USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

INSTITUTO DE MEDICINA LEGAL
Facultade de Medicina

Estudio multidisciplinar de la variabilidad del ADN mitocondrial en poblaciones humanas

VANESA ÁLVAREZ IGLESIAS
2008

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FACULTADE DE MEDICINA

INSTITUTO DE MEDICINA LEGAL

ESTUDIO MULTIDISCIPLINAR DE LA VARIABILIDAD DEL ADN MITOCONDRIAL EN POBLACIONES HUMANAS

Memoria que presenta para optar al grado de doctor:

Vanesa Álvarez Iglesias

Santiago de Compostela, Mayo de 2008

El Doctor D. Ángel Carracedo Álvarez, Catedrático de Medicina Legal de la Facultad de Medicina de la Universidad de Santiago de Compostela, y el Doctor D. Antonio Salas Ellacuriaga, Investigador Ramón y Cajal en la Universidad de Santiago de Compostela,

CERTIFICAN:

Que la presente memoria que lleva por título **"Estudio multidisciplinar de la variabilidad del ADN mitocondrial en poblaciones humanas",** de la licenciada en Biología por la Universidad de Santiago de Compostela *Vanesa Álvarez Iglesias*, 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 19 de Mayo de 2008.

Prof. Dr. Ángel Carracedo Álvarez

Dr. Antonio Salas Ellacuriaga

Vanesa Álvarez Iglesias

AGRADECEMENTOS:

Ao final, vai resultar que o apartado máis difícil de escribir dunha tese de doutoramento é o apartado de agradecementos, xa que despois de tanto tempo son moitos os que contribuiron de moitos xeitos distintos ao traballo que se presenta aqui, asi que espero non esquecer a ninguén.

Gustaria-me comenzar agradecendo aos meus directores de tese Ángel Carracedo e Antonio Salas. Grazas Ángel por terme dado a oportunidade de formar parte deste gran grupo e apoiarme todo este tempo. Toño, pensei moitas frases pero creo que o que mellor resume estes anos é dicir sinxelamente que, sen ti este traballo hoxe non existiria.

Quero agradecer tamén a todas as persoas que forman parte do Insitituto de Medicina Legal e o Departamento de Anatomía Patolóxica e Ciéncias Forenses.

E como non podía ser doutra maneira, especial mención para todos os meus compañeiros de laboratório: grazas á Maviki polo apoio, o interese e a confianza durante estes anos. Grazas ás que xa estabades aquí cando cheguei: Meli, Mari, Paula, Bea Sobrino e Bea Quintáns, din os primeiros pasos con todas vós, grazas porque aínda hoxe sego atopando respostas cando vos dou a lata, e como non, grazas pola vosa amizade.

Tamén aos que chegastes despois: Chris, Álex, Fonde, Raquel, Nuria, Eva, Francesca, María Cerezo, Manuel, Cata, Ana (grazas porque dende que chega-ches a mitocóndrias teñen outra cor e son moito máis divertidas),Luis, Ana Freire, Rocío, Alberto, Yarimar, Ángela e María Dosil. Espero non ter esquecido a ninguén…

Grazas aos que pasastes e pasades por aquí de cando en vez enchendo de voces distintas as paredes deste laboratório.

E como non, grazas aos meus amigos e a miña família, por axudarme, apoiarme e ser os incondicionais en todas as situacións da miña vida:

Ás miñas duas avoas. Á que non está, por ensinarnos o que é a forza e a enerxia e porque non podo evitar un sorriso cando penso a ledicia que lle supoñian todos os meus proxectos, e á outra por ser como é e porque a palabra dozura está resumida na sua persoa.

Á miña família de Ourense e, como non, á miña família de Canabal.

A Emi e a Patri, porque con vos, o parentesco nunca está claro : ¿família de primeiro ou de cuarto grado? ¿amiga ou xa família? Nunca está claro onde encuadrar-vos e supoño que iso xa o di todo. Grazas polo voso cariño e por estar sempre, nos momentos divertidos e nos momentos de menos ledícia.

Grazas a todos os meus amigos.

E por último, e neste caso aos máis importantes, grazas Mario por acompañarme todos estes anos e porque non teño palabras que resuman o ánimo o apoio e o cariño de todo este tempo. E como non moitísimas grazas aos meus pais e a miña irmá Alba por ser o verdadeiro motor destes anos, polo apoio e a incondicionalidade. Se a alguén lle debo ter chegado até aquí é a vos.

AOS MEUS PAIS

ÍNDICE

ABREVIATURAS

A: Adenina

ADN/DNA: Ácido desoxirribonucleico

ADNmt: ADN mitocondrial

ADNn: ADN nuclear

ARN: Ácido ribonucleico

BSA: Bovine Serum Albumine (Albúmina de suero bovino)

C: Citosina

CRS: Cambridge Reference Sequence (Secuencia de Referencia de Cambridge)

ddNTP: Di-deoxinucleótido trifosfato

dNTP: Deoxinucleótido trifosfato

DTT: Ditotreitol

EDNAP group: Eureopean DNA profiling group

G: Guanina

Hg: Haplogrupo

HVI: Hypervariable region I (Región hipervariable I)

HVII: Hypervariable region II (Región hipervariable II)

ISFG: International Society for Forensic Genetics (Sociedad Internacional de Genética Forense)

IUB: International code for Unresolved Bases

kb: kilobase

ml: Mililitro

mM: Milimolar

ng: Nanogramo

nM: Nanomolar

Pb: Pares de bases

PCR: Polymerase Chain Reaction (reacción en cadena de la polimerasa)

pmol: picomol

rCRS: revised Cambridge Reference Sequence (Secuencia de Referencia de Cambridge revisada)

RFLP: Restriction Fragment Lenght Polymorphism (Polimorfismos basados en la longitud de los fragmentos de restricción)

r.p.m.: Revoluciones por minuto

SDS: Dodecil sulfato de sodio

SNP: Single Nucleotide Polymorphism (polimorfismo nucleotídico único)

SSCP: Single Strand Conformational Polymorphism (Polimorfismo conformacional de cadena única)

STR: Short Tandem repeat (Repetición en tandem corta)

T: Timina

µg: Microgramo

µl: Microlitro

µM: Micromolar

INTRODUCCIÓN

1. La mitocondria: generalidades

Las mitocondrias fueron descubiertas en el siglo XIX, son uno de los orgánulos celulares más conspicuos del citoplasma y tienen como función principal la producción de la energía que necesita la célula. Están presentes y repartidas de modo uniforme en todas las células, tanto vegetales como animales y se hallan en continuo movimiento.

Pueden ser observadas al microscopio y varían considerablemente en tamaño y forma dependiendo del tipo y estado metabólico de la célula, pero en general, son elipsoides de aproximadamente 0.5µm de diámetro y 10µm de largo. El número de mitocondrias de una célula depende de la función de ésta, así una célula eucarionte típica contiene más de 2000 mitocondrias, lo que ocupa alrededor de la quinta parte del volumen celular; esta cantidad es necesaria por ser la central energética de la célula. Si bien las células con demandas de energía particularmente elevadas, como las musculares, tienen muchas más mitocondrias ya que desarrollan trabajos intensos, frente a células poco activas, como por ejemplo las epiteliales.

2.1. Estructura de las mitocondrias

Las mitocondrias están delimitadas por una doble membrana: la primera se denomina *membrana mitocondrial externa***.** Ésta es lisa y contiene porinas, proteínas que forman poros no específicos, lo que permite el libre paso de moléculas mayores de 10 kD entre la mitocondria y el citosol.

A continuación se encuentra la *membrana mitocondrial interna***,** está compuesta por 75% de proteínas y 25% lípidos, por lo que es solamente permeable a O_2, CO_2 y H_2O. Además presenta numerosas invaginaciones o pliegues que se denominan *crestas mitocondriales*, su

número varía con la actividad respiratoria de cada tipo particular de célula, debido a que las enzimas que llevan a cabo el transporte de electrones y la fosforilación oxidativa están embebidas en esta membrana.

Como consecuencia de la presencia de esta doble bicapa lipídica en el interior de las mitocondrias se diferencian dos espacios. El más externo es el *espacio intermembrana*, que es el espacio que delimita la dos membranas mitocondriales, su importancia radica en que permite una concentración de protones en contra de gradiente electroquímico, esencial para la síntesis de ATP.

Por otro lado, está la *matriz mitocondrial*, delimitada exclusivamente por la membrana mitocondrial interna. La matriz es de carácter gelatinoso, formada casi por un 50% de agua, contiene una elevada concentración de enzimas solubles para el metabolismo oxidativo, así como substratos, cofactores e iones inorgánicos. También contiene la maquinaria genética propia del organelo, ADN, ARN y ribosomas que participan en la síntesis de algunos componentes proteicos mitocondriales, el resto de proteínas son importadas del citoplasma después de su expresión desde el núcleo.

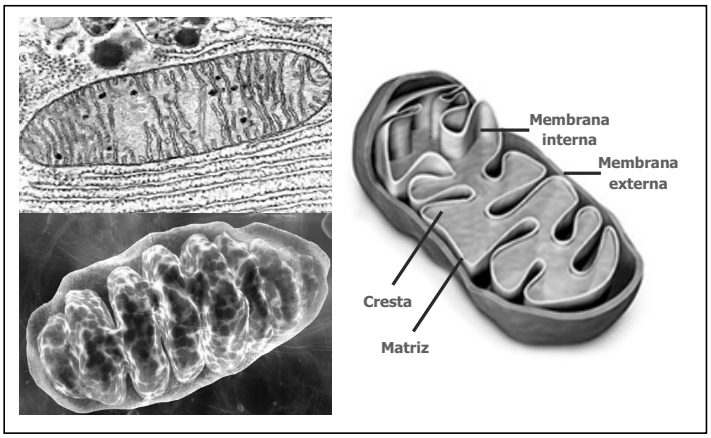


Figura 1: Morfología y estructura de la mitocondria.

2.2. Origen de las mitocondrias: la teoría endosimbionte

Las mitocondrias tienen muchas semejanzas con las bacterias. Se pueden destacar: la forma y tamaño, la presencia de ADN bicatenario y circular en la matriz mitocondrial, su reproducción por fisión binaria, así como el proceso respiratorio que llevan a cabo, el cual es muy semejante al que se observa en bacterias aerobias actuales. Estas pruebas llevaron a Lynn Margulis a proponer en 1976 la teoría endosimbionte, que permite explicar el origen de los orgánulos de doble membrana: mitocondrias y cloroplastos, así como explicar el hecho de que no participen en el tráfico vesicular como el resto de orgánulos de membrana simple. Según esta teoría el origen de las mitocondrias modernas sería a partir de la endosimbiosis de bacterias aerobias con antiguos procariotas o eucariotas anaerobios, esto habría ocurrido hace aproximadamente 1500-700 millones de años. Los nutrientes abastecidos por la célula anaerobia y consumidos por la bacteria, fueron presumiblemente reembolsados con creces a la primera, debido a la alta eficiencia metabólica oxidativa de la segunda. Este mutuo beneficio hizo que la célula invasora llegara a formar parte integral del organismo mayor, acabando por convertirse en parte de ella: la mitocondria. A lo largo de la historia común la mayor parte de los genes mitocondriales han sido transferidos al núcleo, de tal manera que la mitocondria no es viable fuera de la célula huésped y ésta no puede serlo sin las mitocondrias.

2.3. Funciones de las mitocondrias

Todas las funciones celulares dependen de un aporte continuo de energía obtenido a partir de la degradación de moléculas orgánicas durante el proceso de respiración celular. La energía liberada durante este proceso se almacena finalmente en forma de moléculas de ATP, que constituyen una reserva de energía rápidamente disponible para todas las funciones metabólicas celulares. Los principales sustratos para la

respiración celular son los azúcares simples y los lípidos, sobre todo glucosa y ácidos grasos.

La mitocondria realiza la mayor parte de las oxidaciones que tienen lugar en la célula y genera la mayor parte de la energía, aproximadamente un 90% en forma de ATP, que se produce en las células animales. Las mitocondrias son orgánulos transformadores de energía, obtienen grandes cantidades de ATP a partir de glucosa almacenada en forma de glucógeno y ácidos grasos almacenados en forma de grasas principalmente. Los pasos iniciales del metabolismo, tanto del glucógeno como de las grasas, se llevan a cabo en el citosol, y son sus productos los que son transportados al interior de las mitocondrias para continuar con su oxidación (Alberts *et al.* 1996).

Así en la matriz mitocondrial se oxidan metabolitos en rutas como el ciclo de los ácidos tricarboxílicos, también denominado ciclo del ácido cítrico o de Krebs (en honor a su descubridor); o la β-oxidación de ácidos grasos. En estas oxidaciones se obtienen además de un número reducido de ATPs, coenzimas reducidos como el NADH o el $FADH_2$, los cuales ceden sus electrones a la cadena de trasporte electrónico, en la cual a través de una serie de reacciones redox y gracias al acoplamiento con la fosforilación oxidativa permite obtener la energía necesaria para la síntesis de ATP.

De forma más detallada, la respiración celular de la glucosa (glucolisis) se inicia en el citosol, donde es degradada parcialmente hasta formar ácido pirúvico, produciendo una pequeña cantidad de ATP. Luego el ácido pirúvico pasa al interior de las mitocondrias donde, tras su transformación en acetil CoA, se incorpora al ciclo de Krebs. En contraste, los ácidos grasos pasan directamente al interior de las mitocondrias tras un proceso de activación en el citosol, ya en la matriz mitocondrial se transforman en acetil CoA a través del proceso conocido como β-oxidación donde también se obtiene una cantidad limitada de

$FADH_2$ y NADH. Cada acetil CoA en el ciclo de Krebs producirá NADH, $FADH_2$, y una pequeña cantidad de GTP que en términos energéticos es equivalente al ATP.

El NADH y $FADH_2$, son las moléculas donadoras de electrones en la cadena de transporte electrónico, ya que como se mencionó anteriormente fueron reducidas durante la glucolisis, el ciclo del ácido cítrico, y la β-oxidación. La cadena de transporte electrónico implica a numerosas enzimas como la NADH-Q reductasa, la citocromo c oxidasa y la citocromo reductasa, que se localizan en la membrana mitocondrial interna. Los electrones de estos dos donadores a través de la cadena de electrones llegan a través de una secuencia de reacciones redox hasta el oxígeno, el cual se reduce para formar agua. Las enzimas que catalizan estas reacciones y que forman la cadena de transporte de electrones tienen la remarcable capacidad de crear simultáneamente un gradiente de protones a través de la membrana mitocondrial interna entre el espacio intermembranoso (donde se concentran protones) y la matriz mitocondrial.

Así la misión de la cadena transportadora de electrones es la de crear un gradiente electroquímico que se utiliza para la síntesis de ATP en el proceso conocido como fosforilación oxidativa. De esta forma podemos deducir la existencia de tres procesos totalmente dependientes:

- Un flujo de electrones desde sustancias individuales.
- Un uso de la energía desprendida de ese flujo de electrones que se utiliza para la translocación de protones en contra de gradiente, por lo que energéticamente estamos hablando de un proceso desfavorable.
- Un uso de ese gradiente electroquímico para la formación de ATP mediante un proceso favorable desde un punto de vista energético.

La hipótesis del acoplamiento quimiosmótico, le valió el premio Nobel de química a Peter D. Mitchell. Ésta explica que la cadena de transporte de electrones y la fosforilación oxidativa están acopladas por el gradiente de protones. El flujo de protones crea un gradiente de pH y un gradiente electroquímico a través de la membrana mitocondrial interna entre el espacio intermembrana y la matriz mitocondrial. Así los protones regresan desde el espacio intermembrana a la matriz mitocondrial a favor de gradiente. Si bien este paso es a través de la ATP sintetasa, la cual actúa como un canal de iones que "devuelve" los protones a la matriz mitocondrial. Durante esta vuelta, la energía es utilizada para la síntesis de ATP, catalizada por el componente F_1 del complejo F_OF_1 ATP sintetasa.

En resumen, debido a que los protones se han bombeado al espacio intermembrana de la mitocondria en contra de gradiente, ahora pueden fluir nuevamente hacia dentro de la matriz mitocondrial y mediante la vía ATP-sintasa, se genera ATP en el proceso. La reacción es:

$$ADP^{3-} + H^+ + P_i \leftrightarrow ATP^{4-} + H_2O$$

Así cada molécula de NADH contribuye a formar entre 2 y 3 moléculas de ATP, mientras que cada $FADH_2$ contribuye a un máximo de 2 moléculas de ATP. Por lo que el balance energético total a partir de una glucosa, en el metabolismo oxidativo, será de 36 ó 38 ATPs en función del tipo de célula ya que usan lanzaderas distintas para el paso del NADH citosólico al interior mitocondrial.

También se pueden citar otras funciones de las mitocondrias como son: la producción de precursores para la síntesis de diversas sustancias como aminoácidos, ácidos grasos, glucosa, etc., así como la síntesis del 10% de proteínas mitocondriales, ya que el resto de proteínas de la mitocondria se codifican por el ADN nuclear y se sintetizan en los ribosomas libres del citosol.

Uno de los últimos descubrimientos en relación a la biología mitocondrial, es el hecho de que este orgánulo juegue un papel importante en la apoptosis celular, un proceso biológico fundamental por el cual la célula muere de una forma programada. Se trata de un colapso celular que comienza con la formación de burbujas en la membrana plasmática y conlleva la fragmentación de proteínas, condensación de la cromatina, disminución del volumen celular y la degradación del ADN. El proceso termina con la formación de cuerpos apoptóticos (fragmentos celulares rodeados de membrana) que evitan la salida de todos estos compuestos al exterior y posible daño a células colindantes. Estos cuerpos apoptóticos son rápidamente fagocitados por macrófagos circulantes evitando la respuesta inflamatoria.

Hay dos vías de activación de la apoptosis:

- ***Vía extrínseca***: en la que intervienen receptores de muerte, como son el TFN, FAS, FADD,..., que se encuentran situados en la membrana citoplasmática y que tras la unión con sus ligandos específicos, llevan directamente a la activación de las caspasas.
- ***Vía intrínseca***: esta es la ruta en la que la mitocondria juega un papel importante. Por diferentes estímulos, el citocromo c oxidado y otras proteínas se liberan desde la mitocondria al citosol, y esta señalización es llevada a cabo por la proteína p53. Una vez fuera de la mitocondria, el citocromo c se une fuertemente a Apaf1, y a ellas se les une la caspasa 9. De esta manera se forma lo que se denomina el apostosoma, que corta la procaspasa 3, e iniciadora final de la apoptosis (Wang *et al.* 2001). Esta vía se puede activar por estrés oxidativo o por daño en el ADN.

Con independencia de la ruta de activación, el proceso culmina con la activación de las caspasas efectoras, que junto con otros ejecutores independientes llevan a la destrucción celular.

2.4. División de las mitocondrias y segregación mitótica

Las mitocondrias no se sintetizan nunca *de novo*, sino que siempre surgen por crecimiento y división de otras ya existentes. Cada orgánulo ha de duplicar su masa y luego dividirse por la mitad un promedio de una vez por cada generación celular, y para ello han de importar lípidos, nucleótidos y proteínas desde el citosol. Se he demostrado que el núcleo juega un papel esencial en la biogénesis mitocondrial la generación de la mitocondria está altamente influenciada por eventos de transducción de señales extra-mitocondriales.

En la mayoría de las células, los orgánulos transformadores de energía se dividen durante la interfase y replican su ADN durante todo el ciclo celular, es decir, que presentan un desfase con respecto a estos mismos procesos en las células.

El número de orgánulos que se presentan en cada célula puede estar regulado de acuerdo al tipo celular y a sus necesidades energéticas en cada momento.

Durante la división celular, las mitocondrias se distribuyen de forma aleatoria entre las células hijas. Si en una célula que se va a dividir existen moléculas de ADNmt diferentes, la proporción de moléculas "mutantes" que pasarán a las células hijas puede variar mucho, y en consecuencia, también el fenotipo. Este fenómeno se conoce como *segregación mitótica,* y puede tener muchas consecuencias, especialmente desde un punto de vista clínico (Dimauro *et al.* 2005; Taylor *et al.* 2005).

2. El ADN mitocondrial

2.1. Características generales

Una de las principales características de las mitocondrias es la de poseer un sistema genético propio y la maquinaria necesaria para sintetizar el ADN, ARN y las proteínas que codifica. Cada mitocondria contiene entre 2 y 10 copias de la molécula de ADN y a su vez cada célula puede contener cientos de mitocondrias, con lo que el número de copias de este genoma en una célula oscilará entre 1000 y 10 000 dependiendo de cada órgano y tejido.

El ADN mitocondrial presenta toda una serie de características propias que lo diferencian del ADN nuclear, que se resumen en la siguiente tabla.

	GENOMA NUCLEAR	**GENOMA MITOCONDRIAL**
Tamaño aproximado	3000 MB	16.6 MB
Número de moléculas distintas	23 en mujeres 24 en hombres	1
Número de copias por célula diploide	2	>1000
Tipo de molécula	Lineal	Circular
Proteínas asociadas	Histonas y proteínas no histonas	Libre de proteínas
Número de genes	22000	37
Densidad génica	~1/40 kb	1/0.45 kb (alta)
ADN repetitivo	~ 20%	Casi no existe
ADN codificante	~2%	90%
Recombinación	>1 por cromosoma y meiosis	No
Tasa de mutación	Baja	Alta
Herencia	Mendeliana, con la excepción del cromosoma Y	Monoparental materna

Tabla 1: Principales diferencias entre el ADN nuclear y el mitocondrial.

2.2. Estructura y organización del ADN mitocondrial

El ADN mitocondrial (ADNmt) al igual que los bacterianos, es una molécula circular, cerrada y de doble cadena. En los seres humanos tiene un tamaño aproximado de 16569 pares de bases mientras que en levaduras contiene cerca de 80000 pares de bases y en plantas varía entre 100 y 2000 kb. La secuencia completa y la organización del ADN mitocondrial humano fue publicada por primera vez en 1981 por Anderson *et al.* y posteriormente revisada por Andrews *et al* en 1999.

Las dos cadenas complementarias tienen una proporción de C y G muy diferente, por lo que tienen un peso molecular muy distinto, existe una hebra rica en bases púricas que es la cadena pesada o H (*heavy*) y la cadena complementaria rica en pririmidinas conocida como cadena ligera o L (*light*). La cadena que se representa por convenio y sobre la que se realiza la numeración del genoma mitocondrial es la cadena L de 1 a 16569.

El ADN mitocondrial presenta una gran economía y su información está muy compactada ya que sus genes no presentan intrones, se disponen uno a continuación de otro sin tramos intermedios no codificantes y algunos incluso llegan a solaparse como los ATPasa6 y ATPasa 8. Los 37 genes que contiene están dentro de una gran *región codificante* que representa el 90% del total de la molécula, 28 de estos genes están codificados por la cadena H y nueve por la L.

De los 37 genes, 22 codifican para ARNt, 2 para ARNr (12S y 16S) y 13 codifican para ARNs mensajeros y por lo tanto para 13 polipéptidos implicados en la cadena de fosforilación oxidativa (OXPHOS): 7 de las 48 subunidades de la NADH deshidrogenasa, para el complejo I, un citocromo b del complejo III, 3 de las 13 subunidades de la citocromo oxidasa c (COX) del complejo IV y 2 ATPasas (ATPasa 6 y 8) del complejo V.

Las demás proteínas necesarias en la mitocondria están codificadas por genes nucleares, se sintetizan en el citoplasma y son transportadas posteriormente al interior de la mitocondria. Se trata tanto de proteínas implicadas en el proceso de la fosforilación oxidativa como de ADN y ARN polimerasas y algunas proteínas ribosómicas mitocondriales.

El 10% restante de la molécula es ADN no codificante, contiene aproximadamente 1.2 kb y es conocida como *región control*. Contiene fundamentalmente los promotores de la transcripción de ambas cadenas, el origen de replicación de la cadena pesada y secuencias de regulación (Anderson *et al.* 1981). También contiene lugares de unión de factores de transcripción, secuencias conservadas relacionadas con el inicio de la replicación de la cadena H y secuencias asociadas a la finalización del bucle de desplazamiento o *D-loop.* El *D-loop* es una región de aproximadamente 1100 pares de bases, visible bajo microscopía electrónica, que se genera durante la replicación de la cadena H por la síntesis de un corto segmento de la hebra pesada, ADN 7S, que permanece unido al ADN molde, desplazando la cadena H y formando una triple hélice (Crews *et al.*1979).

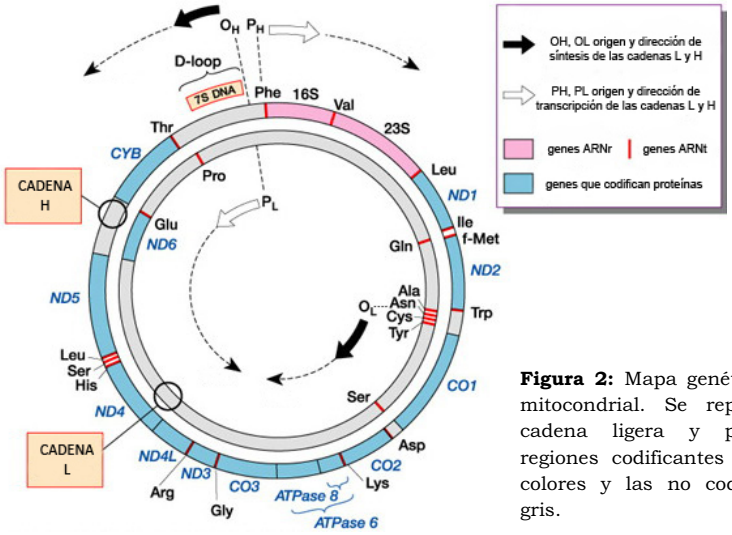


Figura 2: Mapa genético del ADN mitocondrial. Se representan la cadena ligera y pesada. Las regiones codificantes en distintos colores y las no codificantes en gris.

La región control es muy polimórfica, presenta dos regiones hipervariables bien caracterizadas, definidas por Vigilant *et al.* en 1989 y conocidas como la *región hipervariable I* (HVI) y la *región hipervariable II* (HVII). La región control se extiende desde la posición 16024 hasta la 16569 y desde la 1 hasta la 576.

Otra característica especial del ADN mitocondrial es que el código genético está alterado respecto al universal ya que 4 de los 64 codones presentan un significado diferente al que posee para la síntesis citoplasmática (Tabla 2). El hecho de que el código genético mitocondrial es casi el mismo en todos los organismos proporciona fuertes evidencias de que todas las células eucariotas evolucionaron a partir de un antepasado común.

Codón	Código Universal	Código mitocondrial
UGA	STOP	Trp
AUA	Ile	Met
AGA AGG	Arg	STOP

Tabla 2: Diferencias entre el código genético universal y mitocondrial.

2.3. Herencia del ADN mitocondrial humano

El ADNmt se hereda con un patrón no mendeliano, durante muchos años, el hecho de que el ADNmt (y las mitocondrias) presentaran una herencia estrictamente materna y ausente de recombinación se ha considerado un dogma, a pesar de que en otros organismos como plantas, hongos, mejillones o peces sí que se produce recombinación del ADN mitocondrial.

Giles *et al.* en 1980 analizando polimorfismos de secuencia mediante RFLPs, observaron que la progenie presentaba siempre la variante materna. Estudios posteriores confirmaron que las mitocondrias del esperma son selectivamente destruidas en el oocito y que el ADNmt paterno es marcado mediante ubiquitinación durante la

espermatogénesis para su destrucción posterior en el oocito (Sutovsky *et al.* 1999, 2000).

Estudios recientes han puesto en duda este dogma, Schwartz *et al.* (2003) observaron en un caso de intolerancia severa al ejercicio que el ADNmt de este paciente era de origen predominantemente paterno. Zsurka *et al* (2005, 2007) llevaron a cabo estudios similares en familias en las que varios miembros presentaban heteroplasmías en músculo esquelético; Kraytsberg *et al.* (2004) encontraron recombinación entre el genoma materno y el paterno de un paciente en aproximadamente el 0.7% del ADNmt total de su tejido muscular. Sin embargo, todos estos estudios no explican por qué no aparecen señales de recombinación en la filogenia del ADNmt o por qué la mayoría de las heteroplasmías ocurren generalmente en solo una o dos posiciones en un perfil en lugar de en más posiciones, lo que sería incompatible con un modelo de contribución mitocondrial patrilineal (Bandelt et al. 2004, 2005). El reanálisis de muchos estudios en los que se habían observado evidencias de recombinación, permitió comprobar errores en los datos así como en los métodos estadísticos empleados (Pakendorf *et al.* 2005) lo que provocó la rectificación de algunos autores (Hagelberg *et al.* 1999, 2000), en otros muchos trabajos el mosaicismo del ADNmt podría deberse a casos de contaminación o mezcla de distintas muestras y no por fenómenos de recombinación (Bandelt *et al.* 2005).

Por lo tanto, parece muy probable que el único caso de herencia paterna del ADNmt detectada en humanos represente un fallo en el normal reconocimiento y eliminación de las moléculas del genoma mitocondrial paterno, lo que es un fenómeno extremadamente raro. De esta forma, al menos por el momento, la herencia materna del ADNmt aún puede ser considerada como la norma, de manera que todos los individuos de un mismo linaje materno exhiben la misma secuencia de ADNmt (salvo en casos excepcionales en donde exista segregación de heteroplasmías en algún individuo del linaje, o mutaciones puntuales a

nivel germinal). Este modo de herencia uniparental y su naturaleza haploide son una de las grandes ventajas del ADNmt, ya que esto permite a los investigadores seguir los linajes hacia atrás en el tiempo, buscando la ascendencia materna de una población, evitando así los efectos de la herencia biparental y la recombinación, que provocan confusión en esta búsqueda (Pakendorf *et al.* 2005).

2.4. Tasa de mutación del ADN mitocondrial

La tasa de mutación depende de la velocidad a la que surgen y se fijan las mutaciones en los linajes, permite introducir una escala temporal en la evolución molecular y así hacer estimas temporales sobre el ancestro común más reciente entre dos linajes, es decir, estimar el tiempo de divergencia entre ellos.

Estudios comparativos entre las secuencias de ADN de diferentes organismos revelaron que la velocidad de substitución nucleotídica durante la evolución ha sido de 5 a 10 veces mayor en los genomas mitocondriales que en los nucleares (Brown *et al.* 1979), lo cual posiblemente se deba a la reducida fidelidad de los sistemas de replicación y/o reparación en el ADNmt. Dentro de la mitocondria se consume más del 90% del oxígeno que entra en la célula para proporcionar la mayor parte de la energía que necesita la célula, esta actividad metabólica va a generar gran cantidad de radicales libres con efecto mutagénico. Frente a este daño oxidativo la molécula de ADNmt no presenta histonas protectoras y además la ADNpolimerasa mitocondrial presenta una escasa actividad correctora con respecto a la nuclear.

La tasa de evolución está en función de dos procesos: la frecuencia con la surgen nuevas mutaciones en la molécula y la probabilidad de que estas nuevas mutaciones se fijen en la población (tasa de fijación). El mecanismo por el cual se produce una mutación y se fija en la línea materna es importante no sólo en la investigación de

enfermedades mitocondriales, sino también en el estudio de las poblaciones humanas (estimas de diversidad de secuencias de ADNmt se utilizan para datar eventos demográficos), y en identificación y casuística criminal en genética forense.

La tasa de mutación estimada de la región codificante es de 2-4% por millón de años (Cann *et al.* 1987; Stoneking *et al.* 1990) mitocondrial, excepto la región control, cuya tasa es superior, aunque es difícil dar una cifra para esta tasa de mutación ya que es un tema muy controvertido y no existe un acuerdo al respecto (Pakendorf *et al.* 2005). Para estimar la tasa de mutación del ADNmt pueden utilizarse aproximaciones distintas: a través del estudio de pedigríes o a través del estudio de la filogenia, en función del sistema utilizado, diferentes autores han obtenido estimas distintas. Estudios realizados a partir de estudios familiares aportan valores bastante diferentes entre sí: Howell *et al.* (1996) obtienen una tasa de $1.08x10^{-5}$ mutación por sitio y generación, mientras que Bendall *et al.* (1996) sitúan esta cifra entre $1.2x10^{-6}$ y $2.7x10^{-5}$ y Parsons *et al.* (1997) en $2.7x10^{-5}$. Comparaciones filogenéticas, basadas a su vez en comparaciones interespecíficas o intraespecíficas, proporcionan estimas más similares entre sí e inferiores a las obtenidas por observaciones directas de mutaciones del ADNmt en familias (Ward *et al.* 1991; Stoneking *et al.* 1992). Ahora el debate se centra en hasta qué punto esto refleja el estado real de los hechos y cuál de ellas usar para estudios de historia poblacional (Pakendorf *et al.* 2005).

Para poder entender estas diferencias entre las estimas basadas en pedigríes y las basadas en estudios filogenéticos es importante considerar que el nivel de estabilidad molecular varía para diferentes sitios a lo largo de la molécula de ADNmt, existen posiciones dentro de la región control conocidas como “hot spots” o “puntos calientes de mutación” cuya tasa de mutación es 4 o 5 veces mayor que la media (Pakendorf *et al.* 2005). Las estimas generacionales abarcan periodos

muy pequeños por lo que puede que las posiciones elegidas sean de este tipo y se sobreestime la tasa de mutación, por el contrario los estudios filogenéticos abarcan periodos evolutivos más amplios y las estimas obtenidas reflejarían valores promedio de todas las posiciones variables de la molécula de ADNmt a lo largo de un linaje, esta estima sería correcta si estos estudios tuviesen en cuenta la tasa de mutación recurrente de los puntos calientes (Sigurgardottir *et al.* 2000). Aunque es importante estimar la tasa de mutación del ADNmt, desde un punto de vista práctico es mucho más relevante tener información sobre las tasas de mutación sitio-específicas.

2.5. Polimorfismos del ADN mitocondrial

El término polimorfismo fue descrito por Ford en 1940 como la presencia en una población de dos o más formas alélicas discretas, de las que la de menor frecuencia no puede mantenerse sólo por mutación recurrente. En términos generales se asume que un locus es polimórfico cuando el alelo más común para este locus tiene una frecuencia inferior al 99%.

Los polimorfismos están distribuidos a lo largo de todo el genoma, y se pueden clasificar como:

- *Polimorfismos de longitud:* producidos por inserciones o delecciones de uno o más nucleótidos. En el ADNmt existen tractos homopoliméricos de una o más bases nucleotídicas. Desde un punto de vista formal, se engloban dentro del grupo de los polimorfismos de longitud aunque desde un punto de vista biológico nada tienen que ver con los polimorfismos que caracterizan a los microsatélites y minisatélites. Estos tractos están localizados en las posiciones 16184-16188/16190-16193, 303-309/311-315 y 568-573 de la región control, con una frecuencia bastante elevada coexisten moléculas con diferente número de inserciones dentro de una misma

mitocondria o en mitocondrias distintas de la misma célula, situación que se conoce como heteroplasmía.

En el ADNmt sólo se ha descrito un microsatélite corto (Bodenteich *et al.* 1992) entre las posiciones 514 y 523 de la región D-loop del genoma mitocondrial. Se trata de un microsatélite dinucleotídico (AC)n con un bajo grado de heterocigosidad. Bodenteich *et al.* 1992 en un estudio de sus frecuencias alélicas en una muestra de 68 individuos sólo encuentran cuatro alelos con una distribución de frecuencias alélicas muy heterogénea, de la misma manera Szibor *et al.* 2007 en un estudio de 2458 individuos sólo encuentran 5 alelos de $(AC)_3$ a $(AC)_7$, y al igual que en el estudio anterior con una distribución de frecuencias muy heterogénea, lo que implica un grado de heterocigosidad muy bajo y según este último autor, heteroplasmías de longitud en este tracto pueden aparecer o desaparecer en pocas generaciones lo que limita su utilidad forense.

- *Polimorfismos de secuencia:* producidos por el cambio de uno o más nucleótidos en una secuencia de ADN. Puede cambiar sólo un nucleótido, de forma que haya dos bases alternativas y cada una de ellas aparece en la población en una frecuencia superior a 1%; a este tipo de polimorfismo se le denomina SNP (Single Nucleotide Polymorphism). Es el tipo de variación más frecuente tanto en el ADN nuclear como en el ADN mitocondrial. La mayoría de estos polimorfismos se encuentran en regiones donde no alteran la función génica, y las transiciones son el tipo predominante de sustitución frente a las tranversiones en una proporción aproximada de 40:1 (Tully *et al.* 2001).

2.6. Heteroplasmías en el ADN mitocondrial

La heteroplasmía se define como el estado en el que un individuo, célula o mitocondria presenta más de un genotipo de ADNmt. En muchos individuos no hay evidencia de heteroplasmía, pero hay evidencias que indican que el ADNmt está mutando constantemente. Como estas mutaciones ocurren al azar, todas las mutaciones adquiridas estarán presentes en muy bajo nivel y puede que no se detecten en una muestra de tejido o de sangre. Así utilizaríamos el término de "homoplasmía" para describir el estado en el que no podemos detectar esas mutaciones adquiridas (Taylor *et al.* 2005). En un principio se pensó que se trataba de un fenómeno muy poco frecuente (Monnat *et al.* 1986), sin embargo trabajos posteriores y técnicas de detección cada vez más sensibles han puesto de manifiesto que la heteroplasmía en células somáticas es más frecuente de lo que en un principio se pensó, y está presente tanto en animales como plantas (Jazin *et al.* 1998; Calloway *et al.* 2000; Kmiec *et al.* 2006). Se tiende a considerar que todos los individuos son heteroplásmicos a algún nivel, sería muy ingenuo pensar que todos los genomas mitocondriales dentro de un individuo, tejido o incluso de una célula son iguales (Bär *et al.* 2000). La heteroplasmía puede observarse en diferentes formas:

- Un individuo puede presentar más de un tipo de ADNmt en un solo tejido.
- Un individuo puede tener un tipo de ADNmt en un tejido y otro en tipo en otro tejido.
- Un individuo puede ser heteroplásmico en un tejido y homoplásmico en otro tejido.

Cuando una célula heteroplásmica se divide, la herencia mitocondrial de las células hijas es una cuestión de azar (Figura 3). Durante la trasmisión se produce un *cuello de botella*, y como resultado

de ello, en la ovogénesis un número reducido de moléculas de ADNmt determinaría el genotipo citoplasmático de la siguiente generación. Después de varios ciclos de división la proporción de ADNmt mutante y normal en una célula puede derivar hacia el mutante puro o hacia el normal puro (homoplasmía) en un proceso conocido como *segregación replicativa*. Este proceso puede suceder durante la replicación de la célula somática o durante la proliferación de las células germinales femeninas.

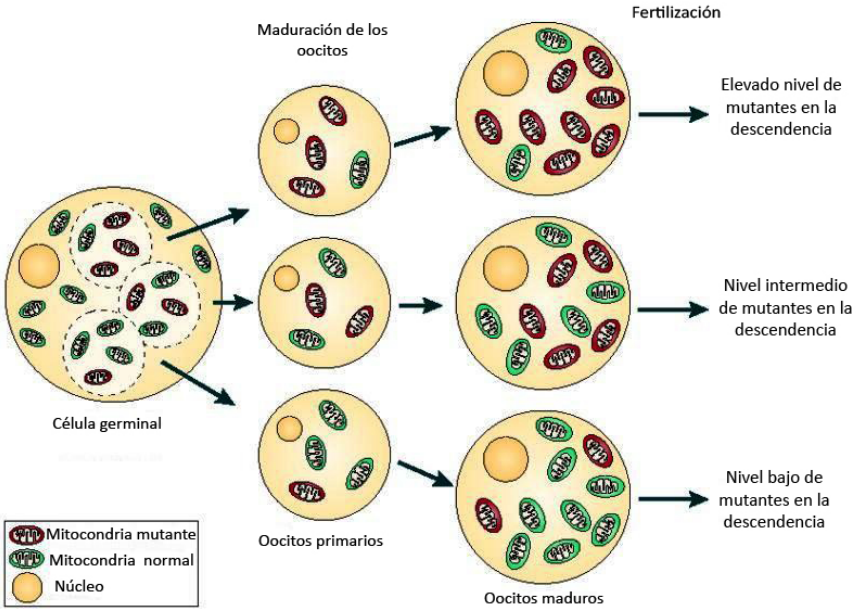


Figura 3: Cuello de botella durante la trasmisión del ADNmt (modificada de Taylor *et al.* 2005).

Existen dos tipos de heteroplasmías: *puntuales* y *de longitud.* Las heteroplasmías puntuales se deben a un único cambio de base y en los primeros años fueron escasamente reportadas debido fundamentalmente a la tecnología de secuenciación utilizada, mucho menos sensible y "limpia" que la usada actualmente. Hoy en día, sin embargo, la optimización de los kits de secuenciación y el desarrollo de otras tecnologías como la minisecuenciación permiten la detección de

muchas heteroplasmías (Tully *et al.* 2004; Cassandrini *et al.* 2006; Álvarez-Iglesias *et al.* 2007) (Figura 4). Las heteroplasmías de longitud son mucho más frecuentes que las puntuales y se deben a la variación en el número de bases en un tracto homopolimérico, cuando se presenta esta circunstancia es muy difícil poder caracterizar la población de variantes de longitud por secuenciación directa y en estos casos es necesaria la secuenciación de las dos cadenas de ADNmt y el uso de primers alternativos para poder determinar el tracto homopolimérico.

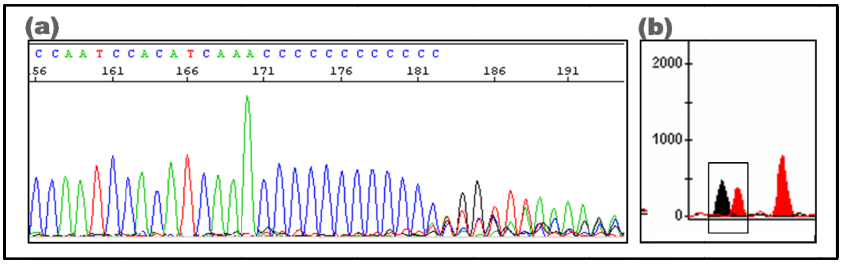


Figura 4: (a) Heteroplasmía de longitud en el tracto homopolimérico de la región hipervariable I detectada por secuenciación directa. (b) Heteroplasmía puntual detectada por minisecuenciación con el kit SNaPshot™.

Cuando se comparan diferentes vestigios biológicos de un mismo individuo o en estudios de parentesco es importante tener en cuenta este fenómeno, que puede tener repercusiones tanto desde un punto de vista forense como clínico.

Muchas enfermedades mitocondriales están caracterizadas por la coexistencia de un tipo de ADNmt normal y un tipo mutado en diferentes proporciones. El porcentaje relativo de la forma mutante va a determinar o contribuir a la expresión de la enfermedad en diferentes tejidos. Es necesario un número mínimo de ADNmt mutados para causar una disfunción mitocondrial, y además existe una correlación entre el grado de heteroplasmía y la severidad del desorden mitocondrial (Schapira, 2006; Dimauro *et al.* 2005).

La aparición de heteroplasmías no invalida el uso del ADNmt en el análisis forense (Bär *et al.* 2000; Tully *et al.* 2004). De hecho, es frecuente observar heteroplasmías en el análisis de pelos y otras muestras.

3. Aplicaciones del ADNmt en genética forense

3.1. Aspectos generales

La genética forense se encuadra dentro de la Medicina Legal, disciplina que tiene como objetivo ayudar a resolver cuestiones legales, tanto en casos criminales como civiles.

Hace apenas 20 años que Jeffreys *et al.* (1985) revolucionaron el panorama de las investigaciones forenses desarrollando la primera técnica de identificación individual y estudio de la paternidad. Se denominó "DNA fingerprint" (huella genética) y fue la primera técnica en utilizar el ADN como herramienta de investigación. Utilizando sondas multilocus analizaban minisatélites situados en diferentes regiones del genoma, obteniendo un patrón de bandas prácticamente específico para cada individuo.

Hasta ese momento la genética forense se basaba en el estudio de marcadores genéticos clásicos, marcadores sanguíneos (sistema ABO) y proteínas séricas que exhibían cierta variabilidad genética, de manera que se tenía una evidencia limitada en la identidad individual. Posteriormente llegó el estudio de enzimas, antígenos eritrocitarios y el sistema HLA (antígenos de los leucocitos humanos). En 1980 Wyman y White introdujeron el concepto de identificación genética a través del estudio de RFLPs (*Restriction Fragment of Length Polymorphism*). El uso de RFLPs de loci VNTR (*Variable Number of Tandem Repeat*) sustituyó en gran medida el uso de los marcadores proteicos clásicos. El ADN presentaba una gran estabilidad química frente a la rápida degradación

de los marcadores clásicos, y su gran variabilidad enseguida lo convirtieron en el objeto de elección.

Poco después del descubrimiento de la "*DNA fingerprint*" la introducción de la PCR (Mullis *et al.* 1987; Saiki *et al.* 1988) permitió el análisis de loci hipervariables, minisatélites y microsatélites (STRs) a partir de vestigios biológicos que no podían ser analizados hasta el momento, debido a la cantidad limitada de ADN que podía obtenerse en la mayoría de los escenarios forenses.

Inicialmente, el estudio de los polimorfismos de ADN se ha centrado mayoritariamente en el análisis de marcadores nucleares. Sin embargo, en muchas ocasiones, especialmente en contextos forenses, las muestras han sido expuestas a condiciones tan adversas que el estudio del ADN nuclear no es posible debido a que está muy degradado o en escasa cantidad o sea incluso inexistente como por ejemplo en los pelos sin bulbo. En estos casos, el ADN mitocondrial puede ser empleado ya que debido a sus características, las probabilidades de éxito en el análisis son muy superiores a las del análisis de marcadores nucleares.

El genoma mitocondrial contiene información de gran utilidad y en ocasiones suficiente para el establecimiento de la identidad o fuente de un determinado espécimen biológico, jugando un papel crucial en la identificación forense.

La gran utilidad del ADNmt en la genética forense se debe fundamentalmente a:

- El ADNmt está presente en un número muy alto de copias por célula, frente a las copias 2 del nuclear (o 1 en el caso del cromosoma Y), lo que va a permitir obtener resultados positivos en muchos casos en los que el ADN nuclear falla como pelos sin bulbo, material esquelético parcialmente degradado, muestras antiguas, etc.

- La elevada tasa de mutación del ADNmt en comparación con el ADN nuclear proporciona un gran apoyo para la identificación de evidencias biológicas.
- La herencia uniparental y la ausencia de recombinación del ADNmt, permiten la comparación de miembros de un linaje que comparten un ancestro común por vía materna, aportando la base para la identificación de evidencias biológicas en casos de restos humanos o de individuos vivos emparentados, como en el famoso caso de la identificación de los restos de la familia Romanov (Gill *et al.* 1994).
- Posibilidad de automatización del proceso, el análisis del ADNmt puede llevarse a cabo mediante la amplificación vía PCR y secuenciación automática, y también a través de un gran número de nuevas tecnologías que han surgido en los últimos años (Sobrino *et al.* 2005).

La utilización del ADNmt en la resolución de casos judiciales forenses es relativamente reciente, el primer caso en el que los resultados del análisis de ADNmt fueron llevados a los tribunales en Estados Unidos fue en Junio de 1996 (FBI Crime Lab) a raíz de un caso de agresión sexual y posterior asesinato en el estado de Tenesse. Coetáneamente en España, el primer análisis se llevó a cabo también en 1996 (*Sumario 2/95, Juzgado de 1ª Instancia e Instrucción de Puente Genil, Córdoba*) en el Instituto de Medicina Legal de Santiago de Compostela. Actualmente, este tipo de análisis se realiza de forma rutinaria en muchos laboratorios del mundo.

3.2. Técnicas de análisis de ADNmt

Los primeros estudios para el análisis de la variabilidad del ADNmt se basaban en la digestión mediante el uso de enzimas de restricción (Denaro *et al.* 1981) y separación electroforética (RFLPs). Se utilizaba fundamentalmente con fines poblacionales, ya que su uso en

forense era muy limitado debido a que se requería gran cantidad de ADN no degradado.

A finales de la década de los años 80 y durante la de los 90, surgieron otras tecnologías para analizar la variabilidad del ADNmt como la SSCP (*Single Strand Corformation Polymorphism)* (Orita *et al.* 1989), esta técnica se basa en la diferente movilidad que presentan las hebras de ADN de una muestra previamente desnaturalizada sobre un gel de acrilamida, en función de la conformación que adopten dichas hebras durante el recorrido electroforético. Es una técnica de *screening* simple y versátil que se utilizó para la detección de mutaciones puntuales. Sin embargo, aunque este abordaje resultaba sencillo, barato y rápido, presentaba como desventaja su difícil estandarización entre laboratorios y además sólo detectaba un 90% de las mutaciones.

Posteriormente, se desarrollaron variaciones de esta técnica, como la RE-SSCP (*Restriction Enzyme-SSCP*) (Barros *et al.* 1997), como su nombre indica, esta técnica combina la digestión enzimática con los SSCPs, tiene la ventaja de poder detectar casi el 100% de las mutaciones, utilizando para la digestión de los fragmentos de ADN varias enzimas de restricción. También surgió otra variante de los SSCPs como la FSSCP-OF (*Fluorescent-SSCP-Overlaping fragments*) (Salas *et al.* 1999), metodología que combina dos estrategias complementarias: amplificación por PCR de fragmentos solapantes y detección mediante tecnología fluorescente. El uso de tecnología fluorescente permite la amplificación multiplex de los fragmentos solapantes y su análisis conjunto en un secuenciador automático. El poder de discriminación de esta técnica también está próximo al 100%.

Sin embargo, el método elegido por la mayoría de los laboratorios fue la secuenciación, ya que así se podía conocer exactamente qué mutación está presente en el ADN.

A finales de los años 70, se publicaron dos métodos de secuenciación distintos: *secuenciación química* de Maxam y Gilber (1977) el método de *secuenciación enzimática* de Sanger *et al.* (1977). Los métodos de secuenciación que se utilizan hoy día mayoritariamente están basados en el método dideoxy o enzimático. Este sistema requiere el uso de 2´-3´-dideoxinucleótido trifosfato (ddNTP), que difieren de los deoxinucleótidos precursores de la síntesis de ADN en la sustitución del grupo 3´-hidroxilo por un hidrógeno. Los ddNTPs son reconocidos igual que los dNTPs por la ADN polimerasa y son adicionados al extremo de la cadena creciente, pero al carecer del grupo 3´-hidroxilo no pueden unirse al nucleósido 5´-trifosfato entrante y por lo tanto, su incorporación supone la interrupción de la síntesis de la cadena de ADN. De esta manera, la inclusión de niveles bajos de dideoxinucleótidos en un medio de replicación permite que la síntesis de ADN termine al azar en cualquier sitio.

El sistema que se utiliza mayoritariamente hoy día, es la *secuenciación cíclica,* combina la técnica de la PCR con el método anteriormente descrito (Carothers *et al.* 1989; Hopgoood *et al.* 1992).

En 1995 Wilson *et al.* llevaron a cabo la validación experimental del uso del ADNmt con fines forenses sobre material degradado, estudiaron los efectos de la contaminación química en sangre y semen, el efecto en la extracción y secuenciación de distintos substratos sobre los que se puede hallar material biológico, así como estudios de validación en pelo sin bulbo, incluyendo la exposición a agentes químicos, el tratamiento cosmético y la contaminación biológica. Los datos que obtuvieron confirmaron que el análisis del ADN mitocondrial por PCR seguido de secuenciación es un método válido y seguro para ser utilizado en la práctica forense.

En los primeros años, la región elegida para amplificar por PCR y posteriormente secuenciar, fue la región hipervariable I (HVI) de la región control, comprendida entre la posición 16024 y 16365, definida

por Vigilant *et al.* 1989. Posteriormente, se incorporó a la rutina del análisis la región hipervariable II (HVII) comprendida entre las posiciones 073 y 340, de manera que pasaban a evaluarse alrededor de 610 pb en lugar de los 340 que se analizaban con la región HVI, y así aumentaba el poder de discriminación. Sin embargo la tendencia es poder analizar cada vez más posiciones, y así en estos últimos años la recomendación es analizar la región control completa desde la posición 16024 hasta la 16569 y desde la 1 hasta la 576.

Sin embargo, la región control tiene un poder de discriminación limitado, no sólo en el contexto forense, sino también desde un punto de vista antropológico o clínico. Esto depende principalmente de la población que se esté estudiando ya que los valores de diversidad y la filogenia varían enormemente entre las distintas poblaciones, por ejemplo, poblaciones del Este de África presentan los mayores niveles de diversidad. Muchos haplogrupos, como por ejemplo el haplogrupo H, que es el más frecuente en Europa, están muy mal definidos en las regiones hipervariables, de manera que perfiles que coincidan en estas regiones no tienen por qué pertenecer necesariamente al mismo haplogrupo ni por supuesto al mismo individuo (Quintáns *et al.* 2004; Álvarez-Iglesias *et al.* 2007).

La situación ideal, sería poder secuenciar siempre genomas completos, sin embargo esto es muy laborioso, requiere mucho tiempo, gran cantidad de muestra y por lo tanto resulta totalmente inviable en la rutina forense. El análisis de SNPs de la región codificante del genoma mitocondrial ha resultado ser una buena estrategia para solventar estos problemas.

3.3. Análisis de SNPs de la región codificante del ADNmt

Los SNPs son el tipo de polimorfismo humano más abundante, resulta complicado estimar su número pero en las distintas bases de datos se han identificado más de cinco millones. Su abundancia

además de su simplicidad y su baja tasa de mutación en relación a otros marcadores han despertado un gran interés en distintos campos como el de la genética forense, el antropológico, y por supuesto también el clínico (Sobrino *et al.* 2005).

Este interés creciente de los últimos años en estos marcadores ha ido paralelo al espectacular desarrollo de nuevas tecnologías para su genotipado que han ido surgiendo, que permiten el análisis de varios marcadores a la vez mediante multiplex, lo que es especialmente interesante en el campo forense. El abanico de tecnologías para el genotipado de SNPs actualmente es enorme, sin embargo las elegidas para el campo forense, teniendo en cuenta las características que presentan muchas veces las muestras, deben cumplir una serie de requisitos como tener una elevada sensibilidad (ya que muy a menudo la calidad y cantidad de las muestras es escasa), fiabilidad, una buena capacidad para analizar reacciones multiplex y ser reproducibles (Sobrino *et al.* 2005).

En el caso del ADNmt el método más popular para el análisis de la variabilidad de la región codificante es la técnica de minisecuenciación, concretamente el uso del kit de SNaPshot™ (Applied Biosystems). En la reacción de minisecuenciación una ADN polimerasa es utilizada para la extensión de un primer complementario a la secuencia nucleotídica y que finaliza en su extremo 3´ en la base anterior a la posición polimórfica, incorporando un único nucleótido trifosfato (ddNTP) marcado fluorescentemente complementario a la base polimórfica. Cada ddNTP está marcado con un marcador fluorescente distinto, y la reacción multiplex puede llevarse acabo porque los productos de minisecuenciación llevan incorporadas colas no homólogas en su extremo 5´ que permiten que puedan ser separados mediante electroforesis en los secuenciadores automáticos (Figura 5).

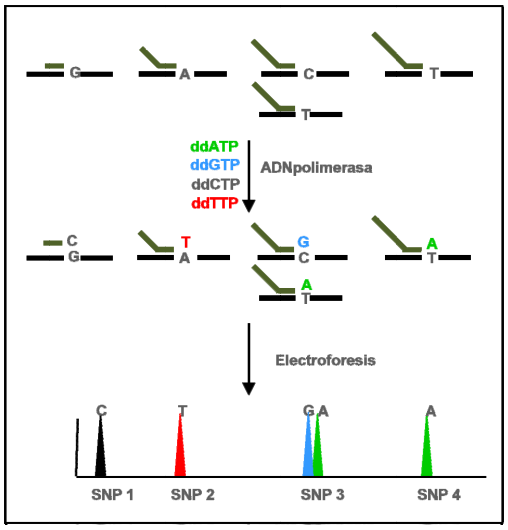


Figura 5: Representación esquemática de la técnica de minisecuenciación con el kit de SNaPshot™. Una ADN polimerasa incorpora un ddNTP marcado fluorescentemente complementario a la base polimórfica a interrogar.

En 1996 Tully *et al.* ya habían descrito esta técnica para genotipar posiciones de la región control, sin embargo no ha sido hasta estos últimos tres años cuando ha aparecido el grueso de trabajos realizados con esta técnica, concretamente con el kit de SNaPshot™. En el año 2003, aparece el primer diseño para genotipar SNPs de la región codificante de los principales haplogrupos caucasoides (Brandstätter *et al.* 2003), en este trabajo analizan 16 polimorfismos en dos reacciones multiplex y se propone como un buen sistema de *screening* previo al análisis por secuenciación de la región control. En el año siguiente, aparecen dos trabajos más, Quintáns *et al.* genotipan 17 SNPs también de haplogrupos caucasoides, pero en este caso lo que se consigue es aumentar el poder de discriminación en el análisis de ADNmt, especialmente en el caso de muestras pertenecientes al haplogrupo H, el más frecuente de la población europea; ese mismo año Vallone *et al.* presentan otro set de 11 SNPs con el mismo fin que los anteriores. Desde estos primeros trabajos hasta la actualidad, se han presentado muchos más ensayos de este tipo para población caucasoide (Grignani

et al. 2005; Wiesbauer *et al.* 2006; Brandstätter *et al.* 2006), también diseños enfocados a poblaciones del Este de Asia y Nativo Americanos (Lee *et al.* 2006; Álvarez-Iglesias *et al.* 2007) y algún trabajo en ADN antiguo (Endicott *et al.* 2006).

El éxito de esta técnica en los últimos años probablemente se deba a que permite interrogar en una única reacción un número considerable de SNPs (Álvarez-Iglesias *et al.* 2007), lo que es muy importante en el contexto forense donde a menudo la cantidad de muestra es limitada, permite amplificar fragmentos de tamaño reducido de manera que se facilita el análisis de muestras degradadas, y por último también cabe destacar que es una técnica rentable frente a otras mucho más costosas como lo DNA *micro-arrays*, ya que la mayoría de los laboratorios tienen acceso a un secuenciador automático.

En 1996, Chee *et al.* describieron el análisis del genoma completo de ADN mitocondrial con *arrays* de ADN, estos chips contenían 135000 sondas complementarias con las 16.6 kb de genoma mitocondrial humano. Más recientemente, en 2006 Sigurdsson *et al.* en vez de abordar el genoma mitocondrial completo interrogan 150 SNPs de la región codificante en una muestra de 265 individuos de distintas poblaciones del mundo, obteniendo un poder de discriminación, según los autores, incluso mayor que analizando 500 pb de la región control.

Esta tecnología, sin embargo no ha tenido el impacto que se esperaba a raíz de los primeros trabajos, probablemente por varias razones: es una tecnología cara, muy pocos laboratorios tienen acceso a ella y no está estandarizada. El uso de los chips que se utilizan para secuenciar el genoma completo también es limitado, seguramente por las mismas razones y porque en realizad no cubren el 100% de la variabilidad observada en todas las poblaciones de manera que pueden quedar sin representar zonas de la filogenia.

3.4. Validación y estandarización de la prueba del ADNmt

En Europa el principal esfuerzo de validación de la prueba del ADNmt se ha venido realizando por el grupo de la *European DNA Profiling/Standarization DNA Profiling Group* (EDNAP/STADNAP) (Carracedo *et al.* 1998; Parson *et al.* 2004; Tully *et al.* 2004). Desde 1997 el Grupo Español y Portugués de la ISFG (GEP-ISFG) ha venido organizando ejercicios de colaboración y demostración de aptitud anualmente (Montesino *et al.* 2007; Crespillo *et al.* 2006; Salas *et al.* 2005b). Estos ejercicios son una herramienta muy útil para evaluar, comparar distintas metodologías, resultados y análisis estadísticos entre distintos laboratorios y los resultados han confirmado que el análisis de ADNmt mediante PCR y secuenciación automática es una herramienta válida y fiable en la identificación forense.

Existen una serie de recomendaciones y criterios establecidos por parte de la EDNAP y de la *DNA Comission* de la ISFG con el fin de estandarizar el proceso del genotipado de ADNmt (Bär *et al.* 2000; Carracedo *et al.* 2000; Tully *et al.* 2001). Estas recomendaciones abarcan todo el proceso del análisis, desde la recogida de la muestra hasta la interpretación de los resultados. En cuanto a la parte metodológica, son aplicables las mismas recomendaciones que en el análisis por PCR de ADN nuclear pero merece especial atención el problema de la contaminación, ya que el análisis del genoma mitocondrial muchas veces ser realiza con un material altamente degradado y escaso. La contaminación con ADN exógeno supone un gran problema en estos casos porque la cantidad de ADN endógeno en la muestra es mínima.

Existen distintos niveles potenciales de contaminación:

- Contaminación con ADN humano. La contaminación puede ser previa a la recogida de la muestra (por ejemplo por contacto con restos adyacentes en enterramientos colectivos). Puede

producirse durante la recogida de la muestra por parte del equipo encargado, esto puede evitarse la mayor parte de las veces usando sencillas medidas de precaución. Y también puede producirse contaminación en el laboratorio durante el proceso de extracción del ADN de la muestra, ya sea con ADN de la persona que está manipulando las muestras o bien por la presencia de ADN humano en el ambiente del laboratorio, y durante la fase de PCR por la transferencia de productos de amplificación generados en reacciones anteriores, bien en el proceso de amplificación o por la contaminación de los reactivos.

- Contaminación con ADN no humano. Bacterias, hongos, insectos, etc. Pueden contaminar los vestigios a analizar como resultado de los procesos de descomposición naturales, ya que durante el proceso de extracción de la muestra el ADN se extrae conjuntamente. Siempre que se utilicen primers específicos humanos esta cantidad de ADN exógeno no debe ser un problema, aunque sí puede causar reacciones competitivas con los primers durante la amplificación favoreciendo la amplificación de productos inespecíficos o incluso inhibir la amplificación de la muestra.

Para evitar estos problemas de contaminación, y asegurar que la contaminación no pueda causar resultados erróneos, la EDNAP y la *DNA Comission* de la ISFG (Bär *et al.* 2000; Carracedo *et al.* 2000; Tully *et al.* 2001) proponen una serie de recomendaciones:

- Las áreas de pre y post-PCR deben estar separadas.
- Las muestras a investigar y las muestras de referencia deben ser analizadas por separado. En lugares en los que el espacio es limitado y no pueden analizarse en espacios diferentes, las

muestras dubitadas deben analizarse antes que las muestras de referencia.

- Se recomienda el uso de pipetas de desplazamiento positivo y de puntas con filtro para los pasos pre-PCR.
- El material de plástico utilizado para las manipulaciones pre-PCR debe estar libre de ADN, un buen tratamiento para esto es la radiación UV. De la misma manera, siempre que sea posible este tratamiento debe ser aplicado a las soluciones utilizadas en los pasos pre-PCR.
- Durante la extracción y la amplificación deben usarse controles negativos con los que se trabajará durante todos los pasos hasta la secuenciación. Así mismo deben usarse también controles positivos durante la PCR.
- Siempre que sea posible y haya suficiente muestra deben hacerse dobles extracciones, debe analizarse la muestra por duplicado, y todas las muestras deben ser secuenciadas para las dos cadenas complementarias.
- Las secuencias del personal de laboratorio que manipule las muestras deben ser analizadas. Es probable que con determinada frecuencia, las secuencias del operador coincidan con las muestras investigadas. En estos casos el resultado no queda invalidado pero alerta al investigador de una contaminación potencial. Para evitar esto, se recomienda siempre que sea posible, que distintas personas determinen por separado los mismos análisis.

Aunque estos criterios están perfectamente validados y aceptados para el análisis de ADNmt mediante secuenciación, es importante establecer nuevos criterios que incluyan el análisis de SNPs de la región codificante de ADNmt.

De la misma manera, se necesitan criterios de nomenclatura y criterios que permitan una correcta interpretación de la prueba del ADNmt. Sin embargo tal y como se describe en los siguientes apartados no existen criterios en muchas ocasiones.

3.5. Nomenclatura

De acuerdo con las recomendaciones del *European DNA profiling* (EDNAP) *group* (Tully *et al.* 2001) la numeración que se utiliza para denominar las mutaciones puntuales, inserciones y delecciones de la región control hace referencia a la secuencia de referencia de Cambridge (Anderson *et al.* 1981). Sin embargo, para denominar estos cambios en el resto de la molécula se recomienda utilizar como referencia la Secuencia de Cambridge revisada (rCRS) (Andrews *et al.* 1999). Se considera siempre la secuencia de la cadena L y sólo se enumeran las bases que difieren respecto a la secuencia de referencia.

Para anotar un cambio de base, se escribe la posición en que se encuentra seguida de la base mutada, por ejemplo la posición 16126 es una T en la secuencia de referencia, si se produce un cambio por una citosina se anotaría 16126C. Sin embargo Salas *et al.* (2005a), recomiendan especificar la base mutada solamente cuando se trata de una transversión.

Si se observa una delección respecto a la secuencia de Anderson, se reporta la posición deleccionada seguida de una "d", por ejemplo 249d.

Las inserciones se designarán con un ".1" seguido de la base insertada, después de la posición que está inmediatamente al lado de la inserción en sentido 5', por ejemplo en la posición 523 es frecuente la inserción de "CA" o incluso de "CACA", en este caso se reportará 523.1C 524.2A 525.3C 526.4A. Cuando las inserciones ocurren dentro de tractos homopoliméricos se desconoce la localización exacta de la

inserción, en estos casos se asume que la inserción ocurre en la posición más alta de ese tramo, por ejemplo una inserción en el tramo homopolimérico entre las posiciones 302 y 310 se anotará como 309.1C, en el caso de que ocurran dos inserciones se designarán 309.1C 309.2C. En estos casos, Salas *et al.* (2005a) prefieren un modo mucho más compacto para reportar las inserciones, de manera que los ejemplos anteriores se anotarían como 523+CACA (en este caso se podría designar como 524+ACAC, si se interpreta que el motivo que se repite es AC en lugar de CA), en el caso de los homopolímeros se anotará 309+C en el caso de que haya una única inserción, 309+CC en el caso de que se inserten 2 y así sucesivamente.

Para describir una heteroplasmía puntual en la que dos bases están presentes en una intensidad aproximadamente igual se recomienda utilizar el código IUB (R=A+G, Y=C+T, K=G+T, M=A+C, S= G+C, W=A+T). Por ejemplo si en la posición 152 aparecen una citosina y una timina se anotará como 152Y. Alternativamente se puede designar como 152 T~C. Si una base está en una proporción substancialmente mayor que la otra se utilizará una notación del tipo C>T. Es importante señalar que siempre que se detecta una heteroplasmía es necesario confirmarla con la secuencia de la cadena L y H, en el caso de que la heteroplasmía no pueda confirmarse con la segunda reacción de secuencia la posición se designará como ambigua "N".

En el caso de las heteroplasmías de longitud se utiliza un planteamiento similar al anterior para definir si la heteroplasmía está presente o no. Esto afecta particularmente a los tractos de policitosinas de HV1 (entre las posiciones 16183 y 16194) y HV2 (entre 302 y 315). Si el número de citosinas puede ser confirmado con la secuencia de las dos cadenas, la nomenclatura sería del tipo 309.1>309.2, sin embargo en el caso de que no sea posible la confirmación, el número de bases deberá ser reportado como ambiguo.

3.6. Criterios de inclusión-exclusión en la interpretación de la evidencia de secuencias de ADNmt

Cuando se lleva a cabo un análisis de ADNmt, muchas veces es para identificar restos a través de individuos relacionados por vía materna, o para ver la asociación entre muestras recogidas en una escena de un crimen y un sospechoso. De manera que podemos estar ante dos situaciones distintas: (i) que la evidencia proceda del sospechoso o de algún individuo relacionado por vía materna, o (ii) que la evidencia y la muestra del sospechoso procedan de dos individuos distintos (o de dos linajes distintos).

La interpretación es sencilla sólo cuando dos perfiles son claramente distintos, con dos o más diferencias mutacionales entre ambas. En este caso se podría *excluir* directamente un individuo como origen de un vestigio biológico.

En cuanto a los criterios de *inclusión,* sólo existe consenso unánime en el supuesto más sencillo, en el caso de que dos secuencias coincidan completamente en todas las posiciones. Sin embargo en muchas ocasiones las diferencias son más sutiles y la interpretación no es tan sencilla. Lo mismo ocurre con la interpretación de las heteroplasmías.

Cuando dos secuencias presentan una heteroplasmía en la misma posición muchos laboratorios consideran estas posiciones como ambiguas (N) o no identificables, de manera que no se consideran y no se adopta un criterio de inclusión. Sin embargo, otros laboratorios consideran que en una situación como la descrita deberían representar información adicional de variación incrementando el poder de la prueba de ADNmt. Así, si una muestra de referencia y una muestra desconocida presentan la misma secuencia, la co-ocurrencia de heteroplasmía en un sitio particular aportaría una evidencia adicional, restringiendo el número de posibles donadores e incrementando la

probabilidad de que la referencia sea la fuente de la muestra desconocida.

Los casos se complican cuando existen más diferencias entre las muestras que se comparan. Por ejemplo, si se comparan dos secuencias en las que una de ellas en una posición presenta heteroplasmía, algunos laboratorios sugieren que dado que las dos secuencias comparten un nucleótido para la posición que presenta la heteroplasmía no se podría excluir la muestra cuestionada como originada por la fuente conocida. Bien es verdad que las heteroplasmías aparecen tan sólo en un porcentaje reducido de veces en la rutina forense, y en muy raras ocasiones se produce más de una diferencia entre tejidos del mismo individuo, por lo que no deberían representar un problema para la identificación forense (Tully *et al.* 2004).

Cuando sólo existe una diferencia clara en una posición entre dos secuencias cuestionadas, muchos laboratorios aplican un razonamiento similar al de la situación anterior y concluyen que no se puede excluir la muestra cuestionada como originada por la fuente conocida, pero otros laboratorios en cambio interpretan este resultado como inconclusivo.

No se deben utilizar criterios demasiado rígidos para la interpretación de resultados de ADNmt, ya que hay que tener en cuenta muchos factores relacionados con la biología del ADNmt como que la segregación y la tasa de mutación tienen distintas velocidades en diferentes tejidos o como las tasas de mutación sitio-específicas.

4. Estudio de la variabilidad mitocondrial en poblaciones humanas

Las características únicas del ADN mitocondrial humano: elevado número de copias, herencia materna, ausencia de recombinación y alta tasa de mutación, han hecho que esta molécula sea la elegida en

muchos estudios relacionados con la historia y evolución de las poblaciones humanas (Pakendorf *et al.* 2005).

La *filogeografía* se encarga del estudio de la distribución geográfica de los distintos linajes mitocondriales. Estudiando los linajes actuales se puede reconstruir la historia siguiendo el principio de coalescencia hasta llegar a un único ancestro materno.

La *filogenia* molecular establece relaciones evolutivas basándose en datos moleculares, bien en secuencias aminoacídicas de proteínas o nucleotídicas a partir de ADN. La escala evolutiva analizada puede encontrarse a diferentes niveles: dentro de una misma población, entre poblaciones de una misma especie o entre especies. Para poder inferir relaciones filogenéticas se exige como norma básica aplicar el *criterio de homología*, de manera que el carácter que se esté comparando tiene que ser homólogo, es decir, que debe ser compartido por ascendencia o ser idéntico por descendencia.

Las relaciones filogenéticas entre un grupo de organismos pueden representarse gráficamente mediante *árboles filogenéticos* que consisten en nodos (unidades taxonómicas) conectados por ramas. Para poder distinguir entre un carácter ancestral y uno derivado es necesario que este tipo de árbol tenga una raíz de la que salen las diferentes ramas o *"outgroups"* (grupos externos).

Existen diferentes procedimientos para la reconstrucción filogenética, pero el más utilizado es el de "máxima parsimonia" o de "evolución mínima": la evolución ocurre parsimoniosamente de manera que los cambios nucleotídicos en todas las posiciones son igualmente probables y reconstrucción filogenética asume la explicación más sencilla de la información, lo que quiere decir que el árbol elegido es el que requiere el mínimo número de eventos mutacionales para explicar la distribución de los caracteres entre los individuos a estudiar.

La mayor parte de los estudios de la variación del ADNmt en las distintas poblaciones humanas se han realizado principalmente por dos métodos: por secuenciación directa de la región control o por la amplificación de la molécula de ADNmt completa en nueve fragmentos solapantes de 1500 a 3000 pb y posterior digestión con un set de 14 enzimas de restricción (Torroni *et al.* 1992; Chen *et al.* 1995).

Los haplotipos obtenidos pueden ser clasificados mediante análisis de parsimonia en diferentes grupos que comparten las mismas mutaciones puntuales, son mutaciones estables y antiguas que forman parte de la región codificante. Los distintos agrupamientos se denominan *haplogrupos* y resultan específicos de determinados grupos poblacionales de manera que facilitan una gran información acerca de las relaciones interpoblacionales. A pesar de la alta tasa de mutación promedio de la región control, existen también posiciones diagnóstico para el establecimiento de haplogrupos. Es importante señalar que a lo largo de la molécula de ADNmt existen ciertas posiciones con una elevada tasa de mutación que aparecen de manera indistinta en diferentes haplogrupos, y también aparecen con frecuencia mutaciones recurrentes que pueden oscurecer y dificultar la interpretación de la filogenia.

En los últimos años, con el desarrollo de tecnologías de alto rendimiento, la secuenciación de genomas completos se ha vuelto común incluso en los estudios poblacionales (Ingman *et al.* 2001; Herrnstadt *et al.* 2002; Yao *et al.* 2006). La filogenia obtenida basada en los datos de genomas completos ofrece una mayor resolución que la obtenida sólo con secuencias de la región control, sin embargo teniendo en cuenta el esfuerzo que supone trabajar con genomas completos, resulta una buena estrategia combinar los datos de la región control con posiciones concretas de la región codificante analizadas tradicionalmente con RFLPs y en los últimos años con otras tecnologías

como la minisecuenciación con SNaPshot (Quintáns *et al.* 2004; Brandstätter *et al.* 2007; Álvarez-Iglesias *et al.* 2007).

La interpretación de todos los resultados del análisis de la molécula de ADNmt ha tenido una consecuencia de gran impacto en el mundo de las ciencias: el origen africano de la molécula de ADNmt, y por extensión, de las poblaciones humanas.

