia del genoma.

**Haplotipo:** combinación de estados alélicos de un grupo de marcadores polimórficos, próximos y con ausencia de recombinación, procedentes de la misma molécula de ADN.

**Heterocromatina:** segmentos del genoma transcripcionalmente inertes y altamente condensados.

**Heteroduplex:** molécula de ADN de doble cadena en la cual una cadena contiene una mutación con respecto a la otra, la cual puede ser detectada por varios métodos.

**Homoplasia:** generación del mismo estado alélico mediante vías independientes.

**Indel:** mutación o polimorfismo implicado en la inserción o delección de secuencia de ADN.

**Locus:** Posición que ocupa un gen o marcador en un cromosoma, el plural es loci

**Match:** lugar en el interior de una molécula de ADN de doble cadena donde las bases son complementarias. En genética forense se refiere a cuando dos perfiles coinciden.

**Matrilocalidad:** patrón de residencia marital en el cual el marido traslada su lugar de residencia al pueblo de su mujer.

**Máxima parsimonia:** es un método de reconstrucción filogenética basado en la búsqueda del árbol filogenético que requiera el menor número de cambios evolutivos para explicar las diferencias entre los OTUs que lo componen.



**Meiosis:** proceso nuclear en células eucariotas diploides que produce gametos con un único miembro de cada par de cromosomas homólogos por núcleo, generando así gametos haploides.

**Migración:** proceso de movimiento de una población o individuo desde un área habitada hacia otra.

**Mismatch:** lugar en el interior de una molécula de ADN de doble cadena donde las bases no son complementarias. Formado in vivo por procesos de mutación, o in vitro como un heteroduplex para detectar una secuencia diferente entre moléculas de ADN.

**Monofilético:** grupo de linajes que comparten un ancestro común.

**Muller's ratchet:** acumulación unidireccional de mutaciones deletéreas por deriva genética en secuencias no recombinantes.

**Mutación:** cualquier cambio en una secuencia de ADN.

**Neandertal (*Homo neanderthalensis*):** es una especie del género *Homo* que habitó Europa y determinadas regiones de Asia occidental aproximadamente desde hace 230.000 años hasta 29.000 años atrás, durante el Paleolítico medio.

**Neolítico:** es el período de la historia del ser humano que se sitúa aproximadamente entre los 9000 y 6000 años de nuestra era y se corresponde con la revolución agrícola.

**Nodo:** posición de un taxon dentro de una filogenia, en el final de una rama evolutiva.

**Ortólogo:** secuencias homologas que han divergido desde la separación de su ancestro común como resultado de un evento de especiación.

**OTU:** Unidad taxonómica operacional.

**Paleolítico:** es el período más largo de la historia del ser humano, se extiende desde hace 2,5 millones de años hasta hace unos 10000 años.

**Parafilético:** linaje que incluye al antepasado común de sus miembros pero no a todos los descendientes de este.

**Patrilocalidad:** patrón de residencia marital en el cual la mujer traslada su residencia al pueblo de su marido. Contrario a matrilocalidad.

**PCR:** amplificación exponencial de una secuencia específica de ADN usando primers oligonucleotídicos específicos, un protocolo de ciclos térmicos y una ADN polimerasa térmicamente activa.

**Pleistoceno:** es el período más antiguo del Cuaternario. Es una época geológica que se encuentra comprendida entre 1,8-1,6 millones y 10.000 años antes del presente.

**Polimorfismo:** La existencia de dos o más variantes (de secuencias de ADN, proteínas, cromosomas, fenotipos) con frecuencias significativas en la población. Para ADN cualquier variante de secuencia con una frecuencia  $\geq 1\%$ .

**Población:** grupo de individuos que puede ser definido en función de determinadas características compartidas, las cuales pueden ser sociales o físicas.

**Primer o Cebador:** molécula de ADN de cadena simple corta (típicamente 18-24 bases), sintetizada químicamente usualmente a partir de una secuencia específica, para usar en la PCR.

**Recombinación:** intercambio de ADN entre miembros de un mismo par cromosómico, generalmente durante la meiosis.

**RFLPs:** polimorfismo identificado por diferencias en las longitudes de sus fragmentos de restricción.

**Selección natural:** contribución diferencial de los individuos a la siguiente generación en base a sus habilidades para sobrevivir y reproducirse.

**Singleton:** un alelo o haplotipo que aparece sólo una vez dentro de la población.

**Slippage replicativo:** errores en la replicación del ADN en el cual la ADN polimerasa se desliza en un tracto repetido, permitiendo la ganancia o pérdida de una unidad repetitiva. Es el principal mecanismo de mutación en microsatélites.

**Sonda:** secuencia de ADN específico de tamaño variable, que puede ser marcada y usada para detectar secuencias blanco específicas en procesos de hibridación.

**Taxón:** unidad evolutiva de investigación.

**Temperatura de melting (T<sub>m</sub>):** temperatura a la cual los puentes de Hidrógeno que unen las dos hebras del ADN se rompen, quedando el ADN en forma de cadena simple. El valor de esta temperatura depende de la longitud y de la composición nucleotídica de la molécula de ADN.

**Transición:** sustitución de una base, en la que una base pirimidínica (C o T) es intercambiada por otra pirimidínica, o una base purínica (A o G) es intercambiada por otra purínica.

**Transversión:** sustitución de una base, en la que una base pirimidínica es sustituida por una purínica o viceversa.

**Variante alélica:** término usado sin excesivo rigor para referirnos a un polimorfismo raro o mutación. Técnicamente implica un cambio con respecto al tipo salvaje.