En 1987 Cann *et al.,* en base a los resultados obtenidos en el estudio de individuos de poblaciones de Asia, Australia, Nueva Guinea, Europa y África, formulan una hipótesis conocida como la *"Eva mitocondrial".* De esta hipótesis conviene señalar tres aspectos principales: (i) todos los tipos de ADNmt actuales se remontan a un único ancestro, (ii) este ancestro vivió en África, y (iii) probablemente vivió hace unos 200 000 años.

Esta interpretación ha sido tremendamente controvertida fundamentalmente debido a la utilización de determinados métodos de reconstrucción filogenética pero, sin embargo, actualmente la hipótesis africana del origen del hombre moderno ha servido de apoyo a un gran número de estudios tanto de análisis de ADNmt como de marcadores nucleares.

La hipótesis formulada por Cann *et al.* (1987) y posteriormente apoyada por otros autores (Chen *et al.* 1995; Horai *et al.* 1990; Ingman *et al.* 2000) está basada en el principio de coalescencia según el cual, asumiendo que hubo un origen único para todos los organismos vivos se deduce que toda la variación en cualquier segmento del ADN mitocondrial (o nuclear) en las generaciones presentes deben proceder en última instancia de un único ancestro que existió en generaciones previas. En el caso del ADNmt, al ser de herencia unilineal materna, esta reconstrucción de linajes es mucho más sencilla y de este plantemiento se puede deducir que este ancestro era femenino (Eva), que no era el único individuo vivo, sino un miembro de la población

para la que el resto de los linajes mitocondriales se fueron perdiendo con el paso de las generaciones, y que no fue la primera mujer que apareció en el planeta, sino que sólo representa el punto para el cual todos los linajes coalescen. Eva sería, en palabras de Allan C. Wilson y Rebecca L. Cann, "*la feliz mujer cuyo linaje permanece*".

El ADNmt no sólo ha contribuido al esclarecimiento del origen del hombre, sino también al conocimiento de las migraciones humanas, como la colonización del Nuevo Mundo, la colonización del Pacífico, las migraciones iniciales a Nueva Guinea y Australia y el asentamiento en Europa (Pakendorf *et al.* 2005). A lo largo de la historia y durante todos estos movimientos poblacionales las mutaciones se han ido acumulando secuencialmente en los linajes de ADNmt a partir de secuencias fundadoras y estos linajes maternos han divergido a medida que las poblaciones humanas colonizaban distintas regiones geográficas en todo el mundo. Por esta razón muchos de los haplogrupos mitocondriales son continente-específicos (Figura 6).

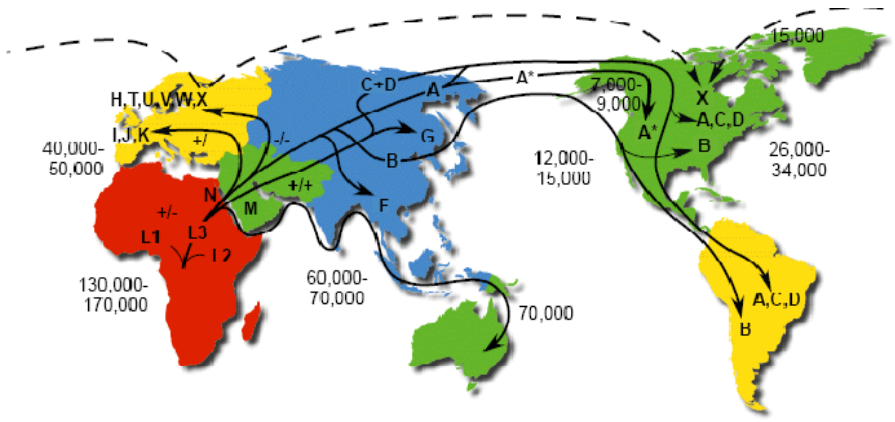


Figura 6: Representación de las migraciones humanas a través del ADNmt (adaptado de Mitomap).

El extenso haplogrupo L con los subgrupos L1, L2 y L3 son característicos de poblaciones africanas, mientras que los haplogrupos

M y N originados en el este de África a partir de L3 se dispersaron hacia Eurasia y el Nuevo Mundo. Los haplogrupos H, I, J, N1b, T, U, V y W son característicos de poblaciones caucasoides, los haplogrupos A, B, C y D se fundaron en Asia y el Nuevo Mundo, y los haplogrupos G, Y y Z predominan en Siberia (Figura 6).

Sin embargo, es importante señalar que el ADNmt es sólo un locus y sólo refleja de forma parcial la historia demográfica matrilineal de las poblaciones. La historia escrita a través de sólo un locus no tiene que reflejar de manera fiel la historia de una población, por lo que los estudios realizados a partir de la variabilidad del ADNmt deben ser complementados con datos de otros marcadores como el cromosoma Y, e idealmente, también con datos autosómicos (Pakendorf *et al.* 2005).

5. Detección de errores en bases de datos mitocondriales

Existe un gran número de publicaciones en la literatura forense alertando de la elevada incidencia de errores en los análisis de ADNmt. Sin embargo no es un tema que afecte sólo al campo forense (Bandelt *et al.* 2004a, 2004b; Yao *et al.* 2004), sino que también existe un buen número de publicaciones alertando errores del mismo tipo en otras áreas de investigación como en estudios genético-poblacionales y en estudios clínicos (Bandelt *et al.* 2001; Salas *et al.* 2005a; Kong *et al.* 2006).

Los errores típicos encontrados en las diferentes áreas se pueden clasificar en cinco categorías (Bandelt *et al.* 2001; Salas *et al.* 2005a):

- Tipo I (cambio de base): se reporta una posición en lugar de otra. Este tipo de error puede producirse por una alineación con la secuencia de referencia o una lectura incorrectas, o también puede ocurrir en el momento de preparar la tabla de resultados. Por ejemplo, puede reportarse la posición

A234G cuando en realidad se trata de la posición A235G, o reportar la posición C295T en lugar de la C285T.

- Tipo II ('*reference bias*'): se omiten variantes respecto a la secuencia de referencia.
- Tipo III (mutaciones fantasma): mutaciones poco comunes aparecen simultáneamente en diferentes linajes del set de datos. Estas mutaciones se producen durante el proceso de secuenciación y/o por el mal uso de software de lectura e interpretación de electroferogramas.
- Tipo IV ('*base misscoring*'): en este tipo de error se reporta mal la letra correspondiente con el nucleótido, se reportan transiciones como transversiones o viceversa. Por ejemplo se reporta una transversión en la posición T489G en lugar de la transición T489C, o se descuida la posición exacta en la que ocurre una *indel,* por ejemplo se reporta la deleción en la posición 248del en lugar de la posición correcta 249del.
- Tipo V (recombinación artefactual): se origina un haplotipo que es el resultado de diferentes fragmentos de más de una muestra. Este tipo de errores se origina porque la mayoría de laboratorios analizan la región en dos o más fragmentos que no se solapan entre sí. Existen muchos pasos en el análisis en el laboratorio en los que se puede producir este error (al hacer distintas PCRs, en el momento de cargar las muestras en el secuenciador, etc.), pero también puede producirse en el momento de transferir los datos a la tabla de resultados o a la base de datos.

Es complicado encontrar un sistema con el que se vayan a evitar todos los errores, pero tomando una serie de precauciones sí se podrían evitar muchos de ellos. En lo que concierne a la parte de trabajo en el

laboratorio, para evitar problemas de contaminación, adoptar las medidas y recomendaciones descritas en el apartado 3.4 puede ser una buena estrategia, también es recomendable utilizar secuenciación de fragmentos solapantes, de manera que se puedan comprobar variantes que deben aparecer en dos segmentos de la misma muestra que se secuencian por separado, y por último, se recomienda optimizar la química de la secuenciación, cambiar o modificar primers para intentar minimizar el efecto del ruido de fondo. En cuanto a la parte de documentación, es importante que la lectura e interpretación de los electroferogramas la realicen por lo menos dos personas, y la lectura automatizada mediante software adecuados debe ir acompañada de una inspección visual de las secuencias (Salas *et al.* 2005a).

Cuando al estudiar una base de datos de ADNmt se observa una sobre-representación de transversiones debe ser motivo de sospecha del conjunto de datos, ya que en general en todo el genoma las transiciones son mucho más frecuentes que las transversiones. De la misma manera las inserciones y deleciones (con la excepción de las posiciones que forman los tractos homopoliméricos en la región control) son eventos poco frecuentes.

Un buen sistema para detectar muchos de estos errores es utilizar una aproximación filogenética (Salas *et al.* 2006), aunque con este método sólo se puede detectar un número limitado de errores que va a depender de la composición haplotípica de la población que se esté estudiando, ya que existen zonas de la filogenia mucho mejor definidas que otras.

El impacto de los errores en las bases de datos de ADNmt, va a ser distinto en los diferentes campos que estudian este genoma. En los estudios de genética de poblaciones los errores van a afectar a cuestiones relacionadas con la evolución del ADNmt, es conocido el debate que existe a cerca de la recombinación del ADNmt y el gran impacto que han tenido errores en su momento publicados y en algunos

casos posteriormente rectificados (Hagelberg *et al.* 1999, 2000). También van a afectar a análisis filogenéticos y demográficos, de manera que se pueden alterar las estimas de determinados eventos poblacionales. Sin embargo, las consecuencias de los diferentes errores en el campo forense pueden ser mucho más graves: pueden producirse falsas exclusiones, y errores en las bases de datos pueden distorsionar la valoración estadística de la prueba del ADNmt (Salas *et al.* 2006).

En el contexto forense, las bases de datos son absolutamente necesarias para interpretar el peso de la evidencia de ADNmt en un contexto criminal o en una identificación.

El peso de la evidencia depende de la frecuencia con la que un determinado perfil aparece en la población de referencia, pero muchas veces los tamaños muestrales de las bases de datos son muy pequeños en relación con la gran variabilidad de la población, puede darse la situación de que la muestra en realidad no sea representativa de la población y las estimas de frecuencias haplotípicas no sean las adecuadas. Para poder estimar estas frecuencias de una manera más apropiada se necesitan grandes bases de datos. La base de datos pública de ADNmt más grande es la de *Scientific Working Group on DNA Analysis Methods* (SWGDAM) que contiene alrededor de 5000 perfiles divididos, el problema es que estos perfiles se encuentran divididos en grupos de datos más pequeños, lo que limita su uso en estimas haplotípicas. La nueva base de datos de ADNmt EMPOP (http://www.empop.org) está diseñada de manera que intenta evitar la adición de perfiles erróneos en la medida de lo posible, e intentará cubrir todas las poblaciones del mundo, actualmente la mayoría de los perfiles incluidos pertenecen a poblaciones de países del oeste de Eurasia.

Otro problema importante a tener en cuenta, relacionado con el uso de las bases de datos mitocondriales, es que cuando se utiliza una

determinada base de datos en un caso criminal, se está asumiendo que la base de datos es representativa de la población de la región, esto normalmente se corresponde con una localización geográfica o con un grupo étnico. Sin embargo, a menudo asumir que el contexto geográfico en el que se comete el crimen está representado adecuadamente en la base de datos es cuestionable.

6. Las enfermedades mitocondriales

6.1. Qué son las enfermedades mitocondriales

Las enfermedades mitocondriales empezaron a describirse en la década de los 60 en pacientes con intolerancia al ejercicio, estudios histoquímicos en tejido muscular de estos individuos revelaron la presencia de acúmulos de mitocondrias, tanto de aspecto normal como anormal. Las áreas con acumulaciones de mitocondrias aparecían de color púrpura y estas fibras de aspecto anormal se denominaron *"ragged-red fibers"* y se convirtieron en un marcador patológico de las miopatías mitocondriales. Sin embargo, pronto fue evidente que en muchos pacientes con RRF la miopatía estaba asociada con signos y síntomas que implicaban al sistema nervioso central y se denominaron encefalomiopatías mitocondriales. Y también pronto quedó claro que la ausencia de RRF no excluía una etiología mitocondrial (Andreu *et al.* 2003; Dimauro, 2004). Desde entonces hasta ahora se ha recorrido mucho camino en el conocimiento de estas enfermedades y aunque actualmente sabemos que las mutaciones de ADN mitocondrial son una causa importante de enfermedad todavía queda por determinar el verdadero impacto de estas mutaciones en la salud humana (Taylor *et al.* 2005).

La mitocondria, como ya se ha mencionado anteriormente, es la principal fuente de energía para todos los tejidos humanos y contiene muchas vías metabólicas: el complejo piruvato deshidrogenasa, el ciclo

de la carnitina, el sistema de la β–oxidación y el ciclo de Krebs. Defectos en cualquiera de estas vías causan enfermedades mitocondriales (Andreu *et al.* 2004). Sin embargo, según Dimauro (2004), por convención el término "enfermedad mitocondrial" se refiere sólo a los desórdenes de la cadena respiratoria mitocondrial o sistema de fosforilación oxidativa (OXPHOS), que engloba la cadena de transporte electrónico y la fosforilación oxidativa que representan el objetivo final del metabolismo oxidativo, que es ser la principal fuente celular de ATP. Esta función exclusiva que la mitocondria desempeña dentro del metabolismo de las células eucariotas es esencial para entender su papel en las enfermedades humanas. La dependencia celular del ATP derivado del sistema OXPHOS hace que la célula sea extremadamente vulnerable al malfuncionamiento mitocondrial y que incluso pequeñas reducciones en la eficiencia de la producción de ATP en determinados tejidos puedan ser suficientes para causar síntomas (McFarland *et al.* 2007).

El sistema OXPHOS es la única ruta celular que está bajo el control tanto del ADN nuclear como del ADN mitocondrial, de manera que las enfermedades mitocondriales humanas engloban tanto mutaciones en el ADN mitocondrial como en el nuclear. Está formado por 5 complejos polipeptídicos (I-V) localizados dentro de la membrana mitocondrial interna.

De las aproximadamente 80 proteínas que forman estos cinco complejos, sólo 13 están codificadas por el ADN mitocondrial, mientras que todas las demás lo están por el nuclear. Los 13 polipéptidos codificados por el ADNmt incluyen 7 de las 48 subunidades del complejo I (ND1-6 y ND4L), 1 de las 11 subunidades del complejo III (citocromo b), 3 de las 13 subunidades del complejo IV (COXI-III) y 2 de las 14 subunidades del complejo V (ATPasa 6 y 8)(McFarland *et al.* 2007)(Figura 7).

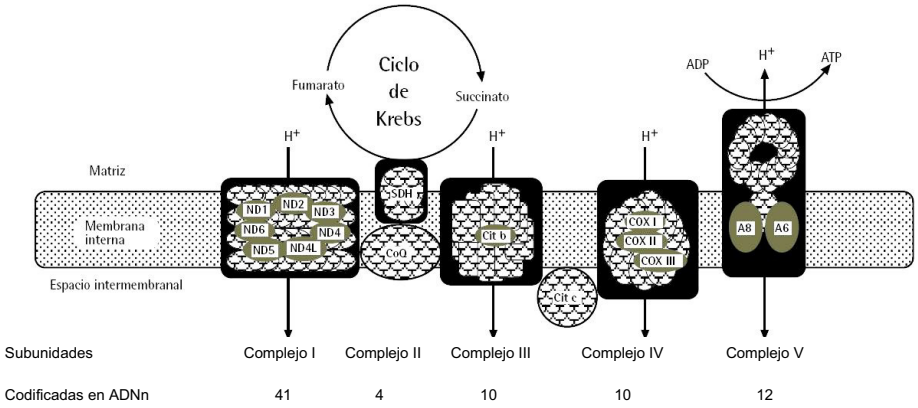


Figura 7: Representación esquemática del sistema OXPHOS y composición de la cadena respiratoria mitocondrial (modificado de Andreu *et al.* 2oo4*).*

Los desórdenes causados por mutaciones en el ADN nuclear son muy numerosos, no sólo porque la mayoría de las subunidades que forman la cadena respiratoria están codificadas por este ADN, sino porque la correcta estructura y funcionamiento de la cadena respiratoria requiere muchos pasos que dependen por completo del núcleo. Así, estas mutaciones podrían clasificarse en cinco grupos (Dimauro, 2004):

a) Mutaciones que afectan a genes que codifican proteínas de la cadena respiratoria, tanto estructurales como de ensamblaje.
b) Defectos en la señalización intergenómica afectarán a factores de los que dependen la homeostasis, la replicación y la trascripción del ADNmt.
c) Mutaciones que afectan al transporte de proteínas codificadas por ADNn desde el citoplasma al interior de la mitocondria.

d) Mutaciones que alteran la membrana interna fosfolipídica de la mitocondria en la que se encuentra la cadena respiratoria.
e) Las mitocondrias se mueven dentro de la célula, se dividen por fisión y se fusionan unas con otras. Alteraciones en estas funciones esenciales también pueden causar enfermedades.

6.2. Mutaciones patogénicas del ADN mitocondrial

Tal y como dice Dimauro en muchos de sus artículos *"la pequeña molécula del ADN mitocondrial presente en todas las células humanas ha resultado ser la verdadera Caja de Pandora de los reordenamientos y las mutaciones patogénicas"*. Desde que en 1988 se reportaron las primeras mutaciones patogénicas del genoma mitocondrial por Holt *et al.* y Wallace *et al.* se han identificado más de 100 mutaciones en esta molécula, y el número de nuevas candidatas aumenta cada día.

Las mutaciones de los 37 genes del ADNmt se pueden agrupar en dos grupos, aquellas que afectan a los genes implicados en la síntesis de proteínas: reordenamientos y mutaciones puntuales en los 2 genes de ARNr y 22 genes de ARNt (ejemplo de estas últimas son MELAS (*mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes*) y MERRF (Epilepsia mioclónica asociada a fibras rojas rasgadas)), y aquellas que afectan a los genes que codifican alguna de las 13 subunidades de la cadena respiratoria, como por ejemplo NARP (neuropatía, ataxia y retinitis pigmentosa) y MILS (síndrome de Leigh heredado por vía materna) (Dimauro *et al.* 2005).

Los reordenamientos incluyen delecciones y duplicaciones. El tamaño de una delección puede variar desde una única base a varias kilobases y puede estar localizada en cualquier parte de la molécula, ocurren de forma heteroplásmica y el grado de heteroplasmía puede variar entre tejidos y a lo largo del tiempo. La mayoría son fenómenos

esporádicos y se cree que ni se heredan de la madre ni se transmiten a la descendencia (Dimauro *et al.*2005). La duplicaciones de ADNmt suelen ser raras y pueden ocurrir solas o asociadas a inserciones, se cree que pueden ser un paso intermedio en la generación de delecciones y no tienen un papel patogénico claro (Schapira, 2006, McKenzie *et al.* 2004). Ejemplos de enfermedades causadas por reordenamientos son el KSS (síndrome de Kearns-Sayre), el PS (síndrome de Pearson), y el PEO (oftalmoplegia externa progresiva).

La relación genotipo-fenotipo en las enfermedades mitocondriales es realmente compleja, ya que la misma mutación puede aparecer asociada a distintos fenotipos, y el mismo fenotipo puede ser el resultado de varias mutaciones diferentes. Además, muchos factores independientes pueden influir en la expresión clínica de una mutación, como la distribución tisular, los niveles de heteroplasmía o que la dependencia de cada órgano del sistema OXPHOS es diferente en cuanto a las necesidades de energía (Schapira, 2006).

Sin embargo, teniendo en cuenta la naturaleza altamente polimórfica del ADNmt, lo que resulta realmente complicado, cuando se detecta un cambio de secuencia de un paciente, es determinar si se trata de un polimorfismo o de una mutación patogénica (Mitchell *et al.* 2005). Así en 2001 Dimauro y Schon proponen un decálogo de criterios para determinar la patogenicidad de una mutación de ADNmt:

- La mutación debe estar ausente en población control de individuos del mismo grupo étnico.
- La mutación debe ocurrir en un sitio conservado evolutivamente y funcionalmente importarte del genoma mitocondrial.
- Heteroplasmía y efecto umbral. La mutación debe ser heteroplásmica: las mutaciones normalmente afectan a alguna pero no a todas las copias del ADNmt de una célula,

un tejido o un individuo, con lo que normalmente el resultado es una mezcla de genomas mitocondriales mutados y no mutados. La expresión clínica de una mutación patogénica de ADNmt está determinada por la proporción relativa de genomas normales y mutados.

La proporción de genomas mutados debe estar en relación con el defecto bioquímico. Dependiendo de la demanda energética de una célula en particular, el nivel de genomas mutados requeridos para producir una expresión fenotípica de la mutación varía (efecto umbral). Este umbral es diferente en tejidos distintos, pero cuando es sobrepasado, la producción de ATP es menor a la demanda de energía.

- Debe haber segregación dentro de la familia: la proporción de genomas mitocondriales mutados debe ser mayor en los tejidos de individuos afectos que en el mismo tejido de miembros de la misma familia no afectos.
- La mutación debe segregar dentro del propio individuo: la proporción de genomas mutados debe ser mayor en los tejidos clínicamente afectados.

Posteriormente, en el año 2005 Dimauro y Davidzon incluyen un criterio más: la mutación debe causar una o varias deficiencias en la cadena respiratoria de los tejidos afectados, o defectos en la síntesis proteica que afecten la respiración celular, que sea demostrable en células cíbridas.

Muchas mutaciones del ADNmt no cumplen muchos de estos criterios, particularmente los referidos a la heteroplasmía y segregación con la enfermedad. Muchas mutaciones pueden aparecer en sitios filogenéticamente conservados y estar ausentes en la población control y sin embargo esto es insuficiente para declararlas patogénicas (McFarland *et al.* 2004a). Otras mutaciones, en cambio, pueden pasar

inadvertidas por el hecho de no encontrarse en un sitio conservado evolutivamente o porque su segregación no pueda demostrarse por estar en familias pequeñas. En cambio, otras variantes neutrales pueden confundirse con mutaciones patogénicas cuando aparecen en pocos individuos relacionados. Todas estas circunstancias ponen de manifiesto la necesidad de considerar otros métodos para definir la patogenicidad de una mutación (McFarland *et al.* 2004b, Mitchell *et al.* 2005).

McFarland *et al.* en 2004 proponen un sistema de puntuación numérica, basado en evidencias funcionales, conservación, frecuencia y presencia de heteroplasmía, (tabla 3) para definir la patogenicidad de mutaciones que afectan al ARNt mitocondrial.

Criterio patogénico		**Puntuación**
Posición evolutivamente conservada	Sin cambio	2
	Cambio único	1
	Más de un cambio	0
Mutación reportada más de una vez	Si	2
	No	0
Presencia de heteroplasmía	Si	2
	No	0
Evidencia histoquímica de enfermedad mitocondrial	Evidencia fuerte	2
	Evidencia débil	1
Defecto bioquímico en los complejos I, III o IV	Si	2
	No	0
Segregación de la mutación con la enfermedad	En un individuo o familia afectados	2
	No segregación	0
Niveles elevados de mutación en fibras COX negativas[1]	Si	3
	No	0
Estabilidad reducida de ARNt mutado o evidencia de patogenicidad en células cíbridas	Si	5
	No	0
Puntuación máxima		20

Tabla 3: Criterios de puntuación de mutaciones de ARNt listadas en MITOMAP.

Con estos criterios de puntuación, los autores analizaron las mutaciones de ARNt listadas en diciembre de 2003 en MITOMAP como

patogénicas. Del total de 91 mutaciones, 66 aparecían como "provisionales", 4 como "poco claras" y otras 4 no estaban claramente descritas ni como mutaciones patogénicas ni como polimorfismos. Algunas posiciones en cambio, aparecían identificadas como posiciones patogénicas y como variantes neutrales. Con el sistema propuesto, la puntuación máxima que puede obtener una mutación es de 20 puntos, todas las mutaciones fueron clasificadas en cuatro grupos:

- Sin duda patogénicas si la puntuación es mayor o igual a 14 puntos.
- Probablemente patogénicas, entre 11 y 13 puntos.
- Posiblemente patogénicas con 7-10 puntos.
- Variantes neutrales si la puntuación obtenida es igual o menor a 6 puntos.

Después de aplicar este sistema de puntuación, el 26% de las mutaciones de ARNt citadas como patogénicas en MITOMAP no se apoyan en evidencias suficientes como para ser consideradas patogénicas y sólo el 16% del total fueron clasificadas como patogénicas sin ninguna duda.

En un planteamiento similar, Mitchell *et al.* (2005) analizan las posiciones clasificadas como patogénicas en MITOMAP en junio de 2004 y algunas posiciones adicionales, en total 50 mutaciones en los genes codificados por ADN mitocondrial del Complejo I, que comprenden aproximadamente 6000 bases del genoma mitocondrial, representan el 38% del ADNmt. Defectos en este complejo suponen la mayoría de las anomalías en pacientes con desórdenes en la cadena respiratoria mitocondrial. Los criterios utilizados y el sistema de puntuación son bastantes similares a los del estudio anterior y se muestran en la tabla 4.

Criterio patogénico		Puntuación
Defecto bioquímico	Demostrado en tejidos afectados	8
	Demostrado en múltiples tejidos	2
Estudios funcionales		7
Mutación reportada más de una vez por laboratorios independientes		5
Presencia de heteroplasmía		5
Segregación de la variante con la enfermedad dentro de una familia		3
Conservación	- 1 si la variante aparece en un mamífero - 1 si la variante aparece en un primate - 1 si hay entre 2 y 4 mamíferos con variantes - 2 si las variantes son de dos clases de aminoácidos - 2 si la variante aparece en bases de datos en ausencia de enfermedad	10
Puntuación máxima		40

Tabla 4: Criterios de puntuación de mutaciones de ADNmt del Complejo I listadas en MITOMAP.

Aplicando este sistema de puntuación a las 50 mutaciones analizadas obtuvieron los siguientes resultados:

- 21 de las mutaciones clasificadas como patogénicas en MITOMAP obtuvieron menos de 10 puntos, con lo que habría pocas evidencias de que fuesen realmente patogénicas. Este sistema de puntuación define los cambios con una puntuación igual o menor que 10 como polimorfismos neutrales.
- 10 fueron clasificadas como "posiblemente patogénicas" con una puntuación entre 10 y 20.
- 3 variantes se clasificaron como "probablemente patogénicas" con puntos comprendidos entre 21 y 29.
- Sólo 16 mutaciones del total de 50 se clasificaron con más de 30 puntos como patogénicas.

6.3. "Novel mutation" no es sinónimo de patogenicidad

Una mutación o un polimorfismo de ADNmt es *novel* si no se ha observado antes, no ha sido reportada en la literatura o no puede ser encontrado en otras fuentes de información disponibles. Sin embargo en la práctica, muchos trabajos que presentan mutaciones nóveles

reducen esta búsqueda a la presencia de la posición en cuestión en la base de datos MITOMAP (http://www.mitomap.org) (Bandelt *et al.* 2006, 2007).

En un estudio de 2006 Bandelt *et al.* escogen un *set* de mutaciones del ADNmt asociadas con dististintas patologías y las comparan con bases de datos públicas de ADNmt: MITOMAP (Brandon *et al.* 2005; http://www.mitomap.org), la base de datos de Uppsala *Uppsala mtDB database* (Ingman and Gyllensten 2006; http://www.gen-pat.uu.se/mtDB/), la base de datos de la región control SWGDAM (http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm); la base de datos *Mitochondrial DNA Concordance database* (Miller *et al.* 1996; http://www.bioanth.cam.ac.uk/mtDNA/index.html), y también en el buscador universal de internet Google.

Además de compararlas con todas las bases de datos citadas, las analizan bajo una perspectiva filogenética teniendo en cuenta la clasificación en haplogrupos y nomenclatura más actualizada.

En esta búsqueda se encuentran con posiciones del genoma mitocondrial que aparecen varias veces en la literatura proclamadas como nóveles, como por ejemplo la transición T9137C que aparece hasta tres veces publicada como novel (De Joanna *et al.* 2000; Máximo *et al.* 2002; Babusiaková *et al.* 2004), además de aparecer referenciada en distintos trabajos de genética de poblaciones asociada a los haplogrupos HV y F2a (Maca-Meyer *et al.* 2001, Kong *et al.* 2003, 2004).

Muchas de las mutaciones que se publican en la literatura como nóveles y patogénicas, muchas veces, además de no ser nóveles no está claro su papel patogénico. En muchas ocasiones con una revisión de la literatura adecuada y con un análisis bajo una perspectiva filogenética, se llega a la conclusión de que muchas de estas mutaciones proclamadas como patogénicas por el simple hecho de afectar al ARN

ribosómico o transferente, o por ser no-sinónimas, son simplemente polimorfismos (Bandelt *et al.* 2006, 2007).

Establecer la patogenicidad de una mutación resulta tal y como se describe en el apartado anterior tremendamente complicado, por lo que debería primar la cautela cuando se asocia una posición determinada del genoma mitocondrial a alguna enfermedad.

6.4. Métodos de genotipado de mutaciones patogénicas

Durante los últimos veinte años se han diseñado distintas técnicas para buscar y genotipar mutaciones patogénicas del genoma mitocondrial. Esta metodología va desde el análisis mediante RFLPs, el análisis de heterodúplex (HDA) o de SSCP, hasta la secuenciación automática de genomas mitocondriales completos (Barros *et al.* 1997; Wallace *et al.* 1999; Salas *et al.* 2001; Dimauro 2007). La principal desventaja de los métodos clásicos de *screening,* tal y como se ha mencionado en apartados anteriores es que muchas mutaciones pueden pasar desapercibidas. La situación ideal sería poder secuenciar el genoma mitocondrial completo de cada paciente sospechoso de padecer una enfermedad mitocondrial, pero al igual que en otros campos como en el de la genética forense, esta estrategia es totalmente inviable en un laboratorio de rutina clínica.

Se han propuesto estrategias novedosas que permiten rastrear el genoma mitocondrial completo, como la endonucleasa mismatch-específica "Surveyor™ Nuclease" (Banwarth *et al.* 2008), que parece ser un sistema bastante eficaz para detectar mutaciones heteroplásmicas en el genoma mitocondrial. Esta estrategia sin embargo, presenta como principal desventaja la imposibilidad de detectar mutaciones homoplásmicas.

Como ya se ha visto en apartados anteriores las técnicas basadas en minisecuenciación han resultado un sistema muy eficaz en otras

áreas de conocimiento como la genética forense o la genética de poblaciones (Crespillo *et al.* 2006; Quintáns *et al.* 2004; Álvarez-Iglesias *et al.* 2007; Brandstätter *et al.* 2007), y recientemente se han publicado trabajos con esta técnica en el ámbito clínico (Filippini *et al.* 2007; Álvarez-Iglesias *et al.* 2008).

Este tipo de metodología debe entenderse sólo como un sistema de *screening,* que no evitará la confirmación de las mutaciones encontradas mediante secuenciación de las dos cadenas del genoma mitocondrial.

JUSTIFICACIÓN Y OBJETIVOS

El estudio de la variabilidad mitocondrial tiene diversas aplicaciones en el ámbito de la genética humana, que van desde la genética forense, la genética de poblaciones, a la genética clínica.

En el ámbito de la genética forense, el genoma mitocondrial se utiliza de forma sistemática en el estudio de restos altamente degradados así como en muestras que no presentan ADN nuclear, tales como los pelos sin bulbo. Los pelos son además un vestigio común en la escena del crimen.

El genoma mitocondrial es el marcador por excelencia en estudios de genética de poblaciones humana (antropología molecular), y existe desde hace más de dos décadas un interés especial por conocer sus patrones de variabilidad que en general reflejan el pasado demográfico de la población. El estudio de la diversidad genética en poblaciones es además de enorme interés en genética forense ya que la existencia de estratificación poblacional puede tener consecuencias importantes en la estima de las frecuencias haplotípicas y por lo tanto en la valoración estadística de la prueba. En genética clínica, la estratificación es también la causa de un alto incremento de falsos positivos en estudios basados en poblaciones (casos-control).

Existe un amplio espectro de mutaciones relacionadas con enfermedades mitocondriales. En el laboratorio de diagnóstico molecular interesa desarrollar técnicas que permitan optimizar el esfuerzo de genotipado y la detección de mutaciones causales. La secuenciación de genomas completos no es una técnica adecuada ya que requiere mucho tiempo y demanda además una alta especialización técnica.

Este documento de tesis se enmarca en este contexto multidisciplinar, que tiene como base el estudio de la variabilidad genética en poblaciones humanas.

OBJETIVOS GENERALES

1. Selección mediante métodos filogenéticos de SNPs codificantes para distintas aplicaciones forenses, médicas y poblacionales.
2. Desarrollo de métodos de genotipado multiplex de SNPs codificantes.
3. Estudio de la variabilidad mitocondrial en sus diversos ámbitos de aplicación en genética humana.

OBJETIVOS ESPECÍFICOS

1. Estudio de la filogenia mitocondrial en su máximo grado de resolución, esto es, el genoma completo, para la correcta selección de SNPs ya sea en cuanto al ámbito de la identificación o en la definición de linajes mitocondriales.
2. Desarrollo de técnicas de genotipado de SNPs mitocondriales en muestras de interés forense con el fin de incrementar el poder de discriminación del test mitocondrial, generalmente asociado al genotipado por secuenciación de las regiones hipervariable I y II.
3. Diseño de multiplexes de SNPs en la región codificante que permitan resolver ramas concretas de la filogenia y asignar con la máxima precisión aquellos perfiles previamente genotipados para la región control.
4. Contribuir al proceso de estandarización y validación del análisis mitocondrial en el ámbito forense.

5. Estudio de la utilidad de la prueba del ADN mitocondrial en casos de mezclas biológicas; especialmente en muestras con contenido limitado de ADN o en mezclas típicamente forenses (semen, saliva, etc.).
6. Valoración de los distintos contenidos de ADN mitocondrial en muestras de distinta naturaleza y su contribución diferencial a las mezclas biológicas.
7. Desarrollo de técnicas de genotipado de SNPs de interés patológico, que permitan reducir sustancialmente el tiempo de genotipado y la eficacia en el laboratorio de diagnóstico molecular.
8. Estudio de la variabilidad mitocondrial en poblaciones humanas, con especial atención a regiones geográficas y poblaciones escasamente caracterizadas y/o linajes mal caracterizados en poblaciones concretas.

MATERIAL Y MÉTODOS

1. Muestras

Para el desarrollo de esta tesis se ha procedido al análisis de marcadores de ADN mitocondrial en un gran número de muestras no emparentadas, con fines genético-forenses, genético-poblacionales y genético-clínicos.

Para la realización de los estudios poblacionales se han empleado:

- ADNs de 282 individuos de distintas comarcas de Galicia, extraídos a partir de muestras recogidas por el Centro de Transfusión de Galicia.
- ADNs de 60 individuos Nativos Americanos de Argentina, extraídos a partir de manchas de sangre.
- ADNs de 61 individuos de Taiwán, extraídos a partir de manchas de sangre.
- ADNs de 102 individuos de Córdoba (Argentina) extraídos a partir de manchas de sangre.
- ADNs de 185 individuos de Colombia extraídos a partir de manchas de sangre.
- ADNs de 90 individuos de El Salvador extraídos a partir de manchas de sangre.
- ADNs de 101 individuos de Cataluña extraídos a partir de manchas de sangre.
- ADNs de 135 individuos de Cantabria extraídos a partir de manchas de sangre.
- ADNs de 8 individuos de País Vasco extraídos a partir de manchas de sangre.

Para la realización de los estudios forenses, se han empleado muestras con diferente origen, pelos, manchas de saliva y mezclas de fluidos biológicos de distintos individuos en diferentes proporciones.

Para el análisis de marcadores clínicos se han empleado ADNs de pacientes recogidos en la Fundación Pública Galega de Medicina Xenómica.

2. Extracción del ADN

2.1. Extracción de ADN a partir de manchas con fenol/cloroformo-alchohol isoamílico

Para la extracción de ADN de manchas de sangre, semen y saliva se ha utilizado un método de extracción con fenol-cloroformo especialmente diseñado para manchas de fluidos biológicos.

- Cortar 1cm^2 de tejido y resuspenderlo 500µl de tampón DLB (Tris-ClH 1M, ClNa 5M, EDTA 0.5M pH 8.0).
- Añadir 50µl de SDS al 10% y 5µl de proteinasa K (20 mg/ml).
- Incubar toda la noche a 56 ºC con agitación suave.
- Añadir 20µl de NaCl 5M y 575µl de fenol: cloroformo: alcohol isoamílico (25:24:1).
- Mezclar por inversión.
- Centrifugar 3 minutos a 12000 r.p.m. y pasar la fase acuosa a otro tubo.
- Añadir 575µl de cloroformo: alcohol isoamílico y mezclar por inversión.
- Centrifugar 3 minutos a 12000 r.p.m.
- Pasar la fase acuosa a otro tubo y añadir 1ml de etanol absoluto frío.
- Mantener durante 15 minutos a –80 ºC y centrifugar 15 minutos a 12000 r.p.m.

- Vaciar el contenido de los tubos y dejar secar el precipitado al aire.
- Resuspender el precipitado en 50 -100 µl agua bidestilada estéril e incubar a 56 ºC con agitación suave entre 2 y 16 horas.
- El rendimiento final es de aproximadamente 35 ng/µl.

2.2. Extracción de ADN de muestras de cabello sin bulbo con resina quelante (Chelex ®)

Para la extracción de ADN de cabello sin bulbo se utiliza Chelex® (Bio Rad, CA, USA), es una resina quelante con alta afinidad por iones metálicos polivalentes. Esta extracción sigue el protocolo descrito por Singer-Sam *et al.* (1998) con algunas modificaciones:

- Añadir a una porción de al menos 2 o 3 cm del pelo 200 µl de Chelex ® al 5% en un tubo *eppendorf* de 1.5 ml.
- Añadir 20 ng de proteinasa K y 7µl de DTT 1M y mezclar suavemente.
- Incubar a 56 ºC entre 30 minutos y 1 hora.
- Agitar en el "Vortex" a alta velocidad durante 5-10 segundos.
- Dar un "spin" en una microcentrífuga durante 10-20 segundos a 10000 -15000 r.p.m.
- Hervir en un baño durante 8 minutos.
- Agitar en el "Vortex" a alta velocidad durante 5-10 segundos.
- Centrifugar durante 3 minutos a 10000 -15000 r.p.m. antes de añadir la muestra a la mezcla de PCR.

2.3. Extracción diferencial de ADN a partir de muestras espermáticas

- Cortar el hisopo e introducir el algodón en un tubo de 15ml estéril.
- Añadir 3ml de H_2O estéril y agitar en el "Vortex" cada 15 minutos durante 2 horas.

- Transferir 1,5ml del volumen del tubo a un tubo *eppendorf* de 2ml.
- Centrifugar el tubo a 12000 r.p.m. durante 5 minutos para colectar la máxima cantidad de material celular. Eliminar el sobrenadante hasta dejar 50µl del precipitado.
- Repetir los pasos anteriores hasta agotar el líquido del tubo que contiene el algodón.
- Añadir 1ml de H_2O estéril al tubo con el algodón y dar un "vórtex".
- Retirar de nuevo el sobrenadante al tubo *eppendorf* y centrifugarlo a 12000 r.p.m. durante 5 minutos.
- Eliminar el sobrenandante y repetir hasta dejar 40-50µl sobre el *pellet*.
- Para aprovechar todo el líquido del algodón colocar una cestilla dentro del tubo *eppendorf* y centrifugar a 12000 r.p.m. durante 10 minutos.
- Desechar la cestilla y el algodón y eliminar sobrenadante hasta dejar 40-50µl sobre el pellet.
- Añadir al pellet 400µl de tampón de extracción y 20µl de proteinasa K (20 mg/ml).
- Mezclar e incubar a 56ºC durante 2 horas.
- Centrifugar 5 minutos a 12000 r.p.m. para precipitar los espermatozoides.
- Trasferir el sobrenadante, que consiste en células femeninas, espermatozoides degradados, etc, a otro tubo *eppendorf* y guardar a 4ºC hasta la purificación del ADN.
- Añadir al pellet de espermatozoides 200 µl de tampón de extracción y centrifugar 10 minutos a 12000 r.p.m.
- Eliminar el sobrenadante y repetir 3 veces los lavados, dejando al final 50µl de volumen.

- Añadir a los espermatozoides 400µl de tampón de extracción, 20µl de proteinasa K y 20µl de DTT 1M.
- Dar un "vortex" e incubar toda la noche a 37ºC con agitación.
- Añadir tanto a la fracción masculina como a la femenina 200µl de fenol: cloroformo: alcohol isoamílico.
- Mezclar por inversión y centrifugar 3 minutos a 12000 r.p.m.
- Retirar el sobrenadante a otro tubo.
- Añadir 200µl de cloroformo: alcohol isoamílico, mezclar, centrifugar 3 minutos a 12000 r.p.m. y retirar el sobrenadante a otro tubo.
- Añadir 500µl de TE a un "Centricon" (MILLIPORE), añadir la fracción masculina y 1ml de TE, por este orden.
- Centrifugar a 1500 r.p.m. durante 30 minutos.
- A la fase acuosa de la fracción femenina se le añaden 2µl de glicógeno (10mg/ml), 40µl de acetato sódico 2M y 1ml de etanol absoluto frío.
- Mezclar por inversión y guardar 30 minutos a -80ºC.
- Centrifugar a 12000 r.p.m. durante 20 minutos, decantar y secar a RT 10-15 minutos.
- Resuspender en 50µl de H_2O estéril o TE.

3. Análisis de la región control

3.1. Amplificación

En la mayoría de las muestras se procedió al análisis de la región control completa (16024-577). En las muestras restantes se analizó la región HV1(16024-16401) dentro de la región control.

La amplificación de la región control se llevó a cabo en dos amplicones, denominados HVS1 (16024-16569) y HVS2 (01-577) respectivamente, de alrededor de 600 pares de bases. Con las siguientes condiciones:

- BSA 0.16µg/µl
- Buffer 1X
- dNTPs 200µM
- Cl_2Mg 1.5mM
- Primer L 0.2µM
- Primer H 0.2µM
- Taq Polimerasa 2.5U
- ADN : 5ng en un volumen final de 25 µl
- Primers HVS1: 15997 L 5´CACCATTAGCACCCAAAGCT-3´
  017H 5´- CCCGTGAGTGGTTAATAGGGT-3´
- Primers HVS2: 16555L 5´- CCCACACGTTCCCCTTAAAT-3´
  599H 5´- TTGAGGAGGTAAGCTACATA-3´

El programa de PCR utilizado en termociclador es el siguiente: 1 minuto a 95ºC seguido de 36 ciclos a 95ºC 10 segundos, 58ºC 30 segundos y 72ºC 30 segundos, seguidos de una extensión final de 10 minutos a 72ºC.

En aquellas muestras en las que no fue posible la amplificación de fragmentos de este tamaño combinamos los primers anteriores con los que se describen a continuación, para obtener amplicones más pequeños:

- 16159L 5´- TACTTGACCACCTGTAGTAC-3´
- 16380L 5´- TCAGATAGGGGTCCCTTGAC-3´
- 16236H 5´- CTTTGGAGTTGCAGTTGATG-3´
- 16401H 5´- TGATTTCACGGAGGATGGTG-3´
- 172L 5´- ATTATTTATCGCACCTACGT-3´
- 370L 5´- CCCTAACACCAGCCTAACCA-3´
- 285H 5´- GGGGTTTGGTGGAAATTTTTTG-3´
- 408H 5´- CTGTTAAAAGTGCATACCGCCA-3´
- 611H 5´-AGTGTATTGCTTTGAGGAGG-3´

Todos los primers utlizados son diseños propios, con la excepción del 15997L (Ward *et al.* 1991), el 16401H y 408H (Vigilant *et al.* 1989) y 599H (Brandstätter *et al.* 2004) (Figura 8).

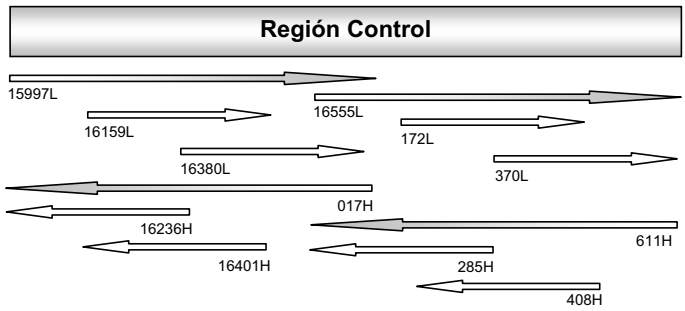


Figura 8: Representación esquemática de los primers utlizados en la amplificación y secuenciación de la región control.

3.2. Purificación post-PCR

La purificación de los productos de PCR se llevó a cabo mediante filtración en placas MultiScreen ® PCR µ96 (MILLIPORE).

Este proceso consiste en una única filtración del producto de PCR, previamente resuspendido en 75 µl de agua bidestilada. Una vez filtrado se recupera el producto que queda sobre la membrana resuspendiéndolo en 20-30 µl de agua bidestilada.

3.3. Reacción de secuenciación

Para la secuenciación de las muestras se emplearon dos kits diferentes: dRhodamine Terminator Cycle Sequencing Kits (AB) y BigDye® Terminator v3.1 Cycle Sequencing Kits (AB), ambos basados en el método de secuenciación de Sanger. Los dos kits incluyen una ADN Polimerasa termoestable modificada, sin actividad exonucleasa 3´-5´, dNTPs y ddNTPs marcados con cuatro fluorocromos distintos.

En el caso de las muestras secuenciadas con dRhodamine, en un volumen final de 5µl se incorporaron 2.5µl de producto de PCR purificado, 2µl de kit y 3.2 pmoles de primer. El programa utilizado en

el termociclador fue: 15 minutos a 95ºC seguidos de 36 ciclos de 15 segundos a 96ºC de desnaturalización, 30 segundos a 55ºC de anillamiento y 30 segundos a 70ºC de extensión; con una extensión final de 10 minutos a 72ºC.

En el caso de las muestras secuenciadas con BigDye®, en un volumen final de 10µl se incoroporaron 3-4µl de producto de PCR purificado, 0.5µl de kit, 2µl de buffer 5X, 3.2 pmoles de primer y H_2O hasta completar el volumen final. El programa utilizado en el termociclador fue: 96ºC durante 1 minuto, seguidos de 35 ciclos de 96ºC durante 10 segundos, 50ºC durante 5 segundos y 60ºC durante 4 minutos.

3.4. Purificación post-secuenciación

Para eliminar los ddNTPs no incorporados durante la reacción de secuencia y así evitar su migración durante la electroforesis, en el caso de las muestras secuenciadas con dRhodamine (AB) se utilizó un protocolo de purificación alcohólica:

- Añadir a las muestras 3.75µl de Cl_2Mg y 13.75µl de etanol absoluto.
- Centrifugar durante 20 minutos a 13000 r.p.m.
- Eliminar al máximo el etanol con una punta de pipeta.
- Añadir 64.75µl de etanol al 70% y centrifugar 15 minutos a 13 r.p.m.
- Eliminar el etanol por decantación primero y a continuación con una punta de pipeta.
- Dejar evaporar por completo el etanol antes de cargarlo en el secuenciador.

En el caso de las muestras secuenciadas con BigDye v3.1 (AB) el protocolo de purificación elegido consiste en una primera purificación enzimática con SAP seguida de una purificación mediante filtración en placa Montage™ SEQ 96 (MILLIPORE):

- Añadir a los 10µl de la reacción de secuencia 1µl de SAP.
- Incubar a 37ºC durante 1 hora y a continuación inactivar el enzima durante 15 minutos a 85ºC.
- Añadir 25µl de la solución de inyección al volumen de la reacción de secuencia y colocar todo el volumen en la placa de filtración.
- Filtrar hasta que todo el volumen desaparece y añadir de nuevo 25µl de la misma solución.
- Filtrar de nuevo y añadir 40µl de solución para hacer un último lavado.
- Añadir 20-30µl de solución a cada pocillo y dejar a temperatura ambiente 5-10 minutos.
- Resuspender suavemente y recuperar todo el volumen a la placa que irá a continuación al secuenciador.

3.5. Electroforesis capilar

Las secuencias purificadas fueron resuspendidas en formamida HiDi™ (AB) en el caso de las muestras secuenciadas con dRhodamine (AB), o en solución de inyección en el caso de las muestras secuenciadas con BigDye (AB), y la electroforesis capilar tuvo lugar en un secuenciador ABI PRISM 3100® Genetic Analyzer (AB) y en un secuenciador ABI PRISM 3130xl® Genetic Analyzer (AB), con las condiciones de los módulos por defecto del secuenciador.

Las secuencias resultantes son analizadas con el programa Sequencing Analysis 5.2 (AB), y alineadas con la Secuencia de Referencia de Cambridge revisada (rCRS) (Andrews *et al.* 1999), utilizando el programa SeqScape 5.2 (AB).

4. Análisis de SNPs de la región codificante

Para el análisis de variantes de la región codificante se diseñaron varias PCR-multiplex para la amplificación conjunta de distintos sets de SNPs, para posteriormente analizarlos mediante la tecnología de

minisecuenciación. En esta técnica una ADN polimerasa es utilizada para la extensión de un primer complementario a la secuencia nucleotídica y que finaliza en su extremo 3´ en la base anterior a la posición polimórfica, incorporando un único nucleósido trifosfato marcado fluorescentemente complementario a la base polimórfica.

4.1. Diseño de primers para multiplex

En todos los diseños cada pareja de primers fue seleccionada de manera independiente utilizando el software de acceso libre, Primer 3 (http://frodo.wi.mit.edu /cgi-bin/primer3/primer3_www.cgi). Todos los primers se eligieron con una temperatura de anillamiento de 60ºC, y además se escogieron amplicones deliberadamente pequeños para favorecer la amplificación de muestras degradadas o con poca cantidad de ADN.

Las secuencias de los oligos elegidos fueron alineadas en la base de datos NCBI (National Centre for Biotechnology Information; http://www.ncbi.nlm.nih.gov/blast/Blast.cgi), para comprobar posibles homologías de todos los primers y asegurar su especificidad.

Finalmente, para comprobar las posibles interacciones entre primers, el conjunto de parejas candidatas se analiza en el software AutoDimer(http://www.cstl.nist.gov/div831/strbase//AutoDimerHomepage/AutDimerProgramHomepage.htm).

4.2. Amplificación de los SNPs

Los SNPs característicos de haplogrupos europeos, inicialmente se amplificaron siguiendo el protocolo descrito en el trabajo publicado (Quintáns *et al.* 2004). Posteriormente se mejoró el diseño permitiendo la incorporación de dos nuevos marcadores, T10463C y A13966G, y se modificaron las condiciones de amplifiación como se describe a continuación.

Todos los SNPs se amplificaron en 14 amplicones en una única reacción de PCR que se llevó a cabo usando 5ng de ADN, 4µl de PCR master mix de QIAGEN kit Multiplex PCR (Qiagen, Hilden, Alemania), 1µl de mix de primers y H_2O hasta completar un volumen final de 10 µl. Las concentraciones finales de cada pareja de primers se indican en la tabla 5. La amplificación se realizó en un termociclador 9700 Thermocycler (AB); después de una preincubación a 95ºC durante 15 minutos, la PCR tiene lugar en 30 ciclos con las siguientes condiciones: 94ºC de desnaturalización durante 30 segundos, 60ºC de anillamiento durante 90 segundos y 72ºC de extensión durante 90 segundos, seguidos de una extensión final a 72ºC durante 15 minutos.

Marcador	**Primer forward y reverse**	**Tamaño (bp)**	**CF (µM)**
3010	CAATAACTTGACCAACGGAACA CGGTCTGAACTCAGATCACGTA	180	0.15
3915/3992	TAGCAGAGACCAACCGAACC GAAGATTGTAGTGGTGAGGGTGT	158	0.15
4216	CTCTACACAACATATTTTGTCACCAAG GGTTTGAGGGGGAATGCTGGAG	195	0.30
4336	GGAGCTTAAACCCCCTTATTTC GATAGGTGGCACGGAGAATTT	80	0.15
4529-4580	CAACCCGTCATCTACTCTACCAT CTTCTGTGGAACGAGGGTTTATT	148	0.30
4769/4793	CCGGACAATGAACCATAACC TGGGTAACCTCTGGGACTCA	118	0.15
7028	CACCGTAGGTGGCCTGACTGGC GTGTAGCCTGAGAATAGGGG	168	0.15
6776	GCTTCCTAGGGTTTATCGTGTG GAGTGTGGCGAGTCAGCTAAA	140	0.15
10398/10400/10463	AAATTGCCCTCCTTTTACCCCTA TGTAAATGAGGGGCATTTGG	224	0.40
10873	CATAATTTGAATCAACACAACCACC GTTAGGGGGTCGGAGGAAAAGGTTG	123	0.10
12308	CTGCTAACTCATGCCCCCATG ATTACTTTTATTTGGAGTTGCACCAAGATT	106	0.40
12705	TGTAGCATTGTTCGTTACATGG AGTTGGAATAGGTTGTTAGCGG	147	0.20
13966	TCACTTTCCTAGGACTTCTAACAGC GGTTAGGTCTAGGAGGAGTAGGG	191	0.20
14766	TCAACTACAAGAACACCAATGACC GGAGGTCGATGATGAGTGG	82	0.40

Tabla 5: Secuencias de los primers utilizados para la amplificación. En la tercera columna se indica el tamaño del amplicón. La última columna indica la concentración final de cada pareja en la reacción.

Para el análisis de los 32 SNPs característicos de haplogrupos Nativo Americanos y del Este de Asia (Álvarez-Iglesias *et al.* 2007), se llevó a cabo una única reacción de amplificación de 18 amplicones con tamaños comprendidos entre 53 y 168 pares de bases. Las condiciones de PCR son iguales a las descritas anteriormente, pero con una temperatura de anillamiento de 58 ºC y con las concentraciones de primers que se indican en la tabla 6.

Marcador	**Primer forward y reverse**	**Tamaño (pb)**	**CF (µM)**
8281-8289del	TAGGGCCCGTATTTACCCTAT AAGAGGTGTTGGTTCTCTTAATCTTT	110	0.5
1119	GCTAAGACCCAAACTGGGATT GTTCTGGCGAGCAGTTTTGT	85	0.5
1719/1736	CCAAACCCACTCCACCTTACT GCGCCAGGTTTCAATTTCTA	88	0.2
2092	TGCCCACAGAACCCTCTAAA TGTCCAAAGAGCTGTTCCTCT	70	0.5
8414	TACTACCGTATGGCCCACCA GGGAGGTAGGTGGTAGTTTGTG	97	0.2
3547/3552	CATCTACCATCACCCTCTACATCA ACCAGGGGGTTGGGTATG	94	0.4
3970/4071	GGCTTCAACATCGAATACGC GAAGTAGGGTCTTGGTGACAAAA	169	0.4
4491	ATTAATCCCCTGGCCCAAC GATGAGTGTGCCTGCAAAGA	57	0.2
4820/4824/4883	TAGCCCCCTTTCACTTCTGA GGCTTACGTTTAGTGAGGGAGA	135	0.4
5417	CCTCAATCACACTACTCCCCATA GGGGTGGGTTTTGTATGTTC	73	0.5
9824	TTTTGTAGCCACAGGCTTCC GAGGAAAGTTGAGCCAATAATGA	55	0.4
10310/10398/14000	CCATGAGCCCTACAAACAACT TGAGTCGAAATCATTCGTTTTG	159	0.5
11959/11969/12007	AACCACGTTCTCCTGATCAAA AATGTGGTGGGTGAGTGAGC	120	0.5
12338/1258	TGGTCTTAGGCCCCAAAAAT GGGGAATTAGGGAAGTCAGG	98	0.4
12705/12714/12771	CCCAAACATTAATCAGTTCTTCAA CTCGGGCGTATCATCAACTG	152	0.4
13759	GGAAGCCTATTCGCAGGATT GTTGTTTGGAAGGGGGATG	65	0.4
14502/14569	GACAACCATCATTCCCCCTA CTATTTATGGGGGTTTAGTATTGATT	125	0.4
15487/15335	CGGCTTACTTCTCTTCCTTCTCT ATATCATTCGGGCTTGATGTG	130	0.5

Tabla 6: Secuencias de los primers utilizados para la amplificación. En la tercera columna se indica el tamaño del amplicón. La última columna indica la concentración final de cada pareja en la reacción.

Para el análisis de las 21 mutaciones se llevó a cabo una única amplificación de 13 amplicones con tamaños comprendidos entre los 86 y los 197 pares de bases. Las condiciones de amplificación fueron iguales a las del multiplex anterior pero con las concentraciones finales de primers descritas en la tabla 7.

Mutación	Primer forward y reverse	Tamaño (pb)	CF (µM)
3243	TATACCCACACCCACCCAAG GGCCATGGGTATGTTGTTAAG	119	0.2
3460	CCGAACGAAAAATTCTAGGC GCGGTGATGTAGAGGGTGAT	154	0.2
3697	GCCTAGCCGTTTACTCAATCC GCCTAGCCGTTTACTCAATCC	95	0.2
3946/3949	TAGCAGAGACCAACCGAACC GAAGATTGTAGTGGTGAGGGTGT	158	0.2
7445	CCCTACCACACATTCGAAGAA TGGCTTGAAACCAGCTTTG	86	0.2
8993	AATGCCCTAGCCCACTTCTT AGGTGGCCTGCAGTAATGTT	140	0.2
9176	AAATCGCTGTCGCCTTAATC TCATTAGGAGGGCTGAGAGG	155	0.3
10158/10191	TCAACACCCTCCTAGCCTTA GGGTAAAAGGAGGGCAATTT	197	0.5
10663	ACACCCACTCCCTCTTAGCC GGCCATATGTGTTGGAGATTG	111	0.3
11777/11778/11832	CACGGGCTTACATCCTCATT GGGGGTAAGGCGAGGTTAG	158	0.2
12706	TGTAGCATTGTTCGTTACATGG AGTTGGAATAGGTTGTTAGCGG	147	0.3
13513/13514	ATTGGCAGCCTAGCATTAGC CAGGGAGGTAGCGATGAGAG	132	0.3
14459/14482/ 14484/14487	CTCCATCGCTAACCCCACTA TTCTGAATTTTGGGGGAGGT	170	0.4

Tabla 7: Secuencias de los primers utilizados para la amplificación. En la tercera columna se indica el tamaño del amplicón. La última columna indica la concentración final de cada pareja en la reacción.

El genotipado de los 71 SNPs de la región codificante de ADNmt que caracterizan las principales ramas filogenéticas dentro del haplogrupo R0, se llevó a cabo en tres reacciones multiplex usando 5ng de ADN, 4µl de PCR master mix de QIAGEN kit Multiplex PCR (Qiagen,

Hilden, Alemania), 1µl de mix de primers y H_2O hasta completar un volumen final de 10 µl. Las concentraciones finales de cada pareja de primers en cada reacción se indican en las tablas 8, 9 y 10.

SNP	Primer forward y reverse	Tamaño (pb)	CF (µM)
709/750	GGCTCACATCACCCCATAAA CGTTTTGAGCTGCATTGCT	161	0.2
2581/2706	GCCTGCCCAGTGACACAT GCTCCATAGGGTCTTCTCGT	195	0.2
3010	CAATAACTTGACCAACGGAACA CGGTCTGAACTCAGATCACGTA	179	0.4
3796/3847	TCAACATTACTAATAAGTGGCTCCTTT GGTTCGGTTGGTCTCTGCTA	135	0.2
4550/4580	CAACCCGTCATCTACTCTACCAT CTTCTGTGGAACGAGGGTTTATT	148	0.2
6253/6296/6365	TGACTCTTACCTCCCTCTCTCC GATGAAATTGATGGCCCCTA	189	0.2
6776	GCTTCCTAGGGTTTATCGTGTG GAGTGTGGCGAGTCAGCTAAA	140	0.4
7337	GGCTCATTCATTTCTCTAACAGC TCCAGGTTTATGGAGGGTTC	110	0.4
10810	GCTAAAACTAATCGTCCCAACA AATTAGGCTGTGGGTGGTTG	97	0.2
12858/12957	CAACACAGCAGCCATTCAAG GAGGCCTAGTAGTGGGGTGA	157	0.2
13708/13759	AACGAAAATAACCCCACCCTA GTTGTTTGGAAGGGGGATG	113	0.2
14365/14470	CCACCCCATCATACTCTTTCA TAGGGGGAATGATGGTTGTC	159	0.3
14766/14770	TCAACTACAAGAACACCAATGACC GGAGGTCGATGATGAGTGG	82	0.2
15218	ACTATCCGCCATCCCATACA GGGCAAGATGAAGTGAAAGG	110	0.4

Tabla 8: Secuencias de los primers utilizados para la amplificación. En la tercera columna se indica el tamaño del amplicón. La última columna indica la concentración final de cada pareja en la reacción.

SNP	Primer Forward y Reverse	Tamaño (pb)	CF (µM)
1438	AACTTAAGGGTCGAAGGTGGA AGGGCCCTGTTCAACTAAGC	66	0.1
2259	TCAAGCTCAACACCCACTACC TGCGGAGGAGAATGTTTTCA	131	0.2
5250/5263	ATTCCATCCACCCTCCTCTC GGTGGGGATGATGAGGCTAT	111	0.6
8869	GGACTCCTGCCTCACTCATTT AAGTGGGCTAGGGCATTTTT	128	0.4
8994	AATGCCCTAGCCCACTTCTT AGGTGGCCTGCAGTAATGTT	140	0.1
9336	GCCATGTGATTTCACTTCCA GTGGCCTTGGTATGTGCTTT	117	0.2
10166/10211	ACTACCACAACTCAACGGCTACA AGGGGTAAAAGGAGGGCAAT	145	0.2
11140	CATTCACAGCCACAGAACTAATCAT GTTCTGGCTGGTTGCCTCAT	99	0.2
11719	CAGCCATTCTCATCCAAACC GCGTTCGTAGTTTGAGTTTGC	113	0.3
12308	CTGCTAACTCATGCCCCCATG ATTACTTTTATTTGGAGTTGCACCAAGATT	106	0.3
12438	CCACCCTAACCCTGACTTCC GTGGATGCGACAATGGATTT	106	0.1
12705	TGTAGCATTGTTCGTTACATGG AGTTGGAATAGGTTGTTAGCGG	147	0.2
13101/13105	CAGCCCTACTCCACTCAAGC TGGGCTATTTTCTGCTAGGG	83	0.2
14869/14872	CAACATCTCCGCATGATGAA AGGCGTCTGGTGAGTAGTGC	104	0.2
15452A	AGACGCCCTCGGCTTACTT GTCGCCTAGGAGGTCTGGTG	78	0.2
15773	CCGCAGACCTCCTCATTCTA CGGATGCTACTTGTCCAATG	81	0.1
15833/15904	CCCTTTTACCATCATTGGACA AAAGGTTTTCATCTCCGGTTT	162	0.2

Tabla 9: Secuencias de los primers utilizados para la amplificación del multipléx número 2. En la tercera columna se indica el tamaño del amplicón. La última columna indica la concentración final de cada pareja en la reacción.

SNP	Primer Forward y Reverse	Tamaño (pb)	CF (µM)
951/961G	TCACACGATTAACCCAAGTCA ACTCAGGTGAGTTTTAGCTTTATTG	87	0.2
3915/3936/3992	TAGCAGAGACCAACCGAACC GAAGATTGTAGTGGTGAGGGTGT	158	0.6
4310/4336	AGCATTCCCCCTCAAACCTA TTTTGGATTCTCAGGGATGG	127	0.4
4727/4745/4769/4793	TCCTTCTAATAGCTATCCTCTTCAACA TGGGTAACCTCTGGGACTCA	154	0.4
7028	CACCGTAGGTGGCCTGACTGGC GTGTAGCCTGAGAATAGGGG	168	0.4
7645	ACATGCAGCGCAAGTAGGTC AAAATGATTATGAGGGCGTGA	90	0.2
8269/8271	TAGGGCCCGTATTTACCCTAT AAGAGGTGTTGGTTCTCTTAATCTTT	110	0.2
8473/8592/8598/8602	CCCAACTAAAAATATTAAACACAAACT GGAGGTGGGGATCAATAGAG	193	0.2
9066/9088/9150	CCTACTCATGCACCTAATTGGA GGCTTACTAGAAGTGTGAAAACGTA	155	0.3
10044	CCGTTAACTTCCAATTAACTAGTTTTG AAGGCTAGGAGGGTGTTGATT	91	0.6
10394	CCATGAGCCCTACAAACAACT TGAGTCGAAATCATTCGTTTTG	159	0.3
13404	TATGTGCTCCGGGTCCATC TGGTGAGGGAGGTTGAAGTG	104	0.2

Tabla 10: Secuencias de los primers utilizados para la amplificación del multiplex número 3. En la tercera columna se indica el tamaño del amplicón. La última columna indica la concentración final de cada pareja en la reacción.

4.3. Purificación post-PCR, reacción de minisecuenciación y purificación post-extensión.

Una vez amplificados y chequeados en geles de poliacrilamida, es necesario purificar los productos de PCR con el fin de eliminar los primers sobrantes y los dNTPs no incorporados durante la amplificación. Esta purificación se llevó a cabo con ExoSAP IT (Amershan Biosciences): a 1µl de producto de PCR se añadió 0,5 µl del enzima. La digestión tiene lugar durante 15 minutos a 37ºC y el enzima se inactiva a 80ºC durante otros 15 minutos.

Tras la purificación se llevó a cabo la reacción de minisecuenciación con el kit SNaPshot ™ (AB). En este kit, el enzima utilizado es la AmpliTaq® FS, que es una enzima termoestable y sin actividad exonucleasa 3´-5´, para evitar que esta actividad pueda llegar a ser mayor que la polimerasa y así degradar el primer. Además incluye ddNTP en lugar de dNTPs, de manera que cuando la polimerasa incorpora un ddNTP se para la reacción. Cada ddNTP tiene asociado un fluorocromo diferente (tabla 11) que permite saber qué base se ha insertado y así conocer la identidad del marcador interrogado. Esto, sumado al uso de sondas de distintos tamaños permite detectar varios marcadores en una única reacción. La variación de los tamaños de estas sondas se consigue con la adicción de colas no homólogas. Así, el tamaño de la sonda definirá la posición del polimorfismo y el ddNTP fluoresecente facilitará la identidad del nucleótido.

ddNTP	Marcador	Color
A	d6G	VERDE
C	dTAMRA™	NEGRO
G	dR110	AZUL
T	dROX™	ROJO

Tabla 11: Fluorocromos utilizados en el kit de SNaPshot que permiten la detección de los ddNTPs en una misma reacción.

Al producto de PCR purificado se añaden 2-4µl de kit, 1µl de mix de sondas y H_2O hasta completar un volumen final de 10µl. Las sondas utilizadas en cada multiplex así como sus concentraciones finales utilizadas se especifican en las tablas que se presentan a continuación. La reacción de minisecuenciación se llevó a cabo en un termociclador 9700 Thermocycler (AB), en 25 ciclos con 96ºC de desnaturalización durante 10 segundos, 50ºC de anillamiento durante 5 segundos, seguidos de una extensión a 60ºC durante 30 segundos.

Tras la reacción de minisecuenciación es necesario eliminar los ddNTPs no incorporados para evitar que migren durante la electroforesis y produzcan ruido de fondo, así todo el producto obtenido fue tratado con 1µl de SAP (Amershan Biosciences) a 37ºC durante 60 minutos, seguidos de 15 minutos a 80ºC que inactivarán el enzima.

SNP	Primer de minisecuenciación[1]	Tamaño (pb)	Alelos	Cadena	CF (µM)
3010	GGAGATGTTGGATCAGGACATCCC	24	G-A	L	0.10
3915	$(gact)_2$TAAGCCTGAGACTAGTTCGGACTC	32	G-A	H	0.20
3992	$(gact)_4$tCCCTATTCTTCATAGCCGAATACA	45	C-T	L	0.30
4216	$(c)_4$TACCACTCACCCTAGCATTACTTATATGA	33	T-C	L	0.20
4336	$(gact)_7$ctGCTTAAACCCCCTTATTTCTAGGAC	55	T-C	L	0.20
4529	CTTTGCAGGCACACTCATCAC	21	A-T	L	0.15
4580	TTACCTGAGTAGGCCTAGAAATAAACAT	28	G-A	L	0.10
4769	$(gact)_3$tGGGCTATTCCTAGTTTTATTGCTATAGC	40	A-G	H	0.30
4793	$(gact)_7$ctACTCAGAAGTGAAAGGGGGC	50	A-G	H	0.40
6776	$(gact)_5$gacCGTGTGTCTACGTCTATTCCTACTGTAAATAT	55	T-C	H	0.30
7028	TACACGACACGTACTACGTTGTAGC	25	C-T	L	0.10
10398	$(gact)_7$ATGAGTGACTACAAAAAGGATTAGACTGA	58	G-A	L	0.50
10400	$(c)_{10}$CGTTTTGTTTAAACTATATACCAATTC	37	G-A	H	0.50
10463	$(gact)_9$ctATGTAAATGAGGGGCATTTGGTA	61	T-C	H	0.30
10873	$(c)_{15}$GTTGTTGTTGATTTGGTTAAAAAATAGTAG	45	A-G	L	0.10
12308	$(c)_{22}$CATTGGTCTTAGGCCCCAA	41	A-G	L	0.40
12705	$(c)_{17}$AACATTAATCAGTTCTTCAAATATCTACTCAT	49	C-T	L	0.30
13966	$(gact)_9$gaCAATCCCCTATCTAGGCCTTCTT	61	A-G	L	0.30
14766	AATGACCCCAATACGCAAAA	20	C-T	L	0.30

Tabla 12: Secuencia de las sondas utlizadas en la reacción de minisecuenciación. [(1)] En minúsculas se señala la cola no específica del primer. La quinta columna denota la cadena sobre la que se produce el anillamiento del primer. CF: concentración final en la reacción de cada sonda.

La minisecuenciación de los SNPs de poblaciones del Este de Asia y Nativo Americanos se realizó en una única reacción con sondas con tamaños comprendidos entre 19 y 67 pares de bases.

SNP	Primer de minisecuenciación[1]	Tamaño (pb)	Alelos	Cadena	CF (μM)
1119	(gact)$_8$gTTAGCCCTAAACCTCAACAGTTAAA	58	T - C	L	0.40
1719	gactgacACTCCACCTTACTACCAGACAACCTTA	34	G - A	L	0.15
1736	CCTTAGCCAAACCATTTACCCA	22	A - G	L	0.40
2092	(gact)$_4$ACCCTCTAAATCCCCTTGTAAATTTAA	43	C - T	L	0.50
3547	CCCGACCTTAGCTCTCACC	19	A - G	L	0.15
3552[a]	(gact)$_{10}$gacCTTAGCTCTCACCATCGC	61	T - A	L	0.40
3970	(gact)$_6$gaGTTTGTGTATTCGGCTATGAAGAATA	52	C - T	H	0.20
4071	gactgaCCGCACTCTCCCCTGAACTCTA	28	C - T	L	0.20
4491	CCTGCAAAGATGGTAGAGTAGATGA	25	G - A	H	0.20
4820	GAGGGGTGCCTTGGGTAAC	19	G - A	H	0.40
4824	CTTTCACTTCTGAGTCCCAGAGGTT	25	A - G	L	0.20
4883	TCACATGACAAAAACTAGCCCC	22	C - T	L	0.20
5417	(gact)$_9$CCATATCTAACAACGTAAAAATAAAATGACA	67	G - A	L	0.40
9bp del	(gact)$_8$gacGTATTTACCCTATAGCACCCCCTCTA	61	C - G[3]	L	0.30
8414	(gact)$_3$gacGCCCACCATAATTACCCCCATA	37	C - T	L	0.20
9824	(gact)$_3$GCACAGGCTTCCACGGACT	31	T - C	L	0.20
10310	(gact)$_7$gaCATGAGCCCTACAAACAACTAACCT	55	G - A	L	0.20
10398	(gact)$_7$gATGAGTGACTACAAAAAGGATTAGACTGA	58	A - G	L	0.40
10400	(c)$_{11}$GTTTTGTTTAAACTATATACCAATTC	37	C - T	H	0.40
11959	(gact)$_5$GTCACTCTCCTACTTACAGGACTCAACAT	49	A - G	L	0.30
11969	(gact)$_4$GACTCAACATACTAGTCACA	46	G - A	L	0.40
12007	(gact)$_4$CCTCTACATATTTACCACAACACAATG	43	G - A	L	0.15
12338	(gact)$_8$GGTGCAACTCCAAATAAAAGTAA	55	T - C	L	0.50
12358	(gact)$_6$GAGGAAGTCAGGGTTAGGGTGG	46	A - G	H	0.30
12705	(c)$_{17}$AACATTAATCAGTTCTTCAAATATCTACTCAT	49	C - T	L	0.30
12714	(gact)$_5$CAGTTCTTCAAATATCTACTCATCTTCCTAAT	52	T - C	L	0.50
12771	(gact)$_2$CAACTGTTCATCGGCTGAGA	28	G - A	L	0.15
13759	(gact)$_4$GTCATTACTAACAACATTTCCCCC	40	G - A	L	0.15
14502	(gact)$_4$GAACCATCATTCCCCCTAAATAAA	40	T - C	L	0.50
14569	(gact)$_2$GATAATAACACACCCGACCACACC	32	G - A	L	0.15
15487T	(gact)$_{11}$GCTGGGTCGCCTAGGAGGTC	64	A - T	H	0.40
15535	(gact)$_2$GATTATACCCTAGCCAACCCCTTAAA	34	C - T	L	0.30

Tabla 13: Secuencia de las sondas utilizadas en la reacción de minisecuenciación. [(1)] En minúsculas se señala la cola no específica del primer. La quinta columna denota la cadena sobre la que se produce el anillamiento del primer. CF: concentración final en la reacción de cada sonda.

La minisecuenciación de mutaciones patogénicas se realizó en una única reacción con sondas con tamaños comprendidos entre 25 y 73 pares de bases. Tal y como se refleja en la siguiente tabla, en tres ocasiones, para genotipar las posiciones 3949, 14484 y 14487 fue necesaria la utilización de primers degenerados debido a la posibilidad de co-existencia de dos bases en algunas posiciones de la secuencia de la sonda utilizada, lo que provocaría dificultades en el momento de anillamiento de la sonda.

Mutación	Primer de minisecuenciación[1]	Tamaño (pb)	Alelos	Cadena	CF (μM)
3243	(gact)$_4$gaACAGGGTTTGTTAAGATGGCAG	40	A-G	L	0.2
3460	(gact)$_2$gGCTACTACAACCCTTCGCTGAC	31	G-A	L	0.2
3697	(gact)$_9$ctAAACTCAAACTACGCCCTGATC	60	G-A	L	0.2
3946	cGAACTAGTCTCAGGCTTCAACATC	25	G-A	L	0.2
3949	(gact)$_7$gaAGTCTCAGGCTTCAACATCRAA	52	T-C	L	0.2
7445	(gact)$_5$TCGAAGAACCCGTATACATAAAATCTAG	48	A-G/C	L	0.2
8993	(gact)$_{12}$gaCCTACTCATTCAACCAATAGCCC	73	T-G/C	L	0.2
9176	(gact)$_{10}$gATCCAAGCCTACGTTTTCACACTTC	67	T-C/G	L	0.2
10158	ACAACTCAACGGCTACATAGAAAAA	25	T-C	L	0.2
10191	(gact)$_4$AGTGCGGCTTCGACCCTATA	36	T-C	L	0.3
10663	gactgCAATATTGTGCCTATTGCCATACTAG	31	T-C	L	0.2
11777	(gact)$_5$gCAAACTACGAACGCACTCACAGT	44	C-A	L	0.2
11778	(gact)$_9$gaGAAGTCCTTGAGAGAGGATTATGATG	64	G-A	H	0.3
11832	(gact)$_8$ gacTCAAACTCTACTCCCACTAATAGCTTTTT	64	G-A	L	0.1
12706	(gact)$_9$GCGGTAACTAAGATTAGTATGGTAATTAGGA	52	T-C	H	0.2
13513	(gact)$_8$TTCCTCACAGGTTTCTACTCCAAA	56	G-A	L	0.1
13514	(gact)$_{10}$gTGCGGTTTCGATGATGTGG	60	A-G	H	0.2
14459	(gact)$_2$gacCTCAGGATACTCCTCAATAGCCATC	36	G-A	L	0.2
14484	(gact)$_3$ATCGCTGTAGTATATCCAAAGACAACVA	40	T-C	L	0.5
14487	(gact)$_7$gCTGTAGTATATCCAAAGACAACCAYCA	56	T-C	L	0.6
14482	(gact)$_{10}$GCACATCGCTGTAGTATATCCAAAGACAAC	70	C-A/G	L	0.6

Tabla 14: Secuencia de las sondas utilizadas en la reacción de minisecuenciación. [(1)] En minúsculas se señala la cola no específica del primer, en las sondas correspondientes a las mutaciones 3949, 14484 y 14487 se utilizan primers degenerados que se señalan siguiendo el código IUB. La quinta columna denota la cadena sobre la que se produce el anillamiento del primer. CF: concentración final en la reacción de cada sonda.

La minisecuenciación de los 71 SNPs pertenecientes al haplogrupo R0 se llevó a cabo en tres reacciones, con las sondas que se describen en las siguientes tablas (tabla 15, tabla 16, tabla 17).

SNP	Primer de minisecuenciación[1]	Tamaño (pb)	Alelos	Cadena	CF
709	$(gact)_3$TTACACATGCAAGCATCCCC	32	G-A	L	0.2
750	CTCTAAATCACCACGATCAAAAGG	24	A-G	L	0.2
2581	TGATTATGCTACCTTTGCACGGT	23	A-G	H	0.2
2706	$(gact)_2$gAGGGTCTTCTCGTCTTGCTGTGT	32	A-G	H	0.2
3010	$(gact)_3$gAACCTTTAATAGCGGCTGCACCAT	37	G-A	H	0.2
3796	$(gact)_2$gaCTAATAAGTGGCTCCTTTAACCTCTCC	37	A-G	L	0.1
3847	$(gact)_5$ATTACTCCTGCCATCATGACCC	42	T-C	L	0.2
4550	$(gact)_{13}$gaGCGCTAAGCTCGCACTGATT	74	T-C	L	0.3
4580	$(gact)_3$gaTTACCTGAGTAGGCCTAGAAATAAACAT	42	G-A	L	0.2
6253	$(gact)_6$gaTGTTCCTGCTCCGGCCTCCACT	48	T-C	H	0.1
6296	$(gact)_6$AACAGGTTGAACAGTCTACCCTCC	48	C-T	L	0.2
6365	$(gact)_7$gaGATGGCCCCTAAGATAGAGGAGAC	54	T-C	H	0.2
6776	$(gact)_6$acCGTGTGTCTACGTCTATTCCTACTGTAAATAT	58	T-C	H	0.3
7337	$(gact)_{10}$TGATTTGAGAAGCCTTCGCTTC	62	G-A	L	0.3
10810	$(gact)_5$gacCAACAATTATATTACTACCATTGACATGACT	54	T-C	L	0.2
12858	$(gact)_8$gacGCAGCCATTCAAGCAATCCTATA	58	C-T	L	0.3
12957	$(gact)_9$CAACAAATAGCCCTTCTAAACGCTAA	62	T-C	L	0.3
13708	$(gact)_{13}$ gaCTACTAAACCCCATTAAAGGCCTG	78	G-A	L	0.3
13759	$(gact)_{10}$TTCTCATTACTAACAACATTTCCCCC	66	G-A	L	0.1
14365	$(gact)_{11}$ gaGTTAGCGATGGAGGTAGGATTGGT	70	C-T	H	0.3
14470A	$(gact)_{13}$gCCTCAATAGCCATCGCTGTAGTATA	78	T-A	L	0.4
14770	$(gact)_{11}$GAATGAGTGGTTAATTAATTTTATTAGGGG	74	C-T	H	0.3
14766	$(gact)_{11}$ gCAATGACCCCAATACGCAAAA	66	C-T	L	0.3
15218	$(gact)_{11}$TCCTCAGATTCATTGAACTAGGTCTG	70	A-G	H	0.3

Tabla 15: Secuencia de las sondas utlizadas en la reacción de minisecuenciación. [(1)] En minúsculas se señala la cola no específica del primer. La quinta columna denota la cadena sobre la que se produce el anillamiento del primer. CF: concentración final en la reacción de cada sonda.

SNP	Primer de minisecuenciación[1]	Tamaño (pb)	Alelos	Cadena	CF
951	CTTTATTGGGGAGGGGGTGAT	21	G-A	H	0.2
961G	(gact)$_{15}$CGTAAAGAGTGTTTTAGATCACCCCC	86	T-G	L	0.06
3915	(gact)$_{2}$GAAGCCTGAGACTAGTTCGGACTC	32	G-A	H	0.2
3936	(gact)$_{14}$gTGCGGCGTATTCGATGTTGAA	78	C-T	H	0.6
3992	(gact)$_{3}$gaCCCTATTCTTCATAGCCGAATACA	38	C-T	L	0.3
4310	TCTGATAAAAGAGTTACTTTGATAGAGTAAATAATAGG	38	A-G	L	0.4
4336	AGGGATGGGTTCGATTCTCAT	21	T-C	H	0.2
4727	(gact)$_{5}$gaTACTCTCCGGACAATGAACCAT	44	A-G	L	0.2
4745	gactAATGAACCATAACCAATACTACCAATCA	32	A-G	L	0.2
4769	(gact)$_{4}$gaACCAATCAATACTCATCATTAATAATCATAAT	50	A-G	L	0.2
4793	(gact)$_{5}$gATAATCATAATAGCTATAGCAATAAAACTAGGAAT	56	A-G	L	0.4
7028	(gact)$_{4}$gacTACACGACACGTACTACGTTGTAGC	44	C-T	L	0.2
7645	(gact)$_{6}$GCTACTTCCCCTATCATAGGAGAGCT	50	T-C	L	0.2
8269	(gact)$_{8}$gacTGAAATAGGGCCCGTATTTACCCTATA	62	G-A	L	0.06
8271T	(gact)$_{15}$GGGCCCGTATTTACCCTATAGC	82	A-T	L	0.3
8473	(gact)$_{6}$gaAAAAATATTAAACACAAACTACCACCTACC	56	T-C	L	0.2
8592	(gact)$_{11}$gacGCCTACCCGCCGCAGTACT	66	G-A	L	0.1
8598	(gact)$_{15}$gacACCCGCCGCAGTACTGATCAT	84	T-C	L	0.4
8602	(gact)$_{13}$ CGCCGCAGTACTGATCATTCTA	74	T-C	L	0.4
9066	(gact)$_{8}$gaGTGTAGAGGGAAGGTTAATGGTTGATAT	62	A-G	H	0.3
9088	(gact)$_{10}$AGTAGAATTAGAATTGTGAAGATGATAAGTGTAG	74	T-C	H	0.2
9150	(gact)$_{12}$gCCTAGAAATCGCTGTCGCCTT	70	A-G	L	0.2
10044	(gact)$_{9}$TTAAGGCGAAGTTTATTACTCTTTTTTGAA	66	A-G	H	0.6
10394	(gact)$_{9}$gaCTGGCCTATGAGTGACTACAAAAAGGATTAGA	70	C-T	L	0.3
13404	(gact)$_{11}$gaATCATCCACAACCTTAACAATGAACAAGATAT	78	T-C	L	0.3

Tabla 16: Secuencia de las sondas utlizadas en la reacción de minisecuenciación. [(1)] En minúsculas se señala la cola no específica del primer. La quinta columna denota la cadena sobre la que se produce el anillamiento del primer. CF: concentración final en la reacción de cada sonda.

SNP	Primer de minisecuenciación[1]	Tamaño (pb)	Alelos	Cadena	CF
1438	$(gact)_{10}$gaGTCGAAGGTGGATTTAGCAGTAAACT	68	A-G	L	0.06
2259	TCCCAAACATATAACTGAACTCCTCA	26	C-T	L	0.2
5250	TTCTTCGATAATGGCCCATTTGGGCA	26	T-C	H	0.2
5263	$(gact)_2$gacGGCTTTTTGCCCAAATGGG	30	C-T	L	0.2
8869	$(gact)_2$gTTATGAGCGGGCACAGTGATT	30	A-G	L	0.6
8994	$(gact)_9$gacGCCTACTCATTCAACCAATAGCCCT	64	G-A	L	0.06
9336	$(gact)_3$gacTTCCACTCCATAACGCTCCTC	36	A-G	L	0.15
10166	$(gact)_4$gaGGCTACATAGAAAAATCCACCCC	41	T-C	L	0.3
10211	$(gact)_4$TCCCCCGCCCGCGTCCCTTT	36	C-T	L	0.15
11140	$(gact)_7$gacCATATTTTATATCTTCTTCGAAACCACACTTAT	64	C-T	L	0.2
11719	$(gact)_7$gaATTCTCATAATCGCCCACGG	50	G-A	L	0.4
12308	$(C)_{24}$ATTGGTCTTAGGCCCCA	41	A-G	L	0.4
12438	$(gact)_6$gCTAACAAAAAAAACTCATACCCCCA	50	T-C	L	0.15
12705	$(gact)_3$gaAACATTAATCAGTTCTTCAAATATCTACTCAT	46	C-T	L	0.2
13101C	$(gact)_7$gaGAAGCGGATGAGTAAGAAGATTCC	54	A-C	H	0.2
13105	$(gact)_4$gaCACTCAAGCACTATAGTTGTAGCAGGA	45	A-G	L	0.2
14869	$(gact)_{13}$gacGCTCACTCCTTGGCGCCTGCCT	77	G-A	L	0.2
14872	$(gact)_{12}$CTCCTTGGCGCCTGCCTGAT	68	C-T	L	0.4
15452A	$(gact)_8$gGCCCTCGGCTTACTTCTCTTC	54	C-A	L	0.4
15773	$(gact)_9$gacCCTGAATCGGAGGACAACCA	59	G-A	L	0.1
15833	$(gact)_9$gacGACAAGTAGCATCCGTACTATACTTCACAACAATC	74	C-T	L	0.2
15904	$(gact)_{10}$TTTCATCTCCGGTTTACAAGACTGGTGTATTA	72	C-T	H	0.1

Tabla 17: Secuencia de las sondas utlizadas en la reacción de minisecuenciación. [(1)] En minúsculas se señala la cola no específica del primer. La quinta columna denota la cadena sobre la que se produce el anillamiento del primer. CF: concentración final en la reacción de cada sonda.

4.4. Electroforesis capilar

Después de la reacción de minisecuenciación tiene lugar la electroforesis en presencia de un marcador interno de tamaño, el GeneScan-120LIZ™ (AB), que consiste en nueve fragmentos marcados fluorescentemente y cuyos tamaños van de 15 a 120 nucleótidos. 1.5µl del producto de minisecuenciación se mezclaron con 0.25 µl del marcador interno y 10 µl de formamida HiDi™ (AB).

En las muestras en las que se analizó el multiplex de SNPs de haplogrupos europeos y en las muestras en las que se analizó el multiplex de SNPs de haplogrupos Nativo Americanos y del Este de Asia, la electroforesis se llevó a cabo en un secuenciador capilar ABI PRISM 3100® Genetic Analyzer (AB) utilizando polímero POP6® (AB), utilizando las condiciones electroforéticas que por defecto trae el propio secuenciador. Los datos resultantes fueron analizados con el programa GeneScan™ Analysis Software versión3.2 (AB).

En las muestras en las que se genotipó el multiplex de mutaciones patogénicas, y en las muestras en las que se genotiparon los tres multiplex del haplogrupo R0, la electroforesis se llevó a cabo en un secuenciador capilar ABI PRISM 3130xl® Genetic Analyzer (AB), utilizando indistintamente los polímeros POP6® y POP7® (AB). Las condiciones electroforéticas también en este caso fueron las del módulo que por defecto trae el secuenciador. Los resultados fueron analizados con el programa GeneMapperTM 3.2 Software (AB).

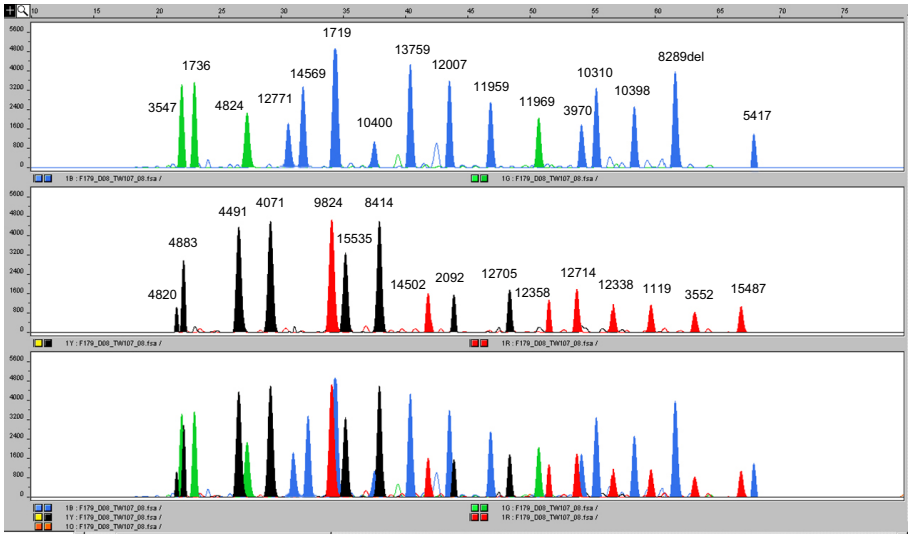


Figura 9: Ejemplo de perfil obtenido en una muestra en la que se analizaron SNPs pertenecientes a poblaciones del Este de Asia. La muestra pertenece al haplogrupo B5b.

RESULTADOS Y DISCUSIÓN

Artículo 1: Typing of mitocondrial coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing

Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A.

(2004) *Forensic Sci Int* 130: 97-111.

Genotipado de SNPs de la región codificante de ADNmt de interés forense y antropológico por minisecuenciación usando SNaPshot

ABSTRACT/RESUMEN:

El desarrollo de nuevas metodologías de alto rendimiento para el análisis de SNPs es una de las áreas más estimulantes de la investigación genética. En este trabajo describimos un ensayo rápido y robusto para genotipar simultáneamente 17 SNPs de la región codificante del ADN mitocondrial (ADNmt), mediante minisecuenciación usando SNaPshot. El SNaPshot es una metodología basada en la extensión de un oligonucleótido sin marcar de una única base con terminadores dideoxi marcados. El set de SNPs implementados en esta reacción multiplex de SNaPshot nos permite asignar los haplotipos del Oeste de Eurasia en sus ramas correspondientes de la filogenia del ADNmt, con especial atención en aquellos haplogrupos que carecen de posiciones diagnóstico en las regiones hipervariables I y II (HVSI/II; los segmentos más comúnmente analizados por secuenciación). Es particularmente interesante el set de SNPs que subdivide el haplogrupo H; el haplogrupo más frecuente en Europa (40-50%) y uno de los peor caracterizados filogenéticamente en las regiones HVSI/II. Además, las posiciones polimórficas seleccionadas en este multiplex incrementan considerablemente el poder de discriminación actual del análisis mitocondrial en el campo forense y también pueden ser usadas como un sistema rápido de "*screening*" previo al análisis completo por secuenciación. Este método ha sido validado en una muestra de 266 individuos y presenta una elevada precisión y robustez, evitando el uso de estrategias clásicas más laboriosas que requieren mucho tiempo (como por ejemplo los RFLPs) y grandes cantidades de ADN.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forensic Science International 140 (2004) 251–257
Forensic Science International
www.elsevier.com/locate/forsciint

Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing

B. Quintáns, V. Álvarez-Iglesias, A. Salas*, C. Phillips, M.V. Lareu, A. Carracedo

Unidad de Genética, Instituto de Medicina Legal, Universidad de Santiago de Compostela, San Francisco s/n, Santiago de Compostela, Galicia 15782, Spain

Received 10 December 2003; accepted 13 December 2003

Abstract

The development of new methodologies for high-throughput SNP analysis is one of the most stimulating areas in genetic research. Here, we describe a rapid and robust assay to simultaneously genotype 17 mitochondrial DNA (mtDNA) coding region SNPs by minisequencing using SNaPshot. SNaPshot is a methodology based on a single base extension of an unlabeled oligonucleotide with labeled dideoxy terminators. The set of SNPs implemented in this multiplexed SNaPshot reaction allow us to allocate common mitochondrial West Eurasian haplotypes into their corresponding branch in the mtDNA skeleton, with special focus on those haplogroups lacking unambiguous diagnostic positions in the first and second hypervariable regions (HVS-I/II; by far, the most common segments analyzed by sequencing). Particularly interesting is the set of SNPs that subdivide haplogroup H; the most frequent haplogroup in Europe (40–50%) and one of the most poorly characterized phylogenetically in the HVS-I/II region. In addition, the polymorphic positions selected for this multiplex reaction increase considerably the discrimination power of current mitochondrial analysis in the forensic field and can also be used as a rapid screening tool prior to full sequencing analysis. The method has been validated in a sample of 266 individuals and shows high accuracy and robustness avoiding both the use of alternative time-consuming classical strategies (i.e. RFLP typing) and the need for high quantities of DNA template.

Keywords: SNaPshot; SNPs; mtDNA; Hypervariable region; Human identification; HVS-I; HVS-II

1. Introduction

In the last few years interest in point mutations also known as single nucleotide polymorphism (SNPs) has increased as a result of their numerous applications in medical genetics [1–3] (i.e. diagnosis of diseases and studies of pharmacological response), in human and evolutionary genetics [3], and also in the forensic field [4]. SNPs located within the non-recombinant part of the Y chromosome (NRY) and in the mitochondrial genome are particularly valuable because they possess uniparental inheritance (paternal and maternal, respectively) and do not suffer recombination; therefore, they are inherited as haplotype-blocks.

In human population studies, this aspect is especially useful because mutations are recorded in the molecule and transferred from generation to generation as a genetic fingerprint of the evolutionary history of the locus. The analysis of this variability allows us to infer the history of human evolution and patterns of migration by studying present lineages ([5,6]; among others).

In forensic genetics, the male-specificity of the Y chromosome and its paternal inheritance make this chromosome especially interesting in cases of mixed stain analysis (male/female cell admixture) and in kinship testing. Similarly, polymorphisms of the mitochondrial DNA genome has been reported as useful for identity testing and, in general, for the analysis of degraded material or samples containing little or no genomic DNA (i.e. skeletal remains and hair shafts) [7–12]. However, the two routinely analyzed mtDNA hypervariable regions (HVS-I/II) provide limited power of discrimination in a forensic context and in many cases, provide

* Corresponding author. Tel.: +34-98-1582327; fax: +34-98-1580336.
E-mail address: apimlase@usc.es (A. Salas).

0379-0738/$ – see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2003.12.005

scant information in evolutionary studies. Since the sequencing of the entire mitochondrial molecule is not practical, the analysis of additional information found in the mtDNA coding region seems to be a good strategy to circumvent this drawback.

Classical methods for studying point mutations in mtDNA and Y chromosome were mainly based on PCR amplification followed by restriction fragment analysis using polyacrilamide gel electrophoresis coupled with silver staining or using agarose gels stained with ethidium bromide. These methods are time-consuming, costly and require large amounts of DNA. During recent years, numerous new methodologies for high-throughput SNP analysis have been developed such as: DNA microarrays, FRET analysis by LightCycler, MALDI-TOF spectrometry, Pyrosequencing, TaqMan probes or Molecular Beacons ([13–15]; among others).

Here, we describe a rapid, robust, and cheap assay to simultaneously genotype 17 mtDNA based SNPs for forensic and anthropological application using a minisequencing reaction and capillary electrophoresis (AB, Applied Biosystems, Foster City, CA, USA). In a comparable study, Brandstätter et al. [16] have recently developed a multiplex reaction for typing a set of coding region SNPs that allow a rapid screening in forensic casework prior to sequencing analysis. The coding region SNPs presented here can be used to increase the discrimination power in forensic analysis after sequencing of both hypervariable regions, as well as for anthropological purposes in order to allocate west European HVS-I/II profiles to a specific haplogroup background. We focus particular attention on SNPs that allow us to resolve the phylogeny in those haplogroups poorly characterized in their HVS-I/II segments.

The technique and the selected markers avoid the use of several time-consuming and expensive strategies when typing individuals of European ancestry for forensic and anthropological purposes. Special attention has been paid to the sub-typing of haplogroup H, which is the most frequent lineage in these populations, and one of the least well characterized phylogenetically when looking at the HVS-I/II segments (the most common mtDNA region analyzed by medical, forensic, and anthropological research groups).

2. Materials and methods

2.1. DNA samples and sequencing

DNA was extracted from fresh blood samples using phenol–chloroform as described by Valverde et al. [17]. DNA was quantified by fluorescent detection with DyNA QuantTM 200 (Amersham Biosciences, Uppsala, Sweden).

A total of 266 individuals from Galicia (the northwest corner of Iberia) were analyzed. The first hypervariable segment (HVS-I) was sequenced from position 16024 to 16400 following conditions described by Salas et al. [18]. A preliminary classification of the sequences into haplogroups was done following phylogenetic criteria according to Richards et al. [19] (see Fig. 1).

2.2. PCR-multiplex

The selected 17 SNPs were combined in two multiplex reactions. Multiplex 1 included a selection of SNPs defining common European haplogroups (see [19] and references therein). Multiplex 2 included exclusively, polymorphisms defining sub-haplogroups inside haplogroup H. The characterization of H sub-lineages is mainly based on the complete sequences available in Herrnstadt et al. [20]. Primers were designed (except for testing position 12308 which used primers previously described by Torroni et al. [21]) in order to adjust the annealing temperatures and amplicon lengths to allow analysis in multiplex reactions (see Table 1). We selected the primers using Primer3 software (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi).

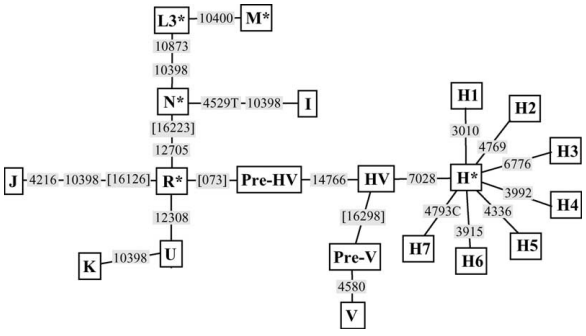


Fig. 1. Network showing the mtDNA skeleton for the mtDNA coding region SNPs selected for the SNaPshot multiplexes. In brackets we indicate some relevant HVS-I/II diagnostic positions.

Table 1
Coding region substitution site (SNP), primer sequence (5′–3′) and final concentration of each primer used for SNP amplifications

SNP[a]	Size (bp)	Forward and reverse primers	FC[b] (μM)
PCR-Multiplex 1			
4216	195	CTCTACACAACATATTTTGTCACCAAG	0.30
		GGTTTGAGGGGGAATGCTGGAG	
4529–4580[c]	148	CAACCCGTCATCTACTCTACCAT	0.30
		CTTCTGTGGAACGAGGGTTTATT	
7028	168	CACCGTAGGTGGCCTGACTGGC	0.15
		GTGTAGCCTGAGAATAGGGG	
10398–10400[c]	224	AAATTGCCCTCCTTTTACCCCTA	0.40
		TGTAAATGAGGGGCATTTGG	
10873	123	CATAATTTGAATCAACACAACCACC	0.10
		GTTAGGGGGTCGGAGGAAAAGGTTG	
12308[d]	106	CTGCTAACTCATGCCCCCATG	0.40
		ATTACTTTTATTTGGAGTTGCACCAAGATT	
12705	147	TGTAGCATTGTTCGTTACATGG	0.20
		AGTTGGAATAGGTTGTTAGCGG	
14766	82	TCAACTACAAGAACACCAATGACC	0.40
		GGAGGTCGATGATGAGTGG	
PCR-Multiplex 2			
3010	180	CAATAACTTGACCAACGGAACA	0.15
		CGGTCTGAACTCAGATCACGTA	
3915–3992[c]	158	TAGCAGAGACCAACCGAACC	0.15
		GAAGATTGTAGTGGTGAGGGTGT	
4336	80	GGAGCTTAAACCCCCTTATTTC	0.15
		GATAGGTGGCACGGAGAATTT	
4769–4793[c]	118	CCGGACAATGAACCATAACC	0.15
		TGGGTAACCTCTGGGACTCA	
6776	140	GCTTCCTAGGGTTTATCGTGTG	0.15
		GAGTGTGGCGAGTCAGCTAAA	

[a] SNPs are termed by position in the mitochondrial genome according to the revised Cambridge reference sequence (rCRS) [22].
[b] FC: final concentration.
[c] Two SNPs are localized in the same amplicon.
[d] Primers reference for this polymorphism: Torroni et al. [21]. Reverse primer was designed by Torroni et al. with a mismatch (underlined) to detect this polymorphism using RFLP.

To test for possible repetitive sequences, primers were aligned with the sequence databases at the National Center for Biotechnology Information (NCBI) using BLAST (http://www.ncbi.nlm.nih.gov/blast/blast.cgi). Co-amplification in the same amplicon is used for those SNPs closely positioned within the mtDNA molecule. The sizes of the PCR products ranged from 80 to 224 bp, which facilitates the analysis of samples that are highly degraded or carrying low quantities of DNA.

We first optimized the concentration of primers, $MgCl_2$ and DNA template. We next performed both multiplexes using 10 ng of DNA template in a 25 μl reaction volume comprising 1X Taq Gold Buffer (Applied Biosystems), 200 μM of each dNTP, 2 mM $MgCl_2$ and 0.5 units of AmpliTaq Gold Polymerase (Applied Biosystems). Primer concentrations for Multiplex 1 are between 0.1 and 0.4 μM (see Table 1), and 0.15 μM of each forward and reverse primer for Multiplex 2. Amplification was carried out in a 9700 Thermocycler (Applied Biosystems). After a 95 °C pre-incubation step for 11 min, PCR was performed in a total of 32 cycles using the following conditions: 94 °C denaturation for 30 s, annealing at 60 °C for 30 s and extension at 72 °C during 1 min, followed by a 15 min of final extension at 72 °C. PCR products were checked by polyacrylamide gel electrophoresis (T9, C5) visualized by silver staining.

2.3. SNaPshot reaction

After amplification, PCR products require purification to remove primers and un-incorporated dNTPs. Post-PCR purification was performed with different protocols varying the concentrations of *ExoI* and SAP (Amersham Biosciences) (data not shown), best results were obtained using a treatment with ExoSapIT (Amershan Pharmacia Biotech): 2.5 μl of PCR product was incubated with 1 μl of ExoSapIT for 15 min at 37 °C followed by 15 min at 80 °C for enzyme inactivation. The minisequencing reaction was performed in a 9700 Thermocycler following the recommendations of the manufacturer (Applied Biosystems): 2 μl of SNaPshot ready reaction mix, 0.2 μM of extension primer for each

Table 2
Minisequencing primer sequences

SNP	Chain	Base change	Size (bp)	Extension primer[a]	FC[b] (μM)
3010	L	G–A	24	CTCGATGTTGGATCAGGACATCCC	0.10
3915	H	G–A	32	TGACTGACTAAGCCTGAGACTAGTTCGGACTC	0.20
3992	L	C–T	45	TGACTGACTGACTGACTGACTCCCTATTCTTCATAGCCGAATACA	0.30
4216	L	T–C	33	CCCCTACCACTCACCCTAGCATTACTTATATGA	0.20
4336	L	T–C	55	CTGACTGACTGACTGACTGACTGACTGACTGCTTAAACCCCCTTATTTCTAGGAC	0.20
4529	L	A–T	21	CTTTGCAGGCACACTCATCAC	0.15
4580	L	G–A	28	TTACCTGAGTAGGCCTAGAAATAAACAT	0.10
4769	H	A–G	40	GACTGACTGACTGGGCTATTCCTAGTTTTATTGCTATAGC	0.30
4793	H	A–C	50	GACTGACTGACTGACTGACTGACTGACTGAACTCAGAAGTGAAAGGGGGC	0.40
6776	H	T–C	55	ACTGACTGACTGACTGACTGACTCGTGTGTCTACGTCTATTCCTACTGTAAATAT	0.30
7028	L	C–T	25	TACACGACACGTACTACGTTGTAGC	0.10
10398	L	A–G	57	GACTGACTGACTGACTGACTGACTGACTATGAGTGACTACAAAAAGGATTAGACTGA	0.50
10400	H	C–T	37	CCCCCCCCCCCGTTTTGTTTAAACTATATACCAATTC	0.50
10873	H	T–C	45	CCCCCCCCCCCCCCCCGTTGTTGTTGATTTGGTTAAAAAATAGTAG	0.10
12308	L	A–G	41	CCCCCCCCCCCCCCCCCCCCCCCCCATTGGTCTTAGGCCCCAA	0.40
12705	L	C–T	49	CCCCCCCCCCCCCCCCCCAACATTAATCAGTTCTTCAAATATCTACTCAT	0.30
14766	L	C–T	20	AATGACCCCAATACGCAAAA	0.30

Second column denotes the chain for primer annealing. Polymorphisms refer to the L strand and follow the standard nomenclature of rCRS [22].
[a] The non-specific primer tail of poly(C) or poly(GACT) is underlined.
[b] FC: final concentration.

SNP and 1 μl of both purified PCR products in a total volume of 7 μl. The reaction mixture was subjected to 25 single base extension cycles of denaturation at 96 °C for 10 s, annealing at 50 °C for 5 s and extension at 60 °C during 30 s.

Extension primers (alternatively termed minisequencing primers) were designed using Primer3. Detection of extension products is based on a combination strategy of colors (using different flourochromes to assay A/G and C/T polymorphism) and extension primer sizes (the length of a primer being modified by the addition of non-homologous tails at the 5′-end). Poly(dC) or poly(dGACT) tails were added at the 5′-end of the specific hybridizing part of the primer depending on the mtDNA variant interrogated. The final sizes ranged between 20 and 57 bp (see Table 2). All minisequencing primers were purified by HPLC to remove incomplete primer synthesis products (Sigma-Genosys Ltd., Pampisford, Cambridgeshire, UK). After minisequencing reactions, a post-extension treatment to remove the 5′-phosphoryl group of the ddNTPs helps to prevent un-incorporated ddNTPs co-migrating with the extended primers and producing high background signal. The final volume (7 μl) was treated with 0.7 μl of SAP (Amersham Biosciences) for 60 min at 37 °C, followed by 15 min at 80 °C for enzyme inactivation.

The minisequencing products (1.5 μl) were mixed with 10 μl of HiDi™ formamide and 0.2 μl of GeneScan-120 LIZ size standard (Applied Biosystems) and electrophoresis was performed on an ABI PRISM 3100® Genetic Analyzer (Applied Biosystems). Minisequencing products were injected electrokinetically for 10 s at 15 kV and electrophoresed for 20 min at 15 kV and 9 μA at 60 °C (default module) in a 36 cm length capillary using the Performance Optimum Polymer 4 (POP4®) with the laser set at a constant power of 9.9 mW. POP6® polymer was also tested in the electrophoresis development. A good resolution was obtained in both cases but the time required with POP4 is shorter. The resulting data was analyzed with GeneScan™ 3.7 Software (Applied Biosystems).

3. Results and discussion

The SNaPshot reaction is based on the dideoxy (ddNTP) single base extension of an unlabeled oligonucleotide at the 3′-end of the base immediately adjacent to the SNP (in the absence of dNTPs in the reaction). Each ddNTP is labeled with different fluorescent dyes and a fifth color is used to label the internal size marker LIZ120. The extended SNaPshot primers used to interrogate different SNPs differ by size and color.

The products of the two multiplexes are combined and minisequenced in a single reaction. Fig. 2 shows an electropherogram for the 17 coding region mtDNA SNPs. With this technology and using a 16 capillary sequencer, 16 samples can be typed for 17 SNPs in less than 25 min, including all the stages of electrophoresis. Thus, 17 SNPs can be analyzed in a maximum number of 920 samples per day.

As expected, the detected size determined by the automated sequencer and the real size of certain products were slightly different. This is due to differences in electrophoretic mobility mainly determined by the length, sequence, and

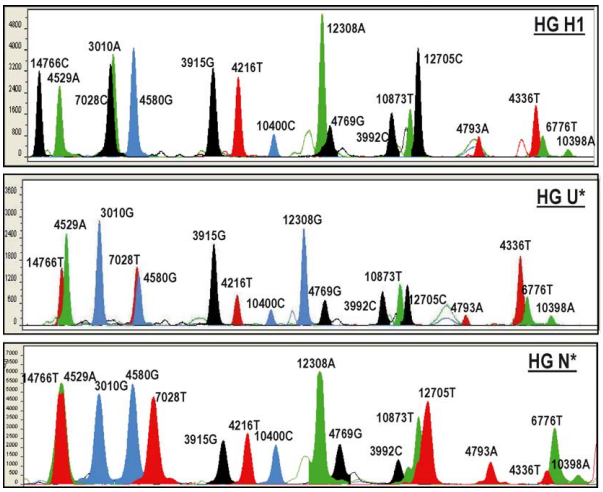


Fig. 2. Electropherograms of 17 SNPs-multiplex assay for three different samples. Polymorphisms refer to the L strand (nomenclature according to the rCRS [22]), including those interrogated using minisequencing primers located in H chain (3915, 10400, 4769, 10873, 4793 and 6776).

the dye used to label the extended primer; the effect of the nucleotide composition being higher in short oligonucleotides than in longer ones. We estimate that to assure correct differentiation of the extended primer, amplicons must differ by at least four nucleotides in length.

Some fragments display stronger fluorescent signal than others in the electropherogram (even after color compensation with the appropriate matrix) due to minisequencing chemistry. Therefore, the same SNP is detected by the automated sequencer at different peak heights depending on the inserted

Table 3
SNP state for the different haplogroups tested using the SNaPshot reaction

rCRS	SNP-HG state																		
	H*	H1	H2	H3	H4	H5	H6	H7	Pre-V	V	HV*	J/T	J	U*	K	R*	M*	N*	I	L3*
3010G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
3915G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G
3992C	C	C	C	C	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
4216T	T	T	T	T	T	T	T	T	T	T	T	C	C	T	T	T	T	T	T	T
4336T	T	T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T
4529A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	I	A
4580G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G
4769A	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
4793A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A
6776T	T	T	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
7028C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T	T	T
10398A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	G	A	G	A	G	G
10400C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C
10873T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T	T	C
12308A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	G	A	A	A	A	A
12705C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	T	T	T
14766C	C	C	C	C	C	C	C	C	C	C	T	T	T	T	T	T	T	T	T	T

The first column indicates the variants corresponding to the rCRS [22]. Nucleotide differences with respect to the rCRS are indicated (underlined) in the next columns according to the profile for each mtDNA haplogroup.

ddNTPs. In general, guanines labeled with dR100 produce a higher signal than other ddNTPs. However, this fact does not unduly affect the readability of the electropherogram.

To test the properties of the set of SNPs implemented in the SNaPshot reaction in a population genetic context, we verified the concordance between HVS-I motifs and coding region SNPs using a phylogenetic approach [19]. In none of the cases did we detect any phylogenetic inconsistence. In addition to this, to test the robustness and accuracy of the minisequencing method, a random set of the samples was typed by using a classical RFLP strategy. There was a complete agreement between SNaPshot and RFLP typing.

The forensic and anthropological usefulness of the multiplex presented here is clearly apparent when looking at the most common European lineage (H; 40–50%). Haplogroup H represents 45% of Galician samples. The use of the SNaPshot reaction allows the sub-typing of this haplogroup into six different sub-lineages with the following frequency profiles: H* (23%), H1 (3010A; 39%), H2 (4769A; 7%), H3 (6776C; 18%), H4 (3992T; 6%), H5 (4336C; 6%), and H6 (3915A; 1%). We did not find any representative of H7 (4793G). The maximum increment in haplotype diversity (HD) corresponds to the most common HVS-I sequence in Europe, the rCRS in HVS-I [22]. There are a total of 44 identical rCRS HVS-I sequences in our Galician sample with the following haplogroup frequency profiles: 18% H*, 41% H1, 5% H2, 20% H3, 14% H4, and 2% HV* (all of them 7028C; HD increase from 0 to 0.737 in this specific set of samples). Another illustrative instance in our dataset is the HVS-I haplotype characterized by a single transition at position 16380, which is shared by four individuals—two harbor substitutions: 7028T-12308G (U*), one: 6776C-7028C (H3), and the other: 7028T-14766T (R*) (Table 3).

The SNaPshot multiplex outlined here provides a rapid, robust, and accurate method for the simultaneous analysis of 17 SNPs on the coding region of the mtDNA genome and increases the discrimination power of current HVS-I/II sequencing.

Acknowledgements

This work was supported by grants from the Ministerio de Educación y Ciencia (DGCYT-P4, BIO2000-0145-P4-02), and Ministerio de Sanidad y Consumo (Fondo de Investigación Sanitaria; Instituto de Salud Carlos III, PI030893; SCO/3425/2002). B.Q. has a research contract with the University of Santiago de Compostela. A.S. is supported by the Isidro Parga Pondal program (Xunta de Galicia).

References

[1] W.E. Evans, M.V. Relling, Pharmacogenomics: translating functional genomics into rational therapeutics, Science 286 (1999) 487–491.

[2] P.C. Gwee, K. Tang, J.M.Z. Chua, E.J.D. Lee, S.S. Chong, C.G.L. Lee, Simultaneous genotyping of seven single-nucleotide polymorphisms in the MDR1 gene by single-tube multiplex minisequencing, Clin. Chem. 49 (2003) 672–676.

[3] M. Jobling, C. Tyler-Smith, New uses for new haplotypes the human Y chromosome, disease and selection, TIG 16 (2000) 356–362.

[4] M. Jobling, A. Pandya, C. Tyler-Smith, The Y chromosome in forensic analysis and paternity testing, Int. J. Legal Med. 110 (1997) 118–124.

[5] A. Salas, M. Richards, T. De la Fé, M.V. Lareu, B. Sobrino, P. Sánchez-Diz, V. Macaulay, A. Carracedo, The making of the African mtDNA landscape, Am. J. Hum. Genet. 71 (2002) 1082–1111.

[6] P. Underhill, G. Passarino, A.A. Lin, P. Shen, M. Mirazon 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.

[7] M.R. Wilson, J.A. DiZinno, D. Polanskey, J. Replogle, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, Int. J. Legal Med. 108 (1995) 68–74.

[8] G. Tully, W. Bär, B. Brinkmann, A. Carracedo, P. Gill, N. Morling, W. Parson, P. Schneider, Considerations by the European DNA profiling (EDNAP) group on the working practices, Forensic Sci. Int. 124 (2001) 83–91.

[9] A. Salas, M.V. Lareu, A. Carracedo, Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report, Int. J. Legal Med. 114 (2001) 186–190.

[10] A. Salas, E.M. Rasmussen, M.V. Lareu, N. Morling, A. Carracedo, Fluorescent SSCP of overlapping fragments (FSSCP-OF): a highly sensitive method for the screening of mitochondrial DNA variation, Forensic Sci. Int. 124 (2001) 97–103.

[11] A. Alonso, P. Martín, C. Albarrán, P. García, D. Primorac, O. Garcíia, L. Fernandez de Simon, J. Garcia-Hirschfeld, M. Sancho, J. Fernández-Piqueras, Specific quantification of human genomes from low copy number DNA samples in forensic and ancient DNA studies, Croat. Med. J. 44 (2003) 273–280.

[12] L. Prieto, M. Montesino, A. Salas, A. Alonso, C. Albarrán, S. Álvarez, M. Crespillo, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A.G. de la Vega, L. Gusmão, C.M. López, M. López-Soto, J.A. Lorente, M. Malaghini, C.A. Martinez, N.M. Modesti, A.M. Palacio, M. Paredes, S.D. Pena, A. Pérez-Lezaun, J.J. Pestano, J. Puente, A. Sala, M. Vide, M.R. Whittle, J.J. Yunis, J. Gómez, The 2000–2001 GEP-ISFG collaborative exercise on mtDNA: assessing the cause of unsuccessful mtDNA PCR amplification of hair shaft samples, Forensic Sci. Int. 134 (2003) 46–53.

[13] A.C. Syvänen, Accessing genetic variation: genotyping single nucleotide polymorphisms, Nat. Rev. Genet. 2 (2001) 930–942.

[14] A. Carracedo, F. Barros, L. Loidi, F. Domínguez, Progress in methodology and standards in European molecular genetics laboratories, Clin. Chim. Acta 278 (1998) 163–169.

[15] 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.

[16] A. Brandstätter, T.J. Parsons, W. Parsons, Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups, Int. J. Legal Med. 117 (2003) 291–298.
[17] E. Valverde, C. Cabrero, R. Cao, M.S. Rodríguez, A. Díez, F. Barros, J. Alemany, A. Carracedo, Population genetics of three VNTR polymorphisms in two different Spanish populations, Int. J. Legal Med. 105 (1993) 251–256.
[18] A. Salas, D. Comas, M.V. Lareu, J. Bertranpetit, A. Carracedo, mtDNA analysis of the Galician population: a genetic edge of European variation, Eur. J. Hum. Genet. 6 (1998) 365–375.
[19] M. Richards, V. Macaulay, E. Hickey, E. Vega, B. Sykes, V. Guida, C. Rengo, D. Sellitto, F. Cruciani, T. Kivisild, R. Villems, M. Thomas, S. Rychkov, O. Rychkov, Y. Rychkov, M. Golge, D. Dimitrov, E. Hill, D. Bradley, V. Romano, F. Cali, G. Vona, A. Demaine, S. Papiha, C. Triantaphyllidis, G. Stefanescu, J. Hatina, M. Belledi, A. Di Rienzo, A. Novelletto, A. Oppenheim, S. Norby, N. Al-Zaheri, S. Santachiara-Benerecetti, R. Scozari, A. Torroni, H.J. Bandelt, Tracing European founder lineages in the near eastern mtDNA pool, Am. J. Hum. Genet. 67 (2000) 1251–1276.
[20] C. Herrnstadt, J.L. Elson, E. Fahy, G. Preston, D.M. Turnbull, C. Anderson, S.S. Ghosh, J.M. Olefsky, M.F. Beal, R.E. Davis, N. Howell, Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups, Am. J. Hum. Genet. 70 (2002) 1152–1171.
[21] A. Torroni, K. Huoponen, P. Francalacci, M. Petrozzi, L. Morelli, R. Scozzari, D. Obinu, M.L. Savontus, D.C. Wallace, Classification of European mtDNAs from an analysis of three European populations, Genetics 144 (1996) 1835–1850.
[22] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 147.

Artículo 2: Mitochondrial DNA error prophylaxis: assessing the causes of errors in the GEP'02-03 proficiency testing trial

Salas A, Prieto L, Montesino M, Albarran C, Arroyo E, Paredes-Herrera MR, Di Lonardo AM, Doutremepuich C, Fernández-Fernández I, de la Vega AG, Alves C, López CM, Lopez-Soto M, Lorente JA, Picornell A, Espinheira RM, Hernández A, Palacio AM, Espinoza M, Yunis JJ, Pérez-Lezaun A, Pestano JJ, Carril JC, Corach D, Vide MC, Álvarez-Iglesias V, Pinheiro MF, Whittle MR, Brehm A, Gomez J.

(2005) *Forensic Sci Int* 148:191-8

Profilaxis de errores de ADN mitocondrial: valoración de las causas de errores en el ejercicio de control de calidad GEP´02-03

ABSTRACT/RESUMEN:

Presentamos el resultado del Ejercicio de Colaboración 2002-2003 de ADN mitocondrial (ADNmt) del Grupo Español y Portugués (GEP) de la Sociedad Internacional de Genética Forense (ISFG). Se enviaron seis muestras diferentes a los laboratorios participantes: cuatro manchas de sangre (M1-M2-M3-M4), una muestra de mezcla de sangre (M5), y dos fragmentos de pelo sin bulbo (M6). La mayoría de los laboratorios presentaron resultados consenso para las manchas de sangre, mejorando ligeramente los resultados de ejercicios de colaboración anteriores. El análisis de la muestra de pelo fue un éxito a pesar de que sólo un número reducido de laboratorios la realizaron. La situación contraria se produjo con el análisis de la muestra de mezcla de sangre (M5), a pesar de esto, los resultados obtenidos con el tipado de la muestra M5 demuestran la idoneidad del ADNmt en el análisis de mezclas. Los errores de edición son el tipo de error más frecuente en los distintos laboratorios. También se han detectado fenómenos de contaminación además de otros problemas menores, como por ejemplo la ausencia de estandarización en la nomenclatura de heteroplasmías puntuales y de longitud, y de *indels*. En la presente edición del ejercicio del GEP-ISFG se ha prestado especial atención a la detección de errores de secuenciación comunes mediante inspección filogenética.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forensic Science International 148 (2005) 191–198
Forensic Science International
www.elsevier.com/locate/forsciint

Mitochondrial DNA error prophylaxis: assessing the causes of errors in the GEP'02–03 proficiency testing trial

Antonio Salas[a,*], Lourdes Prieto[b], Marta Montesino[b], Cristina Albarrán[c], Eduardo Arroyo[d], Miguel R. Paredes-Herrera[e], Ana Ma Di Lonardo[f], Christian Doutremepuich[g], Isabel Fernández-Fernández[h], Alberto González de la Vega[i], Cíntia Alves[j], Carlos M. López[k], Manolo López-Soto[l], José A. Lorente[m], Antònia Picornell[n], Rosa M. Espinheira[o], Alexis Hernández[p], Ana Ma Palacio[q], Marta Espinoza[r], Juan J. Yunis[s], Anna Pérez-Lezaun[t], José J. Pestano[u], Juan Carlos Carril[v], Daniel Corach[w], Ma Conceiçao Vide[x], V. Álvarez-Iglesias[a], M.F. Pinheiro[y], Martín R. Whittle[z], Antonio Brehm[A], Josefina Gómez[B]

[a]*Unidad de Genética, Facultad de Medicina de la Universidad de Santiago de Compostela, Instituto de Medicina Legal, A Coruña, Galicia-Spain*
[b]*Comisaría General de Policía Científica, Sección de Biología-ADN, Madrid, Spain*
[c]*Departamento de Madrid, Sección de Biología, Instituto de Toxicología, Madrid, Spain*
[d]*Toxicología y Legislación Sanitaria, Facultad de Medicina, Universidad Complutense, Madrid, Spain*
[e]*Departamento de Barcelona, Sección de Biología, Instituto Nacional de Toxicología, Barcelona, Spain*
[f]*Banco Nacional de Datos Genéticos, Hospital Dr. C.G. Durand, CF Buenos Aires, Argentina*
[g]*Laboratoire D'He'matologie, Bordeaux, France*
[h]*DataGene, Sondika, Bizkaia, Spain*
[i]*ADF TecnoGen, SL, Madrid, Spain*
[j]*Instituto de Patología e Inmunología, Molecular de la Universidad de Porto, Porto, Portugal*
[k]*Jefatura de Investigación y Criminalística, Dirección General de la Guardia Civil, Laboratorio de ADN, Madrid, Spain*
[l]*Sección de Biología, Departamento de Sevilla, Instituto de Toxicología, Sevilla, Spain*
[m]*Departamento de Medicina Legal, Laboratorio de Identificación Genética, Universidad de Granada, Granada, Spain*
[n]*Laboratori de Genètica, Institut Universitari d'Investigacions en Ciències de la Salut i Departament de Biologia, Universitat de les Illes Balears, Spain*
[o]*Serviço de Genética e Biologia Forense, Delegação de Lisboa do Instituto, Nacional de Medicina Legal, Lisboa, Portugal*
[p]*Departamento de Canarias, Instituto Nacional de Toxicología y Ciencias Forenses, La Cuesta, Santa Cruz de Tenerife, Spain*
[q]*Centro de Análisis Genéticos C.A.G.T., Zaragoza, Spain*
[r]*Departamento de Ciencias Forenses, Unidad de Genética Forense, Poder Judicial, Costa Rica, Spain*
[s]*Servicio Médico Yunis Turbay y Cía, Santafé de Bogota, DC, Colombia, Spain*
[t]*Unitat de Biología Evolutiva, Universidad Pompeu Fabra, Barcelona, Spain*
[u]*Servicio de Genética Forense, Universidad de Las Palmas de Gran Canaria, Las Palmas, Spain*
[v]*Departamento de Genética Forense, GENÓMICA, S.A.U., Madrid, Spain*
[w]*Servicio de Huellas Digitales Genéticas, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Buenos Aires, Argentina*
[x]*Servicio de Biología Forense, Instituto de Medicina Legal, Coimbra, Portugal*
[y]*Instituto de Medicine Legal, Servicio de Biología Forense, Oporto, Spain*
[z]*Genomic Engenharia Molecular LTD A, Sao Paulo, Brazil*

[*] Corresponding author. Tel.: +34-981-582-327; fax: +34-981-580-336.
E-mail address: apimlase@usc.es (A. Salas).

doi:10.1016/j.forsciint.2004.06.008

[A]Centre of Biological and Geological Sciences, University of Madeira, Campus of Penteada, 9000 Funchal, Portugal
[B]Instituto de Toxicología, Unidad de Garantía de Calidad, Madrid, Spain

Received 16 April 2004; received in revised form 8 June 2004; accepted 9 June 2004
Available online 25 August 2004

Abstract

We report the results of the Spanish and Portuguese working group (GEP) of the International Society for Forensic Genetics (ISFG) Collaborative Exercise 2002–2003 on mitochondrial DNA (mtDNA) analysis. Six different samples were submitted to the participating laboratories: four blood stains (M1–M2–M3–M4), one mixture blood sample (M5), and two hair shaft fragments (M6). Most of the labs reported consensus results for the blood stains, slightly improving the results of previous collaborative exercises. Although hair shaft analysis is still carried out by a small number of laboratories, this analysis yielded a high rate of success. On the contrary, the analysis of the mixture blood stain (M5) yielded a lower rate of success; in spite of this, the whole results on M5 typing demonstrated the suitability of mtDNA analysis in mixture samples. We have found that edition errors are among the most common mistakes reported by the different labs. In addition, we have detected contamination events as well as other minor problems, i.e. lack of standarization in nomenclature for punctual and length heteroplasmies, and indels. In the present edition of the GEP-ISFG exercise we have paid special attention to the visual phylogenetic inspection for detecting common sequencing errors.

Keywords: mtDNA; Standardization; Collaborative exercise; Heteroplasmy; Hair shafts; Sequence errors; Phylogenetic

1. Introduction

Since 1997, the Spanish and Portuguese Working Group (GEP) of the International Society for Forensic Genetics (ISFG) included mtDNA profiling as part of the GEP proficiency testing program [1–5]. Along the years, the number of participating labs has progressively increased.

Here we review the results of the sixth mtDNA trial carried out by the GEP-ISFG group corresponding to the period 2002–2003. This exercise shows an important progress in standardization and reliability on the mtDNA analysis, as well as an increasing interest for the GEP-ISFG group to answer important questions concerning mtDNA forensic caseworks. Particularly, in the present exercise, we point out the convenience of using visual phylogenetic inspection of the final sequence report as a prophylactic tool to detect sequencing errors.

Two different cases were included in the GEP proficiency testing trial: a forensic and a paternity case. It is important to note that not all the labs in the group are interested in criminalistic casework, and this is the main reason why mtDNA is generally performed by a small number of labs (27/85). As part of the present exercise, five blood stains (M1, M2, M3, M4, and M5), and two hair shaft fragments (M6) were distributed to all participants. M1 and M2 stains were taken from a father and a mother, respectively. M5 consisted of a blood stain made of a mixture 2:1 from a woman maternally unrelated with the rest of the donors and blood from the M4 donor. Information on the characteristics of M3, M4, M5 and M6, was not provided to the participants till the completion of the exercise. The posed questions for the whole exercise were: to examine the possibility of relationship of M3 and M4 donors with the mother (M1) and the father (M2), and determine if M1, M2, M3, and M4 donors could have been the biological source of M5 and the two hair shafts fragments (M6) as well. Therefore, M1 to M4 were submitted for the paternity exercise, whilst M5 and M6 were part of the forensic simulated case; M6 was submitted exclusively for mtDNA typing.

2. Materials and methods

2.1. Samples

The Quality Control Office (Unidad de Garantía de Calidad, Instituto Nacional de Toxicología, Ministerio de Justicia, Madrid, Spain) provided all the participants with a total of five bloodstains as well as two hair shafts from the same individual (two fragments of approximately 3 cm long). M1, M2, M3, and M4 were prepared by applying 100 μl of whole blood onto Whatman surface (Whatman Bioscience) and were air-dried before distribution. The exercise simulated a traffic accident with several victims. The case consisted of identifying the son of a mother and father (the donors of M1 and M2, respectively) among the donors of M3 and M4. The two hair shafts fragments were submitted in order to evaluate their

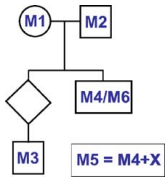


Fig. 1. Diagram showing the genetic relationships between the samples analyzed in 2002–2003 GEP-ISFG Collaborative Exercise. An X indicates a donor not maternally related to M1, M2, M3, M4/M6.

potential biological sources as coming from M1, M2, M3, and M4 donors.

Posterior to the completion of the analysis and submission of the results to the Quality Control Office, labs were informed about the origin of M3, M4, M5 and M6. M4 donor was the son of M1 and M2. M3 donor was a grandson of the donors of M1 and M2 (and maternal nephew of the M4 donor). Blood stain M5 was prepared as a mixture of two blood samples at proportion 2:1, taken from a woman unrelated to M1, M2, M3, M4, and M6, and from the same donor of M4. The mtDNA profile of this unrelated individual was also determined independently of the rest of the exercise in order to evaluate the a posteriori admixture-profile of sample M5 in the whole quality control. Hair shaft fragments were taken from the donor of M4. Fig. 1 displays a diagram showing the familial relationships among samples.

Following the routine established in previous trials, all labs were given an anonymous number and were requested to fill in a questionnaire with all the technical details related with the analysis.

2.2. DNA extraction and amplification

For the bloodstains samples and hair shafts, phenol–chloroform was used by the majority of laboratories (66 and 79%, respectively) usually followed by microcon or centricon-100 purification, and 26 and 16% (respectively) used a Chelex-100 extraction. Most of the labs used the primers described by Wilson et al. [6] and Vigilant et al. [7] for the amplification of the mtDNA HVS-I and HVS-II segments.

2.3. Sequencing

All the labs used automated sequencers (mainly ABI systems of Applied Biosystems: ABI377, ABI3 10, and ABI 3100). Most of the participants used the same primers as those used for the amplification of HVS-I and HVS-II. Rhodamine or BigDye terminators, and Thermosequenase were the most common chemistries used for sequencing.

Table 1
Participation of laboratories in the mtDNA exercise

Samples	Number of labs that analyzed mtDNA	Number of labs that reported the consensus result (%)
M1, M2, M3, M4	21[a]	18 (∼86)
M1, M2,M3, M4, M5	18	8 (∼44)
M1, M2, M3, M4, M6	18	13 (∼72)
M1–M6	13	6 (∼46)

[a] Although 27 labs partially analyzed these samples, only 21 of them reported HVS-I and HVS-II.

3. Results and discussion

3.1. MtDNA sequence results

A total of 27 labs out of 85 (∼32%) reported partial or complete mtDNA results for the samples submitted. This level of participation was slightly higher than in the previous edition of the quality control (30%).

HVS-I and HVS-II sequence results of the blood samples M1, M2, M3, and M4, were submitted by 21 out of 27 labs (∼77%; 89% submitted HVS-I or HVS-II). However, the number of labs reporting results for both hypervariable regions decreased significantly when considering M5 and M6 samples: only 18 labs reported results for M1, M2, M3, M4, and M6, 18 labs reported results for M1–M5, whereas only 13 reported results for the whole set of samples (M1–M6) (see Table 1). A few labs did not consider mtDNA the most appropriate marker for DNA profiling of mixture samples (in fact, this M5 mixture was perfectly resolved by most participating labs by using STRs markers); this explains in part the low participation of labs in mtDNA typing of sample M5.

The inspection of Table 1 shows that 18 out of 21 labs (∼86%) reported the consensus HVS-I/II result for M1–M2–M3 and M4, while only 6 out of 13 (∼46%) reported consensus result for the six samples submitted. M5 sample was responsible for this low rate of success: only ∼44% (8/13) were consensus results. Analysis of the hair shafts (M6) was significantly more successful: 13 out of 19 labs (∼68%; when contrasting this frequency with the ones from Table 1, note that (Table 2), there is one lab that did not analyze M2, but shows results for M6). This clearly contrasts the low rate of success yielded by the previous edition of the GEP mtDNA exercise due to the low quantity and quality of DNA present in those hair samples (see [5]).

Interestingly, we have detected that a high number of labs reported HVS-I/II consensus results for those blood samples belonging to the same lineages, that is, 20/27 M1, 21/27 M3, and 21/27 M4. However, a minor number of them (18 out of 27) reported consensus results concerning the only sample (M2) that belonged to a different mtDNA lineage than the one shared by M1, M3, and M4. Note that the information related to the degree of familial relationship could be

Table 2
Non-consensus results for the five blood samples (M1, M2, M3, M4, M5) and the two hair shafts fragments (M6) analyzed in the 2002–2003 edition of the GEP-ISFG quality control

	M1–M3–M4	M2	M5	M6
Consensus	HVS-I: 16189C 16256T 16270T 16362C HVS-II: 73G 185A 204C 263G 309.1C 315.1C	HVS-I: 16189C 16265C HVS-II: 152C 263G 309.1C 315.1C	HVS-I: 16051G/A 16189C 16256C/T 16270T 16362T/C HVS-II: 73G 146C/T 150T/C 185G/A 204T/C 263G 309.1C 315.1C	HVS-I: 16189C 16256T 16270T 16362C HVS-II: 73G 185A 204C 263G 309.1C 315.1C
Lab 1	HVS-I: 16189C 16256T 16270T HVS-II: =	HVS-I: – HVS-II: –	HVS-I: – HVS-II: –	HVS-I: 16189C HVS-II: 309.1C 310C/T 315.1C
Lab 2	HVS-I: = HVS-II: 73G 185A 204C 263G 309.1C	HVS-I: = HVS-II: 152C 263G 309.1C	HVS-I: – HVS-II: 73G 146C + T 150C + T 185A + G 204C + T 263G 309insC	HVS-I: – HVS-II: –
Lab 3	HVS-I: 16189C HVS-II: 185A 204C 263G 309.1C 315.1C	HVS-I: 16189C HVS-II: 152C 309.1C	HVS-I: 16189C HVS-II: 146C/T 150T/C 185A/G 204C/T 263G 309.1C 315.1C	HVS-I: 16189C HVS-II: 185A 204C 263G 309.1C 315.1C
Lab 4	HVS1: – HVS-II: 185A 204C 263G 309.1C 315.1C	HVS-I: – HVS-II: 152C 303 + 1 309 + 1	HVS-I: – HVS-II: 146C 185A 204C 263G 303 + 1C 311 + 1C	HVS-I: – HVS-II: –
Lab 5	HVS-I: 16189C 16256A 16270A 16362G HVS-II: =	HVS-I: 16189C 16265G HVS-II: 152C 263G	HVS-I: 16051A/G 16189C 16256A/G 16270A 16362A/G HVS-II: =	HVS-I: 16126C 16186T 16189C 16294T HVS-II: 73G 263G 309insC 315insC
Lab 6	HVS-I: = HVS-II: =	HVS-I: = HVS-II: 152 263 309.1C 315.1C	HVS-I: – HVS-II: –	HVS-I: – HVS-II: –
Lab 7	HVS-I: = HVS-II: =	HVS-I: = HVS-II: 152C 263G 309.1C 310C/T 315.1C	HVS-I: – HVS-II: –	HVS-I: – HVS-II: –
Lab 8	HVS-I: = HVS-II: =	HVS-I: = HVS-II: =	HVS-I: = HVS-II: 73G 146C/T 150T/C 185C/T 204T/C 263G 309.1C 315.1C	HVS-I: 16189C 16193.1C 16216G 16223T HVS-II: –
Lab 9	HVS-I: = HVS-II: =	HVS-I: = HVS-II: =	HVS-I: 16051G 16189C 16256T 16270T 16362C HVS-II: 73G 185A 204C 263G 309.1C 315.1C	HVS-I: 16189C 16193.1C 16256T 16270T 16362C HVS-II: =
Lab 10	HVS-I: = HVS-II: =	HVS-I: = HVS-II: =	HVS-I: 16048R 16189C 16256Y 16270T 16362Y HVS-II: =	HVS-I: – HVS-II: –

Table 2 (*Continued*)

	M1–M3–M4	M2	M5	M6
Lab 11	HVS-I: = HVS-II: =	HVS-I: = HVS-II: =	HVS-I: = HVS-II: 73G 146T/C 150C/T 185A 204C 263G 309.1C 315.1C	HVS-I: = HVS-II: =
Lab 12	HVS-I: = HVS-II: =	HVS-I: = HVS-II: =	HVS-I: 16051G > A 16189C 16235A ∼ C 16256T > C 16270T 16362T > C HVS-II: 73G 146C > T 150T > C 185G > A 204T ∼ C 263G 309.1C 315.1C	HVS-I: = HVS-II: =
Lab 13	HVS-I: – HVS-II: 185A 204C 263G 309.1C 315.1C	HVS-I: = HVS-II: 152C 263G 309.1C 315.1C	HVS-I: – HVS-II: HVS-II: 146C 185A 204C 263G 303 + 1C 311 + 1C	HVS-I: – HVS-II: –
Lab 14	HVS-I: = HVS-II: =	HVS-I: = HVS-II: =	HVS-I: 16051G 16189C 16256C/T 16270T 16273A/G16362C/T HVS-II: =	HVS-I: = HVS-II: 73G 185A/G 204C 263G 309.1C 315.1C

The results are shown as reported to the Unidad de Garantía de Calidad. This allows to illustrate some problems concerning to, i.e. difference in nomenclature, especially at heteroplasmies and insertions at homopolimeric stretches. Note that labs reported exactly the same sequence errors for samples M1, M2, and M4 (the three donors belonged to the same matrilineage; see text): (=) indicates "in agreement with the consensus sequence"; (–) indicates "not reported".

indirectly obtained by all the labs (as part of the same quality control exercise) through the analysis of autosomal markers. In conclusion, it seems that the labs are more confident in their mtDNA report when it is supported on the basis of consensus results obtained through the analysis of different samples belonging to the same mtDNA lineage. In support of this argument is the fact that some labs repeated exactly the same sequence mistakes (some of them concerning the omission of polymorphisms positioned at the sequences ends) in the three samples belonging to the same lineages (Table 2; see errors in samples M1, M3, and M4, reported by labs number 1, 2, 3, 4, 5 and 13). This approach is not adequate in the forensic genetic practice since each electropherogram must be independently evaluated. Therefore, this fact was informed and discussed during the GEP-ISFG meeting in order to aware the participants about avoiding this type of errors in future trials.

A lower number of labs submitted complete HVS-I/II results for samples M5 (18/27) and M6 (19/27), and the success rate was significantly worse in the case of sample M5: only 8 out of 18 participants (∼44%) reported the consensus sequence for HVS-I/II segments. This is a poor result, especially if we compare it with the success reached in the analysis of the rest of samples for HVS-I/II: 75–77% for M1, M2, M3, and M4, and 72% for M6.

In the present edition of the control, we have also discussed the advantages of analyzing the complete control region in the forensic casework and in future quality controls, since many polymorphisms of forensic interest are positioned out of the HVS-I and HVS-II segments commonly analyzed. This obviously would lead to increase the discrimination power of the mtDNA test. Nevertheless, we also considered that this issue should be discussed in terms of updating guidelines for mtDNA testing, perhaps in the context of the ISFG.

3.2. Assessing the causes of error

Visual inspection of the whole mtDNA GEP-ISFG report has allowed us to assess the cause of most of the non-consensus results (Table 2). Most of the labs performed both hypervariable regions (HVS-I and HVS-II) and it was not detected a significant correlation between the hypervariable region studied and the number or type of errors.

3.2.1. Edition mistakes

The most common mistakes are due to errors during edition process. Table 2 testifies for several examples: (a) lab 2 omitted the common insertion 315.1C in HVS-II in all the samples; (b) in samples M1, M3 and M4, lab 5 edits the H chain variants instead of the L chain ones in HVS-I segments after position 16189, that is, 16256A 16270A 16362G variants instead of 16256T 16270T 16362C. This is probably due to the fact that this lab performed two separated

amplifications of HVS-I segment (a practical choice when length heteroplasmy occurs at the homopolymeric track around 16189) and just H reading of the 3′ HVS-I segment was done; in M2, error at position 16265C in lab 5 probably responds to the same explanation; (c) in M5, lab 8 edits 185C/T instead of 185G/A; (d) lab 10 edits an erroneous three base pair shift in HVS-I of M5 sample: position 16048R instead of 16051G/A, etc.

3.2.2. Electrophoregrams quality at the 3′ and 5′ ends of the sequences

Also striking is the fact that some labs do not report certain positions, although using primers that would permit to do so. This is probably due to the fact that these labs have some problems in reading and interpreting the 5′ extremes of their sequences (where the quality of the electrophoregram, close to the sequence primer, is usually poorer). This deficiency can be properly corrected by modifying some critical step of the sequence protocol (i.e. purification of the PCR fragments) and performing forward and reverse sequence of the same amplicons.

In the present control, this mainly affects variants at sites 16362 (i.e. lab 1 in M1, M3, and M4 samples) and 73 (i.e. lab 3, 4 and 13 in M1, M3, and M4 samples).

3.2.3. Contamination

We have detected several instances of contamination. Lab 3 reported the same HVS-I sequence in all the samples (16189C), but different (although with errors) HVS-II profiles. The most plausible explanation for this result is contamination of HVS-I amplification primers. All the results reported by this lab contained some additional mistake; therefore, in the GEP-ISFG meeting we stressed the need for an urgent solution to these critical blunders. Lab 5 provides a clear instance of contamination of the hair shaft sample (M6): a genuine U5 profile of M6 sample (‘‘16189C 16256T 16270T 16362C 73G 185A 204C 263G 309.1C 315.1C’’) has been accidentally contaminated with a biological source harbouring the following profile ‘‘16126C 16186T 16189C 16294T 73G 263G 309insC 315insC (Table 2)’’, which clearly identifies a sequence belonging to haplogroup T. In addition, labs 1 and 8 show two additional likely instances of contamination of sample M6.

3.2.4. Nomenclature deficiencies

Table 2 reveals a common problem concerning mtDNA sequence nomenclature. There are a number of labs that do not use the consensus nomenclature recommended by the ISFG mtDNA guidelines [8] to describe punctual heteroplasmies and length variants.

We have detected a common mistake in the report of HVS-II homopolymeric variation. Position 310 in HVS-II is phylogenetically stable and should not mutate frequently. However, when length heteroplasmy affects the homopolymeric C track 303–309, the electropherogram displays a point heteroplasmy-like pattern at this position 310. Thus, a 310C/T pattern (i.e. lab 1) is really a length heteroplasmy provoked by the combination of molecules of different homopolymeric lengths.

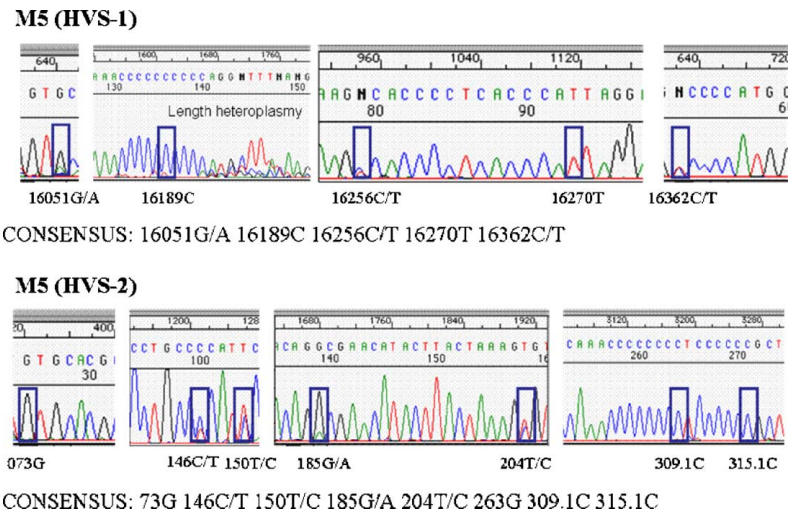


Fig. 2. Different segments of the electrophoregrams corresponding to the mixture sample M5.

3.3. Evaluating mtDNA as a marker to detect sample mixtures

Sample M5 consisted of a mixture of two different blood samples at 2:1 proportions.

- M4 donor provided with the minor component of the mixture: ''16189C 16256T 16270T 16362C 073G 185A 204C 263G 309.1C 315.1C''.
- A maternally unrelated donor to M1, M2, M3, and M4 represented the major component of the mixture and carries the following profile ''16051G 16189C 16270T 073G 146C 150T 263G 309.1C 315.1C''.

Then, the expected profile of such mixture is 16051G > A 16189C 16256C > T 16270T 16362T > C for HVS-1 and 73G 146C > T 150T > C 185G > A 204T > C 263G 309.1C 315.1C for HVS-II. Due to the complexity of the electropherogram, the following HVS-II profile ''16051G/A 16189C 16256C/T 16270T 16362T/C and 73G 146C/T 150T/C 185G/A 204T/C 263G 309.1C 315.1C'' was accepted as M5 consensus (Fig. 2).

Many labs did not detect the admixture in M5, and many other reported several edition mistakes (Table 2). Only one lab reported the existence of the expected profile as indicated above; that is, indicating the correct proportions of both the major and the minor component of M5. This lab used a phylogenetic approach to interpret the electrophoretic pattern of M5: the admixture profile could be explained as a product of an admixture of two mtDNA haplotypes, both belonging to a typical western European haplogroup, namely U5. However, it is important to keep in mind that the power of the mtDNA genome for the determination of mixed forensic samples depends on the phylogenetic nature of the participating profiles in the mixture. In addition, the typing of additional suspected diagnostic positions (coding or non-coding polymorphisms) can help to further define the participating lineages.

Moreover, two other labs reported the consensus result for both HVS-I and HVS-II (as indicated above), this time without providing information on the proportion of the different components (that is, no information on the intervening haplotypes, but a report which consists of a list of the polymorphisms detected in the electrophoregrams). Some other labs identify an admixture but did not report the consensus result.

In conclusion, since three labs reported the consensus result (and taking into account that this fact is unlikely by chance), we interpret that mtDNA is an appropriate marker for detecting DNA admixtures. However we recognize that mtDNA is not the best marker for this purpose, but still can be useful in those contexts where autosomal or Y chromosome STRs marker cannot be typed (degraded samples or low copy number samples). We find additional support to the suitability of mtDNA for mixture DNA typing in Szibor et al. [9], although they used restriction enzyme analysis instead of sequencing typing.

Results on M5 sample have also shown that there were different ways of reporting heteroplasmies (or heteroplasmy-like variants). We encourage the labs to use the recommended nomenclature as proposed by the DNA Commission of the International Society for Forensic Genetics [8].

4. Final remarks

Most of the errors detected in the GEP 2002–2003 mtDNA exercise could be assigned to the following categories: (a) deficient electrophoregrams; (b) edition errors; (c) contamination; (d) deficiencies in terms of detection of heteroplasmies; and (e) other minor problems, i.e. nomenclature.

As for other markers, it is interesting to highlight the fact that the worst and the most evident mistakes were concentrated in a few labs.

Apart from these problems, in general we noted that the technical state of the mtDNA analysis has reached a high level of reliability. Although the present exercise did not yield significant progress in terms of number of errors, we have noticed that most of the labs have gained experience. We have also stressed the fact that an important percentage of mistakes could have been easily corrected, such as edition errors, improving significantly the whole results of the GEP exercise. In the present edition we have made an important effort in order to show the usefulness of the phylogenetic approach for mtDNA prophylaxis: the cause of all the mistakes reported in Table 2 could be detected by direct visual—in cases phylogenetic—inspection of the mtDNA profiles. More details concerning phylogenetic approaches for forensic profilaxis can be seen in [10–12; among others]. In general, the Quality Control Program has again proved to be extremely valuable to the GEP-ISFG group in order to address important questions concerning mtDNA test.

Acknowledgements

The CYTED Program through the network RIGEMA-MEF (Subprogram IIIC) partly supported these Collaborative Exercises. We are also indebted to Ángel Carracedo for critical reading the manuscript and useful comments.

References

[1] J. Gómez, A. Carracedo, A review of the collaborative exercises of the Spanish and Portuguese ISFH working group, In: A. Carracedo, B. Brinkmann, W. Bär (Eds.), Advances in Forensic Haemogenetics, vol. 6, Springer, Berlin, 1996, pp. 695–699.

[2] J. Gómez, M.S. Rodriguez-Calvo, C. Albarrán, A. Amorím, J. Andradas, C. Cabrero, R. Calvet, D. Corach, M. Crespillo, C. Doutremepuich, O. García, H. Geada, M. Gene, S. Jimenez,

J.A. Lorente, S.M. Marques-Santos, B. Mártinez-Jarreta, M. Martínez de Pancorbo, F. Montes, J.M. Ruiz de la Cuesta, P. Sanz, M.F. Terra-Pinheiro, M.C. Vide, A. Carracedo, A review of the collaborative exercises on DNA typing of the Spanish and Portuguese ISFH working group, Int. J. Legal Med. 110 (1997) 273–277.

[3] J. Gómez, A. Carracedo, 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 (2000) 21–30.

[4] A. Alonso, A. Salas, C. Albarran, E. Arroyo, A. Castro, M. Crespillo, A.M. Di Lonardo, M.V. Lareu, C.L. Cubrija, M.L. Soto, J.A. Lorente, M.M. Semper, A. Palacio, M. Paredes, L. Pereira, A.P. Lezaun, J.P. Brito, A. Sala, M.C. Vide, M. Whittle, J.J. Yunis, J. Gómez, Results of the 1999–2000 collaborative exercise and proficiency testing program of mitochondrial DNA of the GEP-ISFG: an inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor, Forensic Sci. Int. 125 (2002) 1–7.

[5] L. Prieto, M. Montesino, A. Salas, A. Alonso, C. Albarrán, S. Álvarez, M. Crespillo, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A. González de la Vega, L. Gusmão, C.M. López, M. López-Soto, J.A. Lorente, M. Malaghini, C.A. Martínez, N.M. Modesti, A.M. Palacio, M. Paredes, A. Pérez-Lezaun, J.J. Pestano, J. Puente, A. Sala, M.C. Vide, M.R. Whittle, J.J. Yunis, J. Gómez, The 2000–2001 GEP-ISFG collaborative exercise on mtDNA: assessing the cause of unsuccessful mtDNA PCR amplification of hair shaft simples, Forensic Sci. Int. 134 (2003) 46–53.

[6] M.R. Wilson, J.A. DiZinno, D. Polanskey, J. Replogle, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, Int. J. Legal Med. 108 (1995) 68–74.

[7] L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A.C. Wilson, African populations and the evolution of human mitochondrial DNA, Science 253 (1991) 1503–1507.

[8] A. Carracedo, W. Bäär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tully, M. Wilson, DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79–85.

[9] M. Szibor, I. Michael, H. Plate, D. Wittig, Krause Identification of the minor component of a mixed stain by using mismatch primer-induced restriction sites in amplified mtDNA, Int. J. Legal Med. 117 (2003) 160–164.

[10] H.-J. Bandelt, P. Lahermo, M. Richards, V. Macaulay, Detecting errors in mtDNA data by phylogenetic analysis, Int. J. Legal Med 115 (2001) 64–69.

[11] H.-J. Bandelt, L. Quintana-Murci, A. Salas, V. Macaulay, The fingerprint of phantom mutations in mtDNA data, Am. J. Hum. Genet. 71 (2002) 1150–1160.

[12] H.-J. Bandelt, A. Salas, S. Lutz-Bonengel, Artificial recombination in forensic mtDNA population databases. Int. J. Legal Med., 2004, in press.

Artículo 3: Results of the 2003-2004 GEP-ISFG collaborative study on mitocondrial DNA: Focus on the mtDNA profile of a mixed semen-saliva stain

Crespillo M, Paredes M, Prieto L, Montesino M, Salas A, Albarran C, Álvarez-Iglesias V, *et al.*

(2006) *Forensic Sci Int* 160: 157-67

Resultados del estudio colaborativo 2003–2004 del GEP-ISFG de AND mitocondrial: evaluación del perfil de ADNmt en manchas mezcladas de semen y saliva

ABSTRACT/RESUMEN:

Presentamos un resumen del séptimo ejercicio de ADN mitocondrial (ADNmt) llevado a cabo por el Grupo Español y Portugués (GEP) de la International Society for Forensic Genetics (ISFG) correspondiente al periodo 2003-2004. Cinco muestras de referencia de mancha de sangre de cinco donantes (M1-M5), una muestra de mezcla de saliva y semen (M6) y una muestra de pelo (M7) fueron enviadas a cada laboratorio participante para el análisis de ADN nuclear (ADNn: STR autosómicos e Y-STR) y de ADNmt. Los laboratorios tenían que averiguar quiénes de las muestras de referencia (M1-M5) eran los donantes de las muestras M6 y M7. Un total de 34 laboratorios presentaron datos de secuencias totales o parciales de las muestras de referencia de sangre (M1-M5), y de la muestra de pelo M7 concluyendo la coincidencia de los perfiles de ADNmt de M5 y M7. El análisis de STRs autosómicos y de cromosoma Y fue la estrategia preferida para investigar los donantes de la mezcla semen-saliva (M6). Los perfiles de ADNn fueron consistentes con una mezcla de saliva del donante (femenino) M4 y semen de M5, siendo el perfil (XY) de semen el componente dominante de la mezcla. En contradicción con el análisis de ADNn, los datos de secuenciación de ADNmt presentaron un resultado mucho más simple: sólo la contribución de saliva (M4) fue detectada, ni con lisis preferencial ni después de una digestión completa del ADN. Algunos laboratorios proporcionaron varias explicaciones para este hallazgo y llevaron a cabo experimentos adicionales para explicar este resultado aparentemente contradictorio. Los resultados señalaron la existencia de cantidades relativas de ADN nuclear y mitocondrial diferentes en saliva y semen. Concluimos que esta circunstancia puede influir sustancialmente en la interpretación de una evidencia de ADNmt en mezclas desequilibradas y como consecuencia dar una falsa exclusión. Durante la reunión anual del GEP-ISFG se planeó un estudio de validación para mejorar la interpretación de ADNmt de diferentes mezclas.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Forensic Science International 160 (2006) 157–167
Forensic Science International
www.elsevier.com/locate/forsciint

Results of the 2003–2004 GEP-ISFG collaborative study on mitochondrial DNA: Focus on the mtDNA profile of a mixed semen-saliva stain

Manuel Crespillo [a,*], Miguel R. Paredes [a], Lourdes Prieto [b], Marta Montesino [b], Antonio Salas [c], Cristina Albarran [d], Álvarez-Iglesias V [c], Antonio Amorin [e], Gemma Berniell-Lee [f], Antonio Brehm [g], Juan C. Carril [h], Daniel Corach [i], Nerea Cuevas [j], Ana M. Di Lonardo [k], Christian Doutremepuich [l], Rosa M. Espinheira [m], Marta Espinoza [n], Felix Gómez [o], Alberto González [p], Alexis Hernández [q], M. Hidalgo [r], Magda Jimenez [s], Fabio P.N. Leite [t], Ana M. López [u], Manuel López-Soto [v], Jose A. Lorente [w], Shintia Pagano [x], Ana M. Palacio [y], José J. Pestano [z], Maria F. Pinheiro [A], Eduardo Raimondi [B], M.M. Ramón [C], Florangel Tovar [D], Lidia Vidal-Rioja [E], Maria C. Vide [F], Martín R. Whittle [G], Juan J. Yunis [H], Julia Garcia-Hirschfel [I]

[a] *Instituto Nacional de Toxicología y Ciencias Forenses, Servicio de Biología, Barcelona, Spain*
[b] *Comisaría General de Policía Científica, Sección de Biología-ADN, Madrid, Spain*
[c] *Unidad de Genética, Facultad de Medicina de la Universidad de Santiago de Compostela, Instituto de Medicina Legal, A Coruña, Spain*
[d] *Instituto Nacional de Toxicología y Ciencias Forenses, Servicio de Biología, Madrid, Spain*
[e] *Instituto de Patología e Imunología Molecular da Universidade do Porto, Porto, Portugal*
[f] *Unitat de Biología Evolutiva, Universidad Pompeu Fabra, Barcelona, Spain*
[g] *Centre of Biological and Geological Sciences, University of Madeira, Funchal, Portugal*
[h] *Departamento de Biología Forense, GENOMICA, S.A.U., Madrid, Spain*
[i] *Servicio de Huellas Digitales Genéticas, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Buenos Aires, Argentina*
[j] *DataGene, Sondita, Biskaia, Spain*
[k] *Banco Nacional de Datos Genéticos, Hospital Dr. C.G. Durand, C.F, Buenos Aires, Argentina*
[l] *Laboratoire D'Hematologie, Bordeaux, France*
[m] *Serviço de Genética e Biología Forense, Delegaçao de Lisboa do Instituto Nacional de Medicina Legal, Lisboa, Portugal*
[n] *Departamento de Ciencias Forenses, Unidad de Genética Forense, Poder Judicial, Heredia, Costa Rica*
[o] *Biopatología Médico-Legal, Madrid, Spain*
[p] *ADF TecnoGen, S.L., Madrid, Spain*
[q] *Instituto Nacional de Toxicología y Ciencias Forense, Servicio de Biología, Santa Cruz de Tenerife, Spain*
[r] *Laboratorio del Departamento de Biología del Servicio de Criminalística de la Guardia Civil, Madrid, Spain*
[s] *Laboratorio de DNA, Instituto Nacional de Medicina Legal y Ciencias Forenses de Santa Fé de Bogotá, Colombia*
[t] *Laboratorio de Pericias, Brazil*
[u] *Departamento de Medicina Legal y Forense, Facultad de Medicina, Univ. Complutense, Madrid, Spain*
[v] *Instituto Nacional de Toxicología y Ciencias Forenses, Servicio de Biología, Sevilla, Spain*
[w] *Laboratorio de Identificación Genética, Departamento de Medicina Legal, Universidad de Granada, Granada, Spain*

* Corresponding author. Tel.: +34 93 317 40 61; fax: +34 93 318 25 30.
E-mail address: manuel.crespillo@mju.es (M. Crespillo).

0379-0738/$ – see front matter
doi:10.1016/j.forsciint.2005.09.005

[x] Lab Biológico de la Dirección Nacional de Policía Técnica de Montevideo, Uruguay
[y] Centro de Análisis Genéticos C.A.G.T., Zaragoza, Spain
[z] Laboratorio de Genética, Facultad de Medicina, Instituto Anatómico Forense, Las Palmas de Gran Canaria, Spain
[A] Instituto de Medicine Legal de Oporto, Servicio de Biología Forense, Oporto, Portugal
[B] PRICAI-Fundación Favaloro, Buenos Aires, Argentina
[C] Laboratori de Genètica, Institut Universitari díInvestigacions en Ciències de la Salut i Departament de Biología, Universitat de les Illes Balears, Spain
[D] Unidad de Polimorfismos Genéticos, Caracas, Venezuela
[E] Laboratorio de Identificación Genética, IMBICE, La Plata, Argentina
[F] Servicio de Biología Forense, Instituto de Medicina Legal, Coimbra, Portugal
[G] Genomic Engenharia Molecular LTDA, Sao Paulo, Brazil
[H] Servicios Médicos Yunis Turbay y Cia, Santafé de Bogotá, DC, Colombia
[I] Instituto Nacional de Toxicología y Ciencias Forenses, Unidad de Garantía de Calidad, Madrid, Spain

Received 19 April 2005; received in revised form 17 September 2005; accepted 17 September 2005
Available online 21 October 2005

Abstract

We report here a review of the seventh mitochondrial DNA (mtDNA) exercise undertaken by the Spanish and Portuguese working group (GEP) of the International Society for Forensic Genetics (ISFG) corresponding to the period 2003–2004. Five reference bloodstains from five donors (M1–M5), a mixed stain of saliva and semen (M6), and a hair sample (M7) were submitted to each participating laboratory for nuclear DNA (nDNA; autosomal STR and Y-STR) and mtDNA analysis. Laboratories were asked to investigate the contributors of samples M6 and M7 among the reference donors (M1–M5). A total of 34 laboratories reported total or partial mtDNA sequence data from both, the reference bloodstains (M1–M5) and the hair sample (M7) concluding a match between mtDNA profiles of M5 and M7. Autosomal STR and Y-STR profiling was the preferred strategy to investigate the contributors of the semen/saliva mixture (M6). Nuclear DNA profiles were consistent with a mixture of saliva from the donor (female) of M4 and semen from donor M5, being the semen (XY) profile the dominant component of the mixture. Strikingly, and in contradiction to the nuclear DNA analysis, mtDNA sequencing results yield a more simple result: only the saliva contribution (M4) was detected, either after preferential lysis or after complete DNA digestion. Some labs provided with several explanations for this finding and carried out additional experiments to explain this apparent contradictory result. The results pointed to the existence of different relative amounts of nuclear and mtDNAs in saliva and semen. We conclude that this circumstance could strongly influence the interpretation of the mtDNA evidence in unbalanced mixtures and in consequence lead to false exclusions. During the GEP-ISFG annual conference a validation study was planned to progress in the interpretation of mtDNA from different mixtures.

Keywords: mtDNA; Standardization; Collaborative exercise; Mixed samples; Control region; SNPs

1. Introduction

The participation in collaborative exercises is a valuable tool to check and compare methodologies, results and statistical analyses among laboratories. At the same time, participation in such exercises is an essential part of control–quality procedures in many laboratories aimed to guarantee the quality of results submitted for a laboratory.

The Spanish and Portuguese working group (GEP) of the International Society for Forensic Genetics (ISFG) comprises forensic genetic laboratories from Spain, Portugal, and several Latin–American countries. With the aim of comparing different laboratory technical strategies and statistical approaches towards a standardization of methodologies and procedures, once a year, the GEP-ISFG organizes a collaborative exercise on DNA profiling [1–6]. During the last 7 years the exercise includes a mtDNA test. The number of participating laboratories in this mtDNA exercise has increased progressively from five in the first 1997–1998 [1] exercise to 34 in the present one.

The 2003–2004 exercise consisted of both a simulated paternity case and a criminal case. In the present edition of the GEP proficiency testing trial, seven samples were submitted to the participating laboratories: five bloodstains from different donors (M1–M5), a mixture sample (M6) and four hair shaft fragments from the same individual (M7). The exercise proposed to investigate the following questions: (a) could the donors of samples M1 and M2 be the children of the donor of M3? (b) could M1 and M2 share the same paternal lineage? (c) could M6 be compatible with a mixture of samples M4 and M5? and finally (d)

could the sample M7 belong to any of the donors of M4 or M5 bloodstains?

We will particularly discuss the mtDNA results obtained with the mixed stain (saliva/semen) that were in apparent contradiction with those obtained with the autosomal STR markers.

2. Materials and methods

Participating laboratories analyzed the samples according with their own lab standard protocols and strategies. By means of a questionnaire, participant laboratories were requested to provide detailed information on the processing of the samples. Table 1 summarizes technical details related to the analysis.

2.1. Samples

The organization and coordination of the GEP-ISFG exercise 2003–2004 was carried out by the Quality Assurance Unit (Instituto Nacional de Toxicología y Ciencias Forenses, Ministerio de Justicia, Madrid, Spain).

Five bloodstains (M1–M5), a body fluid mixture sample (M6) and four hair shaft fragments (M7) were submitted to all participants. Bloodstains were prepared by applying 100 μl of whole blood from healthy donors onto absorbent Whatman paper (Whatman Bioscience) and air-dried before distribution. To prepare the mixed sample M6, 100 μl of saliva were applied onto Whatman paper and, after air-dried, an additional aliquot of 50 μl of a 1:20 dilution of semen was added.

The final report edited by the coordinators included the compilation of data and the methodological strategies employed by all participants, as well as additional information about the origin of the samples: (a) M1 and M2 bloodstains belonged to two brothers, (b) M3 donor was the biological mother of M1 and M2 donors, (c) M4 and M5 samples were donated by unrelated individuals, a woman and a man, respectively, (d) the mixed stain M6 consisted of a mixture of saliva from the same donor of M4 and semen from the donor of M5 and finally, (e) the hair shafts M7 belonged to the donor of M5.

2.2. DNA extraction and amplification

The amplification and sequencing strategies are summarized in Table 1. To obtain DNA from bloodstains and hair (70 and 64%, respectively) most laboratories used a classical phenol–chloroform procedure followed in some cases (37%) by an additional purification and concentration step using Microcon-100 or Centricon-100 centrifugal filter devices (Millipore, Billerica, MA, USA). DNA extraction using chelating resin such as Chelex-100 (Bio-Rad, CA, USA) was used by 20% of the laboratories. The remaining participants carried out alternative methods or commercial extraction kits such as FTA purification reagent (GIBCO-BRL), QIAmp (Qiagen, Hilden, Germany) or DNA IQ (Promega Corporation, Madison, WI, USA).

Three out of 19 participants that submitted mtDNA data for mixed stain M6 used the preferential lysis procedure [7]. All the participants amplified separately both hypervariable regions I and II (HVRI and HVRII). The primers described previously by Wilson et al. [8] were used by a high number of laboratories (65%), while 20% of participants employed the primers described by Vigilant et al. [9]. AmpliTaq Gold (AB, Applied Biosystems, Foster City, CA, USA) was the most frequently DNA polymerase used in the exercise. In addition, we observed a great variety of PCR protocols. Purification protocols of PCR products previous sequencing reaction involved the use of several commercial devices (see Table 1).

2.3. Sequencing

The majority of the laboratories applied dye terminator cycle sequencing; ∼75% of the laboratories used BigDye chemistry (AB) for cycle sequencing, whereas 23% used dichloro-rhodamine terminators (AB). Only one participant (Lab 15) used dye primers. All laboratories were consistent with the use of primers used during HVRI and HVRII amplifications in all the samples. Most of the laboratories (97%) used the automated DNA sequencers of Applied Biosystems (ABI 310, ABI 3100, ABI 377). Electrophoresis was performed using both, slab-gels and capillaries, but the preferred analytical method was the capillary electrophoresis (73%).

3. Results and discussion

GEP-ISFG laboratories reporting partial or complete mtDNA results increased from 18% in 1998 to 37% in 2004 (34 out 93). Table 2 summarizes the level of participation for the different samples included in the exercise, as well as the consensus rates obtained per sample, and Table 3 shows the sequences reported by each laboratory obtained for the different samples. The GEP-ISFG working group considers that a sequence result is consensual when there are at least five participant laboratories involved and at least 70% of them report exactly the same result.

3.1. MtDNA sequence data from bloodstains M1–M5

The results are summarized in Table 3. A high rate of coincidence was observed: 97% in M1–M2–M3, 93% in M4, and 94% in M5. Five laboratories reported some level of heteroplasmy at the polycytosine region of HVRII in samples M1, M2, M3 and M5.

Two types of errors were found:

(a) *Edition mistakes*: lab 1 omitted the common insertion 315.1C and reported base shift in M4 and M5 samples. Until now the design of the collaborative exercise, not

Table 1
Amplification and sequencing strategies used by the participants

Lab n°	HVRI/ HVRII primers	Edited	Polymerase	N° cycles blood/hair	Post-PCR purification	Sequencing chemistry	DNA sequencing
1	Other	15983–16410/117–426	Taq polymerase (invitrogen)	31/31	Wizard (Promega)	BigDye term (AB)	ABI 310
2	[8]	16028–16365/73–340	AmpliTaqGold (AB)	32/36	Microspin (Amersham)	Dichlo-rhod (AB)	ABI 3100
3	[9]	16022–16410/56–510	FIREPol (Gentaur)	40/40	Centricon-100 (Millipore)	BigDye term (AB)	ABI 310
4	[9]	16033–16385/30–360	AmpliTaqGold (AB)	30/36	Microcon-100 (Millipore)	Dichlo-rhod (AB)	ABI 310
5	Other	16024–16569/1–580	Taq polymerase (invitrogen)	36/36	Microspin (Amersham)	Dichlo-rhod (AB)	ABI 3100
6	Other	16023–16400/50–410	AmpliTaqGold (AB)	38/38	Microspin (Amersham)	BigDye term (AB)	ABI 310
7	[8]	16024–16325/73–340	AmpliTaqGold (AB)	38/36	Exo-SapIT (USB)	BigDye term (AB)	ABI 310
8	[8]	16024–16365/73–340	AmpliTaqGold (AB)	36/36	Centricon-100 (Millipore)	Dichlo-rhod (AB)	ABI 377
9	[8]	16110–16401	Taq polymerase (invitrogen)	35/35	‘QIAquick (Qiagen)	BigDye term (AB)	ABI 377
10	[8]	16024–16365/73–340	AmpliTaq (AB)	30/36	Microcon-100 (Millipore)	BigDye term (AB)	ABI 310
11	[8]	16024–16310/73–340	AmpliTaqGold (AB)	38/40	QIAquick (Qiagen)	BigDye term (AB)	ABI 3730
12	[8]	16000–16401/58–480	AmpliTaq (AB)	32/32	Centricon-100 (Millipore)	BigDye term (AB)	ABI 310
13	[8]	16024–16365/73–340	AmpliTaqGold (AB)	36/36	Microcon-100 (Millipore)	BigDye term (AB)	ABI 377
14	[8]	15998–16391/49–407	AmpliTaq (AB)	Non-specified	Microcon-100 (Millipore)	BigDye term (AB)	ABI 3100
15	[8]	16024–16365/72–340	Non-specified	35/35	Wizard (Promega)	Dye primers (non-specified)	Visible Gentics
16	[8]	16024–16383/66–319	EcoTaq polymerase (ecogen)	30/30	Microcon-100 (Millipore)	BigDye term (AB)	ABI 3100
17	[8]	16025–16365/73–340	AmpliTaqGold (AB)	30/35	Microcon-100 (Millipore)	Dichlo-rhod (AB)	ABI 310
18	[8]	16024–16365/73–240	AmpliTaqGold (AB)	36/36	Microcon-100 (Millipore)	BigDye term (AB)	ABI 310
19	Other	16070–16400/80–340	Non-specified	39/39	Non-specified	BigDye term (AB)	ABI 310
20	[8]	16024–16365/73–340	Biotools polymerase (biotools)	36/36	UltracleanPCR (MoBio)	BigDye term (AB)	ABI 310
21	[8]	16024–16365/73–340	AmpliTaq (AB)	36/36	Microcon-100 (Millipore)	Dichlo-rhod (AB)	ABI 310
22	[8]	16024–16365/73–340	Biotools polymerase (biotools)	36/36	Bioclean columns (Biotools)	BigDye term (AB)	ABI 310
23	[8]	16033–16391/57–408	AmpliTaqGold (AB)	30/35	QIAquick (Qiagen)	BigDye term (AB)	ABI 310
24	[8]	16024–16365/73–340	AmpliTaqGold (AB)	32/36	Microcon-100 (Millipore)	BigDye term (AB)	ABI 310
25	[8]	16024–16365/72–340	AmpliTaqGold (AB)	36/32	QIAquick (Qiagen)	BigDye term (AB)	ABI 377
26	[8]	15997–16395	Taq polymerase (invitrogen)	Non-specified	Microcon-100 (Millipore)	BigDye term (AB)	ABI 377

Table 1 (*Continued*)

Lab n°	HVRI/ HVRII primers	Edited	Polymerase	N° cycles blood/hair	Post-PCR purification	Sequencing chemistry	DNA sequencing
27	[8]	16075–16362/111–340	Taq polymerase (fermentas)	Non-specified	Exo-SapIT (USB)	BigDye term (AB)	ABI 310
28	[8]	16024–16365/73–340**	AmpliTaqGold (AB)	36/36	Microspin (Amersham)	BigDye term (AB)	ABI 3100
29	[8]	16024–16365/73–340	AmpliTaqGold (AB)	35/35	Microcon-100 (Millipore)	Dichlo-rhod (AB)	ABI 310 y 377
30	Other	16011–16370/55–349	Non-specified	32/32	Wizard (Promega)	BigDye term (AB)	ABI 377
31	[8,9]	16017–16380/49–409	DyNAzyme (finnzymes)	30	QIAquick (Qiagen)	BigDye term (AB)	ABI 310
32	[9]	16023–16400/50–430	Biotools polymerase (Inm Diag)	Non-specified	QIAquick (Qiagen)	BigDye term (AB)	ABI 3100
33	[8]	15978–16419/29–429	AmpliTaq (AB)	Non-specified	Centricon-100 (Millipore)	BigDye term (AB)	ABI 310
34	[8]	16024–16383/73–340	TaqDNA polimerase (Promega)	35/35	Microspin (Amersham)	BigDye term (AB)	ABI 3100

include the submission of raw data therefore is difficult to clarify if the errors happen during edition process or when the laboratories submitting the report (typographical error).

(b) *Nomenclature deficiencies*: there are a number of laboratories that do not use the nomenclature recommended by the ISFG mtDNA guidelines [10,11] to describe point heteroplasmy and length variants (see lab 28 in Table 3). In addition, 45% of laboratories reported the double HVRII C-insertion in the homopolymeric tract simply as ''309.2C'' instead of ''309.1C, 309.2C''.

3.2. mtDNA sequence data from a saliva/semen mixture (M6)

In comparison with the number of participating laboratories for the M1–M5 samples, the analysis of the mixture M6 (saliva from M4 donor/semen from M5 donor) was lower (19 labs). This low participation could be due, in part, to the fact that many laboratories do not consider mtDNA as an appropriate marker for DNA mixtures. In fact, in routine forensic casework mtDNA typing analysis is not frequently employed for DNA typing from body fluid stains. Moreover, many laboratories only participate in the paternity exercise and, normally, just analyse blood samples.

Table 2
Number of participants in the 2003–2004 GEP-ISFG exercise

	No. of participant labs	Results	
		Consensus	Non-consensus
M1	30	29 (97%)	1 (3%)
M2	30	29 (97%)	1 (3%)
M3	30	29 (97%)	1 (3%)
M4	30	28 (93%)	2 (7%)
M5	31	29 (94%)	2 (6%)
M6	19	13 (68%)	6 (32%)
M7	28	23 (82%)	5 (18%)

All laboratories carried out autosomal STRs typing in this sample. Some of them extracted the DNA by complete digestion and others by preferential lysis. In both cases, the consensus result was an unequal DNA mixture (3:1) compatible with the M4 and the M5 genetic profiles (data not shown), but the male component was by far the predominant one. This is a logical finding since there are more cells in semen than in saliva fluid. The laboratories performing Y-STR analyses only obtained a single haplotype which matched those obtained from M5.

Therefore, the expected consensus mtDNA sequence of the mixed stain M6, should have been 16,266Y 263G 309.1C 315.1C including length heteroplasmy in HVRII (tracts 303–309). Nevertheless, mtDNA sequencing analysis of the saliva/semen mixture M6 produced an unexpected consensus result (13/19 laboratories ∼70%) because only the saliva profile (M4) was detected, either after preferential lysis or complete DNA digestion. Six (30%) laboratories reported the expected mixed mtDNA sequence result. Analyzing the participant's data related to extraction, amplification and sequencing strategies, no methodological differences were observed between these two groups. Therefore, the absence of additional information such as detailed data of extraction procedure, microscopic analysis (if performed) or electropherograms turns difficult to infer the cause why several laboratories detected both components of the mixture. In relation to this question, at the annual GEP-ISFG 2004 meeting the members expressed the utility and the need

Table 3
Results reported by the 34 participating laboratories

	M1	M2	M3	M4	M5	M7
CS	16298C 195C 263G 309.1C 309.2C 315.1C	16298C 195C 263G 309.1C 309.2C 315.1C	16298C 195C 263G 309.1C 309.2C 315.1C	263G 315.1C	16266T 263G 309.1C 315.1C	16266T 263G 309.1C 315.1C
Lab						
1	–	–	–	262G 309.1C	16265T 262G 309C 309.1T 309.2C	16265T 262G 309C 309.1T 309.2C
2	CS	CS	CS	CS	16266T 263G 309.1C 309.2C 315.1C	16266T 263G 309.1C 315.1C
3	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	CS	CS	263G 315.1C
4	CS + 309.3C	CS + 309.3C	CS + 309.3C	CS	16266T 263G 309.1C 309.2C 315.1C	16266T 263G 309.1C 309.2C 315.1C
5	CS	CS	CS	CS	CS	CS
6	CS	CS	CS	CS	CS	CS
7	16298C 195C 263G 309.2C 315.1C	16298 C 195C 263G 309.2C 315.1C	16298 C 195C 263G 309.2C 315.1C	CS	CS	CS
8	CS	CS	CS	CS	CS	CS
9	–	–	–	CS[a,b]	CRS[a,c]	CS[a,c]
10	CS	CS	CS	CS	CS	CS
11	CS	CS	CS	263G 309.1C 315.1C	CS	CS
12	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	CS	CS	CS
13	CS	CS	CS	CS	CS	–
14	CS	CS	CS	CS	CS	CS
15	16298C[a]	CS	CS	–	–	–
16	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	CS	CS	263G 315.1C
17	CS	CS	CS	CS	CS	CS
18	CS	CS	CS	CS	CS	CS
19	CS	CS	CS	CS	CS	–
20	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	CS	CS	CS
21	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	CS	CS	CS
22	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	CS	CS	–
23	CS	CS	CS	CS	CS	CS
24	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	CS	CS	CS
25	CS	CS	CS	CS	CS	CS
26	16298Ca	16298Ca	16298Ca	–	16266T[a]	16266T[a]
27	–	–	–	CS	CS	16266T[a]

Table 3 (*Continued*)

	M1	M2	M3	M4	M5	M7
28	16298C 195C 263G 309CC -CCCC 315C	16298C 195C 263G 309CC-CCCC 315C	16298C 195C 263G 309CC-CCCC 315C	263G 315C ins	16366T 263G 309C-CC 315C	16189T + C 16223C + T 16266C + T 16278C + T 16294C + T 16309A + G 146T + C 152T + C 19 T + C 263A + G 309C–CC 315 C
29	CS	CS	CS	CS	CS	CS
30	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	CS	CS	CS
31	CS	CS	CS	–	–	–
32	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	CS	CS	263G 309.1C 315.1C
33	–	–	–	CS	CS	CS
34	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	16298C 195C 263G 309.2C 315.1C	–	–	–
n	29/30	29/30	29/30	28/30	29/31	23/28

CS: consensus sequence for the reading frames 16024–16365 in HVRI and 73–340 for HVRII; (–) result not reported.
[a] Only HVRI was analyzed and/or reported.
[b] (16149–16394) edition.
[c] (16149–16402) edition.

to attach electropherograms for confirmation of the results and evaluation, in such cases where discrepancies were observed.

There was only a single difference (at site 16,266) between M4 and M5 samples. With the aim of finding additional polymorphisms that might make it possible to discriminate the M4 and M5 samples, laboratories 5 and 25 analyzed other D-Loop segments outside HVRI and HVRII, namely nucleotide positions 16,159 to 17 and 340 to 576. However, these samples showed the same sequence profile at this fragment (16,519C). Therefore, this information was not useful to distinguish between the two samples. In order to clarify this supposed contradictory result between mitochondrial and nuclear DNA several laboratories proposed alternative hypothesis and carried out additional experiments.

3.2.1. Hypothesis 1: primer binding site mutation in the mtDNA from the semen

Laboratories 5 and 25 performed HVRI amplifications using different primer sets (L15997/H16395; L16159/H16395; L16159/H00017) in order to know whether a mutation in the primer binding sites in the sperm DNA could explain the paradox. Nevertheless, all analysis yielded the same result in M6.

3.2.2. Hypothesis 2: mutation in 16,266 nucleotide position in the mtDNA from semen

Lab 25 reproduced the same scenario of the M6 stain using other saliva and semen donors different from the ones used in the collaborative exercise (100 μL of saliva from donor A plus 50 μL of semen diluted 1:20 from donor B). In agreement with the result obtained with M6 sample, only the saliva haplotype was detected, thus indicating that the problem was not caused by the semen of M5 donor.

3.2.3. Hypothesis 3: lower mtDNA copy number in semen than in saliva

(i) Test 1: PCRs from serial dilutions of semen and saliva DNAs. Lab 25 analyzed semen and saliva fluids separately. DNA was extracted from 10 μl of semen from a donor different from the one in the exercise and from 10 μl of saliva from another donor. The amount of nuclear DNA was quantified by slot blot hybridization (Quantiblot, AB) and the results showed that the nuclear DNA concentration from saliva was lower than the DNA concentration from semen. This could explain the results obtained in M6 by autosomal STR analyses, where the male component was the predominant one.

In order to estimate the amount of mitochondrial DNA in each fluid this laboratory carried out a semi-quantitative PCR consisting of HVRI amplifications using decreasing serial DNA dilutions from the semen and saliva fluids. The results were contrary to those obtained with nuclear DNA. As indicated in Fig. 1, HVRI amplicons were detected up to 0.2 pg/μL nuclear DNA from saliva, while no PCR product was detected in the ten times more concentrated extract from semen (2 pg/μL). This finding clearly suggested that the amount of mitochondrial DNA in saliva was higher than in semen.

(ii) *Test 2*: analysis of mtDNA coding SNPs by SNaPshot technology, a more sensitive method than sequencing.

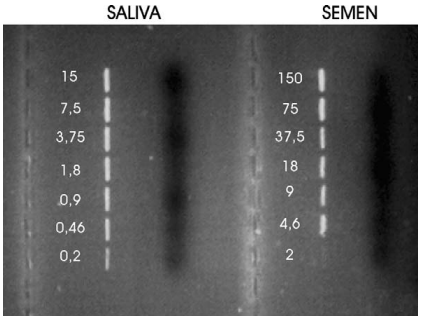


Fig. 1. HVRI amplifications from serial dilutions of saliva and semen DNAs. Numbers indicate the nuclear DNA concentrations in pg/μl.

Lab 5 investigated 15 nucleotide positions in the mitochondrial coding region (3010, 3915, 3992, 4216, 4336, 4529, 4580, 4769, 4793, 6776, 7028, 10,398, 10,400, 10,873 and 14,766) by two PCR multiplexes and SNaPshot methodology in samples M4, M5 and M6 following the protocol of Quintáns et al. (2004). The electropherograms showed a difference at position 3010 between the M4 and M5 samples corresponding to different haplogroup affiliations of samples M4 (haplogroup H1) and M5 (haplogroup H*) (see Fig. 2).

The analysis of M6 showed the expected results, namely, a G/A nucleotide mixture at site 3010, being the adenine the predominant variant (which corresponds with to the saliva component). The same lab also performed the analysis of the 3010 position in a singleplex PCR reaction in order to confirm this result. Therefore, this methodology proved to be very useful because it is highly sensitive in comparison with sequencing. This test allowed confirming the predominant mtDNA component of the saliva in the mixture sample.

3.3. Hair (M7)

Twenty-eight laboratories reported complete or partial HVRI/II mtDNA results for the hair shaft sample. Twenty-four out of 28 laboratories submitted results for both hypervariable regions. The analysis of hair shaft yielded a high rate of success since the majority of them (23 out of 28 laboratories) reported the consensus sequence 16266T 263G 309.1C 315.1C (see Table 3). Lab 1 reported the same erroneous base assignment (due to a base shift) as for the bloodstains (see above), while three laboratories (3, 16 and 32) reported a sequence lacking the 16,266 polymorphism (note that a mutation reversion at the site [e.g. heteroplasmic in the donor sample M7] is unlikely since this site is not considered to be a mutational hotspot [12]); inspection of the original sequence electropherograms from the corresponding laboratories will help to corroborate this potential error. Finally, lab 28 reported a sequence that was clearly contaminated with a sample belonging to haplogroup L2a1 [13,14].

The majority of the participating laboratories agreed that the donor of the sample M5 could not be excluded as the contributor of the hair (M7), rejecting M4 as a potential contributor. In agreement with their artifactual results, two laboratories (3 and 16), concluded that M4 was the donor of the sample M7.

Strikingly, only four laboratories considered the possibility of a mutational event in the hair sample at position 16,266. Several studies [4,15,16] have demonstrated that sequence heteroplasmy is a phenomenon commonly observed in hair shafts. Nevertheless, when reporting their conclusions, only four laboratories mentioned the issue. Following the recommendations of the ISFG DNA Commission, a single difference between an unknown sample and a reference sample cannot exclude a common maternal origin. Although 16,266 is not considered a mutation hotspot (see above; [12]), M4 could not be excluded as the source of M7. Besides, the recommendation of analyzing several fragments of a single hair has been also considered by the DNA Commission in order to evaluate possible heteroplasmies. In the present exercise only two laboratories reported to have analyzed more than one fragment independently. Lab 5 rejected the hypothesis of common maternal origin on the basis of an additional nucleotide difference at position 3010 between samples M4 and M5–M7 (see Table 4).

Table 4
Results obtained by lab 5 in mtDNA coding region SNPs

Sample	SNP profile
M1	7028T 14766C 4529A 4580G 10400C 4216T 10873T 3010G 3915G 3992C 4336T 4769G 4793A 6776T
M2	7028T 14766C 4529A 4580G 10400C 4216T 10873T 3010G 3915G 3992C 4336T 4769G 4793A 6776T
M3	7028T 14766C 4529A 4580G 10400C 4216T 10873T 3010G 3915G 3992C 4336T 4769G 4793A 6776T
M4	7028C 14766C 4529A 4580G 10400C 4216T 10873T **3010A** 3915G 3992C 4336T 4769G 4793A 6776T
M5	7028C 14766C 4529A 4580G 10400C 4216T 10873T **3010G** 3915G 3992C 4336T 4769G 4793A 6776T
M6	7028C 14766C 4529A 4580G 10400C 4216T 10873T **3010A** > **G** 3915G 3992C 4336T 4769G 4793A 6776T
M7	7028C 14766C 4529A 4580G 10400C 4216T 10873T **3010G** 3915G 3992C 4336T 4769G 4793A 6776T

Note that position 3010 is the only difference between M4 and M5.

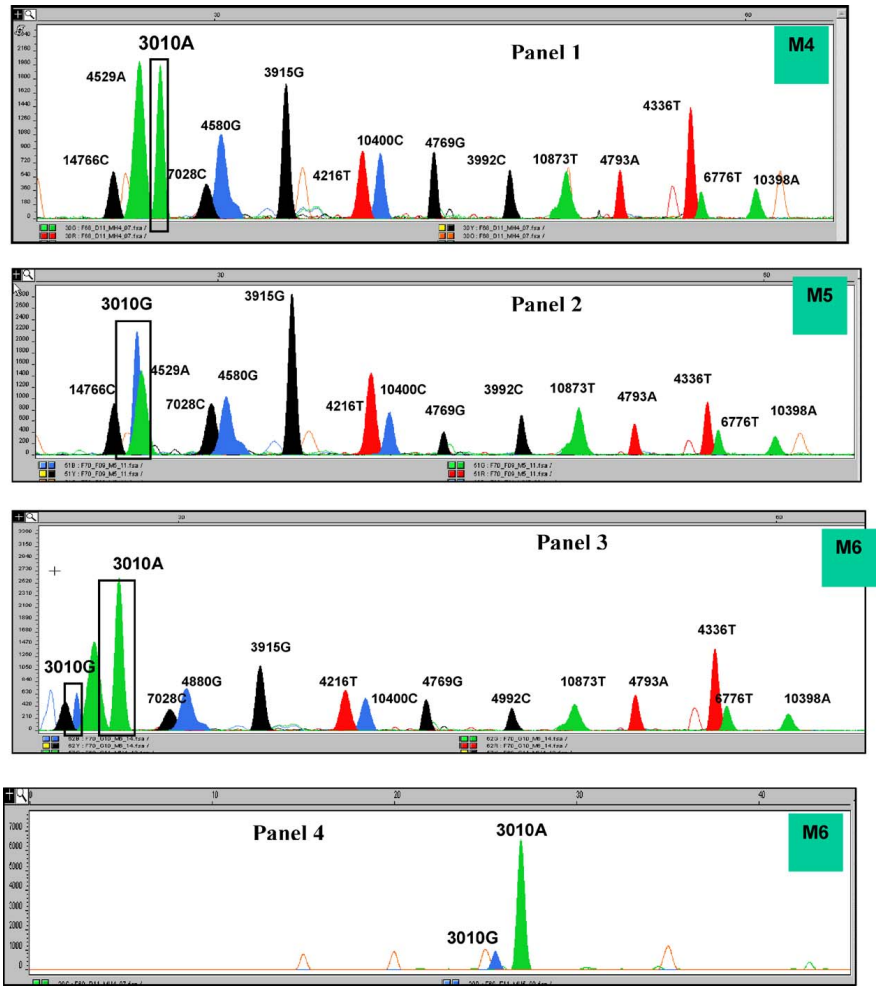


Fig. 2. Coding mtSNPs analysis of samples M4–M6 samples by SNaPshot minisequencing reaction. Panels 1–3: multiplex analysis of M4–M6 samples, respectively. Panel 4: single-plex analysis of M6 sample.

4. Final remarks

The results of the GEP 2003–2004 mtDNA exercise showed again an improvement in quality with respect to previous editions. The number of participating laboratories has slightly increased with respect to the last exercise [6], but this circumstance has not involved a decrease in the quality of the results. The main causes of error were documentation mistakes and nomenclature deficiencies. Contamination or sample mix-ups were also the cause of a few errors. Parti-

cularly important were the good results obtained for the hair shafts (M7) since most of the laboratories (only one exception) reported the consensus results. It is worth mentioning that a scrutiny of the results using a phylogenetic approach [6,12,16–19] could easily recognize most of the causes of errors.

A remarkable aspect constitutes the fact that a significant proportion of laboratories do not follow the international nomenclature recommended to describe the presence of insertions in the homopolymeric track of HVRII. This mistake was already acknowledged during the previous edition of the exercise and once again we emphasize the need of using a common nomenclature and encourage the laboratories to report the data following the recommended nomenclature of the ISFG.

Once again, the quality exercise here reported indicates that additional efforts are needed to improve mtDNA evidence interpretation. For instance, only few laboratories did take into account the fact that there was a single difference between sample M4 and M5–M7 sequences (16,266 nucleotide site). The mtDNA evidence interpretation is not always easy and depends on several factors such as tissue type (e.g. blood–blood or blood–hair comparisons), number of specimens involved (e.g. single hair–several hair analysis), the sequences themselves (e.g. full match, a single or few differences in hotspot positions, a single difference in a mutational stable position, sequence or length heteroplasmy, etc.) and the biological relationship between the samples under analysis (relatives or two samples from a single individual). Although the nucleotide site 16,266 is not a recognized hotspot, it would be desirable to observe additional nucleotide differences in order to safely conclude that samples M4 and M7 come from a different biological source. Here, the useful role of coding region variants seems to be of primordial importance.

The result obtained with the mixture M6 sample is also noteworthy. Although the expected result was a mixture sequences belonging to M4 and M5, the consensus result yield exclusively the M4 profile for both HVRI and HVRII. The mtDNA from sperm was not detected. Different experiments carried out by several laboratories contributed to partially solve the puzzle: the existence of different relative amounts of nuclear and mtDNAs in saliva and semen. Depending on tissue type the number of mtDNA molecules in somatic cells is variable, ranging from about 250 to about 1700 mtDNA copies [20,21]. On the other hand, the number of mtDNA copies in each spermatozoon is in the range of 50–100 [22,23]. The exclusive use of mtDNA for mixture samples analysis could in this case lead to a false exclusion.

The results obtained when analyzing the female and male fractions of the differential extractions are also of interest. Theoretically, when a preferential lysis is performed, the mitochondria located in the mid-pieces of spermatozoon and those situated in the female epithelial cells remain in the first (or female) fraction, whereas the heads of the sperm (without mitochondria) are carried to the second fraction.

Therefore, in this particular case, after preferential lysis, the expected result should be a mixture of M4 and M5 profiles in the first fraction and the lack of any mitochondrial amplicon in the second fraction. Many laboratories detected the mitochondrial female component in the male fraction, which would indicate that the lysis was not carried out correctly. As previously described [24], three possible scenarios could be expected in the second fraction: no mtDNA sequence detected, the presence of (only) the female profile, or the possible amplification of an insert of mtDNA present in the nuclear genome.

The analysis of the results obtained for the M6 sample turned out to be extremely interesting. A GEP-ISFG collaborative study was planned and is now in progress aimed to the complex interpretation of mtDNA when analyzing different body fluid mixtures.

Finally and concerning to mixture interpretation, it seems evident that the inferior participation in the analysis of the M6 sample shows the majority idea that mtDNA in not the marker of choice in the resolution of profiles in mixed samples (saliva:semen). STRs are most clearly the method of choice. Nevertheless, in a number of cases, it could be interesting to know the mitochondrial DNA (mtDNA) haplotypes that contributed to the mixture (e.g. degraded reference samples, exclusion of a maternal relationship between the victim and suspect in rape cases. . .) and in this type of cases we consider that the interpretation should be making with extreme care.

Acknowledgments

We would like to express our deep gratitude to Josefina Gomez. Instituto Nacional de Toxicología y Ciencias Forenses, Dep. Madrid, by the intense and efficient work made in the organization and coordination of this exercise during the last years. We also wish to thank Antonio Alonso from Instituto Nacional de Toxicología y Ciencias Forenses. Madrid for his many helpful comments and suggestions, and finally our gratitude goes to the anonymous donors of samples.

References

[1] J. Gómez, A. Carracedo, A review of the collaborative exercises of the Spanish and Portuguese ISFH working group, in: A. Carracedo, B. Brinkmann, W. Bär (Eds.), Advances in Forensic Haemogenetics, vol. 6, Springer, Berlin, 1996, pp. 695–699.

[2] J. Gómez, M.S. Rodriguez-Calvo, C. Albarrán, A. Amorím, J. Andradas, C. Cabrero, R. Calvet, D. Corach, M. Crespillo, C. Doutremepuich, O. García, H. Geada, M. Gene, S. Jimenez, J.A. Lorente, S.M. Marques-Santos, B. Mártinez-Jarreta, M. Martinez de Pancorbo, F. Montes, J.M. Ruiz de la Cuesta, P. Sanz, M.F. Terra-Pinheiro, M.C. Vide, A. Cariacedo, A review of the collaborative exercises of the Spanish and Portuguese ISFH working group, Int. J. Legal Med. 110 (1997) 273–277.

[3] J. Gómez, A. Carracedo, The 1999 collaborative exercises and proficiency test program on DNA typing of the Spanish and Portuguese of the International Society for Forensic Genetics (GEP-ISFH), Forensic Sci. Int. 114 (2000) 21–30.
[4] A. Alonso, A. Salas, C. Albarran, E. Arroyo, A. Castro, M. Crespillo, A.M. Di Lonardo, M.V. Lareu, C.L. Cubrija, M.L. Soto, J.A. Lorente, M.M. Semper, A. Palacio, M. Paredes, L. Pereira, A.P. Lezaun, J.P. Brito, A. Sala, M.C. Vide, M. Whittle, J.J. Yunis, J. Gómez, Results of the 2000 collaborative exercise and proficiency testing program of mitochondrial DNA of the GEP-ISFG: an inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor, Forensic Sci. Int. 125 (2002) 1–7.
[5] L. Prieto, M. Montesino, A. Salas, A. Alonso, C. Albarrán, S. Álvarez, M. Crespillo, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A. González de la Vega, L. Gusmão, C.M. López, M. López-Soto, J.A. Lorente, M. Malaghini, C.A. Martínez, N.M. Modesti, A.M. Palacio, M. Paredes, S.D.J. Pena, A. Pérez-Lezaun, J.J. Pestano, J. Puente, A. Sala, M.C. Vide, M.R. Whittle, J.J. Yunis, J. Gómez, The 2001 GEP–ISFG Collaborative Exercise on mtDNA: assessing the cause of unsuccessful mtDNA PCR amplification of hair shaft simples, Forensic Sci. Int. 134 (2003) 46–53.
[6] A. Salas, L. Prieto, M. Montesino, C. Albarran, E. Arroyo, M.R. Paredes-Herrera, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A. González de la Vega, C. Alves, C.M. López, M. López-Soto, J.A. Lorente, A. Picornell, R.M. Espinheira, A. Hernández, A.M. Palacio, M. Espinoza, J.J. Yunis, A. Pérez-Lezaun, J.J. Pestano, J.C. Carril, D. Corach, M.C. Vide, V. Alvarez-Iglesias, M.F. Pinheiro, M.R. Whittle, A. Brehm, J. Gómez, Mitochondrial DNA error prophylaxis: assessing the causes of errors in the GEP'02–03 proficiency testing trial, Forensic Sci. Int. 148 (2005) 191–198.
[7] P. Gill, A.J. Jeffreys, D.J. Werrett, Forensic application of DNA extraction ''fingerprints'', Nature 318 (1985) 577–579.
[8] M.R. Wilson, J.A. DiZinno, D. Polanskey, J. Replogle, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, Int. J. Legal Med. 108 (1995) 68–74.
[9] L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A.C. Wilson, African populations and the evolution of human mitochondrial DNA, Science 253 (1991) 1503–1507.
[10] A. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Shneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tully, M. Wilson, DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79–85.
[11] G. Tully, W. Bär, B. Brinkmann, A. Carracedo, P. Gill, N. Morling, W. Parson, P. Schneider, Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles, Forensic Sci. Int. 124 (2001) 83–91.
[12] H.-J. Bandelt, L. Quintana-Murci, A. Salas, V. Macaulay, The fingerprint of phantom mutations in mitochondrial DNA data, Am. J. Hum. Genet. 71 (2002) 1150–1160.
[13] A. Salas, M. Richards, T. De la Fé, M.V. Lareu, B. Sobrino, P. Sánchez-Diz, V. Macaulay, A. Carracedo, The making of the African mtDNA landscape, Am. J. Hum. Genet. 71 (2002) 1082–1111.
[14] A. A.Salas, M. Richards, M.V. Lareu, R. Scozzari, A. Coppa, A. Torroni, V. Macaulay, A. Carracedo, The African diaspora: mitochondrial DNA and the Atlantic slave trade, Am. J. Hum. Genet. 74 (2004) 454–465.
[15] A. Salas, M.V. Lareu, A. Carracedo, Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report, Int. J. Leg. Med. 114 (2001) 186–190.
[16] G. Tully, S.M. Barritt, K. Bender, E. Brignon, C. Capelli, N. Dimo-Simonin, C. Eichmann, C.M. Ernst, C. Lambert, M.V. Lareu, B. Ludes, B. Megav, W. Parson, H. Pfeiffer, A. Salas, P.M. Schneider, E. Staalstrom, Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts, Forensic Sci. Int. 140 (2004) 1–11.
[17] H.-J. Bandelt, A. Salas, C. Bravi, Problems in the FBI mtDNA database, Science 305 (2004) 1402–1404.
[18] H.-J. Bandelt, A. Salas, S. Lutz-Bonengel, Artificial recombination in forensic mtDNA population databases, Int. J. Legal. Med. 118 (2004) 267–273.
[19] Y.G. Yao, C.M. Bravi, H.-J. Bandelt, A call for mtDNA data quality control in Forensic science, Forensic Sci. Int. 141 (2004) 1–6.
[20] E.D. Robin, R. Wong, Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells, J. Cell Phys. 136 (1988) 507–513.
[21] K.L. Veltri, M. Espiritu, G. Singh, Distinct genomic copy number in mitochondria of different mammalian organs, J. Cell Phys. 143 (1990) 160–164.
[22] F. Ankel-Simons, J.M. Cummins, Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution, Proc. Natl. Acad. Sci. 93 (1996) 13859–13863.
[23] L. Bromhan, A. Eyre-Walker, N.H. Smith, J. Maynard Smith, Mitochondria Steve: paternal inheritance of mitochondria in humans, Trend Ecol. Evol. 18 (2003) 1–3.
[24] M.M. Holland, T.J. Parsons, Mitochondrial DNA sequence analysis — validation and use for forensic casework, Forensic Sci. Rev. 11 (1999) 21.

Artículo 4: Analysis of body fluid mixtures by mtDNA sequencing: An inter-laboratory study of the GEP-ISFG working group

Montesino M, Salas A, Crespillo M, Albarran C, Alonso A,

Álvarez-Iglesias V, et al.

(2007) *Forensic Sci Int* 168: 42-56

Análisis de mezclas de fluidos biológicos mediante secuenciación de ADNmt: estudio colaborativo del grupo de trabajo del GEP-ISFG

ABSTRACT/RESUMEN:

El grupo de trabajo de ADN mitocondrial (ADNmt) del GEP-ISFG (Grupo Español y Portugués de la International Society for Forensic Genetics) llevó a cabo un ejercicio inter-laboratorio consistente en el análisis de patrones de secuenciación de ADNmt en manchas de mezclas (saliva/semen y sangre/semen). Las mezclas se prepararon con saliva o sangre de un donante femenino y tres diluciones diferentes de semen (puro, 1:10 y 1:20) con el objetivo de simular un caso forense. Todos los laboratorios extrajeron el ADN mediante lisis preferencial y amplificaron y secuenciaron la región hipervariable I (HVS-I) de ADNmt. STRs autosómicos y de cromosoma Y también fueron analizados para comparar los resultados de ADN nuclear y de ADNmt del mismo extracto de ADN. También se analizó una muestra preparada con semen procedente de un individuo vasectomizado. Los resultados fueron razonablemente consistentes entre los laboratorios para las primeras fracciones, pero no para las segundas, para las que algunos laboratorios reportaron problemas de contaminación. En las primeras fracciones, los haplotipos femeninos y masculinos fueron detectados en general en aquellas muestras preparadas con semen sin diluir. Sin embargo en la mayoría de las mezclas preparadas con semen diluído sólo se detectó el haplotipo femenino, sugiriendo que el número de copias de ADNmt por célula es menor en semen que en saliva o sangre. Aunque el nivel de detección del componente masculino disminuyó en concordancia con el grado de dilución del semen, la pérdida de señal no fue consistentemente uniforme en cada electroferograma. Además, se observaron diferencias entre las mezclas preparadas con fluidos de diferentes donantes y también entre los distintos fluidos. Concluimos que las características particulares de cada mancha pueden influir de manera importante en la interpretación de la evidencia de ADNmt en mezclas forenses, pudiendo dar en algún caso una falsa exclusión.

Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International 168 (2007) 42–56
Forensic Science International
www.elsevier.com/locate/forsciint

Analysis of body fluid mixtures by mtDNA sequencing: An inter-laboratory study of the GEP-ISFG working group

M. Montesino [a], A. Salas [b], M. Crespillo [c], C. Albarrán [d], A. Alonso [d], V. Álvarez-Iglesias [b], J.A. Cano [e], M. Carvalho [f], D. Corach [g], C. Cruz [h], A. Di Lonardo [i], R. Espinheira [h], M.J. Farfán [j], S. Filippini [i], J. García-Hirschfeld [d], A. Hernández [k], G. Lima [l], C.M. López-Cubría [e], M. López-Soto [j], S. Pagano [m], M. Paredes [c], M.F. Pinheiro [l], A.M. Rodríguez-Monge [a], A. Sala [g], S. Sóñora [m], D.R. Sumita [n], M.C. Vide [f], M.R. Whittle [n], A. Zurita [k], L. Prieto [a,*]

[a] *Policía Científica, Madrid, Spain*
[b] *Unidad de Genética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela and Centro Nacional de Genotipado (CeGen), Hospital Clínico Universitario, Galicia, Spain*
[c] *Instituto Nacional de Toxicología y Ciencias Forenses, Barcelona, Spain*
[d] *Instituto Nacional de Toxicología y Ciencias Forenses, Madrid, Spain*
[e] *Dirección General de la Guardia Civil, Madrid, Spain*
[f] *Instituto de Medicina Legal de Coimbra, Portugal*
[g] *Servicio de Huellas Digitales Genéticas, Universidad de Buenos Aires, Argentina*
[h] *Instituto de Medicina Legal de Lisboa, Portugal*
[i] *Banco Nacional de Datos Genéticos, Buenos Aires, Argentina*
[j] *Instituto Nacional de Toxicología y Ciencias Forenses, Sevilla, Spain*
[k] *Instituto Nacional de Toxicología y Ciencias Forenses, Canarias, Spain*
[l] *Instituto de Medicina Legal de Porto, Portugal*
[m] *Dirección Nacional de Policía Técnica, Montevideo, Uruguay*
[n] *Genomic Engenharia Molecular Ltda, São Paulo, Brazil*

Received 24 March 2006; received in revised form 15 June 2006; accepted 17 June 2006
Available online 8 August 2006

Abstract

The mitochondrial DNA (mtDNA) working group of the GEP-ISFG (Spanish and Portuguese Group of the International Society for Forensic Genetics) carried out an inter-laboratory exercise consisting of the analysis of mtDNA sequencing patterns in mixed stains (saliva/semen and blood/semen). Mixtures were prepared with saliva or blood from a female donor and three different semen dilutions (pure, 1:10 and 1:20) in order to simulate forensic casework. All labs extracted the DNA by preferential lysis and amplified and sequenced the first mtDNA hypervariable region (HVS-I). Autosomal and Y-STR markers were also analysed in order to compare nuclear and mitochondrial results from the same DNA extracts. A mixed stain prepared using semen from a vasectomized individual was also analysed. The results were reasonably consistent among labs for the first fractions but not for the second ones, for which some laboratories reported contamination problems. In the first fractions, both the female and male haplotypes were generally detected in those samples prepared with undiluted semen. In contrast, most of the mixtures prepared with diluted semen only yielded the female haplotype, suggesting that the mtDNA copy number per cell is smaller in semen than in saliva or blood. Although the detection level of the male component decreased in accordance with the degree of semen dilution, it was found that the loss of signal was not consistently uniform throughout each electropherogram. Moreover, differences between mixtures prepared from different donors and different body fluids were also observed. We conclude that the particular characteristics of each mixed stain can deeply influence the interpretation of the mtDNA evidence in forensic mixtures (leading in some cases to false exclusions). In this sense, the implementation of preliminary tests with the

* Corresponding author at: Comisaría General de Policía Científica, Laboratorio de ADN, C/Julián González Segador s/n, 28043 Madrid, Spain. Tel.: +34 915822321; fax: +34 915822541.
E-mail addresses: lourditasmt@ya.com, mitocondrial.adn@policia.es (L. Prieto).

doi:10.1016/j.forsciint.2006.06.066

aim of identifying the fluids involved in the mixture is an essential tool. In addition, in order to prevent incorrect conclusions in the interpretation of electropherograms we strongly recommend: (i) the use of additional sequencing primers to confirm the sequencing results and (ii) interpreting the results to the light of the phylogenetic perspective.

Keywords: mtDNA mixture; Preferential lysis; Sequencing; HVS-I; Haplotype; mtDNA phylogeny

1. Introduction

The analysis of mixtures of fluids is routine practise in forensic casework usually related to sexual assault cases. These analyses are generally performed using preferential lyses followed by STR genotyping. In a number of cases, however, the mtDNA analyses can be the unique strategy (e.g. when analyzing degraded or low copy number samples). In other cases, establishing the mtDNA haplotypes of the contributors can add information to the legal investigation (e.g. to exclude a maternal relationship between victim and suspect in rape cases, or when trying to obtain some information about the (geographical) origin of the mtDNA carried by the unknown offender [1]).

Theoretically, when a preferential lysis is performed on semen mixed with other fluids (e.g. saliva, blood, or vaginal fluid), DNA from the non-spermatic cells remains in the first fraction, while the nuclear DNA (nDNA) from the spermatozoa remains in the second one. This is due to the fact that in the spermatic nuclei there are rich disulfide bond proteins [2], which give relative resilience (compared to epithelial and other cells) against the enzymatic treatment employed during DNA extraction [3]. Therefore, if the preferential lysis is effective, the first fraction should contain a mixture of male (from non-spermatic cells and the mid-pieces of spermatozoa) and female mtDNAs, whereas the second fraction should not retain any mtDNA from mitochondria (note however that some mtDNA inserts from the nuclear genome [4] could be interpreted as real mtDNA).

On the other hand, the number of mtDNA copies varies depending on the cell type [5,6]. It is unknown to what extent this fact could affect the detection of minor components in unbalanced mixtures. During the 2004 GEP-ISFG mtDNA proficiency exercise [7], a mixture stain (saliva from a female and semen diluted 1:20) was studied and the mtDNA sequencing analysis yielded an unexpected consensus result: only the HVS-I/HVS-II saliva haplotype was detected, while the male autosomal STR profile was predominant. Hence, the use (exclusively) of mtDNA analysis could in this case lead to a false exclusion. Several additional experiments were performed in order to clarify these apparent contradictory results. The results of these experiments pointed to the existence of different relative amounts of nuclear and mitochondrial DNAs in saliva and semen [7].

Forensic labs have demonstrated to have a great deal of experience in analysing nDNA when performing preferential lysis, but very little in mtDNA [8]. In order to shed light on the mtDNA patterns originated when analyzing mixtures of semen with other body fluids, the mtDNA-working group of the GEP-ISFG carried out the present inter-laboratory study.

2. Materials and methods

The stains were prepared using mixtures of fluids from three healthy couples (which hereafter will be referred to as couples 1, 2 and 3), each one made-up with samples from a male and a female donor. For each couple, the males donated their semen while the female provided the saliva and blood. Samples were prepared in the Policía Científica DNA lab in Madrid (Spain) by mixing saliva or blood with the same volume of semen. In order to simulate forensic casework, in each case the samples were prepared using three semen dilutions in saline buffer: pure, 1:10 and 1:20 (see Table 1). The fresh fluids were mixed in a laminar-flow hood, shaken, and subsequently, 100 μl of the mixture were

Table 1
Samples analysed in this inter-laboratory study

Female/male number pair	Female saliva/semen mixtures	Female blood/semen mixtures
Samples analysed by participating labs		
1	50 μl of saliva + 50 μl of pure semen	50 μl of blood + 50 μl of pure semen
	50 μl of saliva + 50 μl of semen 1/10	50 μl of blood + 50 μl of semen 1/10
	50 μl of saliva + 50 μl of semen 1/20	50 μl of blood + 50 μl of semen 1/20
2	50 μl of saliva + 50 μl of pure semen	50 μl of blood + 50 μl of pure semen
	50 μl of saliva + 50 μl of semen 1/10	50 μl of blood + 50 μl of semen 1/10
	50 μl of saliva + 50 μl of semen 1/20	50 μl of blood + 50 μl of semen 1/20
3	50 μl of saliva + 50 μl of pure semen	50 μl of blood + 50 μl of pure semen
	50 μl of saliva + 50 μl of semen 1/10	50 μl of blood + 50 μl of semen 1/10
	50 μl of saliva + 50 μl of semen 1/20	50 μl of blood + 50 μl of semen 1/20
Samples analysed by coordinating lab		
4 (Female 3 + vasectomized male)	50 μl of saliva + 50 μl of pure semen	50 μl of blood + 50 μl of pure semen
	50 μl of saliva + 50 μl of semen 1/10	50 μl of blood + 50 μl of semen 1/10
	50 μl of saliva + 50 μl of semen 1/20	50 μl of blood + 50 μl of semen 1/20

Table 2
HVS-I haplotypes of donors and the number of labs that analysed each couple

Couple	Donor	HG	Haplotype	No. of labs
1	Female	T	16294T 16304C	9
	Male	H	16362C	
2	Female	K	16192T 16224C 16284G 16286G 16311C 16362C	9
	Male	H	CRS[a]	
3	Female	H	16092C 16257T 16293G 16311C	8
	Male	J	16069T 16126C 16193T 16278T 16311C	
4	Female	H	16092C 16257T 16293G 16311C	1
	Male	J	16069T 16126C 16300G	

HG, the most likely haplogroup assignation.
[a] Cambridge reference sequence [27].

applied onto bloodstain cards (Whatman) and air-dried. All samples were produced as a single lot. The HVS-I haplotypes of each donor are shown in Table 2.

Previously, one lab tested couple 1 mixtures. Saliva and blood from the female donor mixed with undiluted semen and with semen diluted 1:5, 1:10, 1:20 and 1:40 were analysed by sequencing. This test was carried out with the aim of selecting proper semen dilutions in order to avoid excessive dilutions, which would prevent the detection of the male components in the mixtures.

Fourteen labs participated in the study. Each lab analysed two couples (12 samples), while each couple was analysed by at least eight labs (see Table 2). No *a priori* information was given to the participants concerning either the mtDNA haplotypes of the original donors or the semen dilutions. The only recommendation given to the labs was to select the central region of each stain for DNA extraction in order to avoid as much as possible heterogeneity in the distribution of the fluids in the stain. The labs were requested to fill in a questionnaire with the technical details of the analysis. Each lab used their routine methods to carry out preferential lysis, cell count, nuclear and/or mtDNA quantification, PCR amplification and automated DNA sequencing (both strands) of the HVS-I region. Several labs also performed autosomal and Y-STRs typing in both fractions (see Table 3). One of the participants additionally carried out the analyses extracting the DNA using common protocols (no preferential lysis).

Finally, the coordinating lab performed two additional tests. In the first test, to assess how the preferential lysis works in unmixed samples, the DNA from each fluid of each donor was extracted independently (without mixing); nDNA in the extracts was then quantified, and mtDNA HVS-I was amplified. In the second test, stains composed of saliva and blood from the female donor of couple 3 mixed with semen (pure, 1:10 and 1:20) obtained from a vasectomized individual (couple 4) (see Tables 1 and 2) were prepared and analysed (preferential DNA extraction, nDNA quantification, PCR amplification and automated sequencing of the region HVS-I).

3. Results

Results concerning cell count and DNA quantification were heterogeneous between labs; nevertheless, each lab reported coherent results depending on the corresponding semen dilution, that is to say, the number of spermatozoa detected as well as the amount of nuclear or mtDNA decreased in proportion to the semen dilution. The differences in DNA quantification between labs could be attributed to stain sub-sampling (amount of sample stain taken for the analysis), or to different DNA extraction efficiency and/or to different DNA quantification methodologies.

Two labs carried out mtDNA quantification by using real time PCR. As shown in Table 4, there are some differences in the results obtained in couple 2 between the two labs. Lab no. 1 amplified a 218 bp fragment located in the HV-II segment, while lab no. 12 performed the PCR in a 287 bp fragment located in the HV-I segment. In addition, both labs used different monitoring methodologies: SYBR-Green I and Taqman probe, respectively. These two different strategies, plus the possibility of differences in the efficiency of the DNA extraction and in the amount of the final extract volume, as well as differences in the stain sub-sampling, might explain the dissimilar amounts of mtDNA copies detected.

Results concerning mtDNA haplotypes are detailed in Tables 5 and 6. In total, 312 HVS-I sequencing analyses

Table 3
Methodology

Lab	First lysis digestion time	Stain	Quantification	Detection	Edited nucleotides	Autosomal/Y-STRs
1	1 h 50 min	Erythrosine 0.5%	mtDNA RT-PCR	ABI377XL	16024–16366	No/no
2	1 h	Christmas tree	Agarose gel	ABI310	–	Yes/yes
3	2 h	Methylene blue	Quantifiler human[a]	ABI310	16033–16391	Yes/yes
4	4 h	Christmas tree	Quantifiler human[a]	ABI310	16024–16365	No/no
5	90 min	–	–	ABI3100 Avant	16050–16400	Yes/yes
6	19 h (37 °C)	Giemsa	Quantifiler human[a]	ABI310	16023–16410	Yes/yes
7	1 h (37 °C)	Korin–Stockis	Quantifiler human[a]	ABI3100 Avant	–	Yes/yes
8	90 min	Christmas tree	Quantiblot[b]	ABI310	16024–16365	No/no
9	2 h	Christmas tree	Quantiblot[b]	ABI310	16025–16365	No/no
10	2 h	Methylene blue	Quantifiler human[a]	ABI3100	–	Yes/yes
11	2 h	Haematoxylin-eosin	Quantiblot[b]	ABI310	16024–16365	No/no
12	2 h	Christmas tree	Nuclear and mtDNA RT-PCR [9]	ABI377	16024–16365	No/no
13	2 h	Papanicolau	Quantiblot[b]	ABI3100	16024–16519	Yes/no
14	2 h	Erythrosine	Quantiblot[b]/Quantifiler[a] human	ABI377	16024–16365	No/no

(–) Not reported.
[a] Quantifiler™ Human DNA quantification kit (Applied Biosystems).
[b] Quantiblot® Human DNA quantitation kit (Applied Biosystems).

Table 4
mtDNA quantification results by using two different RT-PCR strategies: (i) amplification of a 218 bp HV2 fragment and detection by SYBR-Green I and (ii) amplification of a 287 bp HV1 fragment and detection by Taqman probe

Couple	Sample	Number of mtDNA copies (μl^{-1})	
		First lysis	Second lysis
(i) Amplification of a 218 bp HV2 fragment			
2	Saliva/pure semen	4.23×10^8	Undetectable
	Saliva/semen 1:10	5.78×10^8	Undetectable
	Saliva/semen 1:20	6.41×10^8	Undetectable
	Blood/pure semen	1.32×10^8	Undetectable
	Blood/semen 1:10	4.08×10^7	Undetectable
	Blood/semen 1:20	1.8×10^7	5.32
3	Saliva/pure semen	5.15×10^{15}	100.7
	Saliva/semen 1:10	3.15×10^8	2.83
	Saliva/semen 1:20	1.12×10^8	Undetectable
	Blood/pure semen	3.6×10^8	0.693
	Blood/semen 1:10	3.6×10^7	5.67
	Blood/semen 1:20	2.06×10^7	Undetectable
(ii) Amplification of a 266 bp HV1 fragment			
1	Saliva/pure semen	8.50×10^5	4.32×10^3
	Saliva/semen 1:10	1.04×10^5	2.98×10^3
	Saliva/semen 1:20	3.38×10^4	9.46×10^2
	Blood/pure semen	1.83×10^6	1.19×10^3
	Blood/semen 1:10	2.26×10^6	3.07×10^3
	Blood/semen 1:20	1.09×10^6	4.58×10^2
2	Saliva/pure semen	1.01×10^6	4.42×10^4
	Saliva/semen 1:10	1.51×10^6	4.02×10^4
	Saliva/semen 1:20	1.32×10^6	6.38×10^4
	Blood/pure semen	1.33×10^6	2.14×10^3
	Blood/semen 1:10	8.55×10^5	1.68×10^3
	Blood/semen 1:20	1.06×10^6	1.38×10^3

were performed; 104 of them yielded a complete or partial female and male mixed profile, 142 analyses yielded only the female haplotype, 3 yielded only the male haplotype, and finally, 63 analyses were inconclusive. Therefore, the female haplotype was detected more frequently than the mixture of haplotypes.

3.1. Comparing the first and second fractions

As expected, results were more consistent in the first fraction DNA extracts than in the second ones (Fig. 1 and Table 7). As shown in Table 6, only 3 analyses gave inconclusive results in first fractions whereas a total of 60 analyses were inconclusive (no amplification, blurred sequences or contaminations) in the second ones. In addition, some participants reported to have reanalysed some samples due to DNA contaminations detected in the second fractions in the first attempt.

When comparing the three semen dilutions of the first fractions (Fig. 1(a)), most of the labs reported a mixture of male and female haplotypes for those samples containing undiluted semen (41 out of 52). In contrast, most labs detected only the female haplotype in samples prepared with diluted semen 1:10 (36 out of 52 analyses) and in samples prepared with dilution 1:20 (38 out of 52). The mixtures of male and female haplotypes detected in both 1:10 and 1:20 diluted semen samples were mainly incomplete mixtures (complete female plus incomplete male). For instance, all the mixtures detected in blood/(semen 1:20) samples were incomplete, that is to say, some of the male haplotype polymorphisms were absent (see Table 5). Therefore, the male component became less evident in proportion to the degree of the semen dilutions.

One laboratory reported only the male haplotype in a sample prepared with undiluted semen (couple 1, see Table 5). This result could be attributed to sub-sampling, probably involving a section of the mixed stain containing (by chance) low number of epithelial cells from saliva and a high number of cells from semen.

Surprisingly, one lab reported a mixture of haplotypes in a sample with semen diluted 1:20 whereas with semen diluted 1:10 only the female haplotype was detected. This only occurred in 2 out of 156 first lysis analyses; a possible explanation for this unexpected result is that samples were mistakenly mixed-up at some step in the analysis process.

In addition, the loss of male nucleotide signals was generally not uniform throughout the electropherograms (see Fig. 2), probably indicating some dependence on the fluorochrome involved and/or the nucleotide sequence of the flanking region.

One lab carried out the analyses by using both preferential and total lyses (Table 8). All DNA extracts from the second fraction yielded blurred electropherograms. The mixture of female and male haplotypes was detected in a higher number of samples when the DNA extraction was performed by total lysis than when the lysis was preferential.

Results concerning autosomal STRs were as expected (Fig. 3). In the first lysis, all labs detected a mixed STR profile in those stains of blood or saliva mixed with undiluted semen. Many labs detected the male component in the second lysis. In cases where the semen was diluted 1:10 or 1:20, the preferential extraction was less useful, with the female autosomal STR profiles frequently obscuring that of the male in second fractions. Therefore, patterns of nDNA from semen were clearly different from those obtained for the mtDNA: DNA extracts yielded a mixture of profiles when analysed by autosomal STR typing, but showed only the female mitochondrial haplotype. Firstly, the capabilities of mixture detection may be different in the two analytical systems (STR fragment analysis *versus* sequencing). Secondly, detecting the minor component of a mixture also depends on the overall concentration of the mixture. And thirdly, differences in relative amounts of nuclear and mtDNAs in semen and in saliva or blood are also possible.

A number of participants did not obtain sufficient signal for one or more Y-STR loci in diluted semen mixtures. As expected, Y-STR alleles were obtained in both first (from non-spermatic male cells) and second (from spermatozoa) fractions. In first fraction, DNA extracts from samples prepared with diluted semen (in which only the female autosomal STR and mtDNA profiles were detected) a partial or complete Y-STR profile was identified. These results demonstrate once again the usefulness of performing Y-STR analyses in forensic

Table 5
First fraction results

Couple	Sample	mtDNA profile	Conclusion	No. of labs
1	Saliva/pure semen	16294T/C 16304C/T 16362C/T	Female + male	7
		16294T/C 16304C/T 16362C	Incomplete female + male	1
		16362C	Only male	1
	Saliva/semen 1:10	16294T/C 16304C/T 16362C/T	Female + Male	4
		16294T/C 16304C/T, 16304C/T 16362C/T	Female + incomplete male	4
		16294T 16304C	Only female	1
	Saliva/semen 1:20	16294T/C 16304C/T 16362C/T	Female + Male	3
		16294T 16304C 16362C/T, 16294T/C 16304C/T	Female + Incomplete male	2
		16294T 16304C	Only female	3
		Blurred	Inconclusive	1
2	Saliva/pure semen	16192T/C 16224C/T 16284G/A 16286G/C 16311C/T 16362C/T	Female + male	1
		16192T 16224C/T 16284G 16286G, 16311C 16362C, 16192T/C 16224C/T 16284G 16286G 16311C 16362C, 16192T 16224C/T 16284G/A 16286G/C 16311C/T 16362C/T	Female + incomplete male	3
		16192T 16224C 16284G 16286G 16311C 16362C	Only female	5
	Saliva/semen 1:10	16192T 16224C 16284G 16286G 16311C 16362C	Only female	9
	Saliva/semen 1:20	16192T 16224C 16284G 16286G 16311C 16362C	Only female	9
3	Saliva/pure semen	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male	7
		16069T/C 16092C/T 16257T/C 16293G/A 16311C	Female + incomplete male	1
	Saliva/semen 1:10	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male	2
		16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G 16311C, 16069T/C 16092C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C, 16092C/T 16257T/C 16293G 16311C	Female + incomplete male	3
		16092C 16257T 16293G 16311C	Only female	1
		Blurred	Inconclusive	1
	Saliva/semen 1:20	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311 C	Female + male	1
		16069T/C 16092C/T 16257T 16293G 16311C, 16069T/C 16092C/T 16126C/T 16193T/C, 16257T/C 16278T/C 16293G 16311C	Female + incomplete male	2
		16092C 16257T 16293G 16311C	Only female	5
1	Blood/pure semen	16294T/C 16304C/T 16362C/T	Female + male	5
		16294T 16304C 16362C/T, 16294 T/C 16304 T/C, 16294T 16304 C/T	Female + incomplete male	3
		16294T 16304C	Only female	1
	Blood/semen 1:10	16294T 16304C	Only female	9
	Blood/semen 1:20	16294T 16304C	Only female	9
2	Blood/pure semen	16192T/C 16224C/T 16284G/A 16286G/C 16311C/T	Female + male	3
		16192T/C 16224C/T 16284G 16286G 16311C 16362 C/T, 16192T/C 16224C/T 16284G/A 16311C/T 16362C/T, 16192T 16224C/T 16284G/A 16286G/C 16311C (edited up to 16350)	Female + incomplete male	3
		16192T 16224C 16284G 16286G 16311C 16362C	Only female	3
	Blood/semen 1:10	16192T 16224C 16284G 16286G 16311C 16362C	Only female	9
	Blood/semen 1:20	16192T 16224C 16284G 16286G 16311C 16362C	Only female	8
		Blurred	Inconclusive	1
3	Blood/pure semen	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male	6
		16069T/C 16126C/T 16193T/C 15257C/T 16278T/C 16293G/A 16311C	Incomplete female + male	1
		16092C 16257T 16293G 16311C	Only female	1
	Blood/semen 1:10	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male	2
		16069T/C 16092C/T 16126C/T 16193T/C, 16257T 16278T/C 16293G 16311C	Female + incomplete male	1
		16092C 16257T 16293G 16311C	Only female	5
	Blood/semen 1:20	16092C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C, 16069T/C 16092C/T 16193T/C 16257T/C 16278T/C 16293G 16311C, 16092C/T 16257T 16278T/C 16293G 16311C	Female + incomplete male	3
		16092C 16257T 16293G 16311C	Only female	5
Total				156

Table 6
Second fraction results

Couple	Sample	Haplotype	Conclusion	Lab
1	Saliva/pure semen	16294T/C 16304C/T 16362C/T	Female + male	4
		16294T 16304C	Only female	1
		16362C	Only male	1
		No amplification	Inconclusive	1
		Blurred	Inconclusive	1
		Contamination	Inconclusive	1
	Saliva/semen 1:10	16294T/C 16304C/T 16362C/T	Female + male	2
		16294T/C 16304C/T	Female + incomplete male	2
		16294T 16304C	Only female	3
		No amplification	Inconclusive	1
		Blurred	Inconclusive	1
	Saliva/semen 1:20	16294 T/C 16304 C/T 16362 T/C	Female + male	1
		16294T 16304C	Only female	5
		No amplification	Inconclusive	1
		Blurred	Inconclusive	1
		Contamination	Inconclusive	1
2	Saliva/pure semen	16192T 16224C/T 16284G 16286G 16311C/T 16362C/T, 16192T 16224C 16284G 16284G/C 16311C 16362C/T	Female + incomplete male	3
		16192T 16224C 16284G 16286G 16311C 16362C	Only female	2
		No amplification	Inconclusive	1
		Blurred	Inconclusive	2
		Contamination	Inconclusive	1
	Saliva/semen 1:10	16192T 16224C 16284G 16286G 16311C 16362C	Only female	5
		No amplification	Inconclusive	1
		Blurred	Inconclusive	2
		Contamination	Inconclusive	1
	Saliva/semen 1:20	16192T/C 16224C/T 16284G/A 16286G/C 16311C/T 16362C/T	Female + male	1
		16192T 16224C 16284G 16286G 16311C 16362C	Only female	5
		No amplification	Inconclusive	1
		Blurred	Inconclusive	1
		Contamination	Inconclusive	1
3	Saliva/pure semen	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male	1
		16069T/C 16092C/T 16126C/T 16193T/C 16257T 16293G 16311C, 16092C/T 16257T 16293G 16311C	Female + incomplete male	3
		16092C 16257T 16293G 16311C	Only female	1
		Blurred	Inconclusive	2
		Contamination	Inconclusive	1
	Saliva/semen 1:10	16069T/C 16092C/T 16193T/C 16257T 16293G 16311C, 16069T/C 16092C/T 16126C/T 16193T/C 16257T 16278T/C 16293G 16311C	Female + incomplete male	2
		16092C 16257T 16293G 16311C	Only female	4
		Blurred	Inconclusive	1
		Contamination	Inconclusive	1
	Saliva/semen 1:20	16069T/C 16092C/T 16257T 16293G 16311C, 16069T/C 16092C/T 16193T/C 16257T 16278T/C 16293G 16311C	Female + incomplete male	2
		16092C 16257T 16293G 16311C	Only female	4
		Blurred	Inconclusive	1
		Contamination	Inconclusive	1
1	Blood/pure semen	16294T/C 16304C/T 16362C/T	Female + male	2
		16294T/C 16304C/T	Female + incomplete male	1
		16362C/T	Incomplete female + male	1
		16294T 16304C	Only female	1
		No amplification	Inconclusive	1
		Blurred	Inconclusive	2
		Contamination	Inconclusive	1
	Blood/semen 1:10	16294T 16304C	Only female	7
		No amplification	Inconclusive	1
		Blurred	Inconclusive	1
	Blood/semen 1:20	16294 T 16304 C/T 16362C/T	Female + incomplete male	1
		16294T 16304C	Only female	6
		No amplification	Inconclusive	1
		Blurred	Inconclusive	1

Table 6 (*Continued*)

Couple	Sample	Haplotype	Conclusion	Lab
2	Blood/pure semen	16192T/C 16284G/A 16286G/C 16362C/T	Female + male	1
		16192T 16224C/T 16284G/A 16286G 16311C/T 16362C	Female + incomplete male	1
		16192T 16224C 16284G 16286G 16311C	Only female	1
		No amplification	Inconclusive	1
		Blurred	Inconclusive	3
		Contamination	Inconclusive	2
	Blood/semen 1:10	16192T/C 16224C 16284G 16286G 16311C 16362C	Female + incomplete male	1
		16192T 16224C 16284G 16286G 16311C 16362C	Only female	3
		No amplification	Inconclusive	2
		Blurred	Inconclusive	1
		Contamination	Inconclusive	2
	Blood/semen 1:20	16192T/C 16224C/T 16284G 16286G 16311C 16362C, 16192T 16224C/T 16284G/A 16286G/C 16311C/T 16362C/T	Female + incomplete male	2
		16192T 16224C 16284G 16286G 16311C 16362C	Only female	3
		No amplification	Inconclusive	2
		Blurred	Inconclusive	1
		Contamination	Inconclusive	1
3	Blood/pure semen	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male	3
		16069T 16126C 16193T 16278T 16311C	Only male	1
		No amplification	Inconclusive	1
		Blurred	Inconclusive	2
		Contamination	Inconclusive	1
	Blood/semen 1:10	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male	1
		16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G 16311C	Female + incomplete male	1
		16092C 16257T 16293G 16311C	Only female	3
		No amplification	Inconclusive	1
		Blurred	Inconclusive	1
		Contamination	Inconclusive	1
	Blood/semen 1:20	16092T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + incomplete male	1
		16092C 16257T 16293G 16311C	Only female	2
		No amplification	Inconclusive	1
		Blurred	Inconclusive	3
		Contamination	Inconclusive	1
Total				156

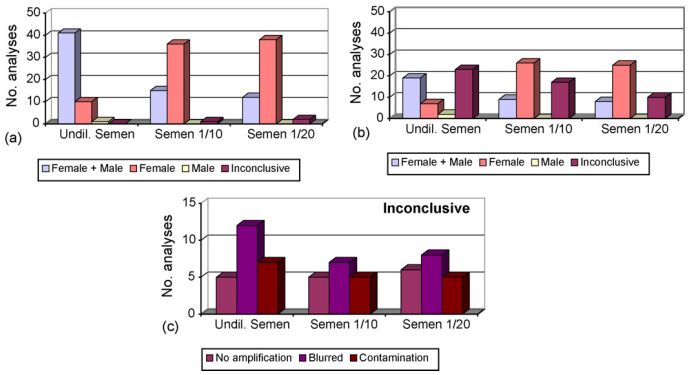


Fig. 1. Haplotype results in the first (a) and second fractions (b and c).

Table 7
Comparison between results obtained in first and second DNA extracts

Result	First fraction	Second fraction
Female + male haplotypes	68	36
Female haplotype	84	58
Male haplotype	1	2
No amplification	0	16
Blurred sequence	3	27
Contamination	0	17

male/female mixtures where the female cell contribution is predominant.

3.2. Comparison of tissues

Concerning the first fraction, no difference between saliva/undiluted semen and blood/undiluted semen mixtures was observed; most of the labs reported the female/male mixture. However, when comparing the samples prepared with 1:10 diluted semen, the female haplotype was mostly detected in blood/semen stains whereas a mixture of haplotypes was detected in half of the saliva/semen stain analyses (Fig. 4(1)). This difference between saliva and blood was still evident in the 1:20 semen samples but not so apparent as in the 1:10 dilutions.

Therefore, it seems that the number of mtDNA copy number per cell may be higher in blood than in saliva.

3.3. Comparison of donors

Results obtained in couples 1 and 3 were similar but very different from those obtained for couple 2 (Fig. 5). In first fractions, most labs detected a mixture of male and female haplotypes from couples 1 and 3 when the samples were prepared with undiluted semen. In contrast, only half of the analyses performed in samples from couple 2 yielded a mixture (Fig. 5(1a)). The differences between both couples 1 and 3 and couple 2 are more pronounced in samples with semen diluted 1:10 since all labs detected only the female haplotype in couple 2, while approximately half of the analyses yielded a mixture of haplotypes in couples 1 and 3 (Fig. 5(1b)). Finally, in samples containing semen diluted 1:20, only the female haplotype was detected in most of the analyses in the three couples, thus minimizing the differences between donors (Fig. 5(1c)). Therefore, our results suggest that different donors contribute different amounts of mtDNA.

Consistent with other results, the high amount of inconclusive data for the second fractions did not allow any clear conclusion to be reached (Fig. 5(2)).

3.4. Semen from a vasectomized individual

The blood and saliva from the female donor of couple 3 were also mixed with semen from a vasectomized individual (couple 4). As indicated in Table 9, in first and second fractions, a mixture of male and female haplotypes was detected in all samples prepared with undiluted semen. In samples with diluted semen only the female haplotype was detected.

In the first fractions, when comparing the results of couple 4 with the ones obtained in couple 3 (both couples with the same female donor), no differences were observed (see Fig. 5(1)). These results point to the fact that there may be no differences between mixtures coming from unvasectomized and vasectomized donors, at least, as far as mtDNA is concerned. We are aware, however, that more mixtures prepared using samples from other vasectomized individuals would be needed in order to derive definitive conclusions.

3.5. Preferential lysis from unmixed fluids

In order to better understand the patterns of preferential lyses in unmixed fluids, one lab carried out differential DNA extractions of blood, saliva and semen from each donor,

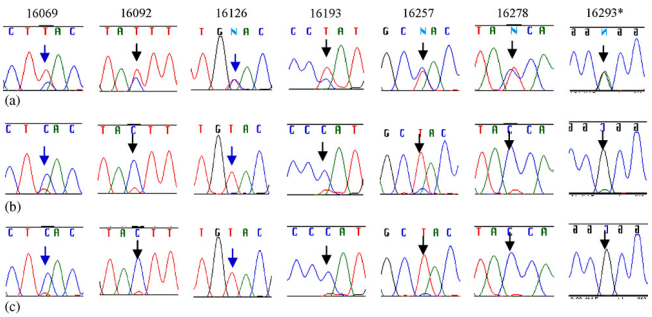


Fig. 2. Electropherograms of blood/semen samples (first fraction) from couple 3: (a) blood/undiluted semen; (b) blood/semen 1:10; (c) blood/semen 1:20. In samples prepared with semen diluted 1:10, the nucleotide position 16,126 does not show any mixture, whereas in the other positions, the male/female mixture is detected. Asterisk (*) indicates sequence with reverse primer.

Table 8
Haplotype results from the lab that performed the analyses with and without preferential lysis

Couple	Sample	Haplotypes	Conclusion
Total lysis			
2	Saliva + undiluted semen	16192T/C 16224C/T 16284G 16284G/C 16311C/T 16362C/T	Female + incomplete male
	Saliva + semen 1/10	16192T/C 16224C 16284G 16286G 16311C 16362C	Female + one male polymorphism
	Saliva + semen 1/20	16192T 16224C 16284G 16286G 16311C 16362C	Only female
	Blood + undiluted semen	16192T/C 16224C/T, 16284G 16284G/C 16311C/T 16362C/T	Female + incomplete male
	Blood + semen 1/10	16192T 16224C 16284G 16286G 16311C 16362C	Only female
	Blood + semen 1/20	16192T 16224C 16284G 16286G 16311C 16362C	Only female
3	Saliva + undiluted semen	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male
	Saliva + semen 1/10	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male
	Saliva + semen 1/20	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male
	Blood + undiluted semen	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male
	Blood + semen 1/10	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male
	Blood + semen 1/20	16069T/C 16092C 16257T 16293G 16311C	Female + one male polymorphism
First fraction			
2	Saliva + undiluted semen	16192T 16224C 16284G 16286G 16311C 16362C	Only female
	Saliva + semen 1/10	16192T 16224C 16284G 16286G 16311C 16362C	Only female
	Saliva + semen 1/20	16192T 16224C 16284G 16286G 16311C 16362C	Only female
	Blood + undiluted semen	16192T 16224C 16284G 16286G 16311C 16362C	Only female
	Blood + semen 1/10	16192T 16224C 16284G 16286G 16311C 16362C	Only female
	Blood + semen 1/20	16192T 16224C 16284G 16286G 16311C 16362C	Only female
3	Saliva + undiluted semen	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male
	Saliva + semen 1/10	Blurred	Inconclusive
	Saliva + semen 1/20	16092C 16257T 16293G 16311C	Only female
	Blood + undiluted semen	16069T/C 16092C/T 16126C/T 16193T/C 16257T/C 16278T/C 16293G/A 16311C	Female + male
	Blood + semen 1/10	16092C 16257T 16293G 16311C	Only female
	Blood + semen 1/20	16092C 16257T 16293G 16311C	Only female

Results of second lyses have not been included because all samples yielded blurred sequences.

quantified the nuclear DNA, and then amplified the HVS-I mtDNA segment, the results of which were checked by agarose gel electrophoresis stained with ethidium bromide.

Results from nuclear DNA quantification are shown in Table 10. Results of HVS-I amplification from blood and saliva are shown in Fig. 6 and those from semen in Fig. 7. nDNA was detected in both first and second fractions (Table 9) and HVS-I amplicons were obtained in both factions (Fig. 6). Neither blood nor saliva are fluids containing cells resistant to the enzymatic action of the DNA extraction chemicals. These results demonstrate that, as expected, preferential lyses performed in saliva and blood were not efficient since second fractions yielded DNA (not all cells were lysed or DNA recovered by the first fraction).

With respect to the analysis of semen samples, it would be expected to detect mtDNA only in the first fraction and nDNA in both fractions (from non-spermatic cells in the first one and from spermatozoa in the second one). The results confirmed this hypothesis. We detected nuclear DNA in both fractions and, as expected, the nDNA concentrations were higher in the second fractions than in the first ones (see Table 10). We carried out HVS-I mtDNA PCRs by using 0.4 ng of DNA as template. MtDNA amplicons were only obtained in first fractions and not in the second ones (see Fig. 7(a) and (b)). These amplicons were sequenced in order to verify that the resulting haplotypes matched the haplotypes of the donors.

In addition, we also performed HVS-I amplifications from second fractions by using increasing amounts (4 and 20 ng) of nDNA template. We only detected a slightly higher signal when the DNA template amount was increased 50 times (see Fig. 7(c) and (d)). The sequencing analysis (in order to determine whether the resulting haplotypes were consistent with the ones from the donors) was not possible due to the low yield of the amplicons. It could be inferred that these amplicons come from nuclear mitochondrial insertions (NUMTs); however, this hypothesis is unlikely to happen (see e.g. Ref. [10]).

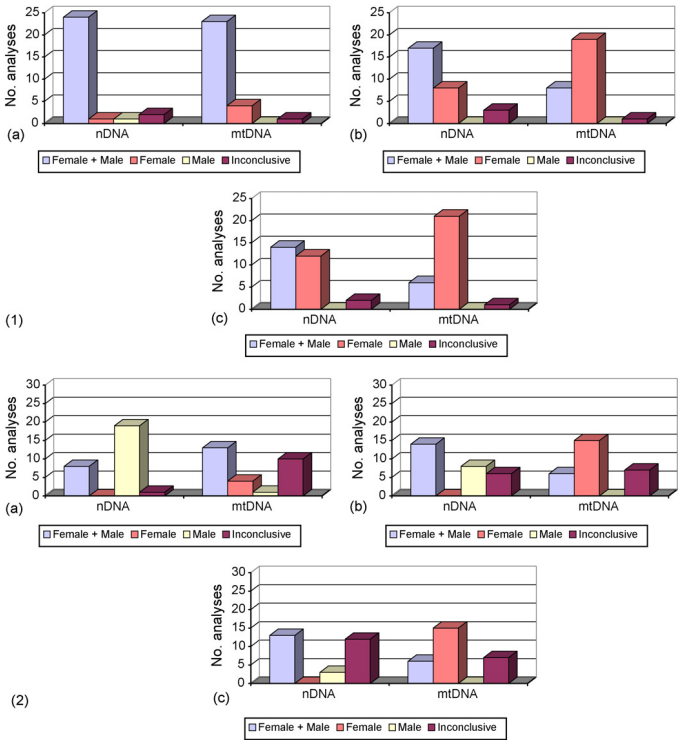


Fig. 3. Comparison between nuclear and mitochondrial results: (1) first fraction; (2) second fraction. (a) Samples prepared with pure semen; (b) samples prepared with semen diluted 1/10; (c) samples prepared with semen diluted 1/20.

4. Discussion

The analysis of forensic mixtures from rape cases using mtDNA is not routine practice in forensic casework but can be necessary when nuclear DNA is absent due to the poor or sub-optimum quality of samples, or when a maternal relationship between the evidence and reference samples is under investigation. In rape cases, labs usually perform preferential lysis in order to try to separate the victim and suspect STR profile. Nevertheless, differential lysis is not a useful tool for separating mtDNA from mixtures since theoretically all mitochondria should remain in first fractions. However, this type of DNA extraction is carried out very often in sexual assault cases. Several labs of the GEP-ISFG have studied blood/semen and saliva/semen mixtures by using the routine methodologies for analysing mixtures (preferential DNA extraction and autosomal and Y-STRs genotyping) and mtDNA (HVS-I automated sequencing).

First, the analyses of autosomal STRs allowed us to verify that, as expected, preferential lyses are less efficient than desirable (specially when semen is diluted) [11], since some labs detected the female nuclear DNA in second fractions. Regarding mtDNA, we found that, although it is impossible to separate the female component from the male one, better and more reliable results were obtained in first fractions than in the second lyses (3 inconclusive results in first fractions and 60 in the second ones). We observed three different types of inconclusive results: no amplification, blurred sequences and/or contamination. The three are expected findings if the lysis is effective. In this case, contamination could be more evident than in samples with high mtDNA content. Consequently, we recommend focusing more on the results obtained in first fractions than on the ones obtained in second fractions when interpreting mixtures from real forensic casework. Note that a good quality electropherogram does not rule out the presence of contamination; therefore, the prevention of

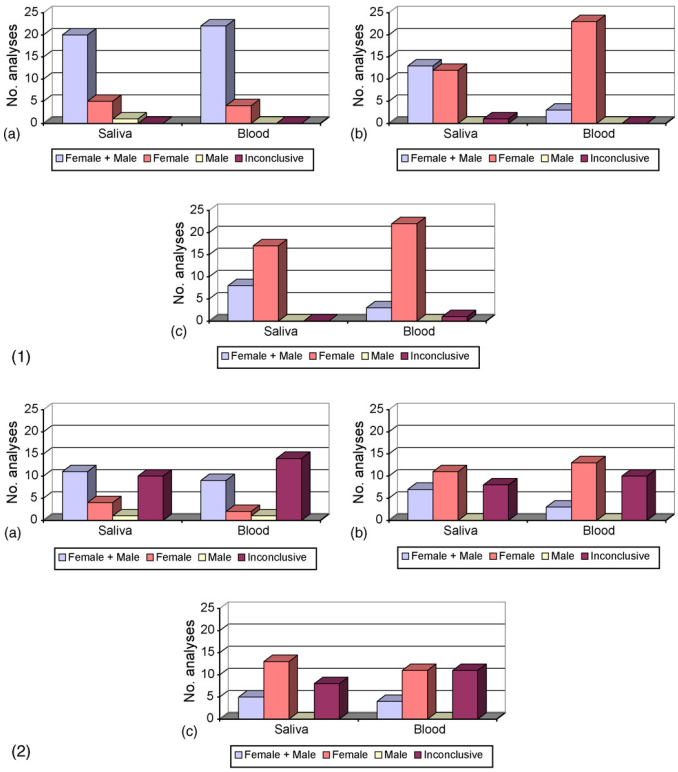


Fig. 4. Haplotype results in the first (1) and second (2) fractions. Comparison between saliva/semen and blood/semen samples. (a) Samples prepared with undiluted semen; (b) samples prepared with semen diluted 1:10; (c) samples prepared with semen diluted 1:20.

contamination should be carefully monitored throughout all the different steps of the process.

Several techniques can be used prior to sequencing in order to separate different mtDNA amplicons, such as cloning, electrophoretic separation of single strand conformational polymorphisms (SSCPs) [12], heteroduplex formation followed by electrophoresis or denaturing high-performance liquid chromatography [13]. Cloning is a technique that is not usually available in forensic labs, and requires expanded sequencing, mainly when individuals are not equally represented in the mixture [14]. SSCP or heteroduplex analyses do not allow you to distinguish all mtDNA variants. However, the separation of female and male nuclear DNAs will improve when using e.g. laser-capture microdissection (LCM) technology [15]. This new tool would enable us to obtain at least the male mtDNA component, and the combination of labelling and LCM the recovery of female mtDNA [16]. However, it is also worth mentioning that this technique may be prone to incidences of undesirable contamination [17].

When comparing the results obtained in first lyses from the three dilutions of semen (pure, 1:10 and 1:20), most labs detected the mixture of haplotypes in samples prepared with undiluted semen but not in those prepared with diluted semen (where only the female haplotype was mainly detected). The results also suggest that 1:10 dilution can lead to the loss of the male signal in the sequence electropherograms. mtDNA analysis of semen is complex because this fluid contains several cell-types (spermatozoa and round cells) at different (unpredictable) concentrations. Several approaches have been undertaken to determine the mtDNA content in spermatozoa [18–21] but so far, there is still no complete agreement on these results, which is probably due to the different methodologies applied. Our results indicate that the amount of mtDNA in semen relative to the amount of mtDNA in saliva or blood could

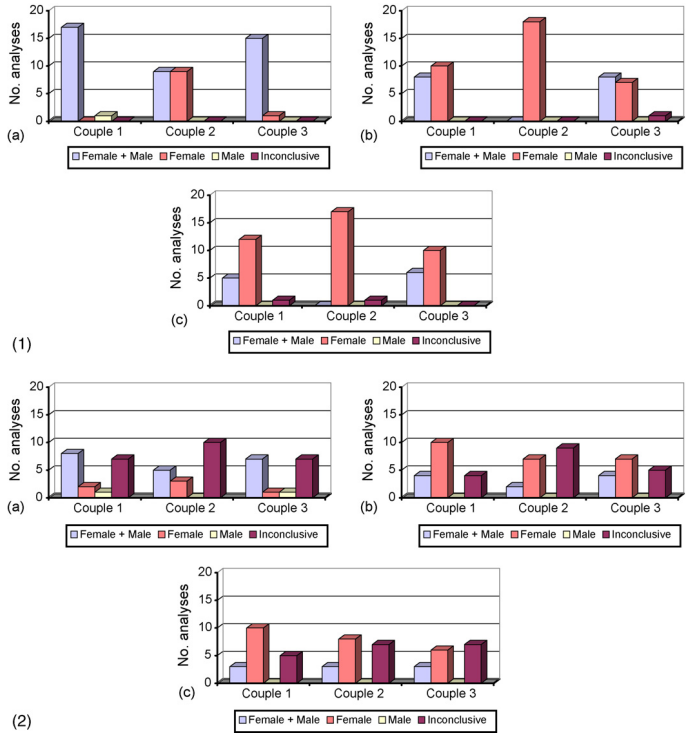


Fig. 5. Haplotype results in the first (1) and second (2) fractions. Comparison between couples. (a) Samples prepared with undiluted semen; (b) samples prepared with semen diluted 1:10; (c) samples prepared with semen diluted 1:20.

Table 9
Haplotype results from couple 4

Sample	Haplotypes	Conclusion
First fraction		
Saliva + undiluted semen	16069T/C 16092C/T 16126C/T 16257T/C 16293G/A 16300A/G 16311C/T	Female + male
Saliva + semen 1/10	16092C 16257T 16293G 16311C	Only female
Saliva + semen 1/20	16092C 16257T 16293G 16311C	Only female
Blood + undiluted semen	16069T/C 16092C/T 16126C/T 16257T/C 16293G/A 16300A/G 16311C/T	Female + male
Blood + semen 1/10	16092C 16257T 16293G 16311C	Only female
Blood + semen 1/20	16092C 16257T 16293G 16311C	Only female
Second fraction		
Saliva + undiluted semen	16069T/C 16092C/T 16126C/T 16257T/C 16293G/A 16300A/G 16311C/T	Female + male
Saliva + semen 1/10	16092C 16257T 16293G 16311C	Only female
Saliva + semen 1/20	16092C 16257T 16293G 16311C	Only female
Blood + undiluted semen	16069T/C 16092C/T 16126C/T 16257T/C 16293G/A 16300A/G 16311C/T	Female + male
Blood + semen 1/10	16092C 16257T 16293G 16311C	Only female
Blood + semen 1/20	16092C 16257T 16293G 16311C	Only female

Table 10
Quantification of nuclear DNA after preferential lysis from unmixed blood, saliva and semen from each donor, in DNA extracts from 0.7 cm × 0.7 cm of stain

Sample	First lysis (ng/μl)	Second lysis (ng/μl)
Blood, female couple 1	1.8	1
Blood, female couple 2	1.4	1.3
Blood, female couple 3	2	2.6
Saliva, female couple 1	0.15	Undetectable
Saliva, female couple 2	0.03	0.03
Saliva, female couple 3	1	0.02
Semen, male couple 1	5.4	81.5
Semen, male couple 2	1.6	55.6
Semen, male couple 3	8.2	106.3

be below the threshold of detection when unbalanced forensic mixtures, that are analysed by standard sequencing procedures, are being studied.

In addition, the lab that coordinated the present study analysed the same types of mixtures using a vasectomized man as the semen donor (couple 4: the same female donor as in couple 3 and a vasectomized male). Results obtained from couple 4 (see Table 9) are not significantly different from those obtained in couple 3 (see Table 5 and Fig. 5(1)). In both couples, the mixture of male and female haplotypes was detected in samples prepared with undiluted semen. As described in Ref. [22], during the last phases of spermiogenesis, the number of mtDNA copies declines in such a way that each mature sperm mitochondrion could contain, on average, only one copy of mtDNA. The main contribution of mtDNA molecules in semen may come from non-spermatic cells which would explain our results. This fact would be consistent with the findings reported elsewhere [23] where, after measurements of mtDNA/β-globin gene ratio by quantitative PCR (qPCR), the authors concluded that the majority of sperm mitochondria are almost totally devoid of mtDNA, and that many spermatozoa probably do not contain any mtDNA at all.

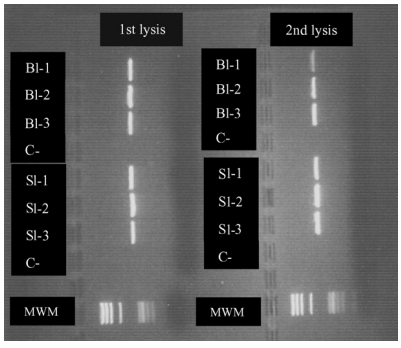


Fig. 6. HVS-I amplification results of blood (Bl) and saliva (Sl) from each female donor (1–3) after preferential lysis. "C−" indicates the negative controls of each DNA extraction. "MWM" indicates molecular weight marker.

Forensic post-coital samples are usually in worse conditions of preservation than fresh semen samples, they are generally diluted (e.g. samples collected through vaginal washings) and mixed with a high quantity of female cells. In addition, these samples are also of sub-optimum quality because most of the time their collection is not carried out immediately after the sexual assault. It is then not surprising to detect only the female mtDNA from forensic mixtures (as occurs in diluted semen samples analysed in the present study). Therefore, as inferred from our results, when mtDNA analysis is undertaken in a rape case, the observation of exclusively the victim's profile is not a full evidence for the absence of a second potential contributor of DNA.

One lab carried out the mixture analysis (couples 2 and 3) using both total and preferential DNA extraction. In this case, the lab carried out both DNA extractions by Chelex-100 [24], a method that usually yields less DNA than the organic one. More haplotype mixtures were observed when the DNA extractions were performed by total lysis rather than by preferential lysis (see Table 8). Therefore, it is possible that the preferential lysis renders less DNA than the none-preferential one. Since it is not possible to separate male and female mtDNA by preferential lysis, in forensic casework it may be more appropriate to perform both types of lyses: the preferential lysis to analyze nDNA and total lysis to analyze mtDNA.

In addition to the mtDNA content in semen, we have seen that the type of fluid in the mixture is also important. We detected both female and male haplotypes in more tests when the semen was mixed with saliva than when it was mixed with blood (Fig. 4). Therefore, it seems that the number of mtDNA copy number (per volume) in blood could be higher than in saliva. This is an important parameter to take into account when evaluating results in actual forensic casework with e.g. samples containing blood from the victim (in addition to vaginal fluid). The male mtDNA may be masked by the high amount of female mtDNA, resulting in false exclusions. The knowledge of the specific type of fluids involved and hence their cellular content and the number of mtDNA molecules per cell is of great interest. The implementation of preliminary tests with the aim of identifying each fluid is an essential tool when a mixture of different fluids is supposed.

We also found that the loss of male mtDNA signal is not uniform in all nucleotide positions (see Fig. 2). The analysis of the mtDNA in mixtures can benefit from a phylogenetic interpretation of the electropherogram profiles [25]. Thus, the phylogenetic perspective could be useful to detect particular diagnostic variants (which may remain undetected) in the mtDNA profile of the contributors, or even to infer the haplotype of the donors (this was in fact the result obtained by one lab in Ref. [7] and have resulted to be useful in real cases; author's unpublished data). Therefore, the presence of e.g. two bases (apparent heteroplasmy) in a stable diagnostic nucleotide position (or in several positions) could support the hypothesis of the presence of two (or more) different haplotypes in the same sample (indicating for instance the minimum amount of contributors to the stain), and even in some cases it could be possible to assign the haplogroup of each contributor (although

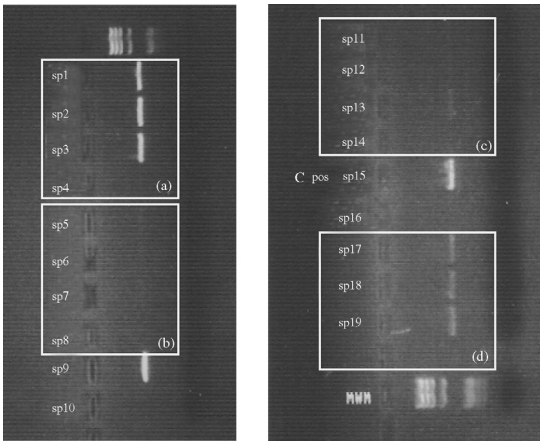


Fig. 7. HVS-I amplification results of semen from each male donor (1–3) after preferential lysis. (a) First fraction amplifications results using 0.4 ng of nuclear DNA (sp1, sp2 and sp3 correspond to the male donors from couples 1 to 3 respectively; sp4 is the DNA extraction negative control). (b) Second fraction amplifications results using 0.4 ng of nuclear DNA (sp5, sp6 and sp7 correspond to the male donors from couples 1 to 3, respectively; sp8 is the DNA extraction negative control; sp9 is an amplification positive control with 0.2 ng; sp10 is a negative control of the amplification). (c) Second fraction amplifications results using 4 ng (10×) of nuclear DNA (sp11–13 correspond to the male donors from couples 1 to 3, respectively; sp14 is a DNA extraction negative control; sp15 is a positive control of the amplification with 0.2 ng; sp16 is a negative control of the amplification). (d) Second fraction amplifications results using 20 ng (50×) of nuclear DNA (sp17–19 correspond to the male donors from couples 1 to 3, respectively).

it could be difficult e.g. to assign private polymorphisms to the particular haplotypes of each contributor) [17].

The technical quality of electropherograms is also an important issue when analyzing mixtures. We strongly recommend using the same guidelines as those for haplotypes showing point heteroplasmies [26]; note that heteroplasmy is in fact a 'natural' mixture. In addition to obtaining electropherograms without background noise in the two strands, it would also be desirable to carry out the sequencing by using additional primers in order to confirm the mixed nucleotide positions. The same approach can be used when one of the haplotypes in the mixture shows a length heteroplasmy and, in fact, there are several primers already designed which allow us to read the electropherograms behind the length heteroplasmy zone. However, the presence of an insertion or deletion polymorphism in one component of the mixture can obscure the sequencing analyses greatly. In these cases, it would be necessary to design new primers in order to prevent overlapping peaks beyond the insertion or deletion.

It is also important to quantify the male and female nDNA in order to facilitate the interpretation of the mixed profile [9]. The present study has allowed us to prove that female fluids mixed with semen dilutions over 1:10 can be very difficult to interpret and that, although better results are obtained when fluid mixtures are balanced, there are variations depending on the type of fluid and the specific donors who contributed the mixture.

5. Conclusions

The analysis of mtDNA from semen mixed with other fluids is not an easy task when using standard sequencing procedures. On the one hand, the technique has some limitations (e.g. different ability of sequencing chemistries to detect point mixtures, low sensitivity in unequal mixtures), and the experience and expertise of individual analysts can be essential for the interpretation of the results. On the other hand, we have seen that these mixtures have special natural characteristics such as the different number of mtDNA copies in different fluids and contributors to the mixture. In addition, the presence of point heteroplasmy, insertions or deletions in one of the haplotypes can complicate the interpretation of the results.

We have demonstrated that the diverse content of mtDNA copies in different body fluids can cause masking of one of the mtDNAs to occur. Furthermore, we may also obtain a set of partial and non-conclusive results, which warns us about the need to be careful when interpreting these results. In order to minimize the number of problems associated with the analysis of mixtures by means of mtDNA standard sequencing, we strongly recommend the identification of the type of fluid in order to know the cell types and the specific content of mtDNA copies. In addition, there are several factors that we must bear in mind: the high risk of contamination, mainly in second fractions; the loss of base signal (maybe stochastically) in some nucleotide positions and not in others; the type of fluids

involved in the mixture; and the possibility of differences in mtDNA content among donors.

Finally, we strongly recommend the use of additional sequencing primers to corroborate the typing data. It is also important to view the results from a phylogenetic perspective [17,28–34] that could help to interpret the mixed sequence patterns and could also prevent undesirable mistakes.

Acknowledgements

We wish to thank the donors of the saliva, blood and semen samples used in this study. We are also very grateful to Fran Álvarez for his technical work, Leonor Gusmão for her encouragement to this project, Milton for his noticiable help, and Ángel Carracedo for his full support and helpful comments.

References

[1] T. Egeland, H.M. Bovelstad, G.O. Storvik, A. Salas, Inferring the most likely geographical origin of mtDNA sequence profiles, Ann. Hum. Genet. 68 (Pt 5) (2004) 461–471.
[2] H.I. Calvin, J.M. Bedford, Formation of disulfide bonds in the nucleus and accessory structures of mammalian spermatozoa during maturation in the epididymis, J. Reprod. Fertil. Suppl. (England) 13 (1971) 65–75.
[3] P. Gill, A.J. Jeffreys, D.J. Werret, Forensic applications of DNA "fingerprints", Nature 318 (1985) 577–579.
[4] M.M. Holland, T.J. Parsons, Mt-DNA sequence analysis—validation and use for forensic casework, Forensic Sci. Rev. 11 (1999) 21.
[5] L. Cavelier, A. Johannisson, U. Gyllensten, Analysis of mtDNA copy number and composition of single mitochondrial particles using flow cytometry and PCR, Exp. Cell Res. 259 (2000) 79–85.
[6] E.D. Robin, R. Wong, Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells, J. Cell Physiol. 136 (1988) 507–513.
[7] M. Crespillo, M.R. Paredes, L. Prieto, M. Montesino, C. Albarrán, V. Álvarez-Iglesias, A. Amorin, G. Berniell-Lee, A. Brehm, J.C. Carril, D. Corach, N. Cuevas, A.M. DiLonardo, C. Doutremepuich, R.M. Espinheira, M. Espinoza, F. Gómez, A. González, A. Hernández, M. Hidalgo, M. Jiménez, F.P. Leite, A.M. López, M. López-Soto, J.A. Lorente, S. Pagano, A.M. Palacio, J.J. Pestano, M.F. Pinheiro, E. Raimondi, M.M. Ramon, F. Tovar, L. Vidal-Rioja, M.C. Vide, M.R. Whittle, J.J. Yunis, J. García-Hirschfeld, Results of the 2003–2004 GEP-ISFG collaborative study on mitochondrial DNA: focus on the mtDNA profile of a mixed semen-saliva stain, Forensic Sci. Int. (2005) Oct 19 (Epub ahead of print).
[8] K.D. Anslinger, B.R. Bayer, B. Rolf, W. Eisenmenger, MtDNA investigations after differential lysis, J. Forensic Sci. 50 (2005) 579–581.
[9] A. Alonso, P. Martín, C. Albarrán, P. García, O. García, L. Fernández de Simón, J. García-Hirschfeld, M. Sancho, C. de la Rúa, J. Fernández-Piqueras, Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies, Forensic Sci. Int. 139 (2004) 141–149.
[10] M. Ricchetti, F. Tekaia, B. Dujon, Continued colonization of the human genome by mitochondrial DNA, PLoS Biol. 2 (2004) E273.
[11] P. Wiegand, M. Schuremkamp, U. Schutte, DNA extraction from mixtures of body fluid using mild preferential lysis, Int. J. Legal Med. 104 (1992) 359–360.
[12] R.M. Myers, V. Sheffield, D.R. Cox, Detecting changes in DNA: ribonuclease cleavage and denaturing gradient gel electrophoresis, in: K.E. Davies (Ed.), Genomic Analysis: a Practical Approach, IRL Press Ltd., Oxford, 1988, pp. 95–139.
[13] G.S. LaBerge, R.J. Shelton, P.B. Danielson, Forensic utility of mitochondrial DNA analysis based on denaturing high-performance liquid chromatography, Croat. Med. J. 44 (2003) 281–288.
[14] J.A. Walker, R.K. Garber, D.J. Hedges, G.E. Kilroy, J. Xing, M.A. Batzer, Resolution of mixed human DNA samples using mitochondrial DNA sequence variants, Anal. Biochem. 325 (2004) 171–173.
[15] K. Elliot, D.S. Hill, C. Lambert, T.R. Burroughes, P. Gill, Use of laser microdissection greatly improves the recovery of DNA from sperm on microscope slides, Forensic Sci. Int. 137 (2003) 28–36.
[16] K. Anslinger, B. Mack, B. Bayer, B. Rolf, W. Eisenmenger, Digoxigenin labelling and laser capture microdissection of male cells, Int. J. Legal Med. 119 (2005) 374–377.
[17] A. Salas, Y.G. Yao, V. Macaulay, A. Vega, Á. Carracedo, H.J. Bandelt, A critical reassessment of the role of mitochondria in tumorigenesis, PloS Med. 2 (11) (2005) E296.
[18] F. Ankel-Simons, J.M. Cummins, Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 13859–13863.
[19] G. Manfredi, D. Thyagarajan, L.C. Papadopoulou, F. Pallotti, E.A. Schon, The fate of human sperm-derived mtDNA in somatic cells, Am. J. Hum. Genet. 61 (1997) 953–960.
[20] L. Bromham, A. Eye-Walker, N.H. Smith, J. Maynard Smith, Mitochondrial steve: paternal inheritance of mitochondria in humans, Trends Ecol. Evol. 18 (2003) 1–3.
[21] C. Díez-Sánchez, E. Ruiz-Pesini, A.C. Lapena, J. Montoya, A. Pérez-Martos, J.A. Enríquez, M.J. López-Pérez, Mitochondrial DNA content of human spermatozoa, Biol. Reprod. 68 (2003) 180–185.
[22] J. Cummins, Mitochondrial DNA in mammalian reproduction, Rev. Reprod. 3 (1998) 172–182.
[23] P. May-Panloup, M.F. Chretien, F. Savagner, C. Vasseur, M. Jean, Y. Malthiery, P. Reynier, Increased sperm mtDNA content in male infertility, Hum. Reprod. 18 (2003) 550–556.
[24] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, Biotechniques 10 (1991) 506–513.
[25] A. Salas, Á. Carracedo, V. Macaulay, M. Richards, H.J. Bandelt, A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics, Biochem. Biophys. Res. Commun. 335 (2005) 891–899.
[26] G. Tully, W. Bär, B. Brinkmann, Á. Carracedo, P. Gill, N. Morling, W. Parson, P. Schenieder, Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles, Forensic Sci. Int. 124 (2001) 83–91.
[27] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 32 (1999) 147.
[28] H.-J. Bandelt, Q.P. Kong, W. Parson, A. Salas, More evidence for non-maternal inheritance of mitochondrial DNA? J. Med. Genet. 42 (2005) 957–960.
[29] H.-J. Bandelt, L. Quintana-Murci, A. Salas, V. Macaulay, The fingerprint of phantom mutations in mitochondrial DNA data, Am. J. Hum. Genet. 71 (2002) 1150–1160.
[30] H.-J. Bandelt, A. Salas, C.M. Bravi, Problems in FBI mtDNA database, Science 305 (2004) 1402–1404.
[31] H.-J. Bandelt, A. Salas, S. Lutz-Bonengel, Artificial recombination in forensic mtDNA population databases, Int. J. Legal Med. 118 (2004) 267–273.
[32] A. Salas, H.-J. Bandelt, V. Macaulay, M. Richards, Phylogeographic investigations: the role of trees in forensic genetics, Forensic Sci. Int., in press.
[33] A. Salas, L. Prieto, M. Montesino, C. Albarrán, E. Arroyo, M.R. Paredes-Herrera, A.M. Di Lonardo, et al., Mitochondrial DNA error prophylaxis: assessing the causes of errors in the GEP'02-03 proficiency testing trial, Forensic Sci. Int. 148 (2005) 191–198.
[34] Y.-G. Yao, A. Salas, C.M. Bravi, H.-J. Bandelt, A reappraisal of complete mtDNA variation in East Asian families with hearing impairment, Hum. Genet. 119 (5) (2006) 505–515.

Artículo 5: Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups

Álvarez-Iglesias V, Jaime JC, Carracedo Á, Salas A.

(2007) *FSI:Genetics* 1-1: 44-55.

SNPs de la región codificante de ADN mitocondrial: haplogrupos del Este de Asia y Nativo Americanos

ABSTRACT/RESUMEN:

Hemos desarrollado una reacción SNaPshot PCR multiplex que contiene 32 SNPs de la región codificante (i) que incrementa el poder de discriminación del tipado de ADN mitocondrial (ADNmt) en la rutina forense, y (ii) que mejora la clasificación de perfiles en haplogrupos de ADNmt tanto en estudios poblacionales como en la investigación clínica. Los SNPs elegidos pertenecen a la filogenia del Este de Asia, incluyendo las ramas derivadas Nativo Americanas. Este multiplex se ha validado mediante genotipado de una muestra del Este de Asia (Taiwan) y de Nativos Americanos (Argentina). Además del tipado de los SNPs de la región codificante, hemos secuenciado la región control completa de todas las muestras. El resultado de este genotipado (región control y SNPs codificantes) no presenta discrepancias con los estudios genético-poblacionales previos (como por ejemplo los basados en secuencias de genomas completos) ni con la filogenia mitocondrial conocida. Observamos que este método es fiable, rápido y rentable en comparación con otras técnicas de genotipado de SNPs en multiplex. Discutimos las ventajas de nuestra selección de SNPs respecto a intentos anteriores, y destacamos la importancia de utilizar la filogenia de ADNmt conocida como marco para la interpretación de perfiles y como herramienta para minimizar los errores de genotipado.

Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics 1 (2007) 44–55
FSI GENETICS
www.elsevier.com/locate/fsig

Coding region mitochondrial DNA SNPs: Targeting East Asian and Native American haplogroups

V. Álvarez-Iglesias [a], J.C. Jaime [a], Á. Carracedo [a,b], A. Salas [a,b,*]

[a] *Unidade de Xenética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, Santiago de Compostela 15782, Galicia, Spain*
[b] *Grupo de Medicina Xenómica, Hospital Clínico Universitario, 15706 Santiago de Compostela, Galicia, Spain*

Received 10 July 2006; received in revised form 15 September 2006; accepted 17 September 2006

Abstract

We have developed a single PCR multiplex SNaPshot reaction that consists of 32 coding region SNPs that allows (i) increasing the discrimination power of the mitochondrial DNA (mtDNA) typing in forensic casework, and (ii) haplogroup assignments of mtDNA profiles in both human population studies (e.g. anthropological) and medical research. The selected SNPs target the East Asian phylogeny, including its Native American derived branches. We have validated this multiplex assay by genotyping a sample of East Asians (Taiwanese) and Native Americans (Argentineans). In addition to the coding SNP typing, we have sequenced the complete control region for the same samples. The genotyping results (control region plus SNaPshot profiles) are in good agreement with previous human population genetic studies (based on e.g. complete sequencing) and the known mtDNA phylogeny. We observe that the SNaPshot method is reliable, rapid, and cost effective in comparison with other techniques of multiplex SNP genotyping. We discuss the advantages of our SNP genotyping selection with respect to previous attempts, and we highlight the importance of using the known mtDNA phylogeny as a framework for SNP profile interpretation and as a tool to minimize genotyping errors.

Keywords: mtDNA; Coding region; HVS-I and HVS-II; SNP; Haplotype; SNaPshot; Phylogeny; Forensics; Population studies; East Asia; Native America

1. Introduction

The control region of the mtDNA genome, in particular, the first and second hypervariable regions (HVS-I/II), has a limited power of discrimination in a forensic, anthropological, and clinical context (e.g. [1]). This depends mainly on the population under study because diversity values and the phylogeny differ between populations (with East Africa showing the highest levels of diversity; e.g. [2,3]). Many haplogroups are poorly defined in both hypervariable segments and, moreover, sharing the same HVS-I profile does not necessarily imply that two mtDNAs belong to the same haplogroup. On the other hand, sequencing complete genomes is tedious, costly, and time-consuming, therefore, unfeasible in routine forensic casework. Analysis of single nucleotide polymorphisms (SNPs) contained in the coding regions of the mitochondrial genome has proved to be a good strategy to address this drawback. In the last few years, interest in SNPs has increased and, simultaneously, numerous strategies have been developed to facilitate mtDNA SNP typing (e.g. [4–9]), which are much more efficient than earlier attempts (e.g. [10,11]). Here we chose SNaPshot minisequencing to type mtDNA coding region SNPs mainly corresponding to the East Asian and Native American part of the mtDNA skeleton. There are some successful precedents in the use of SNaPshot for mtDNA SNP typing (and for autosomal SNPs as well; see e.g. [12]); nevertheless, most of the previous studies were focused on the West European mtDNA phylogeny [4,6,8,9,13]. Recently, Umetsu et al. [5] has developed four multiplex SNaPshot assays for "... haplogrouping of East Asian populations" using "Amplified Product-Length Polymorphism" (APLP).

We here show a single SNaPshot multiplex design that allows the discrimination power of the mtDNA typing to be increased in forensic casework when analyzing East Asian [14] and Native American populations, even when the whole control

* Corresponding author at: Unidade de Xenética Forense, Universidade de Santiago de Compostela, Galicia, Spain. Tel.: +34 981 582327; fax: +34 981 580336.
E-mail address: apimlase@usc.es (A. Salas).

1872-4973/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2006.09.001

region has been already genotyped on the same samples. This multiplex could alternatively be use as a screening tool in forensic casework. An important application in the field of population genetics is that the SNPs selected allow allocating mtDNAs to particular haplogroups not clearly defined with the control region alone. Finally, this method is also amenable for medical case–control association studies (e.g. [15]) or mtDNA research in tumorigenesis (e.g. [16–18]).

2. Material and methods

2.1. Samples

We have collected samples from general population from Taiwan (Chinese Taiwanese; $N = 61$) and Native Americans from Argentina, in particular, Coyas from the Jujuy and Salta Provinces ($N = 60$). Informed consent has been obtained for all the samples used in the present study and the study has been approved by the Ethical committee of the University of Santiago de Compostela.

2.2. SNP selection

In order to optimize the amount of information covered by a specific number of SNPs, we need to consider the whole body of the known mtDNA phylogeny. Thus, we have used the information contained in more than 150 complete mtDNA genomes from Asia and America, basically the data reported in Refs. [19–21]. We have also used control region information from more than 10,000 sequencing profiles collected from the literature. This has allowed us to select those SNPs that, in combination with control region data (mainly the standard HVS-I and HVS-II regions), cover the main haplogroup branches of the Asian and American phylogeny. This SNP selection preceded the publication of the data (more than 672 complete genomes) and the phylogeny recently reported by Ref. [22], but critically re-constructed in Figs. 1 and 2 of Ref. [14] (see below); the consequences of that are discussed below.

2.3. SNaPshot multiplex primer design

The primers both for PCR amplification and the minisequencing reaction were designed to have an annealing temperature around 60 °C using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3). The sequence databases at the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) were interrogated using the online BLAST tool to test the primers against possible repetitive sequences and sequence homologies.

AutoDimer (http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm) was used to test for potential hairpin structures and primer–dimer problems.

2.4. PCR multiplex amplification previous to SNaPshot reaction

The amplicons are deliberately designed to be small in order to facilitate the analysis of samples that are highly degraded or with low amounts of DNA [6,23]. The SNPs were PCR amplified in 18 amplicons with sizes ranging from 53 and 168 bp. Some amplicons encompass several SNPs (Table 1). We performed multiplex reaction using 5 ng of DNA template and PCR master mix of QIAGEN Kit Multiplex PCR (Qiagen, Hilden, Germany), amplification primers and their final concentrations are in Table 1. Amplification was carried out in a 9700 Thermocycler (Applied Biosystems, Foster City, CA, USA). After a 95 °C pre-incubation step for 15 min, PCR was performed in a total of 30 cycles using the following conditions: 94 °C denaturation for 30 s, annealing at 60 °C for 90 s and extension at 72 °C during 90 s, followed by 15 min of final extension at 72 and 4 °C until removed from thermocycler. PCR

Table 1
Coding region SNP, primer sequences (5′–3′) and final concentration (FC) for PCR amplification

SNP[a]	Primer forward	Primer reverse	Size (bp)	FC (μM)
1119	gctaagacccaaactgggatt	gttctggcgagcagtttttgt	84	0.5
1719/1736	ccaaacccactccaccttact	gcgccaggtttcaatttcta	87	0.2
2092	tgcccacagaaccctctaaa	tgtccaaagagctgttcctct	69	0.5
3547/3552	catctaccatcaccctctacatca	accaggggggttgggtatg	93	0.4
3970/4071	ggcttcaacatcgaatacgc	gaagtagggtcttggtgacaaaa	168	0.4
4491	attaatccccctggcccaac	gatgagtgtgcctgcaaaga	56	0.2
4820/4824/4883	tagccccctttcacttctga	ggcttacgtttagtgagggaga	135	0.4
5417	cctcaatcacactactccccata	ggggtgggttttgtatgttc	72	0.5
8281–8289del	tagggcccgtatttaccctat	aagaggtgttggttctcttaatcttt	109	0.5
8414	tactaccgtatggcccacca	gggaggtaggtggtagtttgtg	96	0.2
9824	ttttgtagccacaggcttcc	gaggaaagttgagccaataatga	53	0.4
10310/10398/14000	ccatgagccctacaaacaact	tgagtcgaaatcattcgttttg	158	0.5
11959/11969/12007	aaccacgttctcctgatcaaa	aatgtggtgggtgagtgagc	120	0.5
12338/12358	tggtcttaggcccccaaaaat	ggggaattagggaagtcagg	97	0.4
12705/12714/12771	cccaaacattaatcagttcttcaa	ctcgggcgtatcatcaactg	151	0.4
13759	ggaagcctattcgcaggatt	gttgtttggaaaggggatg	64	0.4
14502/14569	gacaaccatcattccccta	ctatttatgggggtttagtattgatt	124	0.4
15487/15335	cggcttacttctcttccttctct	atatcattcgggcttgatgtg	129	0.5

[a] Some SNPs are co-amplified (separated by "/") in the same amplicon; see Section 2.

products were checked by polyacrylamide gel electrophoresis (T9, C5) visualized by silver staining.

2.5. Minisequencing SNaPshot reaction

Previous to minisequencing reaction, PCR products are treated with ExoSAP-IT (Amersham Biosciences, Uppsala, Sweden) to remove excess of primers and un-incorporated dNTPs: 3 μl of PCR product was incubated with 1.5 μl for 15 min at 37 °C followed by 15 min at 80 °C for enzyme inactivation. Minisequencing reaction is performed using SNaPshot™ Kit (AB). Primers are sized between 19 and 67 bp; the length primers were modified with the addition of non-homologous tails poly(dC) or poly(dGACT) at the 5′-end (Table 2). The minisequencing reaction was carried out in a total volume of 10 μl comprising 3.5 μl of the SNaPshot™ Kit (AB), 1.5 μl PCR product, 1.5 μl of extension primers mix (final concentrations are between 0.1 and 0.5 μM) (Table 2), and water up to 10 μl. The reaction was performed in a 9700 Thermocycler (AB) following the recommendations of the manufacturer: 25 cycles of denaturation at 96 °C for 10 s, annealing at 50 °C for 5 s and extension at 60 °C during 30 s. Un-incorporated ddNTPs are eliminated with a treatment with SAP (Amersham Bioscences). The final volume (10 μl) was treated with 1 μl of SAP for 60 min at 37 °C followed by 15 min at 80 °C for enzyme inactivation.

2.6. Electrophoretic SNaPshot conditions

The minisequencing products (1.5 μl) were mixed with 10 μl of HiDi™ formamide (Amersham Bioscences) and 0.25 μl of GeneScan-120 LIZ (AB) and capillary electrophoresis was undertaken on an ABI PRISM 3100® Genetic Analyzer (AB). Minisequencing products were injected electrokinetically for 10 s at 15 kV and electrophorezed for 20 min at 15 kV and 9 μA at 60 °C in a capillary using the Performance Optimum Polymer 6 (POP6®) with the laser set at a constant power of 9.9 mW. Resulting data were analyzed using GeneScan™ 3.7 Software (AB).

2.7. PCR and sequencing reaction of the control region. Quality assessment

All samples have been amplified and sequenced for the complete control region, namely, from 16024 to 577 (then

Table 2
Coding region SNP, primer extension sequences (5′–3′) and final concentration (FC) for minisequencing reaction

SNP_ID	SNP	Extension primer[a]	Length	Base change	Strand[b]	FC (μM)
#1	1119	*(gact)$_8$*gttagccctaaacctcaacagttaaa	58	T > C	L	0.40
#2	1719	*gactgac*actccaccttactaccagacaacctta	34	G > A	L	0.15
#3	1736	ccttagccaaaccatttaccca	22	A > G	L	0.40
#4	2092	*(gact)$_4$*accctctaaatccccttgtaaatttaa	43	C > T	L	0.50
#5	3547	cccgaccttagctctcacc	19	A > G	L	0.15
#6	3552A	*(gact)$_{10}$*gaccttagctctcaccatcgc	61	T > A	L	0.40
#7	3970	*(gact)$_6$gag*tttgtgtattcggctatgaagaata	52	C > T	H	0.20
#8	4071	*gactgac*gcgcactctcccctgaactcta	28	C > T	L	0.20
#9	4491	cctgcaaagatggtagagtagatga	25	G > A	H	0.20
#10	4820	gaggggtgccttgggtaac	19	G > A	H	0.40
#11	4824	ctttcacttctgagtcccagaggtt	25	A > G	L	0.20
#12	4883	tcacatgacaaaaactagcccc	22	C > T	L	0.20
#13	5417	*(gact)$_9$*ccatatctaacaacgtaaaaataaaatgaca	67	G > A	L	0.40
#14	8281–8289del	*(gact)$_8$gac*gtatttaccctatagcacccctcta	61	C > G[c]	L	0.30
#15	8414	*(gact)$_3$ga*cgcccaccataattacccccata	37	C > T	L	0.20
#16	9824	*(gact)$_3$g*cacaggcttccacggact	31	T > C	L	0.20
#17	10310	*(gact)$_7$gac*atgagcccctacaaacaactaacct	55	G > A	L	0.20
#18	10398[d]	*(gact)$_7$*tatgagtgactacaaaaaggattagactga	58	A > G	L	0.40
#19	10400[d]	*(c)$_{11}$*gttttgtttaaactatataccaattc	37	C > T	H	0.40
#20	11959	*(gact)$_5$g*tcactctcctacttacaggactcaacat	49	A > G	L	0.30
#21	11969	(gact)$_4$ctacttacaggactcaacatactagtcaca	46	G > A	L	0.40
#22	12007	*(gact)$_4$*cctctacatatttaccacaacacaatg	43	G > A	L	0.15
#23	12338	*(gact)$_8$g*gtgcaactccaaataaaagtaa	55	T > C	L	0.50
#24	12358	*(gact)$_6$gag*gaagtcagggttagggtgg	46	A > G	H	0.30
#25	12705[d]	*(c)$_{17}$*aacattaatcagttcttcaaatatctactcat	49	C > T	L	0.30
#26	12714	*(gact)$_5$c*agttcttcaaatatctactcatcttcctaat	52	T > C	L	0.50
#27	12771	*(gact)$_2$c*aactgttcatcggctgaga	28	G > A	L	0.15
#28	13759	*(gact)$_4$gat*cattactaacaacatttcccc	40	G > A	L	0.15
#29	14502	*(gact)$_4$g*aaccatcattccccctaaataaa	40	T > C	L	0.50
#30	14569	*(gact)$_2$gat*aataacacacccgaccacacc	32	G > A	L	0.15
#31	15487T	*(gact)$_{11}$g*ctgggtcgcctaggaggtc	64	A > T	H	0.40
#32	15535	*(gact)$_2$gat*tatacctagccaaccccttaaa	34	C > T	L	0.30

[a] Italic indicates the non-homologous tails.
[b] Strand refers to the targeted DNA chain for SNP genotyping.
[c] Note that the 9 bp deletion (8281–8289del) is interrogated as a C to G change.
[d] This probes were designed in Ref. [6].

covering the classical hypervariable regions HVS-I and HVS-II). For practical purposes, in what follows, HVS-I comprises the control region segment from position 16024 to 16569, while HVS-II expands from positions 1 to 577.

PCR amplification was carried out in a thermocycler GenAmp PCR System 9700 (AB) using primers 15997L-5′-CACCATTAGCACCCAAAGCT-3′ and 017H-5′-CCCGTGAGTGGTTAATAGGGT-3′ for HVS-I and 16555L-5′-CCCACACGTTCCCCTTAAAT-3′ and 599H-5′-TTGAGGAGGTAAGCTACATA-3′ for HVS-II. For those samples that could not be amplified for large fragments, we have combined the previous primers with other alternative ones: 16401H-5′-TGATTTCACGGAGGATGGTG-3′, 16380L-5′-TCAGATAGGGGTCCCTTGAC-3′, 285H-5′-GGGGTTTGGTGGAAATTTTTTG-3′, 172L-5′-ATTATTTATCGCACCTACGT-3′. Primer 611H-5′-AGTGTATTGCTTTGAGGAGG-3′ was also used in some particular cases. Additional primers were required for sequencing: 16236H-5′-CTTTGGAGTTGCAGTTGATG-3′, 16254L-5′-CACATCAACTGCAACTCCAAA-3′, 370L-5′-CCCTAACACCAGCCTAACCA-3′, and 408H-5′-CTGTTAAAAGTGCATACCGCCA-3′. Note that some primers were reported in previous studies: 15997L [24], 16401H and 408H [25], and 599H [26]; while the others were designed for the present study. dRhodamine chemistry (AB) was used for cycle sequencing. Each template was systematically sequenced in both forward and reverse directions. Sequencing was carried out in a capillary electrophoresis ABI 3100. Additional PCR and sequencing analytical details can be provided under request.

Mutations are referred with respect to the revised Cambridge Reference Sequence or rCRS [27]. Haplogroup classification was carried out according to the most updated nomenclature ([14]; and see references therein). We follow the 'forensic style' for nomenclature as indicated in Ref. [28], but with the modifications considered in Ref. [29].

Phylogenetic inconsistencies usually arise due to laboratory artifacts and documentation errors [1,17,29–37]. We have examined all our sequencing results and/or SNaPShot genotypes within the framework of the known mtDNA tree. In order to minimize as much as possible data errors, unexpected results (phylogentic inconsistencies) in the control or the coding region (SNaPShot genotypes) were confirmed by repeating the extraction, PCR amplification, and (also in the case of SNaPShot genotypes) by sequencing (forward and reverse) the corresponding coding region fragments.

3. Results and discussion

3.1. Rationale of the SNP selection

We have developed a single multiplex SNaPshot reaction for simultaneously genotyping 32 selected coding region mtDNA SNPs representing main and derived branches of the East Asian and Native American mtDNA phylogeny. Many of these SNPs were selected in order to cover sub-clades that are not clearly defined by control region polymorphisms; therefore, they are important for haplogroup assignment and for increasing the discrimination power even when the HVS-I/II segments or the whole control region (the most common regions analyzed by forensic and population geneticists) is known.

The rationale of our design (see Fig. 1) contrasts in several aspects with the previous attempt of Ref. [5]. For instance, the whole East Asian haplogroup B is covered by the well-known 9 bp deletion 8281–8289d in Umetsu et al. [5]; however, in addition to this deletion, we decided to incorporate SNPs that cover some of the various haplogroup B sub-branches, such as B4b′d (C15535T), B4b (G4820A), and B2 (A3547G). Note that other B sub-clades are well supported by control region diagnostic sites, for instance, B5b1 (C16111T, C16234T, and

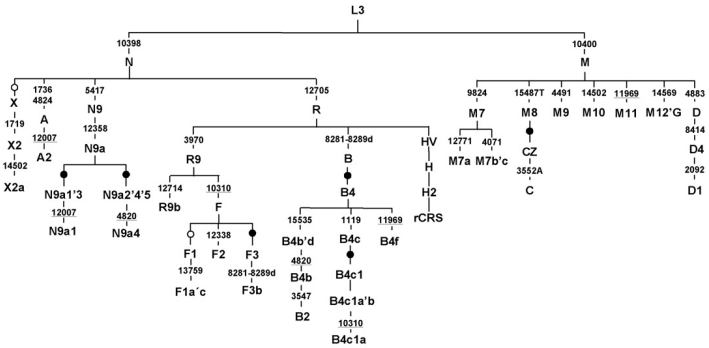


Fig. 1. SNaPshot of the East Asian and Native American mtDNA phylogeny defined by the SNPs incorporated in our multiplex design (or/and by control region variants and other coding SNPs that are not indicated in the tree). For the nomenclature we follow the most recent update of the East Asian phylogeny [14]. Parallel mutations are underlined. Some mutations define other minor branches (mainly) at the twigs of the East Asian phylogeny (for instance, A12358G additionally characterizes the motif that lead to D4b2b2 and A10398G that characterizes macro-haplogroup N is also recurrent in many other different branches (six hits in Figs. 1 and 2 of Ref. [14])) that are not indicated in the present figure. A filled dot indicates the existence of control region site(s) defining the corresponding branch, whereas an empty circle indicates the existence of coding region sites (but not control region SNPs) defining this branch (but not covered by the present SNaPshot design).

A16463G) or B5b2a (T195C, C16278T, and T16324C), or B4f (A200G, C16168T, T16172C, T16249C, T16325C, and G16390A), whereas we are aware that some other sub-clades would need further SNP genotyping effort (e.g. B4a* is defined by the coding region variants T5465C, and G9123A, while some other B4a sub-clades are supported by control region variants [14]). For some branches, such as the one that lead to haplogroup D4, we prioritized C8414T instead of the highly mutational unstable polymorphism G3010A [38] as done in Ref. [5]. Note that G3010A is a good candidate to increase the discrimination power of the mtDNA test (e.g. in forensic casework) because of its high mutability [1,38,39]; however, it is not the best candidate for haplogrouping. Another example, while Umetsu et al. [5] incorporated a respectable number of SNPs (eight out of their 36 selected SNPs; ~22%) defining West European haplogroups (such as haplogroup K, T, V, etc., of their multiplex "D"; Table 1; [5]) or even one sub-Saharan basal one (indicated as diagnostic of L1/2 by Ref. [5]), we here prioritized East Asian SNPs and those leading to the Native American branches (directly derived from the East Asian phylogeny) such as haplogroup A (A1736G and A4824G), haplogroup C (T3552A), and haplogroup X (T14502C).

It is worth mentioning that haplogroup assignation in Ref. [5] enters in conflict with the contemporary (and obviously the most updated) mtDNA phylogeny from East Asia due to the fact that they followed the slightly defective nomenclature employed by Ref. [22]; see Ref. [14] for a critical discussion. Thus, for instance, polymorphism A11959G leads to haplogroup D2 [19], not to M12 as indicated in their Table 1 [5]; the M12 label has been recently used to 'baptize' a sister branch of haplogroup G (with which it shares mutation G14569A, defining haplogroup M12′G). Then, the following diagnostic sites on top of M12′G: C4170T, T5580C, A12030G, G12372A, T14727C, A15010G, T16172C, C16234T, and C16290T, lead to haplogroup M12. Another example, transition C3970T defines haplogroup F according to Ref. [5] but it actually characterizes R9 [21], which embraces haplogroup F (first introduced as R9 in Ref. [40]) and R9b (first named R10 by Ref. [41]).

As mentioned above, both the [5] design and the one presented here do not fully cover the whole East Asian phylogeny. For instance, there are many 'twigs' of the mtDNA tree (sensu [20]) that are not defined by control region variants and would need the use of some extra coding SNPs; this is the case of e.g. B4a1a1a, B4a1a2, etc. [14].

In our present multiplex design, there are also some recurrent mutations appearing in different parts of the tree, thus playing a double role to allocate control region profiles to particular haplogroups; this is the case of for instance A10398G (B4c1c, B5, R11, N, Y, etc.) or T9824A (M7, D4b2; note that there are not control region diagnostic sites for these sub-clades). These branches are not indicated in Fig. 1, but can be contrasted with Figs. 1 and 2 of Ref. [14].

Therefore, thanks to the preliminary studies of Refs. [19–21,40–48]; among others and the recent studies on complete sequencing by Refs. [22,14], we have a better knowledge of the East Asian and Native American mtDNA phylogeny. This higher resolution 'invites' to improve (using e.g. a supplementary SNaPshot multiplex reaction) mtDNA SNP coverage by designing a supplementary multiplex reaction that would allow to fully cover this continental section of the mtDNA tree.

3.2. Population mtDNA profiles and phylogeny

Our SNaPshot multiplex has been validated in a sample of Taiwanese and Native Americans from Argentina. Fig. 2 shows two examples of SNaPShot patterns.

We have obtained SNaPshot and control region sequencing profiles for a total of 121 mtDNAs. We have detected 58 HVS-I, 51 HVS-II, and 60 different control region (HVS-I/II) haplotypes in Taiwanese ($N = 60$), while 49 HVS-I, 51 HVS-II, and 60 different control region (HVS-I/II) Native American haplotypes ($N = 61$; Table 3). These results indicate the convenience of sequencing the complete control region instead of analyzing the classical HVS-I region solely or HVS-I in combination with the classical HVS-II [1,29]. As in other parts of the worldwide phylogeny, many discriminatory polymorphisms within the control region are outside the classical HVS-I and HVS-II segments (G499A, G513A, T16519C, etc.). The genotyping of coding region SNPs does not substantially improve the discrimination provided by the control region in our particular dataset since the maximum is nearly achieved with the analysis of the whole control region alone, but it is undoubtedly of great help to allocate mtDNA sequencing profiles to specific haplogroups. In larger datasets, the utility of our selected coding region SNPs as a tool to increase the discrimination power in forensic casework would be more evident. The later can be predicted using East Asian complete genomes taken from the literature or databases (author's unpublished data).

There are few interesting phylogenetic features that deserve some comment (see data in Table 3). Sample #TW-38 belongs to haplogroup M12. This mtDNA harbors in HVS-II a characteristic profile including variants T125C, T127C, and C128T; at least T125C and T127C are probably part of the basal diagnostic motif of M12. This motif (T125C–T127C) is also found in samples #GD7825 and #Hani57 of Ref. [14] and, in other previous studies, with C128T on top [20,40,49]. #TW-38 also carries T318C which has also been detected in various different M12 profiles [20,40], this variant is probably part of the motif of some minor M12 sub-clade. In addition, the evidences seem to indicate that T16172C, as shown in Ref. [14], is probably not part of the basal motif of M12.

The mtDNA of #Tw-08 carries a peculiar HVS-I profiles which likely corresponds with some unknown haplogroup R minor branch.

Samples #CO-18, #CO-20, #CO-11, #CO-44, and #CO-50, all of them belonging to haplogroup B2, lack variant A73G, a characteristic that probably point to a new sub-branch of B2 in Native Americans. Some other B2 profiles also lack some variant defining their basal motif (#TW-21 and #TW45 lack C16223T while #CO-30 lacks A235G).

The #Tw-51 belongs to F1a and carries an interesting profile in HVS-II: T52C G53A G54C 71delG A73G C150T 249delA

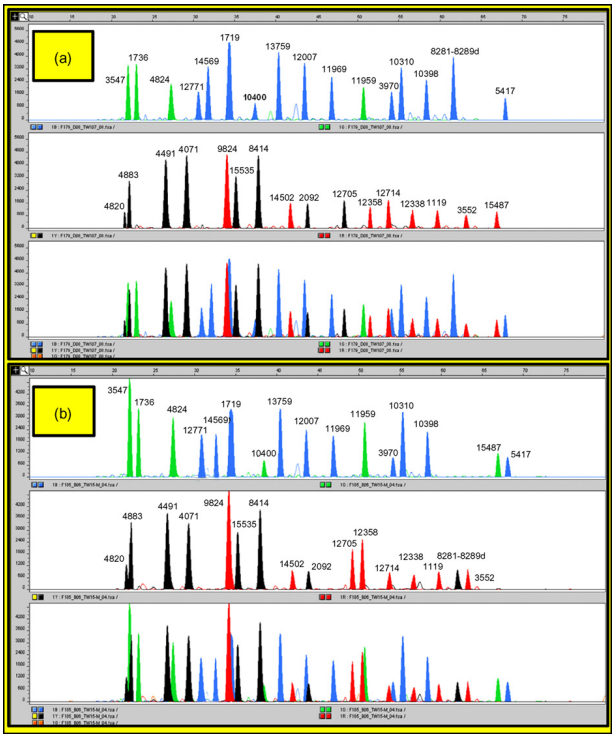


Fig. 2. Example of SNaPshot profiles for a sample belonging to haplogroup B5b (2A) and another belonging to M8a2a (2B).

263 309 + C 315 + C T318C 523-524delAC. Deletion of position 71 is a rare event that was found basically within B4b1 (in the Korean dataset of Ref. [50] and the 'Hispanic' and 'Chinese' datasets of Ref. [51]). Moreover, it was found sporadically in some other lineage (e.g. haplogroup K in sample #163 of Ref. [52]). In combination with variants T52C G53A G54C, 71delG defines probably a new branch of F1a that can tentatively named here as F1a4 (consistently with the nomenclature of Ref. [14]). Characteristic of this sub-clade is also transition T318C (as in #CHN.ASN.000097 and #CHN.ASN.000330 of Ref. [51]).

The Argentinean haplogroup C1 sample #CO-58 carries an interesting 6 bp deletion in HVS-II ranging from position 105 to 110, and creating the following sequence pattern 100-GAGCACCCT-115. We observed in our database three instances of this deletion: #USA.335.000133, #USA.-HIS.000378, and #USA.HIS.000499 in Ref. [51]; the last one also within haplogroup C1. Note that this deletion is a different mutational event (and probably not phylogenetically related) to the characteristic deletion of the Chibchan Native Americans [53] or the Ngöbe of Ref. [54] (also in e.g. the 'Hispanic' of Ref. [51]), since the later occurs within haplogroup A2, from position 106 to position 111, then generating a different patter 100-GAGCCCCCT-115 (note also that deletion 105-110del can occur outside C1; see in e.g. #USA.HIS.000378 within L2b1).

Samples #TW-23, #TW-30, and #TW-62 all carry T16362C (though this branch seems to lack it; see Ref. [14]); anyway, T16362C is an unstable position that mutates back several times along the mtDNA phylogeny, in this particular example, within haplogroup M9.

The Taiwanese #TW-39 belongs to F2a2 since it carries C16291T on top of the F2 diagnostic motif (including T12338C; [14]) and transversion T16092A [42]; a sub-clade with representatives in China, Taiwan, and Korea, as well as Asian descendants in America [51].

Samples #CO-33 and #CO-40 belong to haplogroup A2 (G12007A), but lacks the control region variant C64T; this seems to indicate the existence of an early split within the phylogeny of A2. Samples #CO-40 and #CO-30 also lack the

Table 3
Control and coding region variants in our sample from Taiwan and Argentina

Sample ID	HVS-I (minus 16000)	HVS-II (263 and 315+C is present in all the samples, with the exception of Tw-48 which lacks 315+C)	1119	1719	1736	2092	3547	3552	3970	4071	4820	4824	4883	4491	5417	8414	9824	10310	10398	10400	11959	11969	12007	12338	12358	12705	12714	12771	13759	14502	14569	15487	15535	9 bp[2]	HG
			T	G	A	C	A	T	C	C	G	A	C	G	G	C	T	G	A	C	A	G	G	T	A	C	T	G	G	T	G	A	C		
TW-41	223 247 290 319 519	073 152 235 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A*
CO-33	092 223 234 256 290 319 325 362 519	073 146 153 200 235 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	A	-	-	T	-	-	-	-	-	-	-	-	A2
CO-45	111 189 193+C 209 223 290 319 362 519	064 073 146 235 309+C 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	A	-	-	T	-	-	-	-	-	-	-	-	A2
CO-09	111 223 287 290 319 362	064 073 146 153 235 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	A	-	-	T	-	-	-	-	-	-	-	-	A2
CO-15	111 223 287 290 319 362	064 073 146 153 235 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	A	-	-	T	-	-	-	-	-	-	-	-	A2
CO-43	111 223 290 311 319 362 463 519	064 073 143 146 153 235 309+C 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	A	-	-	T	-	-	-	-	-	-	-	-	A2
CO-40	111 223 290 311 319 362 512 547	073 146 153 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	A	-	-	T	-	-	-	-	-	-	-	-	A2
CO-25	182C 183C 189 223 256 260 288 290 319 362	064 073 146 153 234 235 309+C 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	A	-	-	T	-	-	-	-	-	-	-	-	A2
CO-30	223 290 319 362 519	064 073 146 153 182 194 195 309+C 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	A	-	-	T	-	-	-	-	-	-	-	-	A2
TW-44	126 223 235 290 319	073 235 523-524d	-	-	G	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A5b
CO-19	051 183C 189 217 360 519	073 152 309+C 499 523-524d	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-12	092 183C 189 193+C 217 311 519	073 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-56	092 183C 189 217 519	073 143 151 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-60	093 183C 189 193+C 217	073 093 195 309+C 499	-	-	-	-	G		-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-26	093 183C 189 193+C 217 519	073 093 195 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-52	111 183C 189 217 270 519	073 146 215 309+C 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-36	111 183C 189 217 512 519	073 185 309+C 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-53	168 183C 189 217 519	073 309+C 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-49	176 182C 183C 189 193+C 217 319 422 519	073 293 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-24	182C 183C 189 193+C 217 519	073 203 204 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-37	182C 183C 189 217 242 519	073 146 153 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-41	182C 183C 189 217 249 519	073 146 195 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-35	182C 183C 189 217 519	073 200 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-55	182C 183C 189 217 519	073 197 207 292 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-22	183C 188 189 217 311 519	073 186 309+C 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-27	183C 188 189 217 354 519	073 146 186 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-16	183C 188 189 217 519	073 241 309+C 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-18	183C 188 189 217 519	309+C 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-20	183C 188 189 217 519	093 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-32	183C 189 193+C 217 292 519	063 064 066T 073 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-42	183C 189 193+C 217 381	073 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-07	183C 189 193+C 217 519	073 204 207 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-47	183C 189 217 286 289 519	073 143 146 199 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-10	183C 189 217 289	073 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-23	183C 189 217 359 519	073 103 146 151 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-28	183C 189 217 519	073 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-29	183C 189 217 519	063 064 073 146 215 309+C 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-34	183d 217 428 465 519	073 309+C 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-59	183d 217 465 519	073 309+C 499	-	-	-	-	G		-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-11	183d 217 519	499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-39	183d 217 519	073 143 210 309+C 499 513 524+ACAC	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-44	183d 217 519	309+C 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-48	183d 217 519	042 073 143 210 309+CC 499 524+AC	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-50	183d 217 519	309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
CO-51	189 217 381 519	073 309+CC 499	-	-	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B2
TW-60	129 182C 183C 189 217 519	073 093 309+CC 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4
TW-26	093 150 182C 183C 189 217 261	073 195 309+CC 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4a'g
TW-54	093 182C 183C 189 217 261 319	073 309+C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4a'g
TW-63	129 182C 183C 189 217 261	073 195 309+CC 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4a'g
TW-40	129 182C 183C189 217 261	073 195 309+CC 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4a'g
TW-37	129 183C 189 217 261	073 150 195 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4a'g
TW-20	129 182C 183C 189 261 519	073 309+CC 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4a'g
TW-16	136 182C 183C 189 217 519	073 207 309+CC 499 523-524d	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	D	B4b1a
TW-55	140 182C 189 217 274 335	073 146 150 309+C	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4c1b
TW-53	092 140 182C 183C 189 217 274 311 335 519	073 146 150	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4c1b
TW-58	182C 183C 189 213 217 261 292 519	073 309+CC 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B4g
TW-49	140 183C 189 243 301C/T 311 319 519	073 103 204 207 309+CC 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B5b
TW-59	140 183C 189 243 311 319 519	073 103 204 207 309+C 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	B5b
TW-52	093 172 223 298 327 519	073 249d 489	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C*
TW-11	189 223 265C 298 317 327 519	073 249d 309+C 489	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C*
CO-17	093 145 169 223 274 298 325 327 519	073 249d 290-291d 309+CC 489	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-03	093 192 223 298 325 327	073 201 249d 290-291d 309+C 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-13	093 192 223 298 325 327	073 249d 290-291d 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-14	093 192 223 298 325 327	073 201 249d 290-291d 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-46	136 223 298 325 327 519	073 146 249d 290-291d 309+C 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-54	169 192A 209 223 298 325 327 519	073 249d 290-291d 309+C 489	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1

CO-38	182C 183C 189 223 298 311 325 327 519	073 249d 290-291d 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-02	223 224 298 325 327	073 249d 290-291d 309+C 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-61	223 298 325 327	073 194 195 204 249d 290-291d 309+C 489 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-58	223 298 325 327	073 105-110d 188 249d 258 290-291d 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-01	223 298 325 327 354 356	073 249d 290-291d 309+C 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-57	223 298 325 327 362	034 073 249d 290-291d 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-21	223 298 325 327 519	073 146 249d 309+C 489 493 523-524d	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-08	223 304 325 327 400 519	073 249d 290-291d 309+CC 489	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	C1
CO-04[1]	223 242 311 325 362	-	-	-	-	T	-	-	-	-	-	-	T	-	-	T	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D1
CO-05	223 325 362 519 527	073 489	-	-	-	T	-	-	-	-	-	-	T	-	-	T	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D1
TW-18	092 223 362 395 519	073 309+C 489	-	-	-	-	-	-	-	-	-	-	T	-	-	T	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D4
TW-34	189 362	073 192 489	-	-	-	-	-	-	-	-	-	-	T	-	-	T	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D4
TW-61	223 242 362	073 309+C 437 489	-	-	-	-	-	-	-	-	-	-	T	-	-	T	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D4
TW-19	129 223 270 362 519	073 146 152 309+CC 489 523-524d	-	-	-	-	-	-	-	-	-	-	T	-	-	T	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D4a
CO-06	223 241 294 301 342 362	073 152 279 309+C 489	-	-	-	-	-	-	-	-	-	-	T	-	-	T	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D4i
TW-29	092 223 362 519	073 309+C 489	-	-	-	-	-	-	-	-	-	-	T	-	-	T	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D4*
TW-32	164 172 182C 183C 189 223 266 362	150 489 523-524d	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D5a2
TW-13	162 184 189 223 291 325 362 519	073 150 309+CC 456 489	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D5b
TW-50	183C 189 362	073 150 456 489	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	D5b
TW-03	124 223 362 390 519	073 309+CC 489	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	E1
TW-45[1]	209 218 304 311	061 062 073 249d 309+CC	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	F4b
TW-51	129 172 242 304 519	052 053 054C 071d 073 150 249d 309+C 318 523-524d	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	F1a
TW-28	172 223 362 519	073 200 249d 309+CC 523-524d	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	F1a
TW-31	129 162 172 304 519	073 249d 523-524d	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	F1a1
TW-07	108 129 162 172 201 304 311 519	073 150 249d 523-524d	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	F1a1a
TW-25	108 129 162 172 304 519	073 249d 309+CC 523-524d	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	F1a1a
TW-33	108 129 162 172 304 519	073 249d 523-524d	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	F1a1a
TW-10	189 232A 249 304 311 519	073 249d 309+CC 523-524d	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	F1b
TW-01	172 183C 189 223 304 355 519	073 249d 309+C	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	F2
TW-21	304 311	073 249d 309+C 513 573+CCCC	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	F2
TW-39	092A 291 304 325	073 199 204 249d 309+C 523-524d	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	F2a
TW-57	203 291 304 311 519	073 249d 309+C	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	F2a3
TW-02	203 291 304 519	073 249d 309+C	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	F2a3
TW-42	075 093 220C 254 298 311 362	073 195 249d 309+C	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	F3b
TW-22	172 223 227 272 278 319 362 519	073 152 309+C 489	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	A	-	-	-	G2a1
CO-31	126 362 482 519	152 239 309+CC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H6a[3]
TW-12	068 183C 189 223 311 519	073 146 150 309+CC 489 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M*
TW-48	183C 189 519	073 146 150 152 310 489 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M*
TW-38	223 234 287 290 362	073 125 127 128 309+C 318 489 523-524d	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	A	-	-	-	M12
TW-09	129 192 223 245 297	073 150 199 207 309+C 489 524+AC	-	-	-	-	-	-	-	T	-	-	-	-	-	-	C	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M7b1
TW-46	129 192 223 297	073 150 199 309+C 489	-	-	-	-	-	-	-	T	-	-	-	-	-	-	C	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M7b1
TW-36	223 295 519	073 199 489 523-524d	-	-	-	-	-	-	-	T	-	-	-	-	-	-	C	-	G	T	-	-	-	-	-	T	-	-	A	-	-	-	-	-	M7c1'2
TW-35	172 223 311 519	073 146 309+C 489 523-524d	-	-	-	-	-	-	-	T	-	-	-	-	-	-	C	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M7b'c
TW-43	172 223 311 519	073 146 309+C 489 523-524d	-	-	-	-	-	-	-	T	-	-	-	-	-	-	C	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M7b'c
TW-14	162 184 189 223 298 319 470 471 473	073 195 309+C 489	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	M8a2a
TW-15	184 189 223 278 298 319 470 471 473	073 152 489	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	M8a2a
TW-30	051 086 185 223 362	073 195 489	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M9
TW-62	223 291 362 519	073 131 309+C 489	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M9
TW-47	223 234 316 362 399	073 150 153 185 189 489 523-524d	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M9a
TW-23	223 234 271 362	073 151 153 309+CC 489	-	A	-	-	-	-	-	-	-	-	-	A	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	-	-	-	M9a
TW-17	092 145 172 223 245 257A 261	073 150 309+C 524+ACAC	-	-	-	-	-	-	-	-	A	-	-	-	A	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	-	-	-	N9a4
TW-08	093 172 174 250 311 357 519	073 146 309+CC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R*
TW-24	189 192 288 304 309 390 519	073 183 309+C 523-524d	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	R9b
TW-27	185 223 260 298	073 152 249d 489 523-524d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	Z
TW-56	185 223 260 298 519	073 152 249d 309+C 489	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	Z

Note: Nomenclature of mutations is as follow: a number indicates a transition refered to the rCRS (see text), a letter after this number indicates a transversion; d: deletion; symbol ‘‘+’’ indicates an insertion. (1) A 9 bp refers to the 9 bp deletion between positions 8281 and 8289. (2) There was not enough DNA of sample #CO-04 to sequence the HVS-II; for the same reason, sample #TW-45 could only be replicated for the sequence range 16024-315. (3) Sample #CO-31 haplogroup affiliation is supported by the additional genotyping of the haplogroup H coding region SNPs reported in Ref. [6]; see text. In Sample ID column, prefix ‘‘CO’’ is for the Argentinean samples, while ‘‘TW’’ refers to the Taiwanese mtDNAs.

haplogroup A diagnostic site A235G, whereas #CO-08 lacks the diagnostic haplogroup C1 variant T16298C. Matches or close matches (one mutational step) to these mtDNAs are relatively common in our database. Given the low resolution of Native American haplogroups, these features point to good candidates for complete genome sequencing.

There are various Native American haplogroup B2 profiles with a characteristic pattern: a rCRS T-variant at position 16189 with a deletion of an adenine at the homopolymeric track between positions 16180 and 16183, denoted as 16183d following the forensic convention (note that T16189C is characteristic of haplogroup B). We have searched the database (e.g. >4000 Native American profiles) for this motif and we have not found any single match. We only observed sporadic instances of 16183d–'16189T' mainly within some sub-Saharan haplogroup. We could hypothesize that this could respond to some problem of nomenclature (sensu [29]), since it affects a problematic mtDNA region; the electropherograms in these profiles are however very clear (since the homopolymeric tract is interrupted by a T at 16189; the rCRS pattern) and does not seem to be prone for ambiguity in the nomenclature. It is then obvious that the most parsimonious way to record this variation around this mtDNA segment is simply 16183d; other alternative nomenclature searches in our database did not yield any match.

Sample #CO-31 was initially classified as belonging to the West European haplogroup H based on control region information. We further explore its haplogroup affiliation by genotyping the coding region SNPs reported in Ref. [6]; namely, G3010A, G3915A, C3992T, T4336C, G4769A, A4793C, T6776C, and C7028T. We tentatively classified this sample as H6a since it carries the coding region variant G3915A (see [4,6,14,38,55] and is consistent with the other genotyped haplogroup H characteristic SNPs (which match the rCRS; [27]). Note that, as mentioned in Ref. [4], G3915A also occurs in other H sub-clades and outside haplogroup H; nevertheless, classification of #CO-31 within H6a is supported by other characteristic variants in the control region (T239C, T16362C, and A16482G).

We have found a good agreement between coding and control region profiles. We have detected only three phylogenetic control/coding region inconsistencies that were further explored by sequencing the forward and reverse strands of the corresponding coding region segments in independent DNA extractions (the control region of these samples were also re-sequenced in order to rule out potential artefactual recombinants). These unexpected variants were also re-confirmed by single-SNP SNaPshot. Mutation G1719A defining branch D6 [14] also appears within haplogroup M9 in sample #TW-23. This is not an unusual mutational event since G1719A occurs many times along the mtDNA phylogeny. Transition A10398G that leads from N to L3 also occurs in the B4a′g Taiwanese sample #TW-26. On the other hand, transition G13759A characterizes F1a′c; here, we detected one instance within M7c3 (#TW-36); this mutation was also observed sporadically within different haplogroup backgrounds.

We detected one point heteroplasmy, 16301C/T, in sample #TW-49. This heteroplasmy was also confirmed by sequencing forward and reverse in different PCR and sequencing reactions of an independent DNA extraction.

Finally, we did not detect any representative of some minor sub-clades (e.g. M7a [G12771A] and M10 [T14502C]).

3.3. Considerations about the minisequencing technique

Based on previous SNaPshot mtDNA studies [4,6,9,56] it seems that genotyping errors induced by the presence of secondary homologies at the level of nuclear DNA (nDNA) are unlikely to have important impact in forensic, anthropological, and clinical studies. However, since NUMTS-mediated artifacts have been observed in some circumstances [57,58], it is convenient to be aware of potential errors, especially when working under sub-optimum conditions (e.g. degraded DNA, low copy number DNA, etc.).

In principle, there are various reasons that would prevent SNaPshot primer designs from genotyping artifacts due to secondary nDNA homologies. First, the relative amount of mtDNA respect to the nDNA generally will help to prevent primer annealing in other DNA than the mtDNA molecule; this could be also true when working with degraded DNA; however, we are aware that this statement would require proper validation. Second, we have observed in previous attempts [4,6] a consistent (phylogenetic) accordance between the expected SNaPshot profile and the HVS-I/II sequencing results, with only sporadic inconsistencies that could also easily been explained by recurrent mutation. Anyway, phylogenetically unexpected patterns can be easily confirmed using single-plex SNaPshot typing, or even better, using more standard techniques (sequencing), such as it was done in the present study. It is important to remember that other apparently well-known common techniques for mtDNA genotyping, such as automatic sequencing, are far to be exempt of problems [1,17,29,30,33–37,59]. In fact, there are not *a priori* reasons to believe that SNaPshot reaction is more prone to artefacts than other techniques in similar conditions of DNA quality, amount of DNA, etc.

On the other hand, there are many advantages of this technique. In brief, it allows to genotype multi-locus mtDNA SNPs with low cost and high efficiency and, more importantly in the forensic field, the sensitivity of the technique seems to be higher than sequencing ([23]; author's unpublished data). Moreover, this genotyping is carried out in a single SNaPshot reaction, avoiding e.g. single-plex genotyping which is more prone to artefactual recombination [1,29,33–35]. An additional advantage is that most of the forensic and population genetics laboratories use automatic sequencers; while only a few of them have access to other more uncommon, expensive, and none properly standardized technologies (e.g. high throughput genotyping platforms, DNA micro-arrays), at least, in the forensic genetics field. Note also that the present SNaPshot design is scalable; therefore, incorporation of new coding region SNPs in our multiplex design is perfectly feasible and can be undertaken as a function of the growing phylogenetic refining of the mtDNA tree.

We also consider worth to systematically incorporate the genotyping of coding region sites in the forensic routine [1];

methodologies such as the SNaPshot reaction shown here could well be part of standardization and multi-centric collaborative exercises (such as those of the GEP-ISFG on mtDNA; [23,32,60,61]) in order to promote the use of coding region SNP genotyping in forensic casework. It would be also important to properly test the ability of the SNaPshot for genotyping degraded DNA or even hair shafts (commonly use in population and forensic genetic studies [62,63]), although preliminary results are promising [23].

3.4. Final remark

Such as it has been several times suggested (e.g. [29,64]), interpretation of mtDNA data should be done under the phylogenetic perspective in order to reduce genotyping and documentation errors. As done in the present study, any phylogenetic inconsistence or unexpected result should be double checked, preferably by sequencing the forward and the reverse strands of the mtDNA fragment under study using if possible different DNA extractions.

Acknowledgements

We would like to thank Yong-Gang Yao and Claudio Bravi for critically checking the data and reading the manuscript. The 'Ramón y Cajal' Spanish programme from the Ministerio de Educación y Ciencia (RYC2005-3), the grant of the Ministerio de Sanidad y Consumo (PI030893); SCO/3425/2002), and the grant from the Fundación de Investigación Médica Mutua Madrileña, given to AS, partially supported this project.

References

[1] A. Salas, H.-J. Bandelt, V. Macaulay, M. Richards, Phylogeographic investigations: the role of trees in forensic genetics, Forensic Sci. Int., doi:10.1016/j.forsciint.2006.05.037, in press.
[2] A. Salas, M. Richards, T. De la Fé, M.V. Lareu, B. Sobrino, P. Sánchez-Diz, V. Macaulay, Á. Carracedo, The making of the African mtDNA landscape, Am. J. Hum. Genet. 71 (2002) 1082–1111.
[3] A. Salas, M. Richards, M.V. Lareu, R. Scozzari, A. Coppa, A. Torroni, V. Macaulay, Á. Carracedo, The African diaspora: mitochondrial DNA and the Atlantic slave trade, Am. J. Hum. Genet. 74 (2004) 454–465.
[4] A. Brandstätter, A. Salas, H. Niederstätter, C. Gassner, Á. Carracedo, W. Parson, Dissection of mitochondrial superhaplogroup H using coding region SNPs, Electrophoresis 27 (2006) 2541–2550.
[5] K. Umetsu, M. Tanaka, I. Yuasa, N. Adachi, A. Miyoshi, S. Kashimura, K.S. Park, Y.H. Wei, G. Watanabe, M. Osawa, Multiplex amplified product-length polymorphism analysis of 36 mitochondrial single-nucleotide polymorphisms for haplogrouping of East Asian populations, Electrophoresis 26 (2005) 91–98.
[6] B. Quintáns, V. Álvarez-Iglesias, A. Salas, C. Phillips, M.V. Lareu, Á. Carracedo, Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing, Forensic Sci. Int. 140 (2004) 251–257.
[7] A. Maitra, Y. Cohen, S.E. Gillespie, E. Mambo, N. Fukushima, M.O. Hoque, N. Shah, M. Goggins, J. Califano, D. Sidransky, A. Chakravarti, The Human MitoChip: a high-throughput sequencing microarray for mitochondrial mutation detection, Genome Res. 14 (2004) 812–819.
[8] A. Salas, B. Quintáns, V. Álvarez-Iglesias, SNaPshot typing of mitochondrial DNA coding region variants, Methods Mol. Biol. 297 (2005) 197–208.
[9] M.D. Coble, R.S. Just, J.E. O'Callaghan, I.H. Letmanyi, C.T. Peterson, J.A. Irwin, T.J. Parsons, Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians, Int. J. Legal Med. 118 (2004) 137–146.
[10] F. Barros, M.V. Lareu, A. Salas, A. Carracedo, Rapid and enhanced detection of mitochondrial DNA variation using single-strand conformation analysis of superposed restriction enzyme fragments from polymerase chain reaction-amplified products, Electrophoresis 18 (1997) 52–54.
[11] A. Salas, E.M. Rasmussen, M.V. Lareu, N. Morling, Á. Carracedo, Fluorescent SSCP of overlapping fragments (FSSCP-OF): a highly sensitive method for the screening of mitochondrial DNA variation, Forensic Sci. Int. 124 (2001) 97–103.
[12] J.J. Sánchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, Á. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724.
[13] P.M. Vallone, R.S. Just, M.D. Coble, J.M. Butler, T.J. Parsons, A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome, Int. J. Legal Med. 118 (2004) 147–157.
[14] Q.-P. Kong, H.-J. Bandelt, C. Sun, Y.-G. Yao, A. Salas, A. Achilli, C.Y. Wang, L. Zhong, C.L. Zhu, S.F. Wu, A. Torroni, Y.-P. Zhang, Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations, Hum. Mol. Genet. 15 (2006) 2076–2086.
[15] S.V. Baudouin, D. Saunders, W. Tiangyou, J.L. Elson, J. Poynter, A. Pyle, S. Keers, D.M. Turnbull, N. Howell, P.F. Chinnery, Mitochondrial DNA and survival after sepsis: a prospective study, Lancet 366 (2005) 2118–2121.
[16] A. Vega, A. Salas, E. Gamborino, M.J. Sobrido, V. Macaulay, Á. Carracedo, mtDNA mutations in tumors of the central nervous system reflect the neutral evolution of mtDNA in populations, Oncogene 23 (2004) 1314–1320.
[17] A. Salas, Y.-G. Yao, V. Macaulay, A. Vega, Á. Carracedo, H.-J. Bandelt, A critical reassessment of the role of mitochondria in tumorigenesis, PLoS Med. 2 (2005) e296.
[18] A. Salas, Y.G. Yao, H.J. Bandelt, Mitochondria: more than mitochondrial DNA in cancer-reply, PLoS Med. 3 (2006) 414–415.
[19] H.-J. Bandelt, C. Herrnstadt, Y.-G. Yao, Q.-P. Kong, T. Kivisild, C. Rengo, R. Scozzari, M. Richards, R. Villems, V. Macaulay, N. Howell, A. Torroni, Y.-P. Zhang, Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats, Ann. Hum. Genet. 67 (2003) 512–524.
[20] T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S.S. Papiha, H.-J. Bandelt, R. Villems, The emerging limbs and twigs of the East Asian mtDNA tree, Mol. Biol. Evol. 19 (2002) 1737–1751.
[21] Q.-P. Kong, Y.-G. Yao, C. Sun, H.-J. Bandelt, C.L. Zhu, Y.-P. Zhang, Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences, Am. J. Hum. Genet. 73 (2003) 671–676.
[22] M. Tanaka, V.M. Cabrera, A.M. González, J.M. Larruga, T. Takeyasu, N. Fuku, L.J. Guo, R. Hirose, Y. Fujita, M. Kurata, K. Shinoda, K. Umetsu, Y. Yamada, Y. Oshida, Y. Sato, N. Hattori, Y. Mizuno, Y. Arai, N. Hirose, S. Ohta, O. Ogawa, Y. Tanaka, R. Kawamori, M. Shamoto-Nagai, W. Maruyama, H. Shimokata, R. Suzuki, H. Shimodaira, Mitochondrial genome variation in eastern Asia and the peopling of Japan, Genome Res. 14 (2004) 1832–1850.
[23] M. Crespillo, M.R. Paredes, L. Prieto, M. Montesino, A. Salas, C. Albarrán, V. Álvarez-Iglesias, A. Amorin, G. Berniell-Lee, A. Brehm, J.C. Carril, D. Corach, N. Cuevas, A.M. Di Lonardo, C. Doutremepuich, R.M. Espinheira, M. Espinoza, F. Gómez, A. González, A. Hernández, M. Hidalgo, M. Jimenez, F.P. Leite, A.M. López, M. López-Soto, J.A. Lorente, S. Pagano, A.M. Palacio, J.J. Pestano, M.F. Pinheiro, E. Raimondi, M.M. Ramón, F. Tovar, L. Vidal-Rioja, M.C. Vide, M.R. Whittle, J.J. Yunis, J. García-Hirschfel, Results of the 2003–2004 GEP-ISFG collaborative study on mitochondrial DNA: focus on the mtDNA profile of a mixed semen-saliva stain, Forensic Sci. Int. 160 (2006) 157–167.
[24] R.H. Ward, B.L. Frazier, K. Dew-Jager, S. Pääbo, Extensive mitochondrial diversity within a single Amerindian tribe, Proc. Natl. Acad. Sci. USA 88 (1991) 8720–8724.

[25] L. Vigilant, R. Pennington, H. Harpending, T.D. Kocher, A.C. Wilson, Mitochondria DNA sequences in single hairs from a southern African population, Proc. Natl. Acad. Sci. USA 86 (1989) 9350–9354.
[26] A. Brandstätter, C.T. Peterson, J.A. Irwin, S. Mpoke, D.K. Koech, W. Parson, T.J. Parsons, Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database, Int. J. Legal Med. 118 (2004) 294–306.
[27] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 147.
[28] Á. Carracedo, W. Bar, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tully, M. Wilson, DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2000) 79–85.
[29] A. Salas, Á. Carracedo, V. Macaulay, M. Richards, H.-J. Bandelt, A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics, Biochem. Biophys. Res. Commun. 335 (2005) 891–899.
[30] Y.-G. Yao, A. Salas, C.M. Bravi, H.-J. Bandelt, A reappraisal of complete mtDNA variation in East Asian families with hearing impairment, Hum. Genet. 119 (2006) 505–515.
[31] Y.-G. Yao, V. Macaulay, T. Kivisild, Y.-P. Zhang, H.-J. Bandelt, To trust or not to trust an idiosyncratic mitochondrial data set, Am. J. Hum. Genet. 72 (2003) 1341–1346, author's reply 1346–1349.
[32] A. Salas, L. Prieto, M. Montesino, C. Albarrán, E. Arroyo, M.R. Paredes-Herrera, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A.G. de la Vega, C. Alves, C.M. López, M. López-Soto, J.A. Lorente, A. Picornell, R.M. Espinheira, A. Hernández, A.M. Palacio, M. Espinoza, J.J. Yunis, A. Pérez-Lezaun, J.J. Pestano, J.C. Carril, D. Corach, M.C. Vide, V. Álvarez-Iglesias, M.F. Pinheiro, M.R. Whittle, A. Brehm, J. Gómez, Mitochondrial DNA error prophylaxis: assessing the causes of errors in the GEP'02–03 proficiency testing trial, Forensic Sci. Int. 148 (2005) 191–198.
[33] H.-J. Bandelt, A. Salas, S. Lutz-Bonengel, Artificial recombination in forensic mtDNA population databases, Int. J. Legal Med. 118 (2004) 267–273.
[34] H.-J. Bandelt, A. Salas, C.M. Bravi, Problems in FBI mtDNA database, Science 305 (2004) 1402–1404.
[35] H.-J. Bandelt, L. Quintana-Murci, A. Salas, V. Macaulay, The fingerprint of phantom mutations in mitochondrial DNA data, Am. J. Hum. Genet. 71 (2002) 1150–1160.
[36] H.-J. Bandelt, Q.-P. Kong, W. Parson, A. Salas, More evidence for non-maternal inheritance of mitochondrial DNA? J. Med. Genet. 42 (2005) 957–960.
[37] H.-J. Bandelt, A. Achilli, Q.-P. Kong, A. Salas, S. Lutz-Bonengel, C. Sun, Y.-P. Zhang, A. Torroni, Y.-G. Yao, Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies, Biochem. Biophys. Res. Commun. 333 (2005) 122–130.
[38] E.-L. Loogväli, U. Roostalu, B.A. Malyarchuk, M.V. Derenko, T. Kivisild, E. Metspalu, K. Tambets, M. Reidla, H.-V. Tolk, J. Parik, E. Pennarun, S. Laos, A. Lunkina, M. Golubenko, L. Barać, M. Peričić, O.P. Balanovsky, V. Gusar, E.K. Khusnutdinova, V. Stepanov, V. Puzyrev, P. Rudan, E.V. Balanovska, E. Grechanina, C. Richard, J.P. Moisan, A. Chaventré, N.P. Anagnou, K.I. Pappa, E.N. Michalodimitrakis, M. Claustres, M. Gölge, I. Mikerezi, E. Usanga, R. Villems, Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia, Mol. Biol. Evol. 21 (2004) 2012–2021.
[39] T. Kivisild, P. Shen, D.P. Wall, B. Do, R. Sung, K.K. Davis, G. Passarino, P.A. Underhill, C. Scharfe, A. Torroni, R. Scozzari, D. Modiano, A. Coppa, P. de Knjiff, M.W. Feldman, L.L. Cavalli-Sforza, P.J. Oefner, The role of selection in the evolution of human mitochondrial genomes, Genetics 172 (2006) 373–387.
[40] Y.-G. Yao, Q.-P. Kong, H.-J. Bandelt, T. Kivisild, Y.-P. Zhang, Phylogeographic differentiation of mitochondrial DNA in Han Chinese, Am. J. Hum. Genet. 70 (2002) 635–651.
[41] Y.G. Yao, Y.P. Zhang, Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal, J. Hum. Genet. 47 (2002) 311–318.
[42] Q.P. Kong, Y.G. Yao, C. Sun, C.L. Zhu, L. Zhong, C.Y. Wang, W.W. Cai, X.M. Xu, A.L. Xu, Y.P. Zhang, Phylogeographic analysis of mitochondrial DNA haplogroup F2 in China reveals T12338C in the initiation codon of the ND5 gene not to be pathogenic, J. Hum. Genet. 49 (2004) 414–423.
[43] Q.-P. Kong, Y.-G. Yao, M. Liu, S.P. Shen, C. Chen, C.L. Zhu, M.G. Palanichamy, Y.-P. Zhang, Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China, Hum. Genet. 113 (2003) 391–405.
[44] Y.G. Yao, Q.P. Kong, C.Y. Wang, C.L. Zhu, Y.P. Zhang, Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in China, Mol. Biol. Evol. 21 (2004) 2265–2280.
[45] Y.G. Yao, X.M. Lu, H.R. Luo, W.H. Li, Y.P. Zhang, Gene admixture in the silk road region of China: evidence from mtDNA and melanocortin 1 receptor polymorphism, Genes Genet. Syst. 75 (2000) 173–178.
[46] Y.G. Yao, L. Nie, H. Harpending, Y.X. Fu, Z.G. Yuan, Y.P. Zhang, Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity, Am. J. Phys. Anthropol. 118 (2002) 63–76.
[47] Y.G. Yao, W.S. Watkins, Y.P. Zhang, Evolutionary history of the mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of east and southeast Asia, Hum. Genet. 107 (2000) 504–512.
[48] Y.-G. Yao, Q.P. Kong, X.Y. Man, H.-J. Bandelt, Y.-P. Zhang, Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA, Mol. Biol. Evol. 20 (2003) 214–219.
[49] A.J. Redd, M. Stoneking, Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations, Am. J. Hum. Genet. 65 (1999) 808–828.
[50] H.Y. Lee, J.E. Yoo, M.J. Park, U. Chung, K.J. Shin, Mitochondrial DNA control region sequences in Koreans: identification of useful variable sites and phylogenetic analysis for mtDNA data quality control, Int. J. Legal Med. 120 (2006) 5–14.
[51] K.L. Monson, K.W.P. Miller, M.R. Wilson, J.A. DiZinno, B. Budowle, The mtDNA population database: an integrated software and database resource for forensic comparison, Forensic Sci. Commun. 4 (2002), no. 2.
[52] S. Finnilä, M.S. Lehtonen, K. Majamaa, Phylogenetic network for European mtDNA, Am. J. Hum. Genet. 68 (2001) 1475–1484.
[53] M. Santos, R. Barrantes, D-loop mtDNA deletion as a unique marker of Chibchan Amerindians, Am. J. Hum. Genet. 55 (1994) 413–414.
[54] C.J. Kolman, E. Bermingham, R. Cooke, R.H. Ward, T.D. Arias, F. Guionneau-Sinclair, Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá, Genetics 140 (1995) 275–283.
[55] A. Achilli, C. Rengo, C. Magri, V. Battaglia, A. Olivieri, R. Scozzari, F. Cruciani, M. Zeviani, E. Briem, V. Carelli, P. Moral, J.M. Dugoujon, U. Roostalu, E.L. Loogvali, T. Kivisild, H.-J. Bandelt, M. Richards, R. Villems, A.S. Santachiara-Benerecetti, O. Semino, A. Torroni, The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool, Am. J. Hum. Genet. 75 (2004) 910–918.
[56] A. Brandstätter, T.J. Parsons, W. Parson, Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups, Int. J. Legal Med. 117 (2003) 291–298.
[57] C.M. Bravi, W. Parson, H.-J. Bandelt, Numts Revisited, Springer-Verlag, Berlin/Heidelberg, 2006.
[58] A. Salas, H.-J. Bandelt, 'Fishing' numts using mutational spectrometry, Mutat. Res., in preparation.
[59] Y.G. Yao, C.M. Bravi, H.J. Bandelt, A call for mtDNA data quality control in forensic science, Forensic Sci. Int. 141 (2004) 1–6.
[60] A. Alonso, A. Salas, C. Albarrán, E. Arroyo, A. Castro, M. Crespillo, A.M. di Lonardo, M.V. Lareu, C.L. Cubria, M.L. Soto, J.A. Lorente, M.M. Semper, A. Palacio, M. Paredes, L. Pereira, A.P. Lezaun, J.P. Brito, A. Sala, M.C. Vide, M. Whittle, J.J. Yunis, J. Gómez, Results of the 1999–2000 collaborative exercise and proficiency testing program on mitochondrial DNA of the GEP-ISFG: an inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor, Forensic Sci. Int. 125 (2002) 1–7.
[61] L. Prieto, M. Montesino, A. Salas, A. Alonso, C. Albarrán, S. Álvarez, M. Crespillo, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A.G. de la Vega, L. Gusmão, C.M. López, M. López-Soto, J.A. Lorente, M. Malaghini, C.A. Martínez, N.M. Modesti, A.M. Palacio, M. Paredes,

S.D. Pena, A. Pérez-Lezaun, J.J. Pestano, J. Puente, A. Sala, M. Vide, M.R. Whittle, J.J. Yunis, J. Gómez, The 2000–2001 GEP-ISFG Collaborative Exercise on mtDNA: assessing the cause of unsuccessful mtDNA PCR amplification of hair shaft samples, Forensic Sci. Int. 134 (2003) 46–53.

[62] A. Salas, M.V. Lareu, A. Carracedo, Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report, Int. J. Legal Med. 114 (2001) 186–190.

[63] G. Tully, S.M. Barritt, K. Bender, E. Brignon, C. Capelli, N. Dimo-Simonin, C. Eichmann, C.M. Ernst, C. Lambert, M.V. Lareu, B. Ludes, B. Mevag, W. Parson, H. Pfeiffer, A. Salas, P.M. Schneider, E. Staalstrom, Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts, Forensic Sci. Int. 140 (2004) 1–11.

[64] H.-J. Bandelt, P. Lahermo, M. Richards, V. Macaulay, Detecting errors in mtDNA data by phylogenetic analysis, Int. J. Legal Med. 115 (2001) 64–69.

Artículo 6: Minisequencing mitochondrial DNA pathogenic mutations

Álvarez-Iglesias V, Barros F, Carracedo A, Salas A

(2008) *BMC Med Genet* 9:26

Minisecuenciación de mutaciones patogénicas de ADN mitocondrial

ABSTRACT/RESUMEN:

Antecedentes: Existe un número de mutaciones de ADN mitocondrial (ADNmt) bien conocidas responsables de enfermedades mitocondriales comunes. Par superar los problemas técnicos relacionados con el análisis de genomas completos de ADNmt se han propuesto una gran variedad de técnicas distintas para poder explorar las mutaciones patogénicas de la región codificante.

Métodos: Proponemos un ensayo de minisecuenciación para el análisis de mutaciones de ADNmt. En una única reacción interrogamos un total de 25 mutaciones patogénicas distribuidas por todo el genoma mitocondrial en una muestra de pacientes sospechosos de padecer una enfermedad mitocondrial.

Resultados: Hemos detectado 11 mutaciones homoplásmicas causales en pacientes sospechosos de la enfermedad de Leber, que fueron confirmadas mediante secuenciación automática estándar. Las mutaciones m.11778G>A y m.14484T>C ocurren con una mayor frecuencia de lo esperado en pacientes de Galicia (noroeste de España) con linajes de haplogrupo J (Test exacto de Fisher, valor de $P < 0.01$). El ensayo funciona bien en experimentos de mezclas de ADNs salvaje: mutante que simulan situaciones heteroplásmicas de muchas enfermedades de ADN mitocondrial.

Conclusiones: Hemos desarrollado un sistema de genotipado mediante minisecuenciación para la búsqueda de las mutaciones patogénicas de ADNmt más frecuentes, que es rápido y a bajo coste. La técnica es robusta y reproducible y puede ser fácilmente implementada en laboratorios clínicos estándar.

BMC Medical Genetics

BioMed Central

Technical advance

Open Access

Minisequencing mitochondrial DNA pathogenic mutations

Vanesa Álvarez-Iglesias[1,2], Francisco Barros[2], Ángel Carracedo[1,2] and Antonio Salas*[1]

Address: [1]Unidade de Xenética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, Galicia, Spain and [2]Fundación Pública Galega de Medicina Xenómica (FPGMX), Hospital Clínico Universitario, Universidad de Santiago de Compostela, Galicia, Spain

Email: Vanesa Álvarez-Iglesias - vaneiml@usc.es; Francisco Barros - apimlbar@usc.es; Ángel Carracedo - apimlang@usc.es; Antonio Salas* - apimlase@usc.es

* Corresponding author

Published: 10 April 2008

BMC Medical Genetics 2008, **9**:26 doi:10.1186/1471-2350-9-26

Received: 26 September 2007

Accepted: 10 April 2008

This article is available from: http://www.biomedcentral.com/1471-2350/9/26

Abstract

Background: There are a number of well-known mutations responsible of common mitochondrial DNA (mtDNA) diseases. In order to overcome technical problems related to the analysis of complete mtDNA genomes, a variety of different techniques have been proposed that allow the screening of coding region pathogenic mutations.

Methods: We here propose a minisequencing assay for the analysis of mtDNA mutations. In a single reaction, we interrogate a total of 25 pathogenic mutations distributed all around the whole mtDNA genome in a sample of patients suspected for mtDNA disease.

Results: We have detected 11 causal homoplasmic mutations in patients suspected for Leber disease, which were further confirmed by standard automatic sequencing. Mutations m.11778G>A and m.14484T>C occur at higher frequency than expected by change in the Galician (northwest Spain) patients carrying haplogroup J lineages (Fisher's Exact test, *P*-value < 0.01). The assay performs well in mixture experiments of wild:mutant DNAs that emulate heteroplasmic conditions in mtDNA diseases.

Conclusion: We here developed a minisequencing genotyping method for the screening of the most common pathogenic mtDNA mutations which is simple, fast, and low-cost. The technique is robust and reproducible and can easily be implemented in standard clinical laboratories.

Background

There are over 100 point mutations putatively associated with human mtDNA diseases [1]; however, only a small percentage of these mutations were properly confirmed http://www.mitomap.org/rimtab2.html. In fact, the pathogenicity of some mutations is still under question due to the fact that the conclusions claimed in a substantial number of studies do not rest on solid grounds or are partially or completely flawed [2-5].

During this last two decades, many different techniques were designed for the screening of mtDNA pathogenic mutations, including restriction fragment length polymorphism (RFLP) analysis, heteroduplex analysis (HDA), single strand conformation polymorphisms (SSCP), etc [6-11]. The main disadvantage of these classical screening methods is that, due to the intrinsic nature of the techniques, many mutations usually pass unnoticed. Novel and more efficient strategies have recently been proposed

BMC Medical Genetics 2008, **9**:26 http://www.biomedcentral.com/1471-2350/9/26

that allow the screening of entire mtDNA genomes; for instance, the mismatch-specific DNA endonuclease "Surveyor™ Nuclease" seems to be particularly useful when the most common pathogenic mutations have been ruled out previously (e.g. [12]). Ideally, the gold standard of sequencing the whole genome would solve the problem; nevertheless, this strategy is labor-intensive and therefore unfeasible for routine molecular diagnosis (taking into account the capacities of standard laboratories). Screening methods are also difficult to standardize and some of them are not suitable for replication assays due to their dependence on for instance electrophoretic conditions.

Minisequencing techniques have demonstrated to be efficient in other fields of research, such as forensic and population genetics [13-18], but also in clinical research [19]. This approach, also known as single nucleotide primer extension, allows the determination of the polymorphic position by DNA polymerase addition of the ddNTP complementary to the base interrogated. The primer anneals therefore to its target DNA immediately adjacent to the SNP under scrutiny. Here we present a minisequencing multiplex design which is a rapid, sensitive and low cost assay for screening mtDNA pathogenic mutations in patients suffering mtDNA disease.

Methods

Subject and SNP selection

Blood sample of 15 patients was sent to the Fundación Pública Galega de Medicina Xenómica located in the Hospital Clínico of Santiago de Compostela (Galicia, Spain). Six of these patients are from Galicia (northwest Spain; [20]), and three other patients are from southeast Spain (in the Mediterranean coast); the regional Spanish provenance of the other two patients is unknown. Four of these patients presented clinical features of an undetermined neuropathy, while the rest of the patients were analyzed under suspicion of Leber disease. Written informed consent was obtained from all patients and the study was approved by the institutional review board of the University of Santiago de Compostela (Spain). All the samples were genotyped for a set of 25 variants, 21 are confirmed pathogenic mutations related to different mtDNA diseases such as MELAS, LHON, Leigh Disease, and NARP. These mutations are distributed all around the mtDNA molecule and 'hit' different mtDNA genes (Table 1). All the selected mutations, with the only exception of

Table 1: Characteristics of the mutations incorporated in our multiplex design

POSITIONS	Base change	Amino Acid change	Locus	Disease[1]
3243	A>G	tRNA Leu (UUR)	MT-TL1	MELAS, DM/DMDT, CPEO, MM
3460	G>A	Ala>Thr	MT-ND1	LHON
3697	G>A	Gly>Arg	MT-ND1	MELAS
3946	G>A	Glu>Lys	MT-ND1	MELAS
3949	T>C	Tyr>Lys	MT-ND1	MELAS
7445	A>G	Ter>Ter	MT-CO1	SNHL
7445	A>C	Ter>Ser	MT-CO1	DEAF
8993	T>G	Leu>Arg	MT-ATP6	NARP
8993	T>C	Leu>Pro	MT-ATP6	NARP, Leigh Disease
9176	T>C	Leu>Pro	MT-ATP6	FBSN, Leigh Disease
9176	T>G	Leu>Arg	MT-ATP6	Leigh Disease
10158	T>C	Arg>Pro	MT-ND3	Leigh Disease
10191	T>C	Arg>Pro	MT-ND3	ESOC, Leigh-like Disease
10663	T>C	Val>Ala	MT-ND4	LHON
11777	C>A	Arg>Ser	MT-ND4	Leigh Disease
11778	G>A	Arg>His	MT-ND4	LHON
11832	G>A	Trp>Ter	MT-ND4	Exercise Intolerance
12706	T>C	Phe>Leu	MT-ND5	Leigh Disease
13513	G>A	Asp>Asn	MT-ND5	MELAS, Leigh Disease
13514	A>G	Asp>Gly	MT-ND5	MELAS
14459	G>A	Ala>Val	MT-ND6	LYDT, Leigh Disease
14482	C>A	Met>Ile	MT-ND6	LHON
14482	C>G	Met>Ile	MT-ND6	LHON
14484	T>C	Met>Val	MT-ND6	LHON
14487	T>C	Met>Val	MT-ND6	Dystonia, Leigh Disease

[1] List of some diseases where these mutations are frequently observed: CPEO: Chronic Progressive External Ophthalmoplegia; DEAF Maternally inherited DEAFness or aminoglycoside-induced DEAFness DM: Diabetes Mellitus; LDYT: Leber's hereditary optic neuropathy and DysTonia; LHON: Leber Hereditary Optic Neuropathy; MELAS: Mitochondrial Encephalomyopathy, Lactic Acidosis, and Stroke-like episodes; MM: Mitochondrial Myopathy; NARP: Neurogenic muscle weakness, Ataxia, and Retinitis Pigmentosa; alternate phenotype at this locus is reported as Leigh Disease; SNHL: SensoriNeural Hearing Loss

m.3243A>G, consist of non-synonymous substitutions (both transitions and transversion, but note that the minisequencing technique could be also suitable for interrogating indels [17]).

Primer design

The primers both for PCR amplification (Table 2) and minisequencing reaction (Table 3) were designed to have an annealing temperature around 60°C using Primer3 software http://frodo.wi.mit.edu/cgi-bin/primer3. The sequence databases at the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) were interrogated using the online BLAST tool to test the primers against possible repetitive sequences and sequence homologies in the autosomal genome. Each primer pair for PCR amplification and each single base extension primer were selected independently and AutoDimer http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm was used to test for potential hairpin structures and primer-dimer problems.

Although the simultaneous occurrence of more than one pathogenic mutation in a single individual is infrequent, it could be possible that the presence of m.14484T>C could interfere with the genotyping of m.14487T>G (due to their physical proximity). Therefore, in order to avoid potential artifacts due to deficient annealing, we have designed a degenerate extension primer containing a mixture of nucleotide C and T at site 14484. The same rationale applies to variation at other two positions. According to our design, m.14482C>G mutation could interfere with the genotyping at position 14484; and the presence of m.3946G>A could alter the genotyping of m.3949T>C; we therefore designed another extension degenerate primer containing C and G at site 14482 and another additional extension degenerated primer carrying G and A at site 3946 (Table 3).

PCR multiplex amplification

The SNPs were PCR amplified in 13 amplicons with sizes ranging from 89 and 196 bp. The amplicons are deliberately designed to be small in order to facilitate the analysis of samples that are highly degraded or with low quantity of DNA. Some amplicons encompass several SNPs (Table 1). We performed multiplex reaction using 5 ng of DNA template and PCR master mix of QIAGEN Kit Multiplex PCR (Qiagen, Hilden, Germany), amplification primers and their final concentrations are in Table 2. Amplification was carried out in a 9700 Thermocycler (Applied Biosystems, Foster City, CA, USA). After a 95°C preincubation step for 15 minutes, PCR was performed in a total of 30 cycles using the following conditions: 94°C denaturation for 30 seconds, annealing at 60°C for 90 seconds and extension at 72°C during 90 seconds, followed by a 15 minutes of final extension at 72°C and 4°C until removed from thermocycler. PCR products were checked by polyacrylamide gel electrophoresis (T9, C5) visualized by silver staining.

Minisequencing reaction

Previous to minisequening reaction, PCR products are treated with ExoSAP-IT (Amershan Bioscences, Uppsala, Sweden) to remove excess primers and un-incorporated dNTPs: 3 μl of PCR product was incubated with 1.5 μl for 15 minutes at 37°C followed by 15 minutes at 80°C for enzyme inactivation. Minisequencing reaction is performed using SNaPshot™ Kit (AB). We modified the length of the primers (between 25 and 76 bp) by the addition of non-homologous tails, poly(dGACT) added at the 5'- end (Table 3). The minisequencing reaction was carried out in a total volume of 10 μl comprising 3 μl of the SNaPshot™ Kit (AB), 1.5 μl PCR product, 1.5 μl of extension primers mix (final concentrations are between 0.1 and 0.6 μM) (Table 3), and water up to 10 μl. The reaction was performed in a 9700 Thermocycler (AB) following the recommendations of the manufacturer: 25 cycles of dena-

Table 2: Amplification primers

POSITIONS	FORWARD	REVERSE	SIZE (bp)	Final Concentration (μM)
3243	tatacccacacccacccaag	ggccatgggtatgttgttaag	118	0.2
3460	ccgaacgaaaaattctaggc	gcggtgatgtagagggtgat	153	0.2
3697	gcctagccgtttactcaatcc	tgagattgtttggggctactgc	94	0.15
3946/3949[1]	tagcagagaccaaccgaacc	gaagattgtagtggtgagggtgt	157	0.2
7445	ccctaccacacattcgaagaa	tggcttgaaaccagctttg	89	0.2
8993	aatgccctagcccacttctt	aggtggcctgcagtaatgtt	140	0.15
9176	aaatcgctgtcgccttaatc	tcattaggagggctgagagg	154	0.3
10158/10191	tcaacaccctcctagcctta	gggtaaaaggagggcaattt	196	0.3
10663	acacccactcccctcttagcc	ggccatatgtgttggagattg	110	0.3
11777/11778/11832	cacgggcttacatcctcatt	gggggtaaggcgaggttag	157	0.2
12706[1]	tgtagcattgttcgttacatgg	agttggaataggttgttagcgg	146	0.2
13513/13514	attggcagcctagcattagc	cagggaggtagcgatgagag	131	0.2
14459/14482/14484/14487	ctccatcgctaaccccacta	ttctgaattttgggggaggt	170	0.4

[1]These primers were reported in [16]

BMC Medical Genetics 2008, **9**:26 http://www.biomedcentral.com/1471-2350/9/26

Table 3: Minisequencing primers

POSITIONS	Extension Primer[1]	Length	Base change	Strand	Final Concentration (μM)[3]
10158	acaactcaacggctacatagaaaaa	25	T>C	L	0.2
3946	Cgaactagtctcaggctttcaacatc	25	G>A	L	0.2
10663	GACTGcaatattgtgcctattgccatactag	31	T>C	L	0.2
3460	$(GACT)_2$Ggctactacaacctttcgctgac	31	G>A	L	0.2
10191	$(GACT)_4$agtgcggcttcgaccctata	36	T>C	L	0.3
14459	$(GACT)_2$GACctcaggatactcctcaatagccatc	36	G>A	L	0.3
14484[2]	$(GACT)_3$atcgctgtagtatatccaaagacaacYa	40	T>C	L	0.5
3243	$(GACT)_4$GAacagggtttgttaagatggcag	40	A>G	L	0.2
11777	$(GACT)_5$Gcaaactacgaacgcactcacagt	44	C>A	L	0.2
7445	$(GACT)_6$tcgaagaacccgtatacataaaaatctag	48	A>G/C	L	0.2
12706	$(GACT)_9$gcggtaactaagattagtatggtaattagga	52	T>C	H	0.2
3949[2]	$(GACT)_{11}$GAagtctcaggcttcaacatcRaa	52	T>C	L	0.2
14487[2]	$(GACT)_7$GctgtagtatatccaaagacaaccaSca	56	T>C	L	0.6
13513	$(GACT)_8$ttcctcacaggtttctactccaaa	56	G>A	L	0.1
13514	$(GACT)_{10}$Ttgcggtttcgatgatgtgg	60	A>G	H	0.2
3697	$(GACT)_9$CTaaactcaaactacgccctgatc	60	G>A	L	0.2
11778	$(GACT)_9$GAgaagtccttgagagaggattatgatg	64	G>A	H	0.3
11832	$(GACT)_8$ GACtcaaactctactcccactaatagtttttt	64	G>A	L	0.1
9176	$(GACT)_{11}$Gatccaagcctacgttttcacacttc	70	T>C/G	L	0.2
14482	$(GACT)_{11}$gccatcgtcgctgtagtatatccaaagacaac	73	C>A/G	L	0.6
8993	$(GACT)_{13}$Gcctactcattcaaccaatagccc	76	T>G/C	L	0.2

[1] Capital letters indicate the segment of the primer belonging to the tail
[2] These primers are 'degenerated' at the indicated position in the second column (see Material and Methods) using the IUB code

turation at 96°C for 10 seconds, annealing at 50°C for 5 seconds and extension at 60°C during 30 seconds. Unincorporated ddNTPs are eliminated with a treatment with SAP (Amershan Bioscences). The final volume (10 μl) was treated with 1 μl of SAP for 60 minutes at 37°C followed by 15 minutes at 80°C for enzyme inactivation.

The minisequencing products (1.5 μl) were mixed with 10 μl of HiDi™ formamide (AB) and 0.25 μl of GeneScan - 120 LIZ (AB) and capillary electrophoresis was undertaken on an ABI PRISM 3130 ×l Genetic Analyzer (AB). The data was analyzed using GeneMapper™ 3.7 Software (AB).

Sequencing reaction

All the 13 amplicons were sequenced in an ABI PRISM 3130 ×l Genetic Analyzer (AB) for all the patients with the aim of corroborating the mutations observed with the minisequencing assay. Technical details regarding sequencing reaction are given in [17].

Results

Only 13 amplicons are needed to interrogate the selected 25 variants. This is because (i) 'neighbouring' variants are genotyped using different minisequencing probes annealing at the same amplicon, and (ii) different minisequencing probes were designed to genotype different pathogenic substitutions occurring at the same mtDNA site. Therefore, for each of the sites 7445, 9176, and 14482, there are two common mtDNA mutations described in the literature as pathogenic, the most common one for each of these three sites was already confirmed (m.7445A>G, m.9176T>C, and m.14482C>G), while the alternative one (m.7445A>C, m.9176T>G, and m.14482C>A) still bears the status of 'provisional' in MITOMAP. In addition, position 8993 has two already confirmed variants, namely, T to C and T to G substitutions.

We did not observe mutations in four patients presenting neurological symptoms. On the other hand, a pathogenic mutation was found for each of the other 11 patients with clinical suspicion of Leber disease (Table 4). All the mutations were confirmed by sequencing and were observed as homoplasmic using both minisequencing and automatic sequencing (Figure 1). Nine patients carried mutation m.11778G>A whereas two patients carried m.14484T>C.

The two carriers of m.14484T>C lived in the same geographic (relatively isolated) village (namely, Burela) in the northern cost of Galicia [20,21]. These samples were submitted at different times to the laboratory and according to the carriers, their families were unrelated. Both patients share also the transition m.12696T>C (observed by sequencing), indicating that these individuals could actually share some recent common ancestor. The m.12696T>C transition is not a frequent variant in Europe and likely constitutes a diagnostic variant of a minor HV1 sub-lineage (although this variant is also recurrent in at least other eight non-European haplogroups, including

BMC Medical Genetics 2008, **9**:26 http://www.biomedcentral.com/1471-2350/9/26

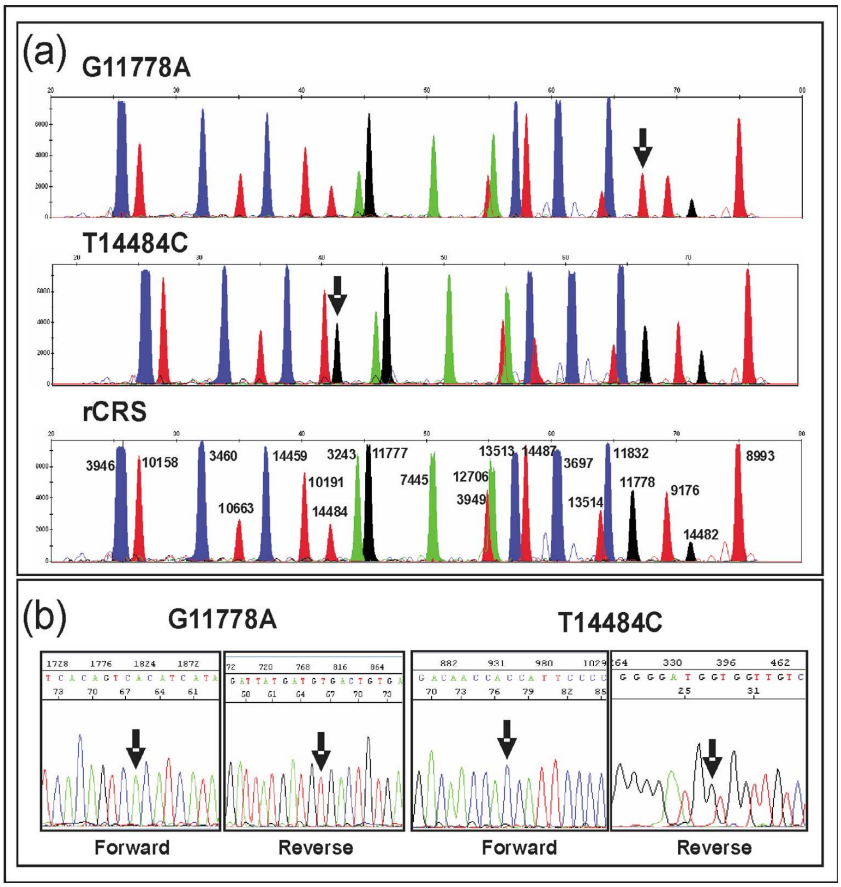


Figure 1
Electropherograms showing (a) two different SNaPshot profiles carrying mutations m.11778G>A and m.14484T>C; rCRS [50] electropherogram is also shown indicating the whole set of mutations tested with the SNaPshot reaction; and (b) forward and reverse sequence electropherograms for these same mutations.

haplogroup J, see below): it appears in at least six haplogroup HV1 complete genomes (see for instance [22,23]). Therefore, the m.14484T>C transition could represent a founder pathogenic mutation responsible of Leber disease in Galicia (as it is the case of a wide spectrum of different non-mtDNA diseases [24,25].

BMC Medical Genetics 2008, **9**:26 http://www.biomedcentral.com/1471-2350/9/26

Table 4: Patient samples analyzed in the present study

		3243	3460	3697	3946	3949	7445	8993	9176	10158	10191	10663	11777	11778	11832	12706	13513	13514	14459	14482	14484	14487	Variants[1]
	rCRS	A	G	G	G	T	A	T	T	T	T	T	C	G	G	T	G	A	G	T	T	T	
#1	I-111	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	
#2	D-112	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	11812
#3	I-138	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	10172
#4	F-393	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	
#5	C-436	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
#6	B-703	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
#7	B-900	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	
#8	B-901	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	
#9	D-992	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
#10	1586	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	
#11	5699	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	12696
#12	5869	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	-	12696
#13	G-495	-	-	-	-	-	-	-	-	-	-		-	A	-	-	-	-	-	-	-	-	
#14	7218	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
#15	9615	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	12705C/T

[1] Variants detected by automatic sequencing (putatively unrelated to the disease outcome). Position 12705 is heteroplasmic C/T.

In order to further investigate this hypothesis, we additionally genotyped all the Galician samples for a set of haplogroup diagnostic mtDNA SNPs following [16] (Table 5). The two carriers of m.14484T>C but also two other Galician samples carrying m.11778G>A could be allocated to haplogroup J (and not to HV1 as it could be inferred by the presence of m.12696T>C). Therefore, the observation of the m.14484T>C mutation in two unrelated patients from the above mentioned small Galician village could just reflect the already reported high incidence of Leber mutations within haplogroup J [26,27]. In fact, we also observe that the frequency of haplogroup J in healthy unrelated Galician individuals is ~14.7% [20,21,28] but it occurs in four out of our six Galician patients (~66.7%); the difference in prevalence is statistically significant (Fisher's Exact test, *P*-value < 0.006)

Other variants were also observed by sequencing analysis in m.11778G>A carriers. For instance, the transition A11812G is diagnostic of haplogroup T2 and G10172A identifies haplogroup J2b; these variants also appear sporadically in other haplogroup backgrounds. In addition, we observe a heteroplasmy at position 12705; transition C12705T is a well-known diagnostic site that (together with T16223C) leads from macro-haplogroup N to R.

Mixtures of two different DNA bearing different SNP profiles were carried out in order to simulate mtDNA heteroplasmy, a common state in mtDNA disease patients. As

Table 5: Coding SNP and haplogroup status of the Galician patients

		3915	3992	4216	4529	4580	4769	4793	6776	7028	10398	10400	10463	10873	12308	12705	13966	14766	HG
	rCRS	G	C	T	A	G	A	A	T	C	A	C	T	T	A	C	A	C	
#3	I-138	-	-	C	-	-	-	-	-	T	G	-	-	-	-	-	-	T	J
#4	F-393	-	-	C	-	-	-	-	-	T	G	-	-	-	-	-	-	T	J
#10	1586	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	H3
#11	5699	-	-	C	-	-	-	-	-	T	G	-	-	-	-	-	-	T	J
#12	5869	-	-	C	-	-	-	-	-	T	G	-	-	-	-	-	-	T	J
#13	G-495	-	-	-	-	-	-	-	-	T	-	-	-	-	G	-	-	T	U

HG = haplogroup

BMC Medical Genetics 2008, **9**:26 http://www.biomedcentral.com/1471-2350/9/26

shown in Figure 2, the minisequencing assay is perfectly able to detect the mixtures at different proportions. According to [29] "*the single most important disadvantage of the SNaPshot is that it seems less accurate than conventional PCR-RFLP analysis for* [detecting] *high levels of mutant mtDNA*". Our results however indicate that the ability of SNaPshot for detecting heteroplasmy is actually equivalent to standard sequencing. Some fluorocromes 'project' more intensity in the electropherogram than other (as it also occurs with some sequencing chemistries), and this should be taking into account when quantifying the real proportions of wild:mutant DNA (heteroplasmy).

Although our primer design take into account the proximity of two pathologic mutations, it is not possible to consider all the potential polymorphisms that could hit within the annealing segments, and therefore alter the PCR efficacy (and in the worst of the cases, produce an artefact related to e.g. the spurious amplification of a NUMT; see below). This is however an universal problem affecting most of the genotyping techniques (including RFLP and sequencing) and therefore, the expert should be always alert under the possibility of false negative results. The experience of previous SNaPshot designs [16,18,30] indicates however that the method is robust under the presence of polymorphisms.

Finally, the multiplex reaction was also tested in 102 control individuals. There were not false positives in this control sample and the SNP-minisequencing profiles were phylogenetically concordant [31,32].

Discussion

MtDNA mutations are responsible for various clinical features, which often make diagnosis a considerable challenge. Since the mitochondria constitute the energetic 'factory' of all nucleated cells, mtDNA diseases affect many tissues with variable clinical outcomes. When searching the Mitomap database for pathogenic mtDNA mutations, many of them harbour the status of 'provisional', while some other stated as 'confirmed'; among the later, some were recently questioned [2,33,34]. We here selected a total of 25 mutations, 21 of them were confirmed as pathogenic in common mtDNA diseases in multiple independent studies and considering different criteria, including the score system proposed by [33] (see also [35]) and also the complementary phylogenetic approach [2]. For instance, we included all the MTND gene mutations evaluated with the complex I pathogenicity scoring system of [33] (see their Table 2) and reaching the 16 top values (≥ 30). Apart from the 21 confirmed mutations included in the design, the remaining four mutations were additionally included because they occur at lower frequencies at four of the initially 21 selected positions.

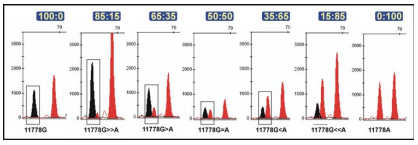


Figure 2
Partial electropherograms showing the performance of the SNaPshot assay with mixtures of wild:mutant DNAs. Numbers in the top indicate the percentages of wild:mutant mixture.

Among the selected mutations, we have the transition m.3243A>G which is a typical cause of MELAS [36] and MIDD [37]; transition m.8993T>G, responsible of a number of patients suffering NARP [38] and MILS [39]; transitions m.3460G>A, m.11778G>A, m.14484T>C, which are common causal mutations in LHON patients [40,41], etc. (see [42] for a review). It can be tentatively said that the selected SNPs actually cover an ample spectra of the fully confirmed causal mutations in the most common mtDNA diseases. For instance, according to [43], the prevalence of m.3243A>G is about 1 in 6,135 in the general (European) population; see also [44]. It is also worth to mention that it is straightforward to add mutations to our initial multiplex design because SNaPshot minisequencing is very flexible in this regard [17].

Conclusion

The minisequencing reaction presented here detected 11 mutations in an easy and straightforward manner in 11 different patients with clinical suspicion of Leber disease. All the mutations and their homoplasmic status were finally confirmed by automatic sequencing. We did not observed false positives in a substantial number of controls. We also simulated heteroplasmic states by mixing wild:mutant DNAs; the minisequencing technique demonstrates a good performance in detecting at least mixtures up to 1:10.

The multiplex reaction shows various advantages in the clinical field with respect to other techniques (see [17] for more details). For instance, the primers were designed in order to obtain amplicons of size ranging 89–196, a feature of special interest in a clinical context where often the laboratories have to deal with suboptimum samples containing low amounts of DNA or highly degraded DNA (paraphine imbibed samples, biopsies, etc.) as it is also the case with forensic samples [13]. Finally, we estimate that the cost of genotyping 25 SNPs is at least 20 times lower than the cost of RFLP genotyping (considering genotyping of one mutation at the time), and it can be done

in less than three hours (including post-PCR purification, minisequencing reaction, purification of minisequencing products, electrophoresis, and documentation) for as many samples as the number of capillaries of the automatic sequencer.

Also important is the fact that minisequencing is robust respect potential artefacts related to false positives at NUMTs (author's unpublished data). Multiplex genotyping also prevents from artificial recombination; the later is a kind of common artefact that more easily arise as more independent fragments are genotyped for the same sample [31,32,45-49].

Finally, the screening method presented here should not be considered as a substitute of complete genome sequencing. The latter should be performed when the minisequencing screening (or the use of other alternative strategies) fails to identify the causal mutation.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

VAI, FB, AC, and AS designed the study, collected and analyzed the data, and wrote the paper.

Availability and requirements

There are over 100 point mutations putatively associated with human mtDNA diseases [1]; however, only a small percentage of these mutations were properly confirmed http://www.mitomap.org/rimtab2.html. The primers both for PCR amplification (Table 2) and minisequencing reaction (Table 3) were designed to have an annealing temperature around 60°C using Primer3 software http://frodo.wi.mit.edu/cgi-bin/primer3. The sequence databases at the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) were interrogated using the online BLAST tool to test the primers against possible repetitive sequences and sequence homologies in the autosomal genome. Each primer pair for PCR amplification and each single base extension primer were selected independently and AutoDimer http/www.cstl.nist.gov/biotech/strbase/AutoDimer Homepage/AutoDimerProgramHomepage.htm was used to test for potential hairpin structures and primer-dimer problems.

Acknowledgements

The 'Ramón y Cajal' Spanish programme from the Ministerio de Educación y Ciencia (RYC2005-3), the grant from the Xunta de Galicia (PGIDIT06PXIB208079PR), and a grant from the Fundación de Investigación Médica Mutua Madrileña awarded to AS supported this project.

References

1. Schapira AH: **Mitochondrial disease.** *Lancet* 2006, **368(9529):**70-82.
2. Bandelt HJ, Yao YG, Salas A, Kivisild T, Bravi CM: **High penetrance of sequencing errors and interpretative shortcomings in mtDNA sequence analysis of LHON patients.** *Biochem Biophys Res Commun* 2007, **352:**283-291.
3. Yao YG, Salas A, Bravi CM, Bandelt HJ: **A reappraisal of complete mtDNA variation in East Asian families with hearing impairment.** *Hum Genet* 2006, **119(5):**505-515.
4. Bandelt HJ, Yao YG, Salas A: **The search of 'novel' mtDNA mutations in hypertrophic cardiomyopathy: MITOMAPping as a risk factor.** *Int J Cardiol* 2007 in press.
5. Kong QP, Bandelt HJ, Sun C, Yao YG, Salas A, Achilli A, Wang CY, Zhong L, Zhu CL, Wu SF, Torroni A, Zhang YP: **Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations.** *Hum Mol Genet* 2006, **15(13):**2076-2086.
6. Wallace DC, Brown MD, Lott MT: **Mitochondrial DNA variation in human evolution and disease.** *Gene* 1999, **238(1):**211-230.
7. Barros F, Lareu MV, Salas A, Carracedo A: **Rapid and enhanced detection of mitochondrial DNA variation using single-strand conformation analysis of superposed restriction enzyme fragments from polymerase chain reaction-amplified products.** *Electrophoresis* 1997, **18(1):**52-54.
8. Salas A, Rasmussen EM, Lareu MV, Morling N, Carracedo A: **Fluorescent SSCP of overlapping fragments (FSSCP-OF): a highly sensitive method for the screening of mitochondrial DNA variation.** *Forensic Sci Int* 2001, **124(2-3):**97-103.
9. Zeviani M, Gellera C, Antozzi C, Rimoldi M, Morandi L, Villani F, Tiranti V, DiDonato S: **Maternally inherited myopathy and cardiomyopathy: association with mutation in mitochondrial DNA tRNA(Leu)(UUR).** *Lancet* 1991, **338(8760):**143-147.
10. DiMauro S: **Mitochondrial DNA medicine.** *Biosci Rep* 2007, **27(1-3):**5-9.
11. Hudson G, Carelli V, Spruijt L, Gerards M, Mowbray C, Achilli A, Pyle A, Elson J, Howell N, La Morgia C, Valentino ML, Huoponen K, Savontaus ML, Nikoskelainen E, Sadun AA, Salomao SR, Belfort R Jr., Griffiths P, Man PY, de Coo RF, Horvath R, Zeviani M, Smeets HJ, Torroni A, Chinnery PF: **Clinical Expression of Leber Hereditary Optic Neuropathy Is Affected by the Mitochondrial DNA-Haplogroup Background.** *Am J Hum Genet* 2007, **81(2):**228-233.
12. Bannwarth S, Procaccio V, Paquis-Flucklinger V: **Surveyor Nuclease: a new strategy for a rapid identification of heteroplasmic mitochondrial DNA mutations in patients with respiratory chain defects.** *Hum Mutat* 2005, **25(6):**575-582.
13. Crespillo M, Paredes MR, Prieto L, Montesino M, Salas A, Albarrán C, Álvarez-Iglesias V, Amorin A, Berniell-Lee G, Brehm A, Carril JC, Corach D, Cuevas N, Di Lonardo AM, Doutremepuich C, Espinheira RM, Espinoza M, Gómez F, González A, Hernández A, Hidalgo M, Jimenez M, Leite FP, López AM, López-Soto M, Lorente JA, Pagano S, Palacio AM, Pestano JJ, Pinheiro MF, Raimondi E, Ramon MM, Tovar F, Vidal-Rioja L, Vide MC, Whittle MR, Yunis JJ, Garcia-Hirschfel J: **Results of the 2003-2004 GEP-ISFG collaborative study on mitochondrial DNA: focus on the mtDNA profile of a mixed semen-saliva stain.** *Forensic Sci Int* 2006, **160(2-3):**157-167.
14. Brandstätter A, Parsons TJ, Parson W: **Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups.** *Int J Legal Med* 2003, **117(5):**291-298.
15. Bandelt HJ, Kivisild T, Parik J, Villems R, Bravi CM, Yao YG, Brandstätter A, Parson W: **Lab-specific mutation processes.** In *Human mitochondrial DNA and the evolution of Homo sapiens Series: Nucleic Acids and Molecular Biology Volume 18*. Edited by: H.-J. Bandelt MRVM. Berlin-Heidelberg , Springer-Verlag; 2006:119–150.
16. Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo : **Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing.** *Forensic Sci Int* 2004, **140(2-3):**251-257.
17. Álvarez-Iglesias V, Jaime JC, Carracedo , Salas A: **Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups.** *Forensic Sci Int: Genet* 2007, **1:**44-55.
18. Brandstätter A, Salas A, Niederstätter H, Gassner C, Carracedo , Parson W: **Dissection of mitochondrial superhaplogroup H using coding region SNPs.** *Electrophoresis* 2007:2541-2550.

19. Filippini S, Blanco A, Fernandez-Marmiesse A, Alvarez-Iglesias V, Ruiz-Ponte C, Carracedo A, Vega A: **Multiplex SNaPshot for detection of BRCA1/2 common mutations in Spanish and Spanish related breast/ovarian cancer families.** *BMC Med Genet* 2007, **8:**40.
20. Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo : **mtDNA analysis of the Galician population: a genetic edge of European variation.** *Eur J Hum Genet* 1998, **6(4):**365-375.
21. Salas A, Lareu MV, Sánchez-Diz P, Calafell F, Carracedo A: **mtDNA hypervariable region II (HVII) sequences in human evolution studies: impact of mutation rate heterogeneity.** *Progress in Forensic Genetics* 2000, **8:**329-331.
22. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A: **The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool.** *Am J Hum Genet* 2004, **75(5):**910-918.
23. Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N: **Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences from the major African, Asian, and European haplogroups.** *Am J Hum Genet* 2002, **70:**1152-1171.
24. Fernández-Marmiesse A, Salas A, Vega A, Fernández-Lorenzo JR, Barreiro J, Carracedo : **Mutation spectra of ABCC8 gene in Spanish patients with Hyperinsulinism of Infancy (HI).** *Hum Mutat* 2006, **27(2):**214.
25. Brage A, Tomé S, Garcia A, Carracedo , Salas A: **Clinical and molecular characterization of Wilson disease in Spanish patients.** *Hepatol Res* 2007, **37(1):**18-26.
26. Brown MD, Sun F, Wallace DC: **Clustering of Caucasian Leber hereditary optic neuropathy patients containing the 11778 or 14484 mutations on an mtDNA lineage.** *Am J Hum Genet* 1997, **60(2):**381-387.
27. Howell N, Oostra RJ, Bolhuis PA, Spruijt L, Clarke LA, Mackey DA, Preston G, Herrnstadt C: **Sequence analysis of the mitochondrial genomes from Dutch pedigrees with Leber hereditary optic neuropathy.** *Am J Hum Genet* 2003, **72(6):**1460-1469.
28. González AM, Brehm A, Pérez JA, Maca-Meyer N, Flores C, Cabrera VM: **Mitochondrial DNA affinities at the Atlantic fringe of Europe.** *Am J Phys Anthropol* 2003, **120(4):**391-404.
29. Cassandrini D, Calevo MG, Tessa A, Manfredi G, Fattori F, Meschini MC, Carrozzo R, Tonoli E, Pedemonte M, Minetti C, Zara F, Santorelli FM, Bruno C: **A new method for analysis of mitochondrial DNA point mutations and assess levels of heteroplasmy.** *Biochem Biophys Res Commun* 2006, **342(2):**387-393.
30. Álvarez-Iglesias V, Salas A, Cerezo M, Ramos-Luis E, Jaime JC, Lareu MV, Carracedo : **Genotyping coding region mtDNA SNPs for Asian and Native American haplogroup assignation.** *Int Congress Series* 2006, **11(1288):**4-6.
31. Bandelt HJ, Kong QP, Parson W, Salas A: **More evidence for non-maternal inheritance of mitochondrial DNA?** *J Med Genet* 2005, **42:**957-960.
32. Bandelt HJ, Salas A, Bravi CM: **Problems in FBI mtDNA database.** *Science* 2004, **305(5689):**1402-1404.
33. Mitchell AL, Elson JL, Howell N, Taylor RW, Turnbull DM: **Sequence variation in mitochondrial complex I genes: mutation or polymorphism?** *J Med Genet* 2006, **43(2):**175-179.
34. McFarland R, Taylor RW, Elson JL, Lightowlers RN, Turnbull DM, Howell N: **Proving pathogenicity: when evolution is not enough.** *Am J Med Genet A* 2004, **131(1):**107-8; author reply 109-10.
35. DiMauro S, Schon EA: **Mitochondrial DNA mutations in human disease.** *Am J Med Genet* 2001, **106(1):**18-26.
36. Goto Y, Nonaka I, Horai S: **A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies.** *Nature* 1990, **348(6302):**651-653.
37. van den Ouweland JM, Lemkes HH, Ruitenbeek W, Sandkuijl LA, de Vijlder MF, Struyvenberg PA, van de Kamp JJ, Maassen JA: **Mutation in mitochondrial tRNA(Leu)(UUR) gene in a large pedigree with maternally transmitted type II diabetes mellitus and deafness.** *Nat Genet* 1992, **1(5):**368-371.
38. Holt IJ, Harding AE, Petty RK, Morgan-Hughes JA: **A new mitochondrial disease associated with mitochondrial DNA heteroplasmy.** *Am J Hum Genet* 1990, **46(3):**428-433.
39. de Vries DD, van Engelen BG, Gabreels FJ, Ruitenbeek W, van Oost BA: **A second missense mutation in the mitochondrial ATPase 6 gene in Leigh's syndrome.** *Ann Neurol* 1993, **34(3):**410-412.
40. Howell N, Bindoff LA, McCullough DA, Kubacka I, Poulton J, Mackey D, Taylor L, Turnbull DM: **Leber hereditary optic neuropathy: identification of the same mitochondrial ND1 mutation in six pedigrees.** *Am J Hum Genet* 1991, **49(5):**939-950.
41. Johns DR, Neufeld MJ, Park RD: **An ND-6 mitochondrial DNA mutation associated with Leber hereditary optic neuropathy.** *Biochem Biophys Res Commun* 1992, **187(3):**1551-1557.
42. Taylor RW, Turnbull DM: **Mitochondrial DNA mutations in human disease.** *Nat Rev Genet* 2005, **6(5):**389-402.
43. Majamaa K, Moilanen JS, Uimonen S, Remes AM, Salmela PI, Karppa M, Majamaa-Voltti KA, Rusanen H, Sorri M, Peuhkurinen KJ, Hassinen IE: **Epidemiology of A3243G, the mutation for mitochondrial encephalomyopathy, lactic acidosis, and strokelike episodes: prevalence of the mutation in an adult population.** *Am J Hum Genet* 1998, **63(2):**447-454.
44. Schaefer AM, Taylor RW, Turnbull DM, Chinnery PF: **The epidemiology of mitochondrial disorders--past, present and future.** *Biochim Biophys Acta* 2004, **1659(2-3):**115-120.
45. Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V: **The fingerprint of phantom mutations in mitochondrial DNA data.** *Am J Hum Genet* 2002, **71(5):**1150-1160.
46. Bandelt HJ, Salas A, Lutz-Bonengel S: **Artificial recombination in forensic mtDNA population databases.** *Int J Legal Med* 2004, **118(5):**267-273.
47. Salas A, Bandelt HJ, Macaulay V, Richards MB: **Phylogeographic investigations: The role of trees in forensic genetics.** *Forensic Sci Int* 2007, **168:**1-13.
48. Salas A, Carracedo , Macaulay V, Richards M, Bandelt HJ: **A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics.** *Biochem Biophys Res Commun* 2005, **335(3):**891-899.
49. Salas A, Yao YG, Macaulay V, Vega A, Carracedo , Bandelt HJ: **A critical reassessment of the role of mitochondria in tumorigenesis.** *PLoS Med* 2005, **2(11):**e296.
50. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N: **Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA.** *Nat Genet* 1999, **23:**147.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2350/9/26/prepub

Artículo 7: The mtDNA ancestry of admixed Colombian population

Salas A, Acosta A, Álvarez-Iglesias V, Cerezo M, Lareu MV, Carracedo Á

(American Journal of Human Biology; in press)

Estudio de la ancestralidad mitocondrial en población Colombiana 'mezclada'

ABSTRACT/RESUMEN:

En un total de 185 individuos de Colombia se secuenció la region hipervariable I (HVS-I) del genoma mitocondrial (ADNmt), y en un grupo de esos individuos se genotipó adicionalmente la región hipervaliable II (HVS-II). La recogida de muestras se realizó teniendo en cuenta la propia autodefinición étnica de los individuos, incluyendo 'Mestizos', 'Mulatos' y 'Afro-Colombianos'. Utilizamos bases de datos con más de 4.300 linajes Nativo Americanos, 6.800 africanos, y 15.600 europeos para comparaciones e inferencias filogenéticas de la población estudiada. Observamos que 'Mulatos' y 'Afro-Colombianos' tienen un componente africano dominante en el ADNmt, miestras que 'Mestizos' presentan predominantemente haplotipos Nativo Americanos. Todas las poblaciones analizadas tienen un índice de diversidad elevado y no reflejan episodios bruscos de deriva genética. Los linajes mitocondriales Nativos Americanos de Colombia probablemente tienen su origen en poblaciones de América Central y América del Sur, mientras que el de los linajes afro-colombianos probablemente esté en poblaciones del oeste, sudoeste y sudeste de África. Nuestros resultados difieren de aquellos obtenidos en otros estudios en terminos de frecuencias de haplogrupos. Esto nos permite conculir que la autodeclaración (como pertenecientes a un determinado grupo 'étnico') no es un buen indicador de etnicidad en nuestras muestras de Colombia, y que, considerar esos grupos pseudo-étnicos genéticamente homogéneos puede ser un factor de riesgo de error tipo I (falsos positivos) en estudios médicos en este país y pueden llevar a una mala interpretación en casos de rutina forense.

AMERICAN JOURNAL OF HUMAN BIOLOGY 00:000–000 (2008)

Original Research Article

The mtDNA Ancestry of Admixed Colombian Populations

A. SALAS,[1,2]* A. ACOSTA,[1] V. ÁLVAREZ-IGLESIAS,[1] M. CEREZO,[1] C. PHILLIPS,[1] M. V. LAREU,[1] AND Á. CARRACEDO[1,2]
[1]*Unidade de Xenética, Instituto de Medicina Legal, Facultade de Medicina, Universidade de Santiago de Compostela, 15782 Galicia, Spain*
[2]*Grupo de Medicina Xenómica, CIBERER, Hospital Clínico Universitario, 15706 Santiago de Compostela, Galicia, Spain*

ABSTRACT A total of 185 individuals from Colombia were sequenced for the first hypervariable region (HVS-I) of the mitochondrial DNA (mtDNA) genome, and a subset of these individuals were additionally genotyped for the second hypervariable segment (HVS-II). These individuals were collected according to their "self-reported ethnicity" in Colombia, comprising "Mestizos," "Mulatos," and "Afro-Colombians." We used databases containing more than 4,300 Native American lineages, 6,800 Africans, and 15,600 Europeans for population comparisons and phylogeographic inferences. We observe that Mulatos and Afro-Colombians have a dominant African mtDNA component, whereas Mestizos carry predominantly Native American haplotypes. All the populations analyzed have high diversity indices and there are no signatures of dramatic genetic drift episodes. Central and South America are the main candidate source populations of the Colombian Native American lineages, whereas west-central, southwest, and southeast Africa are the main original mtDNA sources for the African Colombian mtDNAs. We found that our results differ from those obtained in other studies for the same "population groups" in terms of haplogroup frequencies. This observation leads us to conclude that (i) self-reported ancestry is not a reliable proxy to indicate an individual's "ethinicity" in Colombia, (ii) our results do not support the use of outmoded race descriptions (Mestizos, Mulatos, etc.) mainly because these labels do not correspond to any genetically homogeneous population group, and (iii) studies relying on these terms to describe the population group of the individual, which then treat them as genetically homogeneous, carry a high risk of type I error (false positives) in medical studies in this country and of misinterpretation of the frequency of observed variation in forensic casework. Am. J. Hum. Biol. 00:000–000, 2008.

The two main linguistic groups that dominated the territory of Colombia during pre-Colombian times were the Chibcha and the Carib. Examples of these Native American groups were the Tayronas living in the Caribbean region and the Muiscas in the highlands near Bogotá, both of which belonged to the Chibcha language family. The Muisca people had one of the most developed political systems in South America, surpassed only by the Incas. These populations constituted mainly hunter-gatherer societies and they traded with one another (and with other cultures living in the Magdalena River valley) by exchanging salt, emeralds, beans, maize, other crops, etc.

When the first Spanish arrived to what is now Colombia, the largest and most widespread culture was the Chibchas; these population groups were concentrated mainly in the highland basins and valleys of the Cordillera Oriental. The colonial period led to a dramatic change in the political and socioeconomic regimes of the indigenous Colombian people. The Spanish settled along the north coast of modern Colombia as early as 1500, but their first permanent settlement at Santa Marta and Cartagena was not established until 1525 and 1533, respectively. The city of Bogotá, founded in 1538, became one of the principal administrative centers of the Spanish possessions in the New World (along with Lima and Mexico City).

The arrival of African slaves for forced labor to New Granada (the old name of Colombia used 1533–1858) would dramatically change the demographic landscape of the region. It is estimated that the city of Cartagena alone received more than 200,000 African slaves destined for the Viceroyalty of Perú (Curtin, 1979). The late 18th century saw progress toward the abolition of slavery as part of a movement for independence from Spain.

Today, more than 50 different indigenous ethnic groups have been described in Colombia, and most of them have preserved their original languages (belonging to the Chibchan and Caribean linguistic families). The census data in the country do not record ethnicity, and so percentages are based on estimates from other sources and can vary from one region to another. In general, it is thought that the Colombian population is the result of a complex process of admixture between Europeans, Africans, and Native Americans to different degrees depending on the region.

In the last 20 years, the analysis of molecular DNA markers has contributed significantly to an understanding of the prehistory and history of human populations. Colombia has been the target of a large number of genetic studies, most of them, however, comprise forensic genetics analysis, e.g. short tandem repeat (STR) databases of forensic use (Bravo et al., 2001; Paredes et al., 2003), but with little focus on the anthropological and demographic issues. One of the first attempts to unravel the ancestry of "Mestizo" and "Afro-Colombians" was performed by Rodas et al. (2003) by analyzing mitochondrial DNA (mtDNA) restriction fragment length polymorphisms (RFLP) variation and characterizing the main Native American, West

This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/1042-0533/suppmat.

Contract grant sponsor: Ministerio de Educación y Ciencia; Contract grant number: RYC2005-3; Contract grant sponsor: Xunta de Galicia; Contract grant number: PGIDIT06PXIB208079PR; Contract grant sponsor: Fundación de Investigación Médica Mutua Madrileña.

*Correspondence to: Antonio Salas, Unidade de Xenética, Instituto de Medicina Legal, Facultade de Medicina, Universidade de Santiago de Compostela, 15782 Galicia, Spain. E-mail: apimlase@usc.es

Received 23 November 2007; Revision received 25 January 2008; Accepted 19 February 2008

DOI 10.1002/ajhb.20783

Published online in Wiley InterScience (www.interscience.wiley.com).

European, and African haplogroups; whereas, Torres et al. (2006) studied the mtDNA RFLP variation in Colombian Native Americans alone. Recently, Bedoya et al. (2006) studied the genetic composition of a sample described as "Hispanics" from Antioquia (Colombia). Previous work from the same group characterized the admixture Amerind/"white" sex bias in a sample from Colombia (Carvajal-Carmona et al., 2000, 2003). Lastly, a recent article by Melton et al. (2007) examined the mtDNA variation of three Chibchan and one Arawak population from north-east Colombia mainly focusing on the demographic expansion of the Chibchan-speakers into South America.

The present study aimed to investigate the mtDNA ancestry of admixed Colombian groups of individuals, including "Mestizos" (the term widely used in America to designate individuals of European and Native American co-ancestry), "Mulatos" (the term used in Colombia, to designate individuals of African and European coancestry), and "Afro-Colombians," by analyzing a sample of 185 individuals and focusing on a phylogeographic approach that complements the previous studies by other authors. One of the main aims of this study is to demonstrate that the use of the ancestry descriptors in Colombia (and in many other American countries), whether self declared or assigned during sample collection, is not supported by the genetic variation we have characterized. It should be noted that, just for clarity, we report the ancestry descriptors used by other studies or those based on self declaration in our own study. Furthermore, we advocate the use of simple continental descriptors of ancestry or coancestry, namely European, African, American, European-American, although the term Native American avoids confusion with African American and European American of near universal use when applied to U.S. population studies.

MATERIALS AND METHODS

Samples

We have analyzed 185 individuals from Colombia, most of them from the South (departments of Cauca and Valle del Cauca). These samples consist of individuals belonging to different "self reported ethnicities": 67 Mestizos, 11 Mulatos, and 95 Afro-Colombians. We also included 10 Native Americans (Páez Indians). Informed consent was given by all participants. The uneven nature of the sample sizes of the population groups studied was related to the logistics of sample collection, which is particularly complex given the size, linguistic diversity and difficulty of travel in the Colombian regions analyzed. The protocol and procedures employed were approved by the review committee of the University of Santiago de Compostela (Spain) where the study was performed.

PCR and sequencing

All samples were amplified and double-strand sequenced for the HVS-I; however, because of limitations related to the amount of DNA available, we could not analyze all samples for the same sequence range; however, most were read for the sequence segment 16024–16569. For the same reason, only a subset of samples could be sequenced for the HVS-II segment. PCR amplification was carried out using the GenAmp PCR sequencing system, as described in the work of Álvarez-Iglesias et al. (2007).

Mutations are referenced with respect to the revised Cambridge Reference Sequence (Anderson et al., 1981; Andrews et al., 1999). We followed a standardized forensic nomenclature system as indicated in the work of Carracedo et al. (2000), but with slight modifications considered in that of Salas et al. (2005a). The data were checked following the phylogenetic principles described in previous works (Bandelt et al., 2002, 2004a,b; Salas et al., 2005a,d, 2007) in order to reduce to an absolute minimum possible sequence artifacts. The final results are presented in Table S1.

Statistical approach and mtDNA nomenclature

For phylogeographic purposes we have collected different available population datasets from the literature. Thus, for the African dataset, we have used nearly the same data reported in the work of Černý et al. (2007), which consists of 6,856 profiles from different regions in the African continent.

For Native American lineages, an HVS-I database that consists of 4,394 profiles was employed. Finally, the European database consists of more than 15,600 HVS-I profiles. More information concerning the literature source of these large datasets will be provided under request to the corresponding author.

Only the HVS-I sequence range from position 16090 to 16365 was used for population comparisons, because this is the common segment available for the different population datasets used. Nomenclature of African haplotypes follows (Salas et al., 2002, 2004) with the updates in the works of Kivisild et al. (2004) and Torroni et al. (2006), and slight modifications reported in that of Černý et al. (2007). For Native American haplogroups, we use the most updated nomenclature reported by Kong et al. (2006), Tamm et al. (2007), and Achilli et al. (2008).

Finally, principal component analysis was carried out using the software Stata v.9 based on haplogroup frequency profiles, and including the main Colombian population groups (Mestizos, Mulatos, and Afro-Colombians) and few other populations representing the three main Colombian source populations (Native Americans, Africans, and Europeans).

RESULTS

Variability of the Native American and African Colombian lineages

For the Native American haplogroups (A2, B2, C1, and D1), 45 haplotypes appeared only once in the whole Colombian sample, five appeared twice, five were found three times, three were found six times, and one was found 11 times (Table S1). A total of 33 haplotype sequences were not found in our American database. Among the observed sequences, 15 match in North America (n = 2,005), nine in Central America (n = 485), and 16 in South America (n = 1,657); but many of them are present in the three main regions or at least in two of them (Table 1).

For the African lineages (56 sequences belonging to haplogroup L), 42 sequences appeared once in the whole Colombian sample, nine sequences were found twice, one haplotype was found three times, and three sequences were found five times (Table S1). Forty-seven out of 56 (~84%) mtDNAs had been previously observed in other studies on other American locations. Colombia shares 18

TABLE 1. Shared haplotypes between Colombian Native American lineages and the main American regions

Colombian Native American lineages (minus 16000)	*n*	AM-C	AM-N	AM-S	Total	HG
111 223 290 319 362	6	45	161	50	262	A2
093 111 223 290 311 319 362	1	–	–	–	1	A2
111 129 192 223 290 319 362	1	–	–	–	1	A2
111 183C 189 223 287 290 319 362	3	–	–	3	6	A2
111 187 223 290 319	1	–	–	–	1	A2
111 187 223 290 319 362	1	25	2	–	28	A2
111 189 193+C 223 290 319 362	1	–	1	–	2	A2
111 192 223 274 290 311 319 362	1	–	–	–	1	A2
111 213 223 290 319	1	–	–	–	1	A2
111 213 223 290 319 356 362	1	–	–	–	1	A2
111 213 223 290 319 362	11	–	–	3	14	A2
111 213 223 319 362	1	–	–	–	1	A2
111 223 247T 290 295 319 362	1	–	–	–	1	A2
111 223 274 290 311 319 362	1	–	–	–	1	A2
111 223 290 300 319 362	1	1	–	–	2	A2
111 223 290 311 319 362	1	2	2	2	7	A2
111 223 290 319	1	3	11	1	16	A2
111 223 290 319 356 362	3	–	4	3	10	A2
093 129 183C 189 217	1	–	–	–	1	B2
175 183 189 217	1	–	–	–	1	B2
182C 183C 189 217 300	1	–	–	–	1	B2
182C 183C 189 217 301 304	1	–	–	–	1	B2
183C 189 193+C 217	1	–	1	4	6	B2
183C 189 217	6	1	14	86	107	B2
183C 189 217 235 319C	1	–	–	–	1	B2
223 295 298 325 327	1	–	6	–	7	C1
223 298 325	1	–	–	2	3	C1
092 176 218 223 298 325 327	1	3	1	–	5	C1
104 174 223 298 325 327	1	–	–	–	1	C1
127 223 295 298 325 327	1	–	–	–	1	C1
128 209 223 298 325 327	1	–	–	–	1	C1
129 223 234 298 325 327	3	–	–	–	3	C1
129 223 274 298 325 327	1	–	–	–	1	C1
169 209 223 298 325 327	1	–	–	–	1	C1
169 216 223 298 325 327	1	–	–	–	1	C1
169 223 298 325 327	3	–	–	–	3	C1
172 223 298 325 327	2	–	–	6	8	C1
185 223 239 298 311 325 327	1	–	–	–	1	C1
185 223 298 311 325 327	1	–	–	–	1	C1
185 223 325 327	2	–	–	–	2	C1
209 223 289 298 325 327	1	–	–	–	1	C1
209 223 298 300 325 327	1	–	–	–	1	C1
209 223 298 325 327	1	–	–	1	2	C1
209 223 298 325 327 357	1	–	–	–	1	C1
209 223 325 327	1	–	–	1	2	C1
209 298 325 327	1	–	–	–	1	C1
216 223 298 325 327 356 362	1	–	–	–	1	C1
223 274 298 325 327	3	–	5	–	8	C1
223 298 325 327	6	19	111	195	331	C1
223 298 325 327 335	2	–	1	1	4	C1
223 325 362	2	5	34	35	76	C1
092 223 311 325 362	1	–	–	–	1	D1
142 188 223 325 362	1	–	–	–	1	D1
142 207 223 325 362	2	–	–	–	2	D1
172 185 192 223 301 342 362	1	–	–	–	1	D1
183C 189 223 270 325 362	1	–	–	–	1	D1
223 232A 325 362	1	–	–	–	1	D1
223 311 325 362	1	–	–	–	1	D1
Total	*99*	*485*	*2005*	*1657*	*949*	

AM-C, Central America; AM-N, North America; AM-S, South America; HG, haplogroup.

L-haplotypes each with west and west-central Africa, 12 haplotypes with southeast, and eight with southwest; the latter two are more surprising considering the relatively lower sample sizes of studies on these regions (Table 2) in comparison with the Atlantic African coast.

Haplogroup composition of the different admixed Colombian groups

The Mestizo population consists of 67 individuals. The majority (97%) carry Native American haplogroups: 40% belong to haplogroup A2, 13% to B2, 34% to C1, and 7% to D1. Two haplotypes are of European origin while only one belongs to the typical sub-Saharan haplogroup L2a. In contrast, the Mulato sample is dominated by L haplotypes (~81%); however, the sample size is low ($n = 11$) and therefore reliable frequency estimation of minor haplogroups is not feasible. The Afro-Colombians ($n = 95$) also carry predominantly L-haplotypes (72.6%), but the Native American component is also very significant (23.2%), comprising mtDNAs belonging to haplogroup C1, the main indigenous component (11%).

TABLE 2. Shared haplotypes between Colombian African lineages the main American and African regions

Colombian African lineages (minus 16000)	*n*	AM-C	AM-N	AM-S	AF-E	AF-N	AF-S	AF-SE	AF-SW	AF-W	AF-WC	Total	HG
129 148 168 172 187 223 230 278 293 311 320	1	–	–	–	–	–	–	–	–	–	–	1	L01
129 148 168 172 187 188G 189 223 230 278 293 311 320	2	1	2	2	5	1	1	22	4	–	11	51	L0a1
126 187 189 223 264 270 278 311	2	1	24	2	–	6	1	2	3	17	6	64	L1b
111 126 187 189 223 239 270 278 293 311	1	–	3	–	–	–	–	–	–	3	5	12	L1b1
126 187 189 223 248 264 270 278 293 311	1	–	–	–	–	–	–	–	–	–	–	1	L1b1
126 187 189 223 264 270 278 284 311	1	–	–	–	–	–	–	–	–	–	–	1	L1b1
126 187 189 223 264 270 278 293 311 355	1	–	–	–	–	–	–	–	–	–	–	1	L1b1
126 187 189 223 264 270 278 293 311 362	1	–	–	–	–	2	–	–	–	7	–	10	L1b1
092 126 184 187 189 223 278 294 301 311 360	1	–	–	–	–	–	–	–	–	–	–	1	L1c
093 129 140 184 187 189 223 278 294 301 311 360	2	–	1	–	–	–	–	–	–	–	–	3	L1c
104 129 163 187 189 223 278 293 294 311 360	1	–	–	–	–	–	–	–	–	–	–	1	L1c
129 163 187 189 223 278 293 294 304 311 360	1	–	–	–	–	–	–	–	–	16	2	19	L1c
129 169 172 187 189 192 223 261 278 293 311 360	1	–	–	–	–	–	–	–	–	–	–	1	L1c
129 183C 189 215 223 278 294 311 360	2	–	2	–	–	–	–	2	–	–	2	8	L1c
129 184 187 189 270 278 293 301 311	1	–	–	–	–	–	–	–	–	–	–	1	L1c
129 187 189 214 234 249 258 274 278 293 294 311 360	1	–	1	–	–	–	–	–	–	–	1–	12	L1c
129 187 189 223 265C 278 286G 294 311 359 360	1	–	1	1	–	–	–	–	2	–	–	5	L1c
129 187 189 223 278 284 293 294 311 360	2	–	5	–	–	–	–	–	–	–	–	7	L1c
129 187 189 223 274 278 293 294 311 360	1	–	1	1	–	–	–	2	–	–	24	29	L1c1a
187 189 223 278 284 293 294 311 360	1	–	–	–	–	–	–	–	–	–	–	1	L1c1a
093 213 223 290 294	1	–	–	–	–	–	–	–	–	–	–	1	L2a
183C 189 192 223 294 309 311	1	–	–	–	–	–	–	–	–	–	–	1	L2a
184 223 278 294	1	–	–	–	–	–	–	–	–	–	–	1	L2a
189 192 223 278 294	2	–	5	–	4	–	–	3	–	2	3	19	L2a
189 223 278 294 362	1	–	–	–	–	–	–	–	–	7	2	10	L2a
192 223 278 294	1	–	–	–	–	–	–	–	–	–	–	1	L2a
223 278 290 294 309	1	–	1	–	–	1	–	–	–	3	–	6	L2a
223 278 294 309	5	3	48	8	4	12	–	1–	1	23	2–	134	L2a
093 189 192 223 278 294 309	1	–	2	–	–	–	–	–	–	2	–	5	L2a1
126 223 278 309	3	–	1	–	–	–	–	–	–	–	–	4	L2a1
189 223 278 294 309	1	2	12	1	26	3	–	2	–	13	6	66	L2a1
193 213 223 239 278 294 309	1	–	2	–	–	–	–	–	–	–	1	4	L2a1
223 264 278	1	–	5	–	–	–	–	–	–	5	3	14	L2a1
215 223 278 286 294 309	1	–	–	–	–	–	–	–	–	–	–	1	L2a1a
114A 129 142 213 223 325 362	1	–	–	–	–	–	–	–	–	–	–	1	L2b1
129 213 223 278 354	1	–	–	–	–	–	–	–	–	–	–	1	L2b1
169 223 264 278	2	–	–	–	–	–	–	–	–	–	–	2	L2c2
093 223 278	1	–	–	–	–	–	–	–	–	5	–	6	L3*?
223 319	1	–	–	–	1	1	–	–	–	–	–	3	L3*?
124 183C 187A 189 223 278 362	1	–	–	–	–	–	–	–	–	–	–	1	L3b
124 223 240 278 362	1	–	–	–	–	–	–	–	–	–	–	1	L3b
124 223 362	1	–	2	–	–	–	–	–	–	–	2	5	L3b
223 240 278 362	1	–	–	–	–	–	–	–	–	–	–	1	L3b
124 166 223	1	–	2	–	–	1	–	–	–	3	7	14	L3d
124 223	5	2	8	1	2	3	–	3	–	2–	12	56	L3d
124 223 291	1	1	3	–	–	–	–	–	–	1	–	6	L3d
093 148 223 265T	1	–	3	1	–	–	–	–	–	–	–	5	L3e
185 223 327	1	–	4	2	–	–	1	7	3	–	–	18	L3e1a
213 223 265 320	1	–	–	–	–	–	–	–	–	–	–	1	L3e3
223 265T	1	–	15	–	5	1	1	1–	2	–	7	42	L3e3
223 264 319	1	–	–	–	–	–	–	–	–	–	–	1	L3e4
209 223 311	2	–	3	1	1	5	–	3	8	2	13	38	L3f
209 218 223 256 292 311	2	–	2	–	–	–	–	2	2	–	–	8	L3f1
209 223 235 292 311	1	–	–	–	–	–	–	–	–	2	–	3	L3f1
209 223 292 295 311	5	1	4	–	–	–	–	–	–	13	–	23	L3f1
Total	*78*	*83*	*1148*	*143*	*835*	*1312*	*264*	*416*	*157*	*1184*	*1202*	*732*	

AM-C, Central America; AM-N, North America; AM-S, South America; AF-E, East Africa; AF-N, North Africa; AF-S, South Africa; AF-SE, Southeast Africa; AF-SW, Southwest Africa; AF-W, West Africa; AF-WC, West-Central Africa.

The African component of the Afro-Colombian sample is less diverse in terms of haplotype diversity than the African component of Mulatos (Table 2), but this difference is not statistically significant because of the large standard error affecting the Mulato's sample (Table 3). Nucleotide diversity is slightly higher, but significantly so, in the African mtDNA component of Afro-Colombians than in the L- lineages observed in Mulatos.

On the other hand, the Native American component of Afro-Colombians is more diverse in terms of haplotype diversity than its counterpart in Mestizo, but again, this difference is not statistically significant (Table 3). The opposite pattern occurs for the nucleotide diversity values.

As expected the nucleotide diversity of the African lineages in Colombians is higher than the values for the Native American mtDNAs. The same occurs for the haplotype diversity, although the differences are not statistically significant (Table 3).

The phylogeographic features of admixed Colombians

The most common Native American haplotypes in Colombians match some of the most common haplotypes in America; and this is particularly true for the root haplotypes of the main Native American haplogroups. There is no evidence of important founder effects in our sample;

TABLE 3. Diversity indices for HVS-I (sequence range 16024–16365) in Colombian populations

	n	*K* (*K*/*n*)	π (SE)	*H* (SE)	*M*
Afro-Colombians	95	71 (74.7)	0.025 (0.001)	0.991 (0.003)	8.4
African component	69	48 (70.3)	0.024 (0.001)	0.985 (0.006)	8.3
Native American component	22	19 (82.6)	0.018 (0.001)	0.984 (0.017)	6.0
Mestizo	67	43 (64.2)	0.021 (0.001)	0.974 (0.009)	7.0
Native American component	64	40 (62.5)	0.020 (0.001)	0.971 (0.010)	6.9
Mulato	11	11 (100.0)	0.021 (0.003)	1.000 (0.039)	7.1
African component	9	9 (100.0)	0.020 (0.003)	1.000 (0.052)	6.9

K = number of different sequences found and percentage of sample size in brackets; π = nucleotide diversity; *H* = haplotype diversity; *M* = average number of pairwise difference.

however, some haplotypes appear at relatively high frequencies in Colombia. For instance, A2 haplotype: C16111T G16213A C16223T C16290T G16319A T16362C occurs 10 times in this sample of Colombia, but only five times in previous studies, three of them in an independent Colombian sample (Torres et al., 2006) and two in the Genographic dataset (https://www3.nationalgeographic.com/genographic/resources.html). The one-step mutation haplotypes carrying transition: T16209C on top of the later haplotype is highly prevalent in Venezuela; of the 12 haplotypes observed in the whole database, 11 are found in the indigenous Guahibo people (Vona et al., 2005). Haplotype C16111T C16223T C16290T G16319A T16362C is observed at high frequency in the Haida from Queen Charlotte Islands (North America), but it is also present in the Xavante Brazilians (Ward et al., 1996) or in the Kuna from Panama (Batista et al., 1995). In the neighboring country of Ecuador (Rickards et al., 1999) we also find the only three matching sequences of the Colombian type: C16111T A16183C T16189C C16223T C16287T C16290T G16319A T16362C (either including or excluding the highly unstable variant A16183C). In the Ngöbe of Panama (Kolman et al., 1995) we detect the majority of matching sequences for haplotype: C16111T C16187T C16223T C16290T G16319A T16362C. Also particularly noticeable is the fact that perfect matches of the A2 haplotype: C16111T C16223T C16290T G16319A are only found in the Inuit from Canada. Finally, we observed identical matches of the haplotype C16111T C16223T C16290T G16319A T16356C T16362C in another independent Colombian sample (Torres et al., 2006) as well as in some regions neighboring Colombia.

It can be said that the Native American component of admixed Colombians is more closely phylogenetically related to northern-central South America (specifically the northern regions that encompass northern Brazil, Ecuador, and Venezuela) than to any other American region.

Most of the African lineages found in Colombia are present in Africa. Only four of them were not previously observed in this continent or in Afro-American samples; however, some of these mtDNAs have well-known one-step mutation representatives in Africa. The most frequent L-haplotypes in Colombia are also common in Africa; for instance, the L1b haplotype characterized by the following substitutions: T16126C C16187T T16189C C16223T C16264T C16270T C16278T T16311C is the root type of haplogroup L1b and was observed 16 times in our Colombian sample and 62 times in Africa, specially in West but also in North, Central, and East at lower frequencies; it is also common in other parts of America (see Fig. 4 in Salas et al., 2002). The L1b1 lineage: T16126C C16187T T16189C C16223T C16264T C16270T C16278T A16293G T16311C T16362C was only observed in three different Atlantic Islands: Azores, Cape Verde, and Canary Islands; note that the two latter locations were important ports for slave transport from Africa to America. Another common haplotype in Colombia is the root type of L2a: C16223T C16278T C16294T A16309G (*n* = 13 in Colombia) which is also widespread in Africa and in other parts of America.

The origin of L1c haplogroup is still uncertain, although Central Africa is the most likely candidate source region (Quintana-Murci et al., 2007; Salas et al., 2002, 2004, 2005b); it is however more difficult to account for the origin of the large proportion of L1c American lineages that do not find matches in present day Africa. Three different L1c haplotypes were observed in Colombia. Haplotype G16129A C16187T T16189C C16223T A16265C C16278T C16286G C16294T T16311C T16359C C16360T is observed in, e.g., Brazil (Silva et al., 2006) and could have originated from Angola, where it has been observed at least twice (Plaza et al., 2004), whereas the one-step mutation derivative: G16129A C16187T T16189C C16223T A16265C C16278T C16286G C16294T T16311C T16360C was observed once in the neighboring African population of Cabinda (Beleza et al., 2005). Haplotype: L1c2 G16129A A16183C T16189C A16215G C16223T C16278T C16294T T16311C T16360C could readily originate from Equatorial Guinea (Mateu et al., 1997), where it was found at high frequency (eight matches). Other continental locations are also good candidates, notably Central Africa is a likely source region because there are few one-step mutation haplotypes located in this area. The origin of the common Colombian L1c mtDNA characterized by substitutions: T16093C G16129A T16140C C16184T C16187T T16189C C16223T C16278T C16294T 16301 T16311C T16360C (*n* = 11) is more enigmatic; it was observed only once in America (Monson et al., 2002).

The L3e haplotype: T16093C C16148T C16223T A16265T was only observed in America in the SWGDAM database (Monson et al., 2002) and in another Colombian population with an important recent African ancestry from the department of Chocó (Bravo et al., 2001; Paredes et al., 2003; Salas et al., 2005c), while there are also three matches in the Genographic database for this haplotype. The L3e1a Colombian profile: C16185T C16223T C16327T was detected several times in Mozambique (Pereira et al., 2001; Salas et al., 2002) but also in Cabinda (Beleza et al., 2005) and Angola (Plaza et al., 2004); these two southwest African locations are good candidate source regions considering their important role in the past as source populations for slavery to America.

The 10 Páez Indian samples analyzed in this study carry Native American mtDNAs; half of them belong to haplogroup A2, four individuals bear haplogroup C1 haplotypes, and one individual carries a B2 mtDNA.

Principal component analysis

The two first principal components (PC1 and PC2) account for 61% of the observed variability (see Fig. 1).

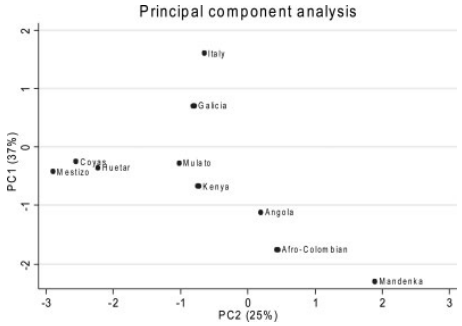


Fig. 1. Scatter plot showing the first two principal component analysis of haplogroup frequency profiles for several populations, including the Colombian samples (Mestizo, Mulato, and Afro-Colombians).

PC1 (37% of the genetic variance) clusters in one pole the sub-Saharan populations of Angola (Plaza et al., 2004) and Mandenka (Graven et al., 1995) together with the Afro-Colombians; on the opposite side of the scatter plot are the European populations of Italy (Tagliabracci et al., 2001) and Galicia, southwest Iberia (Salas et al., 1998); the rest of the populations stand in between (see Fig. 1). The PC2 (25% of the genetic variance) clearly separates the Native American populations of Coyas (Álvarez-Iglesias et al., 2007) and Huetar (Santos and Barrantes, 1994) together with the Colombian Mestizos, from the rest of the populations (see Fig. 1). Finally, the small Mulato sample is closely located to the Kikuyu (Watson et al., 1996), showing mtDNA affinities to sub-Saharan Africa (see Fig. 1). As expected, the main determinants of the PC1 pattern are in general the whole set of African (L-haplogroups) and European lineages, while the main determinants of PC2 are the Native American haplogroups A2, B2, C1, and D1.

DISCUSSION

We have dissected the different mtDNA ancestries of various Colombian admixed "self-reported" population groups from a predominantly phylogeographic perspective.

The Native American component of Colombians is more closely related to Central America and northern South America. Some lineages have high frequencies in Colombia, but we did not observe evidence for the action of genetic drift in these populations. This is also manifested by the high values of the diversity indices computed (Table 3), demonstrating that an important amount of autochthonous Native variation in many admixed populations has survived the demographic impact of the western European colonization (notably in the case of Colombia from Spain).

Several past studies have attempted to disentangle the complex African contribution to America arising from the slave trade, either on a continental scale (Salas et al., 2002, 2004, 2005b), or focusing on contributions on more regionalized variation (Alves-Silva et al., 2000; Bravi et al., 1997; Carvajal-Carmona et al., 2000; Ely et al., 2006; Parra et al., 1998; Salas et al., 2005c). We show that the main African component in Colombia appears to be West/west-central, southeast, and southwest Africa, in agreement with previous findings (Salas et al., 2004, 2005b) and historical records. The amount of variation in the African component is also comparable with that found in the original African continent; this reflects the large number of slaves forcibly moved to Colombia and other American regions which constitutes a large effective population size.

The PC analysis accounts for a substantial percentage of the variability (61%) and helps to summarize the global haplogroup composition of the admixed Colombian populations here analyzed: Mulatos and Afro-Colombians cluster together with other sub-Saharan samples, whereas the Mestizo sample locates close to other Native American groups reflecting the high percentage of L-lineages in the two former samples and the large amount of Native American haplogroups in the latter.

An important aspect of this study is the possibility of detailed evaluation of self-reported ethnicity, used during the sampling process, as a proxy for genetic ancestry. In a previous study (Rodas et al., 2003) it was reported that the frequency of L lineages in five different Afro-Colombian population samples ranged from 21.4% (Quibdó) to 52.5% (Providencia). In our Colombian sample, we found that 72.6% of the lineages belong to some L-haplogroup; this pattern differs significantly both from the Quibdó (Fisher's exact test; $P < 0.0000$) and from the Providencia sample (Fisher's exact test; $P = 0.0203$). The frequency of the Native American component in the Mestizo also differs significantly in both studies, 78% in Rodas' study (2003) *versus* ~97% in our Mestizo sample (Fisher's exact test; $P = 0.0004$). We therefore observe that independently collected samples of the same self-reported population affiliations from Colombia can lead to quite different haplogroup spectra. This has been observed to occur in the study of other admixed populations from other regions or from the use of population descriptors that have a larger cultural than geographical point of reference such as "Hispanic" (Salas et al., 2007). Our results therefore do not support the use of the terminology still found in many population studies that are outmoded and inaccurate (e.g. Mestizo), lack a correctly defined geographic description (e.g. Caucasian rather than European), or have a predominantly cultural or linguistic definition (e.g. Hispanic). Our study and numerous others clearly illustrate that, when the genetic markers used have high resolution for characterizing the ancestry of an individual, these terms incorrectly denote membership of a homogenous group. Individuals belonging to different and variable mtDNA genetic ancestries may use self-descriptive terms that fail to reflect these differences, and studies that rely on such descriptions run the risk of spurious conclusions based on presumed genetically homogeneous groups.

Assumptions about the homogeneity of a group of individuals that are grouped into improperly defined populations can have consequences for forensic genetics and association studies. Undetected population stratification has an important bearing in forensic casework if an expert consulting a database wrongly assumes that the populations (lacking correctly defined genetic ancestry) are genetically homogenous. Such a consideration was highlighted by a recent review of the widely used SWGDAM forensic database (Salas et al., 2007). In case–control association studies, undetected population substructure is a well-known cause of type I error (Mosquera-Miguel et al.,

2008). Often cases and control subjects are matched by recombining individuals ethnically (e.g. both groups are Hispanics) in order to prevent the inflation of association values caused by differences in ancestry between each study group (Salas and Carracedo, 2007). However, we wish to conclude by emphasizing that the only way to prevent the risk of spurious findings in the above situations is to properly define the population in terms of genetic ancestry and levels of admixture using the most informative markers and appropriate sampling designs.

LITERATURE CITED

Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, Salas A, Torroni A, Bandelt HJ. 2008. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. PLoS ONE 3(3):e1764.

Álvarez-Iglesias V, Jaime JC, Carracedo Á, Salas A. 2007. Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. Forensic Sci Int: Genet 1:44–55.

Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt H-J, Pena SD, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461.

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge Reference Sequence for human mitochondrial DNA. Nat Genet 23:147.

Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150–1160.

Bandelt H-J, Salas A, Bravi CM. 2004a. Problems in FBI mtDNA database. Science 305:1402–1404.

Bandelt H-J, Salas A, Lutz-Bonengel S. 2004b. Artificial recombination in forensic mtDNA population databases. Int J Legal Med 118:267–273.

Batista O, Kolman CJ, Bermingham E. 1995. Mitochondrial DNA diversity in the Kuna Amerinds of Panama. Hum Mol Genet 4:921–929.

Bedoya G, Montoya P, Garcia J, Soto I, Bourgeois S, Carvajal L, Labuda D, Alvarez V, Ospina J, Hedrick PW, et al. 2006. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. Proc Natl Acad Sci USA 103:7234–7239.

Beleza S, Gusmão L, Amorim A, Carracedo Á, Salas A. 2005. The genetic legacy of western Bantu migrations. Hum Genet 117:366–375.

Bravi CM, Sans M, Bailliet G, Martinez-Marignac VL, Portas M, Barreto I, Bonilla C, Bianchi NO. 1997. Characterization of mitochondrial DNA and Y-chromosome haplotypes in a Uruguayan population of African ancestry. Hum Biol 69:641–652.

Bravo ML, Moreno MA, Builes JJ, Salas A, Lareu MV, Carracedo A. 2001. Autosomal STR genetic variation in negroid Choco and Bogota populations. Int J Legal Med 115:102–104.

Carracedo A, Bar W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, et al. 2000. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85.

Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, Ruiz-Linares A. 2003. Genetic demography of Antioquia (Colombia) and the central valley of Costa Rica. Hum Genet. 112:534–541.

Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, Ospina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, et al. 2000. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. Am J Hum Genet 67:1287–1295.

Černý V, Salas A, Hájek M, Žaloudková M, Brdička R. 2007. A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome. Ann Hum Genet 71:433–452.

Curtin D. 1979. The Atlantic slave trade 1600-1800. In: Ajayi JFA, Crowder M, editors. New York: Columbia University Press.

Ely B, Wilson JL, Jackson F, Jackson BA. 2006. African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. BMC Biol 4:34.

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L. 1995. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75:752–770.

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F. 1995. Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá. Genetics 140:275–283.

Kong Q-P, Bandelt H-J, Sun C, Yao Y-G, Salas A, Achilli A, Wang CY, Zhong L, Zhu CL, Wu SF, et al. 2006. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum Mol Genet 15:2076–2086.

Mateu E, Comas D, Calafell F, Pérez-Lezaun A, Abade A, Bertranpetit J. 1997. A tale of two islands: population history and mitochondrial sequence variation of Bioko and São Tomé, Gulf of Guinea. Annals of Hum Genet 61:507–518.

Melton PE, Briceno I, Gomez A, Devor EJ, Bernal JE, Crawford MH. 2007. Biological relationship between Central and South American Chibchan speaking populations: evidence from mtDNA. Am J Phys Anthropol 133:753–770.

Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B. 2002. The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Sci Commun4.

Mosquera-Miguel A, Álvarez-Iglesias V, Vega A, Milne R, Cabrera de León A, Benitez J, Carracedo Á, Salas A. 2008. Is mitochondrial DNA variation associated with sporadic breast cancer risk? Cancer Res 68:623–625.

Paredes M, Galindo A, Bernal M, Ávila S, Andrade D, Vergara C, Rincón M, Romero RE, Navarrete M, Cardenas M, et al. 2003. Analysis of the CODIS autosomal STR loci in four main Colombian regions. Forensic Sci Int 137:67–73.

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, et al. 1998. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 63:1839–1851.

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata M-J, Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Annals Hum Genet 65:439–458.

Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo Á, Comas D. 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. Hum Genet 115:439–447.

Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, Sica L, Mouguiama-Daouda P, Comas D, Tzur S, et al. 2007. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. Proc Natl Acad Sci USA 105: 1596–1601.

Rickards O, Martinez-Labarga C, Lum JK, De Stefano GF, Cann RL. 1999. mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations. Am J Hum Genet 65:519–530.

Rodas C, Gelvez N, Keyeux G. 2003. Mitochondrial DNA studies show asymmetrical Amerindian admixture in Afro-Colombian and Mestizo populations. Hum Biol 75:13–30.

Salas A, Bandelt HJ, Macaulay V, Richards MB. 2007. Phylogeographic investigations: the role of trees in forensic genetics. Forensic Sci Int 168:1–13.

Salas A, Carracedo Á. 2007. Studies of association in complex diseases: Statistical problems related to the analysis of genetic polymorphisms. Rev Clin Esp 207:563–565.

Salas A, Carracedo Á, Macaulay V, Richards M, Bandelt H-J. 2005a. A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. Biochem Biophys Res Commun 335:891–899.

Salas A, Carracedo Á, Richards M, Macaulay V. 2005b. Charting the ancestry of African Americans. Am J Hum Genet 77:676–680.

Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo Á. 1998. mtDNA analysis of the Galician population: a genetic edge of European variation. Eur J Hum Genet 6:365–375.

Salas A, Richards M, De la Fé T, Lareu MV, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo Á. 2002. The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111.

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465.

Salas A, Richards M, Lareu MV, Sobrino B, Silva S, Matamoros M, Macaulay V, Carracedo A. 2005c. Shipwrecks and founder effects: divergent demographic histories reflected in Caribbean mtDNA. Am J Phys Anthropol 128:855–860.

Salas A, Yao Y-G, Macaulay V, Vega A, Carracedo Á, Bandelt H-J. 2005d. A critical reassessment of the role of mitochondria in tumorigenesis. PLoS Med 2:e296.

Santos M, Barrantes R. 1994. D-loop mtDNA deletion as a unique marker of Chibchan Amerindians. Am J Hum Genet 55:413–414.
Silva WA, Bortolini MC, Schneider MP, Marrero A, Elion J, Krishnamoorthy R, Zago MA. 2006. MtDNA haplogroup analysis of black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade. Hum Biol 78:29–41.
Tagliabracci C, Turchi C, Buscemi L, Sassaroli C. 2001. Polymorphism of the mitochondrial DNA control region in Italians. Int J Legal Med 114:224–228.
Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, et al. 2007. Beringian standstill and spread of Native American founders. PLoS ONE 2:e829.
Torres MM, Bravi CM, Bortolini MC, Duque C, Callegari-Jacques S, Ortiz D, Bedoya G, Groot de Restrepo H, Ruiz-Linares A. 2006. A revertant of the major founder Native American haplogroup C common in populations from northern South America. Am J Hum Biol 18:59–65.
Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J. 2006. Harvesting the fruit of the human mtDNA tree. Trends Genet 22:339–345.
Vona G, Falchi A, Moral P, Calo CM, Varesi L. 2005. Mitochondrial sequence variation in the Guahibo Amerindian population from Venezuela. Am J Phys Anthropol 127:361–369.
Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, Santos RV. 1996. Mitochondrial DNA polymorphism in 3 Brazilian Indian tribes. Am J Hum Biol 8:317–323.
Watson E, Bauer K, Aman R, Weiss G, von Haeseler A, Pääbo S. 1996. mtDNA sequence diversity in Africa. Am J Hum Genet 59:437–444.

Artículo 8: Gender bias in the multi-ethnic genetic composition of Central Argentina (Córdoba)

Salas A, Jaime JC, Álvarez-Iglesias V, Carracedo Á

(Journal of Human Genetics; in press)

Sesgo por género en la composición genética multi-étnica de Argentina Central (Córdoba)

ABSTRACT/RESUMEN:

Se genotipó en una muestra de la región central de Argentina (Córdoba) la region hipervariable I (HVS-I) y un set de SNPs de la region codificante del AND mitocondrial (mtDNA) (N = 102) y se comparó con los datos de STRs de cromosoma Y (Y-STRs; N = 100), previamente genotipados en los mismos individuos. Además recopilamos una base de datos con más de 4,000, 6,800, and 12,000 secuencias de HVS-I de origen Nativo Americano, Africano sub-Sahariano, y Europeo respectivamente. Como referencia para los perfiles de Córdoba de Y-STRs se utilizó la base de datos YHRD. El componente Nativo Americano es altamente prevalente en la parte materna de la muestra (~41%) en contraste con la contribución al cromosoma Y de la parte paterna (~2%), lo que indica un fuerte sesgo en el género durante los procesos de colonización y mestizaje ocurridos en la historia reciente de Argentina, lo cual está en consonacia con los hechos históricos. La aportación demográfica de esclavos africanos en Córdoba fue muy elevada en el siglo XVIII (~40% del total de la población), pero decreció drásticamente en pocas décadas; los escasos linajes subsaharianos de cromososma Y y de ADNmt observados en nuestra muestra encajan por lo tanto, con los hechos cronológicos. El componente Europeo del cromosoma Y de Córdoba (97%) contrasta con el 57% observado en el ADNmt, y esto también refleja la inmigración experimentada en Argentina durante el principio del siglo pasado, especialmente provenientes de Italia y España.

J Hum Genet
DOI 10.1007/s10038-008-0297-8

ORIGINAL ARTICLE

Gender bias in the multiethnic genetic composition of central Argentina

A. Salas · J. C. Jaime · V. Álvarez-Iglesias · Á. Carracedo

Received: 20 February 2008 / Accepted: 15 April 2008

Abstract A sample of central Argentina (Córdoba) was genotyped for the first hypervariable region (HVS-I) plus a set of coding region mitochondrial DNA (mtDNA) single nucleotide polymorphisms (SNPs) ($N = 102$) and compared with a data set of Y-chromosome short tandem repeats (Y-STRs; $N = 100$) previously genotyped in the same individuals. We additionally compiled a database containing more than 4,000, 6,800, and 12,000 HVS-I sequences of Native American, sub-Saharan African, and European origin, respectively. The Y-Chromosome Haplotype Reference Database (YHRD) was used as a reference for the Y-STR profiles from Córdoba. The Native American component is highly prevalent on the maternal side (~41%) in contrast to the Y-chromosome paternal contribution (~2%), indicating a strong gender bias in the colonization and admixture processes that occurred in the recent history of Argentina, in agreement with historical records. The demographic input of African slaves in Córdoba was very high in the eighteenth century (~40% of the total population) but decreased dramatically after a few decades. Therefore, the minor traces of sub-Saharan Y-chromosome and mtDNA lineages observed in our sample fit with these historical records. The European Y-chromosome component of Córdoba (~97%; in contrast to the 57% observed in the mtDNA side) also mirrors the substantial immigration experienced by Argentina during the beginning of the last century, predominantly from Italy and Spain.

Keywords Coding region · SNP · Haplotype · Native American · Phylogeny

A. Salas · J. C. Jaime · V. Álvarez-Iglesias · Á. Carracedo
Unidade de Xenética, Instituto de Medicina Legal,
Facultad de Medicina,
Universidad de Santiago de Compostela,
15782 Galicia, Spain

A. Salas · V. Álvarez-Iglesias · Á. Carracedo
Grupo de Medicina Xenómica,
Hospital Clínico Universitario,
Universidad de Santiago de Compostela,
15706 Galicia, Spain

A. Salas (✉)
Unidade de Xenética Forense,
Universidade de Santiago de Compostela,
Galicia, Spain
e-mail: apimlase@usc.es

Introduction

The present population of central Argentina (Córdoba) is the result of a complex amalgamation of different cultures and populations with different genetic ancestries. Today, various Native American groups exist that inhabit different regions of Córdoba. A small population of Comechingones (actual population size of ~5,000) still live in a relatively isolated region of the mountain ranges of Córdoba, whereas the Olongastas, considered a subgroup of the Diaguitas (~6,000 individuals), live at lower altitudes in the northwest of the province. Other groups, such as the Sanavirones in the northeast and the Pampeanos (formerly occupying the low flat areas of the Argentinean humid *pampa*), have since become extinct. Although some of these populations have preserved their folklore and other cultural traditions relatively well, none have retained their original languages and now speak Spanish exclusively.

During colonial times, the use of slave labor for agricultural development became an economic necessity for Spanish colonialists. The relatively low density of Native Americans coupled with the resistance of these population

Journal : Large 10038 | Dispatch : 5-5-2008 | Pages : 13
Article No. : 297 | LE | TYPESET
MS Code : 297 | CP | DISK

groups to Spanish acculturation and slavery led to the introduction of very large numbers of sub-Saharan Africans to the region. Although the demographic impact of Africans on the autochthonous population in other Latin American countries such as Colombia and Brazil was extremely high and still persists in modern populations (Alves-Silva et al. 2000; Beleza et al. 2005; Ely et al. 2006; Parra et al. 2001; Salas et al. 2005b, c; 2004), their real demographic impact in Argentina is a debatable topic. Some scholars maintain that a very significant number of African slaves were forcedly moved to Argentina, in particular to regions of intense agricultural activity: according to Victoria-Gomes (2002), in 1778 the population of Africans in Córdoba was ∼44%, whereas in 1887 the official percentage of Africans was reduced to ∼1.8%. The reasons for such a dramatic decrease remain uncertain; Victoria-Gomes (2002) pointed to epidemic causes (e.g., yellow fever) and colonial fights against neighboring populations (the number of Africans recruited in the army was disproportionately large compared with other ethnic groups).

During the nineteenth century, Argentina experienced a large-scale immigration of Europeans, notably from Italy and Spain (but including UK and Germany), which dramatically changed the demography of the country, mainly in urban areas. There are other well-known minorities in this region, including Jews (escaping persecution in the Second World War), Arabs, Armenians, and Japanese. Since Córdoba is one of the most important industrial centers of the country, over the past 50 years, it has attracted numerous immigrants from all over the country and from neighboring countries (i.e., Bolivia, Paraguay, etc.).

Finally, none of the existing Argentinean Native American groups live a completely isolated existence, and the degree of admixture, in particular with individuals of primary European ancestry, seems to be high in some populations. However, the magnitude of this admixture and the final demographic impact of African slaves in the region has not been established.

We collected a sample of central Argentina (Córdoba) to estimate the different main ancestries that contributed to the present population. We analyzed mitochondrial DNA (mtDNA) variation by analyzing the first hypervariable region (HVS-I) and a set of mtDNA coding region single nucleotide polymorphisms (SNPs). The degree of admixture on the maternal side was also contrasted with analysis of a set of Y-chromosome short tandem repeats (Y-STR) markers genotyped in the same individuals (Fondevila et al. 2003), firstly by inferring the haplogroup status of these Y-STR profiles, and secondly by searching these profiles in the worldwide Y Chromosome Haplotype Reference Database (YHRD).

Material and methods

Samples

We collected 102 healthy unrelated individuals from the province of Córdoba in central Argentina. Informed consent was given by all participants. The protocol and procedures employed were reviewed and approved by the review committee of the University of Santiago de Compostela, Spain, where genotyping was carried out. All persons gave their informed consent prior to inclusion.

PCR amplification and sequencing

All samples were amplified and sequenced (forward and reverse) for the HVS-I, analyzing the sequence range 16024–16400. Polymerase chain reaction (PCR) amplification was performed with GeneAmp 9700 thermocyclers, and sequencing analysis was performed as previously described (Álvarez-Iglesias et al. 2007). All samples were additionally genotyped for a set of ten SNPs following (Quintáns et al. 2004). Mutations are referred to the revised Cambridge Reference Sequence (rCRS) (Andrews et al. 1999). We followed a standardized forensic framework for nomenclature, as indicated in Carracedo et al. (2000) but with slight modifications considered in Salas et al. (2005a) concerning insertions. Data were checked following the phylogenetic principles described previously (Bandelt 1994; Bandelt et al. 2004a, b; Salas et al. 2007) to avoid sequence artifacts as much as possible.

Databases and statistical analysis

For phylogeographic purposes, we collected different available population data sets from the literature. Thus, for African haplotypes, we used roughly the same database reported in Černý et al. (2007), which consists of 6,856 profiles from different regions of the African continent. For Native American lineages, an HVS-I database that consists of 4,086 available sequences in the literature was employed, and for the western European database, we considered >12,000 publicly available profiles.

Only the HVS-I segment was used for population comparison, in particular, the 16090–16365 sequence range, as this is the common segment for the different populations used. Nomenclature of African haplotypes follows Salas et al. (2002, 2004), with updates in Černý et al. (2007), Kivisild et al. (2004), and Torroni et al. (2006). For Native American haplogroups, we use the most updated nomenclature from Bandelt et al. (2003) and Kong et al. (2006), and for the European profiles, Achilli et al. (2005, 2004), Loogväli et al. (2004), and Sun et al. (2006), among others. DnaSP 4.10.3 software (Rozas et al. 2003) was used for computation of different diversity indices.

Journal : Large 10038 | Dispatch : 5-5-2008 | Pages : 13
Article No. : 297 | LE | TYPESET
MS Code : 297 | CP | DISK

Principal component analysis was performed using Stata 9.1 (http://www.stata.com/).

Y-chromosome haplogroup inference

The number of occurrences of particular profiles in a worldwide database roughly indicates their most natural geographical origin. Minimal Y-chromosome haplotypes for 100 individuals were reported in Fondevila et al. (2003) and submitted to the YHRD (http://www.yhrd.org/index.html). We searched the YHRD for each of these profiles (YHRD; release 22) with particular focus on the main European source populations of Spain and Italy together with the major continental regions (e.g., Europe, Latin America, etc.), keeping the population grouping scheme provided by the YHRD. Haplogroup pPredictor (https://home.comcast.net/~hapest5/index.html) was used for inferring haplogroup status of Y-STR profiles using flat a priori probabilities.

Results and discussion

Phylogeography of mtDNA HVS-I profiles

The central Argentinean population has two well-differentiated mtDNA ancestral components: ~57% of the lineages are of European origin (predominantly from western Europe), and ~41% are typically Native American (Fig. 1). Only two individuals carried mtDNA of sub-Saharan provenance. About 31% of European lineages belong to haplogroup H, but there are representatives of other common European haplogroups (HV0, I, J, etc.), broadly reflecting the haplogroup spectra of a typical European population (Table 1). At the haplotypic level, the degree of molecular resolution does not allow the phylogeographic allocation of these European lineages to particular regions of Europe; in fact, most of these HVS-I profiles in Córdoba occur across the whole of Europe. There are only few outlined exceptions; for instance, we did not find matches for the haplogroup H profile C16111T T16209C C16270T. Some lineages have some more restricted geographical occurrence. For example, there are seven matches for the haplogroup W sequence C16173T C16223T C16292T T16325C T16352C; curiously, all of them are in Romania (Brandstätter et al. 2007) and Georgia (Quintana-Murci et al. 2004).

With regard to the Native American component, we observed representatives of the four main haplogroups, A2 (~8%), B2 (~5%), C1 (~14%), and D1 (~14%). Five out of nine of the A2 haplotypes match with the basal A2

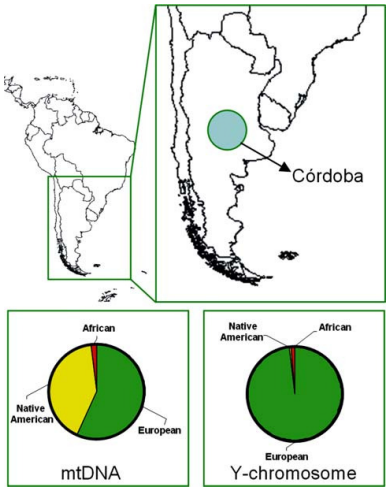


Fig. 1 Location of the sample analyzed and main ancestry of Y–chromosome and mitochondrial DNA lineages

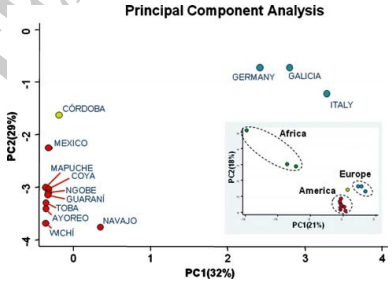


Fig. 2 Principle component analysis (PCA) based on mitochondrial DNA (mtDNA) haplogroup frequencies. The main PCA plot includes several European and Native American samples plus the one from Córdoba analyzed in this study, whereas the nested small PCA plot includes three additional sub-Saharan African samples. References for population samples are as follows: **a** Africa (*green dots*): Mozambique (Salas et al. 2002), Angola (Plaza et al. 2004), Cabinda (Beleza et al. 2005); **b** Europe (*blue dots*): Italy (Bini et al. 2003), Galicia from Spain (Salas et al. 1998), and Germany (Lutz et al. 1998); and **c** America (*red dots*): Mexico (Green et al. 2000), Mapuche from Chile (Moraga et al. 2000), Guarani Kaiowá from Brazil (Marrero et al. 2007), Coya (Alvarez-Iglesias et al. 2007), Toba and Wichí from Argentina (Cabana et al. 2006), Navajo (Monson et al. 2002), Ayoreo from Bolivia and Paraguay (Dornelles et al. 2004), Ngobe from Panamá (Kolman et al. 1995), Córdoba (this study)

Journal : Large 10038 | Dispatch : 5-5-2008 | Pages : 13
Article No. : 297 | LE | TYPESET
MS Code : 297 | CP | DISK

Table 1 First hypervariable region (HVS-I) and mitochondrial DNA (mtDNA) single nucleotide polymorphism (SNP) profiles in 102 Argentineans from Córdoba

Sample ID	HVS-I (minus 16000)	HG	Start	End	mtDNA SNPs									
					4216T	4529A	4580G	7028C	10398A	10400C	10873T	12308A	14766C	14766C
1	051 086 092 111 223 290 319	A2	16024	16400	–	–	–	T	–	–	–	–	T	T
2	092 111 172 218 223 290 319 362	A2	16024	16390	–	–	–	T	–	–	–	–	T	T
3	111 172 223 290 319 362	A2	16024	16390	–	–	–	T	–	–	–	–	T	T
4	111 172 223 290 319 362	A2	16024	16390	–	–	–	T	–	–	–	–	T	T
5	111 223 290 319 362	A2	16024	16400	–	–	–	T	–	–	–	–	T	T
6	111 223 290 319 362	A2	16024	16400	–	–	–	T	–	–	–	–	T	T
7	111 223 290 319 362	A2	16024	16400	–	–	–	T	–	–	–	–	T	T
8	111 223 290 319 362	A2	16024	16390	–	–	–	T	–	–	–	–	T	T
9	111 223 290 319 362	A2	16024	16383	–	–	–	T	–	–	–	–	T	T
10	093 183C 189 217 359	B2	16024	16390	–	–	–	T	–	–	–	–	–	T
11	142 183C 189 217	B2	16024	16400	–	–	–	T	–	–	–	–	–	T
12	153 183C 189 217	B2	16024	16390	–	–	–	T	–	–	–	–	–	T
13	173 182C 183C 189 217	B2	16024	16390	–	–	–	T	–	–	–	–	–	T
14	183C 189 217	B2	16024	16390	–	–	–	T	–	–	–	–	–	T
15	**051** 093 223 **259 293C** 298 325 327 357	C1	16024	16380	–	–	–	T	G	T	C	–	T	T
16	051 129 223 298 325 327	C1	16024	16370	–	–	–	T	G	T	C	–	T	T
17	**071** 223 298 311 325 327 **368**	C1	16024	16390	–	–	–	T	G	T	C	–	T	T
18	092 223 298 325 327	C1	16024	16383	–	–	–	T	G	T	C	–	T	T
19	092 223 298 325 327	C1	16024	16400	–	–	–	T	G	T	C	–	T	T
20	154 223 298 325 327	C1	16024	16390	–	–	–	T	G	T	C	–	T	T
21	183C 189 223 298 325 327	C1	16024	16390	–	–	–	T	G	T	C	–	T	T
22	223 234 298 325 327	C1	16024	16390	–	–	–	T	G	T	C	–	T	T
23	223 298 325 327	C1	16024	16390	–	–	–	T	G	T	C	–	T	T
24	223 298 325 327	C1	16024	16400	–	–	–	T	G	T	C	–	T	T
25	223 298 325 327	C1	16024	16390	–	–	–	T	G	T	C	–	T	T
26	223 298 325 327	C1	16024	16400	–	–	–	T	G	T	C	–	T	T
27	223 298 325 327 335	C1	16024	16400	–	–	–	T	G	T	C	–	T	T
28	223 298 325 327 **368**	C1	16024	16390	–	–	–	T	G	T	C	–	T	T
29	093 187 189 209 223 325 362	D1	16024	16383	–	–	–	T	G	T	C	–	T	T
30	126 223 325 362	D1	16024	16380	–	–	–	T	G	T	C	–	T	T
31	126 223 325 362	D1	16024	16394	–	–	–	T	G	T	C	–	T	T
32	187 189 209 223 325 362	D1	16024	16383	–	–	–	T	G	T	C	–	T	T
33	187 189 209 223 325 362	D1	16024	16390	–	–	–	T	G	T	C	–	T	T
34	189 223 325 362	D1	16024	16390	–	–	–	T	G	T	C	–	T	T
35	190 223 325 362	D1	16024	16390	–	–	–	T	G	T	C	–	T	T
36	223 242 311 325 362	D1	16024	16390	–	–	–	T	G	T	C	–	T	T
37	223 242 311 325 362	D1	16024	16400	–	–	–	T	G	T	C	–	T	T
38	223 242 311 325 362	D1	16024	16400	–	–	–	T	G	T	C	–	T	T
39	223 242 311 325 362	D1	16024	16390	–	–	–	T	G	T	C	–	T	T
40	223 242 311 325 362	D1	16024	16390	–	–	–	T	G	T	C	–	T	T
41	223 242 311 325 362	D1	16024	16383	–	–	–	T	G	T	C	–	T	T
42	223 242 311 325 362	D1	16024	16383	–	–	–	T	G	T	C	–	T	T
43	093	H	16024	16383	–	–	–	–	–	–	–	–	–	–

Journal : Large 10038 Dispatch : 5-5-2008 Pages : 13
Article No. : 297 LE TYPESET
MS Code : 297 CP DISK

Table 1 continued

Sample ID	HVS-I (minus 16000)	HG	Start	End	mtDNA SNPs									
					4216T	4529A	4580G	7028C	10398A	10400C	10873T	12308A	12705C	14766C
44	189	H	16024	16384	–	–	–	–	–	–	–	–	–	–
45	189	H	16024	16390	–	–	–	–	–	–	–	–	–	–
46	189	H	16024	16400	–	–	–	–	–	–	–	–	–	–
47	189	H	16024	16390	–	–	–	–	–	–	–	–	–	–
48	304	H	16024	16380	–	–	–	–	–	–	–	–	–	–
49	304	H	16024	16390	–	–	–	–	–	–	–	–	–	–
50	362	H	16024	16390	–	–	–	–	–	–	–	–	–	–
51	111 209 270	H	16024	16390	–	–	–	–	–	–	–	–	–	–
52	172 192	H	16024	16390	–	–	–	–	–	–	–	–	–	–
53	189 301	H	16024	16390	–	–	–	–	–	–	–	–	–	–
54	193 219 362	H	16024	16383	–	–	–	–	–	–	–	–	–	–
55	293 311	H	16024	16390	–	–	–	–	–	–	–	–	–	–
56	rCRS	H	16024	16400	–	–	–	–	–	–	–	–	–	–
57	rCRS	H	16024	16395	–	–	–	–	–	–	–	–	–	–
58	rCRS	H	16024	16390	–	–	–	–	–	–	–	–	–	–
59	rCRS	H	16024	16383	–	–	–	–	–	–	–	–	–	–
60	rCRS	H	16024	16383	–	–	–	–	–	–	–	–	–	–
61	068 288	H8	16024	16380	–	–	–	–	–	–	–	–	–	–
62	172 311	HV	16024	16384	–	–	–	T	–	–	–	–	–	–
63	298	HV0	16024	16400	–	–	–	T	–	–	–	–	–	–
64	298	HV0	16024	16390	–	–	–	T	–	–	–	–	–	–
65	298	HV0	16024	16380	–	–	–	T	–	–	–	–	–	–
66	298	HV0	16024	16384	–	–	–	T	–	–	–	–	–	–
67	093 221 298	HV0	16024	16383	–	–	–	T	–	–	–	–	–	–
68	189 **235** 298	HV0	16024	16390	–	–	–	T	–	–	–	–	–	–
69	218 239A 298	HV0	16024	16390	–	–	–	T	–	–	–	–	–	–
70	093 129 223 391	I	16024	16380	–	–	–	T	G	–	–	–	T	T
71	129 148 192 223 294	I	16024	16383	–	T	–	T	G	–	–	–	T	T
72	129 192 223	I	16024	16383	–	–	–	T	G	–	–	–	T	T
73	069 093 126	J	16024	16390	C	–	–	T	G	–	–	–	–	T
74	069 126	J	16024	16380	C	–	–	T	G	–	–	–	–	T
75	069 126	J	16024	16380	C	–	–	T	G	–	–	–	–	T
76	069 126 183C 189	J	16024	16380	C	–	–	T	G	–	–	–	–	T
77	069 126 189	J	16024	16390	C	–	–	T	G	–	–	–	–	T
78	069 126 193 278	J	16024	16380	C	–	–	T	G	–	–	–	–	T
79	069 126 274	J	16024	16400	C	–	–	T	G	–	–	–	–	T
80	069 126 145 231 261	J1a	16024	16400	C	–	–	T	G	–	–	–	–	T
81	069 126 145 231 261	J1a	16024	16390	C	–	–	T	G	–	–	–	–	T
82	069 126 145 222 256 261 278	J1b	16024	16400	C	–	–	T	G	–	–	–	–	T
83	093 224 234 311	K	16024	16400	–	–	–	T	–	–	–	G	–	T
84	224 234 311	K	16024	16390	–	–	–	T	–	–	–	G	–	T
85	224 293 311	K1	16024	16390	–	–	–	T	G	–	–	G	–	T
86	224 311	K1	16024	16400	–	–	–	T	G	–	–	G	–	T

Springer

Journal : Large 10038 Dispatch : 5-5-2008 Pages : 13
Article No. : 297 LE TYPESET
MS Code : 297 CP DISK

Table 1 continued

Sample ID	HVS-I (minus 16000)	HG	Start	End	mtDNA SNPs 4216T	4529A	4580G	7028C	10398A	10400C	10873T	12308A	12705C	14766C
87	224 311	K1	16024	16383	–	–	–	T	G	–	–	G	–	T
88	224 311	K1	16024	16390	–	–	–	T	G	–	–	G	–	T
89	126 187 189 193 223 264 270 278 293 311	L1b1	16024	16390	–	–	–	T	G	–	C	–	T	T
90	129 163 187 189 209 223 278 293 294 311 360	L1c1	16024	16386	–	–	–	T	G	–	C	–	T	T
91	126 172 294 304	T	16024	16383	C	–	–	T	–	–	–	–	–	T
92	086 126 163 189 294	T1	16024	16380	C	–	–	T	–	–	–	–	–	T
93	126 163 186 189 294	T1	16024	16390	C	–	–	T	–	–	–	–	–	T
94	126 163 186 189 294	T1	16024	16391	C	–	–	T	–	–	–	–	–	T
95	126 163 186 189 294	T1	16024	16384	C	–	–	T	–	–	–	–	–	T
96	126 294 296 304	T2	16024	16390	C	–	–	T	–	–	–	–	–	T
97	192 256 270 362	U5a	16024	16390	–	–	–	T	–	–	–	G	–	T
98	239 256 270	U5a	16024	16390	–	–	–	T	–	–	–	G	–	T
99	146 162 342	U8a	16024	16380	–	–	–	T	–	–	–	G	–	T
100	173 223 292 325 352	W	16024	16400	–	–	–	T	–	–	–	–	T	T
101	223 292	W	16024	16383	–	–	–	T	–	–	–	–	T	T
102	111 189 223 278	X	16024	16400	–	–	–	T	–	–	–	–	T	T

A dash in the SNP columns indicates a match with the revised Cambridge Reference Sequence (rCRS) (Andrews et al. 1999)

haplotype, which is common across the American continent. There are four instances of the basal C1 haplotype in Córdoba, also common in populations living at different continental latitudes. In contrast, the basal HVS-I haplotype of D1 (T16362C T16325C) was not detected in our sample. However, we found 11 identical matches of the most common D1 haplotype C16223T C16242T T16311C T16325C T16362C, ten of them in other Argentinean populations: four matches in the Mapuches (Ginther et al. 1993) and the rest in the Pilagá and Wichí (Cabana et al. 2006) and in the Coyas (Álvarez-Iglesias et al. 2007). Two other matches were also observed in the Genographic database (https://www3.nationalgeographic.com/genographic/resources.html). There are other interesting matches for the D1 haplotypes of Córdoba. Haplotype C16187T T16189C T16209C C16223T T16325C T16362C appears four times in the database, three of them in the aboriginal and in the general Chilean population (Horai et al. 1993; Moraga et al. 2000) and one in the ancient Kaweskar DNA samples (Patagonia-Tierra del Fuego; South of Argentina, and Chile) studied by García-Bour et al. (2004).

As shown in Table 2, for the sequence range 16090–16385, there are 24 (out of 42) different Native American haplotypes in Córdoba. Most of them ($N = 17$) are only observed once, four sequences appear twice, two sequences occur five times, and one haplotype appears seven times (C16223T C16242T T16311C T16325C T16362C). Sequence (S) and nucleotide (π) diversities and average number of nucleotide diversity (M) are quite high for the Native American component ($S = 0.948$; $\pi = 0.01767$; $M = 6.04$) in contrast to the average values for the continental Native American component ($S = 0.945$; $\pi = 0.01519$; $M = 3.43$), indicating that the admixture process with Europeans was relatively gradual, allowing the preservation of a significant part of the Native American original gene pool in today's general population of Córdoba. There are 13 identical matches between the Native American lineages and South America: nine in central America and ten in North America. However, six haplotypes are found across America (North, Central, and South) (Table 2). Moreover, there are nine haplotypes with only one representative in Córdoba that were still not observed in our large database.

We only detected two sequences of sub-Saharan origin belonging to haplogroups L1b1 and L1c1. For the L1b1 sequence (see Table 1), we did not find exact matches, but there are some one-step-mutation neighbors—such as A16166G on top of T16126C C16187T T16189C C16193T

Journal : Large 10038 | Dispatch : 5-5-2008 | Pages : 13
Article No. : 297 | ☐ LE | ☐ TYPESET
MS Code : 297 | ✔ CP | ✔ DISK

Table 2 Córdoba first hypervariable region (HVS-I) Native American sequences (haplogroups A2, B2, C1, and D1) matches in the main American continental regions (North, Central, South)

HVS-I sequences	Haplogroup	North ($N = 2,005$)	Central ($N = 485$)	South ($N = 1,596$)	Córdoba ($N = 42$)	Total ($N = 4,128$)
092 111 172 218 223 290 319 362	A2	–	–	–	1	1
092 111 223 290 319	A2	–	–	–	1	1
111 172 223 290 319 362	A2	1	1	–	2	4
111 223 290 319 362	A2	161	45	50	5	261
093 189 217 359	B2	–	–	–	1	1
142 189 217	B2	–	–	–	1	1
153 189 217	B2	–	–	–	1	1
173 189 217	B2	–	–	–	1	1
189 217	B2	24	38	90	1	153
092 223 298 325 327	C1	–	1	1	2	4
093 223 259 293A 298 325 327 357	C1	–	–	–	1	1
129 223 298 325 327	C1	5	3	10	1	1
154 223 298 325 327	C1	–	–	1	1	2
189 223 298 325 327	C1	4	–	1	1	6
223 234 298 325 327	C1	3	–	–	1	4
223 298 311 325 327	C1	5	4	4	1	14
223 298 325 327	C1	111	19	192	5	327
223 298 325 327 335	C1	1	–	1	1	3
093 187 189 209 223 325 362	D1	–	–	–	1	1
126 223 325 362	D1	–	–	–	2	2
187 189 209 223 325 362	D1	–	–	4	2	6
189 223 325 362	D1	3	7	8	1	19
190 223 325 362	D1	–	–	1	1	2
223 242 311 325 362	D1	–	1	5	7	13
Totals		318	119	368	42	829

Only the sequence range 16090–16365 was considered here. Length variation around 16189 and variants 16182C and 16183C were omitted from the comparison

C16223T C16264T C16270T C16278T A16293G T16311C, which is found in Cabinda (Beleza et al. 2005)—in the Bakaka of south Cameroon (Coia et al. 2005) but also in the Tacuarembó from Uruguay. The typical central African L1c1 lineage G16129A A16163G C16187T T16189C T16209C C16223T C16278T A16293G C16294T T16311C C16360T also appears at moderate frequencies in Cabinda (Beleza et al. 2005) and other Latin-American countries.

Principal component analysis

We carried out a principal component analysis (PCA) based on mtDNA haplogroup frequencies and using several European, sub-Saharan African and two Native American data sets as population references (see Fig. 2 legend). PC1 (which accounts for 21% of the variability) primarily separates African from non-African populations, whereas PC2 (18%) basically splits European from Native American samples (see nested plot in Fig. 1). The plot also shows a clear-cut heterogeneity pattern in the African samples. In contrast, the Native American groups are tightly grouped in a single cluster, reflecting the fact that most of them essentially carry Native American lineages. A second round of PC analysis (the two first PCs accounting for 61% of the variability) was carried out excluding the African samples (main plot in Fig. 2). PC1 (32%) shows in one pole the European samples and in the other extreme the Native Americans ones. The European component of the Córdoba sample (57%) is clearly reflected in the PC2 (29%), as it is also the case for the Mexican data set (Green et al. 2000), which also has an important European mtDNA background ($\sim 15\%$).

The most likely ancestry of Y-STR profiles

Inference of the haplogroup status based on the Bayes approach implemented in Haplogroup Predictor showed

Journal : Large 10038 Dispatch : 5-5-2008 Pages : 13
Article No. : 297 LE TYPESET
MS Code : 297 CP DISK

Table 3 Inference of haplogroup status of Y-chromosome short tandem repeat (Y-STR) profiles in central Argentineans and number of matches in the Y-Chromosome Haplotype Reference Database (YHRD)

Sample	Y-STRs								No.	HG	Prob.	Worldwide	Europe	Spain	Italy	Latin-Am.	Asia	Africa
ID[a]	319	385	3891	3891I	390	391	392	393				$N = 5,0867$	$(N = 26,305)$	$(N = 1,708)$	$(N = 2,285)$	$(N = 7,035)$	$(N = 11,439)$	$(N = 2,750)$
01	14	11–14	14	30	24	11	13	13	1	R1b	99.9	313	154	47	9	114	–	4
02	14	12–14	14	30	24	11	13	14	1	R1b	86.7	2	1	1	–	1	–	–
03	14	11–14	13	30	24	11	13	13	3	R1b	100	196	105	21	11	55	1	6
04	15	11–14	13	28	24	10	11	13	1	R1a	89.1	3	–	–	–	3	–	–
05	15	14–16	13	29	23	11	13	13	1	I1b2a	33.9	1	–	–	–	1	–	–
06	14	11–14	13	29	23	10	13	13	1	R1b	100	189	139	11	8	29	–	5
07	14	11–14	13	29	24	11	13	13	4	R1b	100	1045	637	126	70	276	3	33
08	14	11–11	13	29	24	11	13	13	1	R1b	100	64	45	4	5	8	1	–
09	15	14–17	14	30	23	10	11	12	1	J2	58.4	4	2	–	–	1	1	–
10	14	13–20	13	30	23	11	11	12	1	J1	97.9	4	–	–	–	2	–	2
11	13	16–18	13	30	25	10	11	13	1	E3b	100	35	32	1	5	3	–	–
12	14	11–14	14	30	24	11	14	13	1	R1b	53.1	14	7	1	1	7	–	–
13	13	16–16	13	30	24	10	11	13	1	E3b	99.9	21	11	3	2	6	–	–
14	14	11–14	13	30	24	10	13	13	1	R1b	99.9	94	47	4	4	19	2	4
15	15	11–14	13	29	24	10	13	12	1	R1b	99.8	8	5	1	–	2	1	
16	14	11–12	13	29	24	11	13	13	1	R1b	99.9	14	7	1	1	5	–	–
17	14	11–14	12	28	23	10	11	13	1	I1a	96.4	2	1	–	1	1	–	–
18	15	14–14	13	29	23	11	13	13	1	R1b	96.3	1	–	–	–	1	–	–
19	17	11–14	13	31	25	11	11	13	1	R1a	100	35	28	–	–	2	–	4
20	14	11–14	13	29	23	11	13	13	3	R1b	100	443	325	20	16	49	1	15
21	14	13–14	12	28	23	10	10	13	1	I1a	99.5	1	–	–	–	1	–	–
22	16	12–14	12	29	22	10	10	12	1	G2	97.6	1	–	–	–	1	–	–
23	15	12–13	13	29	23	10	13	13	1	R1b	99.6	1	–	–	–	1	–	–
24	14	11–15	14	30	24	11	13	13	1	R1b	100	93	49	19	6	18	15	2
25	14	12–14	13	29	23	10	13	13	1	R1b	99.5	22	17	2	–	4	–	–
26	14	11–14	13	28	24	10	14	13	1	R1b	98.4	5	3	–	1	1	–	–
27	13	13–15	14	30	23	10	13	12	1	J1	84.1	1	–	–	–	1	–	–
28	16	13–17	12	28	25	10	11	12	1	J2	98.6	4	1	–	–	2	1	–
29	14	12–14	13	29	24	11	13	13	2	R1b	100	126	79	17	9	27	6	
30	17	12–12	14	29	25	9	10	13	1	I1b1	100	1	–	–	–	1	–	–
31	14	14–16	12	30	22	10	10	15	1	G2	99.4	1	–	–	–	1	–	–
32	14	12–13	13	29	24	11	13	13	1	R1b	99.8	33	20	2	4	11	–	2
33	14	11–14	13	29	24	11	13	14	1	R1b	99.3	57	32	5	4	18	–	1
34	15	14–15	12	29	23	10	10	15	1	G2	98	1	–	–	–	1	–	–

Journal : Large 10038 | Dispatch : 5-5-2008 | Pages : 13
Article No. : 297 | LE | TYPESET
MS Code : 297 | CP | DISK

Table 3 continued

Sample	Y-STRs									No.	HG	Prob.	Worldwide	Europe	Spain	Italy	Latin-Am.	Asia	Africa
ID[a]	319	385	389I	389II	390	391	392	393					*N* = 5,0867	(*N* = 26,305)	(*N* = 1,708)	(*N* = 2,285)	(*N* = 7,035)	(*N* = 11,439)	(*N* = 2,750)
35	14	10–13	12	28	24	10	14	13	1	R1b	85.5	1	–	–	–	1	–	–	–
36	17	12–13	13	28	23	10	11	13	1	I1b1	99.8	3	1	1	–	1	1	–	–
37	15	12–16	12	29	22	10	11	13	1	G2	90.4	3	–	–	–	2	–	–	–
38	14	13–13	13	30	23	11	11	12	1	J1	89.2	3	–	–	–	2	1	–	–
39	15	11–15	13	29	24	10	13	12	1	R1b	99.4	2	–	–	–	1	–	–	–
40	14	11–16	13	29	24	11	13	13	1	R1b	100	45	23	3	4	16	–	–	–
41	14	12–13	14	30	23	10	11	12	1	J1	52.3	1	–	–	–	1	–	–	–
42	14	10–10	13	30	24	12	12	13	1	R1b	100	1	–	–	–	1	–	–	–
43	14	11–14	14	30	23	11	13	13	1	R1b	99.6	101	55	3	4	25	–	13	–
44	17	13–16	12	28	25	10	11	13	1	J2	56.4	2	1	–	–	1	–	–	–
45	16	10–14	13	30	25	10	11	13	1	R1a	100	83	77	–	–	1	4	–	–
46	17	11–13	13	30	25	10	11	13	1	R1a	99.8	10	8	–	1	1	1	–	–
47	15	13–16	13	29	23	9	11	13	1	J2	87.5	7	5	1	–	2	–	–	–
48	14	11–11	14	31	24	11	11	13	1	R1a	82.1	1	–	–	–	1	–	–	–
49	15	14–14	12	29	22	10	13	14	1	G2	99.4	1	–	–	–	1	–	–	–
50	15	13–14	14	31	23	11	11	12	1	J2	85.3	7	–	–	–	1	–	–	–
51	14	11–14	13	31	24	10	13	14	1	R1b	97.9	5	1	–	–	4	–	–	–
52	17	11–12	13	28	23	10	11	13	2	I1b1	99.8	6	4	3	–	2	–	–	–
53	15	11–14	13	29	24	11	13	13	1	R1b	99.6	148	92	12	17	46	–	3	–
54	17	12–12	13	28	24	10	13	13	1	I1b1	100	1	–	–	–	1	–	–	–
55	17	9–18	13	30	21	10	11	15	1	E3a	95.8	1	–	–	–	1	–	–	–
56	14	13–15	13	29	24	10	13	13	1	R1b	66	4	2	–	–	1	–	–	–
57	15	11–12	13	29	24	11	13	13	1	R1b	99.5	5	–	–	–	4	–	–	–
58	14	13–15	13	30	23	10	11	12	1	J1	95	39	22	1	2	13	3	1	–
59	14	11–14	13	30	24	12	13	13	1	R1b	100	11	4	–	–	7	–	–	–
60	13	16–18	12	29	24	10	11	12	1	E3b	98.5	1	–	–	–	1	–	–	–
61	14	11–15	13	29	27	11	13	13	1	R1b	89.2	1	–	–	–	1	–	–	–
62	13	15–18	13	30	24	10	11	13	1	E3b	99.9	39	28	2	3	7	–	–	–
63	15	11–14	13	28	24	11	13	13	1	R1b	99.8	11	4	3	–	5	–	–	–
64	15	13–16	13	29	23	9	11	12	1	J2	95.8	59	41	2	7	8	4	–	–
65	13	16–18	13	30	23	10	11	13	1	E3b	99.9	20	12	–	6	4	1	1	–
66	13	11–14	14	30	25	12	13	13	1	R1b	99.6	1	–	–	–	1	–	–	–
67	14	12–12	13	29	24	12	13	13	1	R1b	99.7	1	–	–	–	1	–	–	–
68	14	11–14	15	31	24	10	13	13	1	R1b	99.9	4	–	–	–	3	–	–	–
69	14	11–15	13	30	24	10	13	13	1	R1b	99.9	33	22	3	1	6	–	1	–

Journal : Large 10038 Dispatch : 5-5-2008 Pages : 13
Article No. : 297 LE TYPESET
MS Code : 297 CP DISK

Table 3 continued

Sample	Y-STRs								No.	HG	Prob.	Worldwide	Europe	Spain	Italy	Latin-Am.	Asia	Africa
ID[a]	319	385	389I	389II	390	391	392	393				*N* = 5,0867	(*N* = 26,305)	(*N* = 1,708)	(*N* = 2,285)	(*N* = 7,035)	(*N* = 11,439)	(*N* = 2,750)
70	14	12–15	13	30	24	11	13	13	1	R1b	99.6	10	4	–	–	4	–	–
71	17	11–14	13	30	25	11	11	13	1	R1a	100	43	35	–	–	1	6	1
72	14	11–13	14	30	24	11	13	13	1	R1b	97.5	21	12	4	–	8	–	–
73	13	16–18	13	29	24	10	11	13	1	E3b	99.9	16	10	–	3	4	1	–
74	15	12–14	12	28	21	10	11	14	1	G2	99.4	1	–	–	–	1	–	–
75	14	11–14	13	28	23	11	13	15	1	R1b	99	1	–	–	–	1	–	–
76	15	13–17	12	29	22	10	11	14	1	G2	98.3	3	2	–	2	1	–	–
77	17	11–13	13	30	25	11	11	13	1	R1a	99.7	22	17	–	–	1	2	1
78	14	16–16	13	30	24	10	15	13	1	Q	93.7	1	–	–	–	1	–	–
79	14	13–17	13	29	23	10	11	12	1	J1	77.9	24	17	2	4	3	3	–
80	14	14–14	12	28	22	10	11	13	1	I1a	99.5	151	138	1	1	7	–	1
81	14	11–14	12	27	24	10	13	13	1	R1b	100	5	2	–	–	2	–	1
82	15	11–14	13	29	25	11	11	13	1	R1a	99.9	44	35	–	–	1	6	–
83	15	14–15	12	29	23	10	11	15	1	G2	96.5	4	–	–	–	4	–	–
84	14	12–18	13	30	23	10	11	12	1	J1	95.9	4	1	–	1	1	2	–
85	14	11–14	12	28	25	10	13	14	1	R1b	99.7	2	1	–	–	1	–	–
86	14	11–15	13	29	24	11	13	13	1	R1b	100	279	179	22	18	44	8	7
87	14	11–14	12	28	24	10	13	13	1	R1b	100	29	16	3	3	10	–	–
88	14	11–16	13	29	24	11	13	14	1	R1b	99.7	3	1	–	–	2	–	–
89	16	14–14	12	28	22	10	11	15	1	G2	96.8	1	–	–	–	1	–	–
90	15	14–15	12	29	22	10	11	14	1	G2	99.7	40	24	2	3	4	2	6
91	14	13–18	13	30	23	10	11	12	1	J1	96.7	44	9	–	1	11	12	8

HG haplogroup, *Prob.* probability

[a] Note that the haplotypes considered here have already been reported in Y-Chromosome Haplotype Reference Database (YHRD) (Fondevila et al. 2003). Sample ID corresponds with those used in Fondevila et al. (2003)

Journal : Large 10038 | Dispatch : 5-5-2008 | Pages : 13
Article No. : 297 | LE | TYPESET
MS Code : 297 | CP | DISK

that ~70% of the profiles yielded a posteriori probabilities >0.98 in a unique and well-defined haplogroup (Table 3). In contrast to the mtDNA variation, most Y-STR lineages (~97%) can be allocated to typical European haplogroups, whereas only a small fraction correspond to lineages of likely Native American (~2%) and African ancestry (~1%) (Fig. 1). The most common European haplogroup is R1b (42%). This lineage is supposed to have spread into the rest of Europe from Iberian and other southern European refugia after the Last Glacial Maximum and today is the most frequent Y-chromosome lineage in Europe. Some other typical Mediterranean clades are also present in our central Argentinean sample. For instance, J2 is of near-Eastern origin but today it is prevalent across Mediterranean coastal regions, including the Iberian Peninsula. Its sister clade, J1, is probably of a more southern European origin and it is now common in the Mediterranean coastal regions. G2 is the most common G lineage in western Europe and makes up 8–10% of several Mediterranean populations (Spain, Italy, Greece, and Turkey).

Y-STRs profiles were also searched in the YHRD to investigate the number of times a particular profile was previously observed in other worldwide populations. This procedure provides additional indication regarding their most natural geographical origin. Strikingly, 25% of the Y-STR profiles do not find a single match in the YHRD. In agreement with their inferred haplogroup status, most haplotypes are more frequently observed in Europe or in populations with important European ancestry (e.g., USA). In agreement with historical documentation, a substantial number of profiles, 29% and 28%, have a higher frequency in Spain and Italy, respectively. For instance, profile 14/11–14/13/30/24/11/13/13 (inferred haplogroup status = R1b) appears three times in Córdoba with 25 matches in Spain (1.5%). The number of profiles that match certain central European samples is relatively high; however, this may just reflect the fact that the YHRD is substantially enriched with profiles obtained from specific areas (the YHRD European data set comprises one third of profiles obtained from Germany and Poland).

Haplotype 27 is of Native American ancestry (a posteriori probability for its haplogroup Q status = 0.94). Profile 5 has a more ambiguous haplogroup allocation (e.g., a posteriori probability of belonging to haplogroup Q = 0.17). When searching the YHRD, we did not find a single match in the whole database; however, a total of seven one-step mutation derivatives are only observed in Latin Americans. Finally, haplotype 55 belongs to the sub-Saharan haplotype E3a (a posteriori probability for its haplogroup status = 0.96), and there are neither matches nor one-step mutation profiles in the YHRD (which could again reflect sampling bias due to the limited presence of sub-Saharan samples in the database).

Final remarks

There is a clear gender bias in the mtDNA and Y-chromosome composition of central Argentina. PCA, based on mtDNA haplogroup frequencies, with the two first principal components accounting for 61% of the variability), together with phylogeographic inferences, clearly indicates the halfway position of the Córdoba population between Europeans and Native Americans. In contrast, most Y-chromosomes in the general population of Córdoba are of European origin, with evidence for an important input from Italy and Spain. This is in agreement with historical records indicating that between 1869 and 1991, the average contribution of Italians and Spanish was 34% and 22% of total newcomers, respectively (source: INDEC, Instituto Nacional de Estadísticas y Censo from Argentina; http://www.indec.gov.ar). The presence of mtDNA and Y-chromosome lineages of sub-Saharan origin in Córdoba is low (<2%) but fits well with the demographic inferences of another study (Victoria-Gomes 2002).

Our results contrast with those obtained for other Argentinean populations. For instance, Dipierri et al. (1998) show that the Native American component in two northwestern Argentinean populations is ~65%, the introgression being more evident on the Y-chromosome side with frequencies of ~28% in Quebrada de Humahuaca and ~64% in San Salvador de Jujuy. On the other hand, the population of La Plata (Argentina) shows a Native American component of ~46% and a paternal contribution of ~11% (Martinez Marignac et al. 2004). In the very isolated Argentinean village of Acuña (Bailliet et al. 2001), mtDNA is mainly Native American, whereas the Y-chromosome part is essentially European. As in Córdoba, all the above-mentioned studies clearly indicate directional mating in Argentina. Moreover, all these studies indicate Argentina shows a clear pattern of population substructure on the specific maternal and paternal genomes, which also corroborates the findings of previous studies based on autosomal STR markers (Toscanini et al. 2006). The forensic field and medical genetic studies will benefit from population studies across the Argentinean territory that would allow detailed knowledge of population structure and its consequences when estimating the weight of forensic haploid evidence (Egeland and Salas 2008) or evaluating the possibility of spurious positive results in medical genetic studies (Salas and Carracedo 2007).

Acknowledgments We thank Chris Phillips for critically reading the manuscript and making useful suggestions. The 'Ramón y Cajal' Spanish programme from the Ministerio de Educación y Ciencia (RYC2005-3). Grants from the Xunta de Galicia (PGIDIT06PXIB208079PR) and Fundación de Investigación Médica Mutua Madrileña awarded to AS partially supported this project.

Journal : Large 10038 | Dispatch : 5-5-2008 | Pages : 13
Article No. : 297 | LE | TYPESET
MS Code : 297 | CP | DISK

References

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogvali EL, Kivisild T, Bandelt H-J, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet 75:910–918

Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti AS, Bandelt H-J, Semino O, Torroni A (2005) Saami and Berbers—an unexpected mitochondrial DNA link. Am J Hum Genet 76:883–886

Álvarez-Iglesias V, Jaime JC, Carracedo Á, Salas A (2007) Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. Forensic Sci Int Genet 1:44–55

Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt H-J, Pena SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Bailliet G, Castilla EE, Adams JP, Orioli IM, Martinez-Marignac VL, Richard SM, Bianchi NO (2001) Correlation between molecular and conventional genealogies in Aicuña: a rural population from Northwestern Argentina. Hum Hered 51:150–159

Bandelt H-J (1994) Phylogenetic networks. Verhandlungen des Naturwissenschaftlichen Vereins Hamburg 34:51–71

Bandelt H-J, Herrnstadt C, Yao Y-G, Kong Q-P, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang Y-P (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Ann Hum Genet 67:512–524

Bandelt H-J, Salas A, Bravi CM (2004a) Problems in FBI mtDNA database. Science 305:1402–1404

Bandelt H-J, Salas A, Lutz-Bonengel S (2004b) Artificial recombination in forensic mtDNA population databases. Int J Legal Med 118:267–273

Beleza S, Gusmão L, Amorim A, Carracedo Á, Salas A (2005) The genetic legacy of western Bantu migrations. Hum Genet 117:366–375

Bini C, Ceccardi S, Luiselli D, Ferri G, Pelotti S, Colalongo C, Falconi M, Pappalardo G (2003) Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). Forensic Sci Int 135:48–52

Brandstätter A, Egyed B, Zimmermann B, Duftner N, Padar Z, Parson W (2007) Migration rates and genetic structure of two Hungarian Ethnic Groups in Transylvania, Romania. Ann Hum Genet 71:791–803

Cabana GS, Merriwether DA, Hunley K, Demarchi DA (2006) Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation. Am J Phys Anthropol 131:108–119

Carracedo A, Bar W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85

Černý V, Salas A, Hájek M, Žaloudková M, Brdička R (2007) A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome. Ann Hum Genet 71(Pt 4):433–452

Coia V, Destro-Bisol G, Verginelli F, Battaggia C, Boschi I, Cruciani F, Spedini G, Comas D, Calafell F (2005) Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. Am J Phys Anthropol 128:678–681

Dipierri JE, Alfaro E, Martínez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO (1998) Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. Hum Biol 70:1001–1010

Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, Salzano FM (2004) Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. Am J Hum Biol 16:479–488

Egeland T, Salas A (2008) Statistical evaluation of haploid genetic evidence. TOForensicSJ 1(1):4-11

Ely B, Wilson JL, Jackson F, Jackson BA (2006) African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. BMC Biol 4:34

Fondevila M, Jaime JC, Salas A, Lareu MV, Carracedo A (2003) Y-chromosome STR haplotypes in Cordoba (Argentina). Forensic Sci Int 137:217–220

García-Bour J, Pérez-Pérez A, Álvarez S, Fernández E, López-Parra AM, Arroyo-Pardo E, Turbón D (2004) Early population differentiation in extinct aborigines from Tierra del Fuego-Patagonia: ancient mtDNA sequences and Y-chromosome STR characterization. Am J Phys Anthropol 123:361–370

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

Green LD, Derr JN, Knight A (2000) mtDNA affinities of the peoples of North-Central Mexico. Am J Hum Genet 66:989–998

Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Molec Biol Evolut 10:23–47

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R (2004) Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75:752–770

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F (1995) Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá. Genetics 140:275–283

Kong Q-P, Bandelt H-J, Sun C, Yao Y-G, Salas A, Achilli A, Wang CY, Zhong L, Zhu CL, Wu SF, Torroni A, Zhang Y-P (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum Mol Genet 15:2076–2086

Loogväli E-L, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk H-V, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barać L, Peričić M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan JP, Chaventré A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustres M, Gölge M, Mikerezi I, Usanga E, Villems R (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol 21:2012–2021

Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 111:67–77

Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, Salzano FM, Bortolini MC (2007) Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. Am J Phys Anthropol 132:301–310

Journal : Large 10038 | Dispatch : 5-5-2008 | Pages : 13
Article No. : 297 | LE | TYPESET
MS Code : 297 | CP | DISK

Martinez Marignac VL, Bertoni B, Parra EJ, Bianchi NO (2004) Characterization of admixture in an urban sample from Buenos Aires, Argentina, using uniparentally and biparentally inherited genetic markers. Hum Biol 76:543–557

Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B (2002) The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Sci Commun 4(2)

Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, Carvallo P (2000) Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. Am J Phys Anthropol 113:19–29

Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C, Sylvester N, Parrish-Gause D, Garvey WT, Jin L, McKeigue PM, Kamboh MI, Ferrell RE, Pollitzer WS, Shriver MD (2001) Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. Am J Phys Anthropol 114:18–29

Plaza S, Salas A, Calafell F, Corte-Real F, Bertranpetit J, Carracedo Á, Comas D (2004) Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. Hum Genet 115:439–447

Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K (2004) Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. Am J Hum Genet 74:827–845

Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo Á (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. Forensic Sci Int 140:251–257

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Salas A, Bandelt HJ, Macaulay V, Richards MB (2007) Phylogeographic investigations: the role of trees in forensic genetics. Forensic Sci Int 168:1–13

Salas A, Carracedo Á (2007) Studies of association in complex diseases: statistical problems related to the analysis of genetic polymorphisms. Rev Clin Esp 207:563–565

Salas A, Carracedo Á, Macaulay V, Richards M, Bandelt H-J (2005a) A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. Biochem Biophys Res Commun 335:891–899

Salas A, Carracedo Á, Richards M, Macaulay V (2005b) Charting the ancestry of African Americans. Am J Hum Genet 77:676–680

Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo Á (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. Eur J Hum Genet 6:365–375

Salas A, Richards M, De la Fé T, Lareu MV, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo Á (2002) The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo Á (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. Am J Hum Genet 74:454–465

Salas A, Richards M, Lareu MV, Sobrino B, Silva S, Matamoros M, Macaulay V, Carracedo Á (2005c) Shipwrecks and founder effects: divergent demographic histories reflected in Caribbean mtDNA. Am J Phys Anthropol 128:855–860

Sun C, Kong QP, Palanichamy M, Agrawal S, Bandelt HJ, Yao YG, Khan F, Zhu CL, Chaudhuri TK, Zhang YP (2006) The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. Mol Biol Evol 23:683–690

Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J (2006) Harvesting the fruit of the human mtDNA tree. Trends Genet 22:339–345

Toscanini UF, Gusmão L, Berardi G, Amorim A, Carracedo Á, Salas A, Raimondi EH (2006) Testing for genetic structure in different urban Argentinian populations. Forensic Sci Int 165(1):35–40

Victoria-Gomes M (2002) La presencia negroafricana en al Argentina. Pasado y permanencia. Boletín digital de la Biblioteca de la Nación Argentina 9:2

Journal : Large 10038 | Dispatch : 5-5-2008 | Pages : 13
Article No. : 297 | LE | TYPESET
MS Code : 297 | CP | DISK

Artículo 9: The genuine mtDNA Native American legacy of El Salvador

Salas A, Lovo-Gómez J, Álvarez-Iglesias V, Cerezo M, Lareu MV, Richards M, Carracedo Á

(Manuscrito en preparación)

El auténtico legado Nativo Americado en el ADNmt de El Salvador

ABSTRACT/RESUMEN:

Hemos secuenciado la region control completa (incluyendo las regiones hipervariables I y II, HVS-I y HVS-II) del genoma mitocondrial (ADNmt) de 90 individuos El Salvador. Hemos llevado a cabo comparaciones entre nuestra población de estudio y otras, utilizando más de 3,800 perfiles de la región control disponiblen en distintas bases de datos públicas. Los resultados revelan un componente predominante Nativo Americano en esta región: el haplogrupo mitocondrial más prevalente en este país es, con mucho el A2 (~90%), en contraste con otras poblaciones de Norte América, Centro o América del Sur. El haplogrupo A2 es muy diverso y presenta una filogenia *star-like* que atestigua una expansión poblacional de hace alrededor de 17,500. Las características filogeográficas del haplogrupo A2 son compatibles con la hipótesis de que actualmente la variabilidad mitocondrial en El Salvador representa en gran medida el componente indígena de los primeros asentamientos de la región. De acuerdo con esta hiopótesis está el escaso componente europeo observado de ADNmt (~5%), y el escaso impacto del tráfico atlántico de esclavos en El Salvador, en contraste con la prevalencia de ADN mitocondriales de África sub-Sahariana en otras poblaciones vecinas de la costa Caribeña.

The genuine mtDNA Native American legacy of El Salvador

Antonio Salas A[1,2*], José Lovo-Gómez J[1,3], Vanesa Álvarez-Iglesias[1], María Cerezo[1], María Victoria Lareu[1], Martin Richards, Carracedo Á[1,2]

[1] *Unidad de Genética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, 15782, Galicia, Spain*

[2] *Grupo de Medicina Xenómica, Hospital Clínico Universitario, 15706, Galicia, Spain*

[3] *Laboratorio de Genética Forense, Instituto de Medicina Legal, Dr. Roberto Masferrer, Corte Suprema de Justicia, San Salvador, El Salvador*

Keywords: mtDNA, HVS-I, HVS-II, control region, haplogroup, Native American, El Salvador, Atlantic slave trade

* Corresponding author: Dr. Antonio Salas, Unidad de Genética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, 15782, Galicia, Spain. Tel: +34–981–582327; Fax: +34–981–580336; E-mail: apimlase@usc.es

Abstract

We have carried out DNA sequencing of the entire control region (including the first and second mtDNA hypervariable segments, HVS-I and II) of the mitochondrial DNA (mtDNA) genome in 90 individuals from El Salvador. We then carried out inter-population comparisons using the available public database containing more than 3,800 control region profiles. The results reveal a predominant Native American component in this region: by far, the most prevalent mtDNA haplogroup in this country (at ~90%) is A2, in contrast with other North, Central and South American populations. Haplogroup A2 is very diverse and shows a star-like phylogeny that testifies to a population expansion ~17,500 years ago. The phylogeographic characteristics of haplogroup A2 are compatible with the hypothesis that the today's mtDNA variability in El Salvador represents to a large extent the indigenous component of the first settlements in the region. Concordant with this hypothesis is the observation of a very limited introgression of Eurasian mtDNAs (~5%), and the very low impact of the Atlantic slave trade in El Salvador, in contrast with the prevalence of sub-Saharan African mtDNA in neighbouring populations from the Caribbean coast.

Introduction

El Salvador lies on the Pacific coast (without an Atlantic seaboard) and it is the smallest of the Central American countries. Most of the country is on a fertile volcanic plateau. It is segmented by two volcanic ranges running roughly west to east, separated by broad, fertile valleys, such as that of the river Lempa. El Salvador was inhabited by Native American groups who were in part descendants of the Aztecs and Toltec of Mexico, such as the Pipil (a Nahua tribe) and the Lenca. These two Native American communities inhabited mainly the western regions, constituting about 60% of the population throughout the colonial era and into the early decades of independence (U.S. Library of Congress: http://www.loc.gov/).

The development of coffee estates led to the slow but continuous dissolution of most of the communal lands of Native villages (U.S. Library of Congress). Thus, the 1930 census, the last to contain the category, designated only 5.6% of the population as “Indian” – although it is not clear what criteria were used in arriving at this figure. Other independent estimates (considering religious activities, distinctive women's dress, language, and involvement in various handicrafts) placed the mid-twentieth-century Indian population at 20% (~400,000 persons). The abandonment of Indian language and customs was hastened by political repression; most natives stopped wearing traditional dress, abandoned the Pipil language, and adopted ladino customs. By 1975 no more than ~1% of the population wore distinctive Indian clothing or followed Indian customs. Nowadays, the official language in El Salvador is Spanish, although Nahua is still spoken among some natives.

Although the American continent has been the target of many forensic and population genetic studies, there are nevertheless many American regions, such as El Salvador, that remain genetically uncharacterized. The mtDNA molecule is commonly used in

anthropological contexts because of particular features (maternal inheritance, lack of recombination and high average mutation rate) that confer great power for phylogenetic and phylogeographic inferences. Many mtDNA studies of Native Americans have, however, been limited to genotyping a handful of mtDNA coding region sites that simply distinguish the four major Native American mtDNA haplogroups, A, B, C and D (generally using RFLP typing); unfortunately, the information provided by these few SNPs is of limited value in forensic and population genetics.

Here we have sequenced the mtDNA control region in a sample from El Salvador in order to investigate to what extent the Native American component has survived the impact of European colonialism and the concomitant influx of African slaves to the Caribbean and Meso-America.

Material and Methods

Sample collection and DNA extraction

We collected 90 samples from healthy unrelated individuals from El Salvador. DNA extraction was undertaken following standard phenol-chloroform protocol. DNA quantification was carried out using DyNA Quant 200 Fluorometer, Hoefer (APB, Uppsala, Sweden).

PCR and sequence analysis

We analyzed the first and second hypervariable segments (HVS-I and HVS-II) of the mtDNA genome. We performed PCR amplifications using a 2700 Thermocycler (Applied Biosystems), using PCR and sequencing primers as reported in (Crespillo et al. 2006). Cycling parameters were 95ºC for 1 min, followed by 36 cycles of 95ºC for 10 sec, 55ºC for 30 sec and 72ºC for 30 sec, and followed by 10 min at 15ºC. We checked amplification products on a polyacrylamide gel visualized by silver staining and purified with Montage (Multiscreen PCR, Millipore Corporation, USA). We performed sequence reaction

products on each strand by means of the ABI Prism dRhodamine Terminator cycle sequencing reaction kit (Applied Biosystems). DNA products were then purified by ethanol precipitation and sequence reaction products analyzed on the ABI Prism 3100 automatic sequencer (Applied Biosystems). We omitted population variation at the hypervariable sites (mainly related to the cytosine homopolymeric track around 310 and the CA-dinucleotide repeat around positions 522) from inter-population comparisons and phylogeograhic analyses. We have used the same primers for amplification and sequencing described in (Álvarez-Iglesias et al. 2007).

Sequences were edited using the numbering system of the revised Cambridge Reference Sequence (Andrews et al. 1999).

Quality checking

Problems with the quality of mtDNA data in forensic, clinical, and population genetic studies are unfortunately rather common; see, for instance, (Bandelt et al. 2002; Bandelt et al. 2004; Salas et al. 2005a; Salas et al. 2005d; Yao et al. 2006). In order to minimize the effects of potential laboratory and documentation errors, the data were read separately by two independent persons in the light of the known phylogeny. We checked phylogenetic inconsistencies by hand with special attention to private or unusual variants (e.g. rare transitions or indels). In some cases, we confirmed the sequences by repeated extraction and sequencing. In addition, to detect potential "phantom mutations" (Bandelt et al. 2002), we also checked the data using the computer program SPECTRA ((Bandelt et al. 2002), available at http://www.stats.gla.ac.uk/~vincent/fingerprint/index.html).

Statistical analysis and population comparison

Haplogroup nomenclature follows the most recently updated version of the Native American phylogeny given in Bravi et al. (Bravi et al. 2007). Diversity indices of HVS-I sequences (haplotype diversity,

nucleotide diversity, average number of pairwise differences) were calculated using Arlequin 3.0 software (Excoffier et al. 2005). Nucleotide and sequence diversity was computed as in Nei (1987). AMOVA (Excoffier et al. 1992) was analyzed using Arlequin 3.0, and the significance of the covariance components associated with the different levels of genetic structure was tested using non-parametric permutation procedure (Excoffier et al. 1992). Comparisons between populations were assessed by F_{ST} distances which were subsequently plotted by multidimensional scaling analysis (MDS) using PROXSCAL technique included in the SPSS 10.0 statistical package.

We estimated median-joining networks of HVS-I sequences using the Network 4.1.1.2 software (Bandelt et al. 1999; Bandelt et al. 1995). Coalescent times were calculated using the ρ statistic with an HVS-I mutation rate of one transition per 20,180 years applied for the sequence range 16090-16365 using (Forster et al. 1996; Saillard et al. 2000).

An mtDNA database of Native American populations was compiled for population comparisons: (i) from North America: Aleuts (Rubicz et al. 2003), Athapaskans, Inupiaq, Yakima (Shields et al. 1993), Chukchi and Siberian Eskimos (Starikovskaya et al. 1998), Bella Coola and Haida (Ward et al. 1993), Nuu-Chah-Nulth (Ward et al. 1991), Cheyenne (Kittles et al. 1999), North Native Americans (various ethnic groups; Bolnick and Smith 2003; Horai et al. 1993; Lorenz and Smith 1994; Malhi et al. 2001; Torroni et al. 1993a; Torroni et al. 1993b), Apache and Navajo (Monson et al. 2002), (ii) from Meso-America: Pima (Kittles et al. 1999), Maya (Horai et al. 1993), Huetar (Santos et al. 1994a), Kuna (Batista et al. 1995), Ngöbe (Kolman et al. 1995), Quiché (Boles et al. 1995), Emberá and Wounan (Kolman and Bermingham 1997), Mexico (Green et al. 2000), Central Native Americans [various ethnic groups] (Torroni et al. 1993a), El Salvador (present study), and (iii) from South America: Native Brazilians and Araucanians or Chileans (Horai et

al. 1993), Equador (Rickards et al. 1999), Embera and Gavião (Ward et al. 1996), Amazonas (Santos et al. 1996), Ayoreo (Dornelles et al. 2004), Chilean Mapuche and Pehueche, Yaghan (Moraga et al. 2000), Argentinian Mapuche (Ginther et al. 1993), Cayapas (Rickards et al. 1999), Xavante, Zoró, and Gavião (Ward et al. 1996), Yanomami (Merriwether et al. 2000; Williams et al. 2002), South Native Americans [various ethinic groups] (Torroni et al. 1993b), Tuacuarembó (Bonilla et al. 2004), Uruguay (Pagano et al. 2005), Guahibo (Vona et al. 2005), Colombia (Horai et al. 1993; Torres et al. 2006), Yuracaré, Trinitario, Movima, and Ignaciano (Bert et al. 2004), and 105 from Arequipa, Tayacaja and San Martin in Peru (Vernesi et al. in press). We also include the data collected from different studies on ancient DNA (García-Bour et al. 2004; Lalueza-Fox et al. 2001; Lalueza-Fox et al. 2003; Monsalve et al. 2002; Moraga et al. 2005; Ribetio-dos-Santos et al. 1996; Stone and Stoneking 1993). In addition, other datasets were additionally used for haplotype matching comparisons (Alves-Silva et al. 2000; Brown et al. 1998; Dornelles et al. 2005; Monson et al. 2002; Ribeiro-Dos-Santos et al. 2006; Smith et al. 1999). In total, 3,843 mtDNAs profiles (mainly HVS-I segments) were compiled for comparison with our sample from El Salvador. Those population samples consisting of less than 15 individuals were only used for haplotype matching between populations. For comparison purposes the common reading frame 16090-16365 of the HVS-I was used.

Results

Summarizing statistics

We observed a total of 55 different HVS-I, 40 different HVS-II, and 76 different combined HVSI/II mtDNA haplotypes in El Salvador (N = 90). Some HVS-I profiles are quite common, such as C16111T T16223C C16290T G16319A T16362T, appearing in 12 mtDNAs, and its one step-mutation 'neighbour' haplotype (16519 on top) appearing 12 times.

As shown in Table 1, El Salvador shows haplotype and nucleotide diversity values slightly lower than those observed in the continental North, South, and other Central American populations.

Phylogeography of Salvadorian Native American mtDNAs

Table 2 shows the full list of control region profiles from El Salvador and their haplogroup allocation. Frequencies of the typical Native American haplogroups A2, B2, and C1 are ~90%, ~2%, and ~3%, respectively. We did not observe any representatives of haplogroup D1 or the Native American branch of haplogroup X, X2a.

Figure 2 shows the frequency distribution of the main mtDNA American haplogroups in different Native American populations and in the three American regions in contrast with the profile for El Salvador. Although haplogroup A2 is at high frequencies in Central America, El Salvador is particularly distinct from the other populations by its extremely high A2 haplogroup frequency. Note also that there exists substantial heterogeneity of haplogroup frequency patterns in America (even between neighbouring populations).

The most prevalent haplogroup in El Salvador by far is A2 (a sub-clade of the Asian haplogroup A), identified by the diagnostic sites C16111T–T16223C–C16290T–G16319A–T16362C in HVS-I, and C64T–A73G–T146C–A153G –A235G–A263G–315+C in HVS-II. There are no solid diagnostic sites in the control region that would allow us to classify A2 sub-lineages from El Salvador. Moreover, several control-region variants regarded as haplogroup diagnostic, such as C64T and A153G, show reversions: complete genome sequence data confirm the existence of multiple back and parallel mutations within haplogroup A2 (Bravi et al. 2007). Although many of them are well known hotspots (e.g. T146C), others seem to behave as hotspots only within A2. For example, position 64 reverts at least seven times within A2 (see Figure 2 in (Bravi et al. 2007)).

The subclade of A2 defined by C16360T, although frequently observed in Central America, especially in the Huetar from Costa Rica (Santos et al. 1994b) and the Ngöbé from Panama (Kolman et al. 1995), is found only in two individuals in the El Salvador sample.

We found only four Native American mtDNAs not belonging to haplogroup A2: two haplogroup B2 and two haplogroup C1 mtDNAs. We did not find any match amongst published data for the B2 sequence #33 that carries the distinctive variant A16269G. The haplogroup C1 sequence #41 carries C16256T; this uncommon variant within haplogroup C1 has been also observed in the Yanomama from Venezuela (Merriwether et al. 2000) and the Zoró from Brazil (Ward et al. 1996).

We did not observe any mtDNAs belonging to the other Native American haplogroups, D1 and X2a.The phylogeny of haplogroup A2 is clearly star-like, indicating an ancient population expansion affecting the topology of the clade, with an estimated age of about 17,400 ± 4,700 YBP.

Non Native American haplotypes in El Salvador

Signals of a European contribution to our sample from El Salvador are limited to three haplotypes (see Table 1): haplotype #96 belongs to haplogroup U2e, with exact matches in several West European locations (e.g. Northwest Spain and Portugal (González et al. 2003)); in Madeira (Brehm et al. 2003), etc. Haplotype #12 can be assigned most plausibly to haplogroup H, while #30 probably belongs to haplogroup T2 and carries T16271C, an unusual variant within haplogroup T; this sequence curiously matches published sequences found in Portugal and Brazil (Alves-Silva et al. 2000; Pereira et al. 2000) but also one single hit in Poland (Malyarchuk et al. 2002).

We detected only one sequence belonging to a typical sub-Saharan haplogroup in El Salvador. It belongs to L0a1a, a sub-clade

highly prevalent in southeast Africa (Salas et al. 2005b; Salas et al. 2002; Salas et al. 2004), where we find exact matches in HVS-I and HVS-II. Exact matches are also found in, for example, the Atlantic African southwest coast, in Cabinda (Beleza et al. 2005), and in the Tongas (Trovoada et al. 2004). Although it is not possible to determine with accuracy the African origin of this haplotype, southeast Africa (Mozambique) is probably the best candidate population source.

Multidimensional scaling analysis (MSA)

MSA was performed on F_{ST} genetic distances between mtDNAs from El Salvador and those of the three main American regions. The plot highlights again the main idiosyncratic feature of El Salvador, the high prevalence of A2 haplotypes. Thus, Salvadorians are more closely related to North Americans along the first dimension of the MSA plot, while the second dimension locates them in between North/Central Americans and the South.

Discussion

El Salvador is the smallest Latin American republic and also the most densely populated. Although historically El Salvador has been home to a culturally diverse mix of peoples, including Native Americans, Africans, and west Europeans, by the 1980s the population of the country was essentially considered to be homogeneous in terms of ethnicity and basic cultural identity. Virtually all Salvadorans speak Spanish, the official language, as their mother tongue, and the vast majority are generally characterized as “mestizos” (or “ladinos”, a term more commonly used in Central America), popularly used to refer to those persons of mosaic geographic ancestry who follow a wide variety of indigenous and “hispanic” customs and habits that over the centuries have come to constitute Spanish-American cultural patterns. In the late 1980s, the ethnic composition of the population was estimated as 89% mestizo, 10% Native American, and 1% “white”

(Consejo Nacional para la Cultura y el Arte (CONCULTURA) 2000). Therefore, in contrast to most other Central American countries, El Salvador no longer possessed an ethnically or linguistically distinct Native American population, although persons of native-like ethnicity or cultural heritage still lived in the western parts of the country. Similarly, there was no ethnically or culturally distinct African-American population as there is in other neighbouring populations (Salas et al. 2005c). However, there is a generally belief that much of the Salvadorian population in the 1980s had a predominantly Native American ancestry (U.S. Library of Congress: http://www.loc.gov/).

The main goal of the present study is to analyze the mtDNA genetic background of El Salvador and to estimate to what extent the ancestral Native American mtDNA component has survived to external demographical influences, especially Europeans and Sub-Saharan Africans.

The present study has revealed interesting features of El Salvador population. Thus, we have observed that, in contrast to the cultural patterns observed in the today's El Salvador population, most of the mtDNA profiles found are typically Native American; haplogroup A2 account for ~90% of the Salvadorian sample.

Correspondingly, the impact of Europeans on the mtDNA pool of El Salvador is very low (~2%). It seems that the Spanish conquerors and more recent European demographic influences did not contribute significantly to the today's genetic composition of El Salvador in the maternal side. This contrasts with the European Y-chromosome contribution to the El Salvador gene pool. According to Lovo-Gómez et al. (2006), about one half in metropolitan areas and two thirds in rural populations of El Salvador belong to non-Native American haplogroups; for instance, the most common Y-chromosome haplogroup in Europe (namely, R1b) is present in El Salvador at 24% in rural areas and 43% in metropolitan regions. Concomitantly, the Native American Y-

chromosome proportion in El Salvador (represented by haplogroup Q3) is about 31-49%.

Therefore, the mtDNA and Y-chromosome variation in El Salvador displays an extreme version of a pattern that is common in other Native American populations (Dipierri et al. 1998): the indigenous female contribution is much higher than the indigenous male contribution. This asymmetrical proportion of females and males to the population gene pool in Native Americans is also consistent with historical documentation.

Also in agreement with the historical record is that the impact of African-American lineages on the mtDNA pool of El Salvador is very low, as indicated by the presence of only one mtDNA of sub-Saharan origin in our sample. The scarcity of the sub-Saharan component strikingly contrasts with the situation on the Caribbean coast, where (as a consequence of the Atlantic slave trade) it is clearly predominant (Salas et al. 2005b; Salas et al. 2002; Salas et al. 2004; Salas et al. 2005c). I this case the Y-chromosome shows the same pattern: no lineages of African-ancestry have been detected in El Salvador (Lovo-Gómez et al. unpublished).

There are no clear signals of recent genetic drift events in El Salvador, in contrast with the patterns observed in other neighbouring Native American populations such as the Ngöbé from Panamá (Kolman et al. 1995) that shows extremely reduced levels of mtDNA diversity (reflecting passage through post-conquest population bottlenecks). Haplogroup A2 is at high frequency in El Salvador (~90% of the sample); its phylogeny is clearly star-like and its age is 17,400 ± 4,700 years. The shape of this phylogeny points to the existence of a prehistoric demographic expansion

In the region that does not clearly fit with the first major population expansion in lower Central America estimated by

archaeologists on the based of the widely distributed and well-dated Clovis culture. The dating based on haplogroup A2 alone could be however overestimated since the phylogeny of A2 could be blurring more than one overlapping demographic expansions.

The most plausible hypothesis compatible with the data is that the Salvadorans have preserved an important imprint of the original indigenous component of the first inhabitants of the region, with scarce subsequent demographic influence from other American or non-American populations. In fact, since we have genotyped samples collected in urban areas we would expect to have an even higher prevalence of the Native American component in more isolated or ethnic groups from the country, as is in fact observed on the Y-chromosome side: the Native American component is higher in the rural than in metropolitan areas.

In contrast to the high impact of the Atlantic slave trade in the Central American Caribbean coast (Salas et al. 2005c), the Pacific side (at least for El Salvador) appears to have preserved its Native American mtDNA heritage intact to the present day. at the same time, this study has also shown that El Salvador shows haplogroup frequency patterns quite different from other modern Native American communities. At the individual haplotype level, El Salvador shows numerous mtDNAs that have never been observed in other American regions, even within Central America. These features provide little support to those that assume (or claim) that “Hispanics” or Native American communities are sufficiently homogeneous to justify the portability of forensic databases from one country to another (e.g. SWGDAM; (Monson et al. 2002)); see (Salas et al. 2006) for a discussion. Stratification also has a key impact on population-based (e.g. case-control) association disease studies since it seems to be the main responsible for the high false positive rate and lack of replication of positive findings in independent studies (Pritchard et al. 2000).

Acknowledgments

We would like to thank Vilma de Aguilar (Ministerio de Salud Pública de El Salvador) for helping with the sample collection. The 'Ramón y Cajal' Spanish programme from the Ministerio de Educación y Ciencia (RYC2005-3), the grant of the Ministerio de Sanidad y Consumo (PI030893; SCO/3425/2002) and Fundación Investigación Médica Mutua Madrileña Automovilística given to AS, partially supported this project.

References

Álvarez-Iglesias V, Jaime JC, Carracedo Á, Salas A (2007) **Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups.** Forensic Sci Int: Genet 1: 44-55.

Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt H-J, Pena SD, Prado VF (2000) **The ancestry of Brazilian mtDNA lineages.** Am J Hum Genet 67: 444-61.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) **Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA.** Nat Genet 23: 147.

Bandelt H-J, Forster P, Röhl A (1999) **Median-joining networks for inferring intraspecific phylogenies.** Mol Biol Evol 16: 37-48.

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) **Mitochondrial portraits of human populations using median networks.** Genetics 141: 743-53.

Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V (2002) **The fingerprint of phantom mutations in mitochondrial DNA data.** Am J Hum Genet 71: 1150-60.

Bandelt H-J, Salas A, Bravi CM (2004) **Problems in FBI mtDNA database.** Science 305: 1402-4.

Batista O, Kolman CJ, Bermingham E (1995) **Mitochondrial DNA diversity in the Kuna Amerinds of Panama.** Hum Mol Genet 4: 921-9.

Beleza S, Gusmão L, Amorim A, Carracedo Á, Salas A (2005) **The genetic legacy of western Bantu migrations.** Hum Genet 117: 366-75.

Bert F, Corella A, Gene M, Perez-Perez A, Turbon D (2004) **Mitochondrial DNA diversity in the Llanos de Moxos: Moxo, Movima and Yuracare Amerindian populations from Bolivia lowlands.** Ann Hum Biol 31: 9-28.

Boles TC, Snow CC, Stover E (1995) **Forensic DNA testing on skeletal remains from mass graves: a pilot project in Guatemala.** J Forensic Sci 40: 349-55.

Bolnick DA, Smith DG (2003) **Unexpected patterns of mitochondrial DNA variation among Native Americans from the southeastern United States.** Am J Phys Anthropol 122: 336-54.

Bonilla C, Bertoni B, Gonzalez S, Cardoso H, Brum-Zorrilla N, Sans M (2004) **Substantial Native American female contribution to the population of Tacuarembo, Uruguay, reveals past episodes of sex-biased gene flow.** Am J Hum Biol 16: 289-97.

Bravi CM, Salas A, Coble M, Kong Q-P, Perego UA, Torroni A, Bandelt H-J (2007) **The phylogeny of Native American complete mtDNA sequences: implications for evolutionary and disease studies.** Am J Hum Genet in press.

Brehm A, Pereira L, Kivisild T, Amorim A (2003) **Mitochondrial portraits of the Madeira and Acores archipelagos witness different genetic pools of its settlers.** Hum Genet 114: 77-86.

Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani F, Wallace DC (1998) **mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America?** Am J Hum Genet 63: 1852-61.

Budowle B, Allard MW, Fisher CL, Isenberg AR, Monson KL, Stewart JE, Wilson MR, Miller KW (2002) **HVI and HVII mitochondrial DNA data in Apaches and Navajos.** Int J Legal Med 116: 212-5.

Burckhardt F, von Haeseler A, Meyer S (1999) **HvrBase: compilation of mtDNA control region sequences from primates.** Nucleic Acids Res 27: 138-42.

Consejo Nacional para la Cultura y el Arte (CONCULTURA) SS (2000) **Biblioteca de Historia Salvadoreña, El Salvador, Historia de sus pueblos, villas y ciudades.**

Crespillo M, Paredes MR, Prieto L, Montesino M, Salas A, Albarrán C, Álvarez-Iglesias V, Amorin A, Berniell-Lee G, Brehm A, Carril JC, Corach D, Cuevas N, Di Lonardo AM, Doutremepuich C, Espinheira RM, Espinoza M, Gómez F, González A, Hernández A, Hidalgo M, Jimenez M, Leite FP, López AM, López-Soto M, Lorente JA, Pagano S, Palacio AM, Pestano JJ, Pinheiro MF, Raimondi E, Ramon MM, Tovar F, Vidal-Rioja L, Vide MC, Whittle MR, Yunis JJ, Garcia-Hirschfel J (2006) **Results of the 2003-2004 GEP-ISFG collaborative study on mitochondrial DNA: focus on the mtDNA profile of a mixed semen-saliva stain.** Forensic Sci Int 160: 157-67.

Dipierri JE, Alfaro E, Martinez-Marignac VL, Bailliet G, Bravi CM, Cejas S, Bianchi NO (1998) **Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina.** Hum Biol 70: 1001-10.

Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, Salzano FM (2004) **Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay.** Am J Hum Biol 16: 479-88.

Dornelles CL, Bonatto SL, De Freitas LB, Salzano FM (2005) **Is haplogroup X present in extant South American Indians?** Am J Phys Anthropol 127: 439-48.

Excoffier L, Laval G, Schneider S (2005) **Arlequin ver. 3.0: An integrated software package for population genetics data analysis**. Evol Bioinformatics Online 1: 47-50.

Excoffier L, Smouse PE, Quattro JM (1992) **Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data.** Genetics 131: 479-91

Forster P, Harding R, Torroni A, Bandelt H-J (1996) **Origin and evolution of Native American mtDNA variation: a reappraisal.** Am J Hum Genet 59: 935-945

García-Bour J, Pérez-Pérez A, Álvarez S, Fernández E, López-Parra AM, Arroyo-Pardo E, Turbón D (2004) **Early population differentiation in extinct aborigines from Tierra del Fuego-Patagonia: ancient mtDNA sequences and Y-chromosome STR characterization.** Am J Phys Anthropol 123: 361-70.

Ginther C, Corach D, Penacino GA, Rey JA, Carnese FR, Hutz MH, Anderson A, Just J, Salzano FM, King MC (1993) **Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes.** Exs 67: 211-9.

González AM, Brehm A, Pérez JA, Maca-Meyer N, Flores C, Cabrera VM (2003) **Mitochondrial DNA affinities at the Atlantic fringe of Europe.** Am J Phys Anthropol 120: 391-404

Green LD, Derr JN, Knight A (2000) **mtDNA affinities of the peoples of North-Central Mexico.** Am J Hum Genet 66: 989-98.

Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) **Peopling of the Americas, founded by four major lineages of mitochondrial DNA.** Molecular Biology and Evolution 10: 23-47

Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, Long JC (1999) **Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck.** American Journal of Physical Anthropology 108: 381-399

Kolman CJ, Bermingham E (1997) **Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama.** Genetics 147: 1289-302

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F (1995) **Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá.** Genetics 140: 275-283.

Lalueza-Fox C, Calderon FL, Calafell F, Morera B, Bertranpetit J (2001) **MtDNA from extinct Tainos and the peopling of the Caribbean.** Ann Hum Genet 65: 137-51.

Lalueza-Fox C, Gilbert MT, Martinez-Fuentes AJ, Calafell F, Bertranpetit J (2003) **Mitochondrial DNA from pre-Columbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean.** Am J Phys Anthropol 121: 97-108.

Lorenz JG, Smith DG (1994) **Distribution of the 9-bp mitochondrial DNA region V deletion among North American Indians.** Hum Biol 66: 777-88.

Lovo-Gomez J, Blanco-Verea A, Lareu MV, Brion M, Carracedo A (2006) **The genetic male legacy from El Salvador.** Forensic Sci Int in press.

Malhi RS, Schultz BA, Smith DG (2001) **Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America.** Hum Biol 73: 17-55.

Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Wozniak M, Miscicka-Sliwka D (2002) **Mitochondrial DNA variability in Poles and Russians.** Ann Hum Genet 66: 261-83.

Merriwether DA, Kemp BM, Crews DE, Neel JV (2000) **Gene flow and genetic variation in the Yanomama as revealed by mitochondrial DNA.** In: Renfrew C (ed) America Past, America Present: Genes and Languages in the Americas and Beyond. McDonald Institute for Archaeological Research, Cambridge, pp 89-124.

Merriwether DA, Kemp BM, Crews DE, Neel JV (2002) **Gene flow and genetic variation in the Yanomama as revealed by mitochondrial DNA.** McDonald Institute for Archaeological Research, Cambridge.

Monsalve MV, Stone AC, Lewis CM, Rempel A, Richards M, Straathof D, Devine DV (2002) **Brief communication: molecular analysis of the**

Kwaday Dan Ts'finchi ancient remains found in a glacier in Canada. Am J Phys Anthropol 119: 288-91

Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B (2002) **The mtDNA Population Database: an integrated software and database resource for forensic comparison.** Forensic Sci Commun 4: no 2.

Moraga M, Santoro CM, Standen VG, Carvallo P, Rothhammer F (2005) **Microevolution in prehistoric Andean populations: chronologic mtDNA variation in the desert valleys of northern Chile.** Am J Phys Anthropol 127: 170-81.

Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, Carvallo P (2000) **Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent.** Am J Phys Anthropol 113: 19-29.

Nei N (1987) Molecular evolutionary genetics. New York: Columbi University Press

Pagano S, Sans M, Pimenoff V, Cantera AM, Alvarez JC, Lorente JA, Peco JM, Mones P, Sajantila A (2005) **Assessment of HV1 and HV2 mtDNA variation for forensic purposes in an Uruguayan population sample.** J Forensic Sci 50: 1239-42.

Pereira L, Prata MJ, Amorim A (2000) **Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation.** Ann Hum Genet 64: 491-506.

Pritchard JK, Stephens M, Donnelly P (2000) **Inference of population structure using multilocus genotype data.** Genetics 155: 945-59.

Ribeiro-Dos-Santos AK, Carvalho BM, Feio-Dos-Santos AC, Santos SE (2006) **Nucleotide variability of HV-I in Afro-descendents populations of the Brazilian Amazon Region.** Forensic Sci Int

Ribeiro-Dos-Santos AK, Santos SE, Machado AL, Guapindaia V, Zago MA (1996) **Heterogeneity of mitochondrial DNA haplotypes in Pre-Columbian Natives of the Amazon region.** Am J Phys Anthropol 101: 29-37.

Rickards O, Martinez-Labarga C, Lum JK, De Stefano GF, Cann RL (1999) **mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native American populations.** Am J Hum Genet 65: 519-30.

Rubicz R, Schurr TG, Babb PL, Crawford MH (2003) **Mitochondrial DNA variation and the origins of the Aleuts.** Hum Biol 75: 809-35

Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) **mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion.** Am J Hum Genet 67: 718-26.

Salas A, Bandelt HJ, Macaulay V, Richards MB (2006) Phylogeographic investigations: **The role of trees in forensic genetics.** Forensic Sci Int in press.

Salas A, Carracedo Á, Macaulay V, Richards M, Bandelt H-J (2005a) **A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics.** Biochem Biophys Res Commun 335: 891-9.

Salas A, Carracedo Á, Richards M, Macaulay V (2005b) **Charting the Ancestry of African Americans.** Am J Hum Genet 77: 676-680.

Salas A, Richards M, De la Fé T, Lareu MV, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo Á (2002) **The making of the African mtDNA landscape.** Am J Hum Genet 71: 1082-111.

Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo Á (2004) **The African diaspora: mitochondrial DNA and the Atlantic slave trade.** Am J Hum Genet 74: 454-65.

Salas A, Richards M, Lareu MV, Sobrino B, Silva S, Matamoros M, Macaulay V, Carracedo Á (2005c) Shipwrecks and founder effects: **Divergent demographic histories reflected in Caribbean mtDNA.** Am J Phys Anthropol 128: 855-860.

Salas A, Yao Y-G, Macaulay V, Vega A, Carracedo Á, Bandelt H-J (2005d) **A critical reassessment of the role of mitochondria in tumorigenesis.** PLoS Med 2: e296.

Santos M, Ward RH, Barrantes R (1994a) **mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica.** Hum Biol 66: 963-77

Santos M, Ward RH, Barrantes R (1994b) **mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica.** Hum Biol 66: 963-977

Santos SE, Ribeiro-Dos-Santos AK, Meyer D, Zago MA (1996) **Multiple founder haplotypes of mitochondrial DNA in Amerindians revealed by RFLP and sequencing.** Ann Hum Genet 60 (Pt 4): 305-19.

Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH (1993) **mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations.** Am J Hum Genet 53: 549-62.

Smith DG, Malhi RS, Eshleman J, Lorenz JG, Kaestle FA (1999) **Distribution of mtDNA haplogroup X among Native North Americans.** Am J Phys Anthropol 110: 271-84.

Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998) **mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World.** American Journal of Human Genetics 63: 1473-1491.

Stone AC, Stoneking M (1993) **Ancient DNA from a pre-Columbian Amerindian population.** Am J Phys Anthropol 92: 463-71.

Torres MM, Bravi CM, Bortolini MC, Duque C, Callegari-Jacques S, Ortiz D, Bedoya G, Groot de Restrepo H, Ruiz-Linares A (2006) **A revertant of the major founder Native American haplogroup C common in populations from northern South America.** Am J Hum Biol 18: 59-65.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993a) **Asian affinities and continental radiation of the four founding Native American mtDNAs.** American Journal of Human Genetics 53: 563-590

Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC (1993b) **mtDNA variation of aboriginal**

Siberians reveals distinct genetic affinities with Native Americans. American Journal of Human Genetics 53: 591-608.

Trovoada MJ, Pereira L, Gusmao L, Abade A, Amorim A, Prata MJ (2004) **Pattern of mtDNA variation in three populations from Sao Tome e Principe.** Ann Hum Genet 68: 40-54.

Vernesi C, Fuselli S, Castri L, Bertorelle G, Barbujani G (in press) **Mitochondrial diversity in linguistic isolates of the Alps: a reappraisal.** Human Biology.

Vona G, Falchi A, Moral P, Calo CM, Varesi L (2005) **Mitochondrial sequence variation in the Guahibo Amerindian population from Venezuela.** Am J Phys Anthropol 127: 361-9.

Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) **Extensive mitochondrial diversity within a single Amerindian tribe.** Proceedings Of the National Academy Of Sciences Of the United States Of America 88: 8720-8724.

Ward RH, Redd A, Valencia D, Frazier B, Pääbo S (1993) **Genetic and linguistic differentiation in the Americas**. Proceedings Of the National Academy Of Sciences Of the United States Of America 90: 10663-10667.

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, Santos RV (1996) **Mitochondrial DNA polymorphism in 3 Brazilian Indian tribes.** Am J Hum Biol 8: 317-323.

Williams SR, Chagnon NA, Spielman RS (2002) **Nuclear and mitochondrial genetic variation in the Yanomamo: a test case for ancient DNA studies of prehistoric populations.** Am J Phys Anthropol 117: 246-59.

Yao Y-G, Salas A, Bravi CM, Bandelt H-J (2006) **A reappraisal of complete mtDNA variation in East Asian families with hearing impairment.** Hum Genet: in press.

Table 1. Native American population mtDNA database.

	N	*H*	*D*	π	*M*
El Salvador	90	49	0.917 ± 0.025	0.013 ± 0.002	3.5
North America	1215	243	0.950 ± 0.003	0.020 ± 0.000	5.1
Meso America	394	142	0.968 ± 0.004	0.023 ± 0.012	6.2
South America	1144	265	0.956 ± 0.003	0.019 ± 0.000	5.3

N = sample size; *H* = number of different haplotypes; *D* =haplotype diversity; π = nucleotide diversity; *M* = average number of nucleotide differences.

Table 2. MtDNA haplotypes in El Salvador.

N	HVS-I reading range (minus 16000)	HVS-II reading range (minus 16000)	HVS-II	HG	HVS-II reading range
1	042 111 223 244 290 319 362 519	024-569	64 73 146 152 153 154 235 263 309+C 315+C 523-524del	A2	021-540
2	051 111 223 290 299 319 362	024-569	64 73 146 153 235 263 315+C 523-524del	A2	021-589
90	111 136 153 223 290 311 319 362	024-560	64 73 146 153 235 263 309+CC 315+C 523-524del	A2	021-595
10	111 136 153 223 290 319 362	024-520	64 73 146 153 235 263 309+CC 315+C 523-524del	A2	021-590
8	111 136 223 290 311 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-560
43	111 136 223 290 311 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
55	111 136 223 290 319 362	024-560	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
62	111 136 223 290 319 362	024-560	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-540
93	111 136 223 290 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
61	111 172 223 290 319 362 519	024-589	64 73 146 153 195 235 263 309+CC 315+C	A2	021-590
83	111 175 223 290 300 319 362	024-569	64 73 153 235 263 309+C 315+C	A2	021-440
77	111 175 290 300 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-560
35	111 181 187 223 290 304 319 362	024-564	64 73 146 153 207 235 263 309+C 315+C	A2	021-320
64	111 182C 183C 189 223 290 319 362	024-549	64 73 146 153 235 263 309+CC 315+C	A2	021-320
67	111 182C 183C 189 223 290 319 362	024-560	64 73 146 153 235 263 315+C 523-524del	A2	021-600
48	111 183C 189 223 290 319 362 381 519	024-560	64 73 146 153 235 263 309+C 315+C	A2	021-310
46	111 187 209 223 290 319 362 371 519	024-560	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-550
40	111 187 223 234 290 319 362 390 519	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-570
54	111 187 223 290 319 362	024-589	64 73 146 153 235 263 309+C 315+C	A2	021-410
6	111 187 223 290 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-560
13	111 187 223 290 319 362	024-520	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-535
17	111 187 223 290 319 362	024-530	64 73 146 153 235 263 315+C 523-524del	A2	021-569
47	111 189 223 259 298 325 327	024-530	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-550
89	111 189 223 274 290 319 362	024-540	64 73 146 153 235 263 309+CC 315+C	A2	021-510
87	111 189 223 290 311 319 362	024-550	64 73 146 153 235 263 292 309+C 315+C 523-524del	A2	021-600
24	111 189 223 290 319 324 362	034-569	64 73 146 153 235 263 309+CC 315+C 523-524del	A2	021-535
32	111 189 223 290 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-535
57	111 189 223 290 319 362	024-560	73 146 152 153 235 263 309+CC 315+C	A2	021-320
75	111 209 223 290 291 319 362 477	024-569	64 73 146 152 153 235 263 315+C 523-524del	A2	021-600
74	111 209 223 290 293C 319 362 519	024-569	64 73 146 153 235 263 309+CC 315+C4 523-524del	A2	021-550
28	111 209 223 290 319 362 519	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
31	111 209 223 290 319 362 519	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-595
99	111 223 234 290 319 362 519	024-569	64 73 146 153 235 263 309+C 315+C 356+C 523-524del	A2	021-600
21	111 223 243 290 299 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
7	111 223 290 299 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-560
37	111 223 290 299 319 362	024-560	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-535

44	111 223 290 299 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
49	111 223 290 299 319 362	024-520	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-570
82	111 223 290 300 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
23	111 223 290 311 319 360 362	024-569	146 153 235 263 309+CC 315+C 523-524del	A2	021-539
26	111 223 290 311 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
4	111 223 290 311 319 362	024-569	64 73 146 153 235 263 315+C 523-524del	A2	021-600
3	111 223 290 319 360 362	024-569	143 146 152 153 204 235 263 309+CC 315+C 523-524del	A2	021-570
59	111 223 290 319 362	024-560	64 73 146 152 153 215 235 263 309+C 315+C 523-524del	A2	021-600
92	111 223 290 319 362	024-569	64 73 146 152 153 235 263 315+C	A2	021-320
38	111 223 290 319 362	024-560	64 73 146 153 235 263 309+C 315+C	A2	021-535
78	111 223 290 319 362	024-520	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
79	111 223 290 319 362	024-530	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
84	111 223 290 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-530
20	111 223 290 319 362	024-549	64 73 146 153 235 263 309+CC 315+C	A2	021-560
29	111 223 290 319 362	024-500	64 73 146 153 235 263 309+CC 315+C 523-524del	A2	021-570
36	111 223 290 319 362	024-560	64 73 153 214 235 263 315+C 523-524del	A2	021-600
63	111 223 290 319 362	024-525	73 146 150 153 235 263 315+C 523-524del	A2	021-560
68	111 223 290 319 362	024-550	n.d	A2	021-430
42	111 223 290 319 362 391	024-569	64 73 146 153 235 263 309+CC 315+C 523-524del	A2	021-600
103	111 223 290 319 362 518 519	024-569	64 73 146 153 235 263 315+C 523-524del	A2	021-550
91	111 223 290 319 362 519	024-560	64 73 146 150 153 174+C 235 263 309+CC 315+C 523-524del	A2	021-320
53	111 223 290 319 362 519	024-569	64 73 146 150 153 235 263 315+C 523-524del	A2	021-600
34	111 223 290 319 362 519	024-569	64 73 146 150 153 235 263 315+C 523-524del	A2	021-600
76	111 223 290 319 362 519	024-550	64 73 146 152 153 235 263 309+C 315+C	A2	021-320
27	111 223 290 319 362 519	024-569	64 73 146 152 153 235 263 309+C 315+C 523-524del	A2	021-589
14	111 223 290 319 362 519	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
51	111 223 290 319 362 519	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
86	111 223 290 319 362 519	021-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
88	111 223 290 319 362 519	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
11	111 223 290 319 362 519	024-569	64 73 146 153 235 263 315+C 523-524del	A2	021-570
9	111 223 290 319 362 519	024-569	73 146 152 153 197 235 263 309+C 315+C 523-524del	A2	021-590
15	111 223 290 319 362 519	024-569	73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
25	111 223 290 319 519	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-544
97	111 261 290 319 362 519	024-550	73 146 152 153 235 263 315+C 523-524del	A2	021-600
45	111 290 311 319 362 391	024-569	64 73 146 153 235 263 315+C 523-524del	A2	021-590
50	111 290 311 319 362 391	024-569	64 73 146 153 235 263 315+C 523-524del	A2	021-589
19	111 290 319 362 391	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
58	153 223 240 290 319 362	024-560	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-590
18	189 223 290 319 362	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-585
56	223 290 311 319 362	024-560	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-530
5	223 290 316 319 362	024-569	64 73 146 153 182 235 263 309+C 315+C 523-524del	A2	021-560

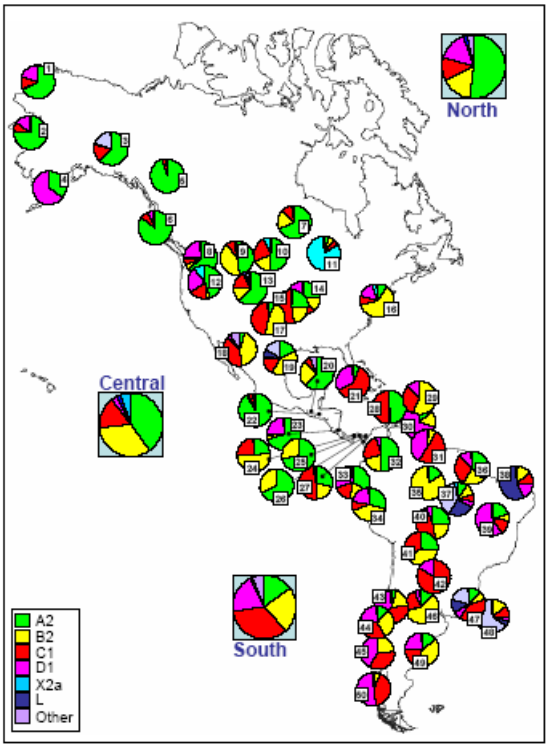
60	223 290 319 352 362	024-560	64 73 146 153 182 235 263 309+C 315+C 523-524del	A2	021-589
70	223 290 319 362	024-545	64 73 146 153 235 263 315+C 523-524del	A2	021-600
100	223 290 319 362 519	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-600
16	92 111 189 223 290 319 362 519	024-569	64 73 146 153 235 263 309+C 315+C 523-524del	A2	021-570
81	93 111 136 223 290 319 324 362	024-569	73 146 235 263 309+C 315+C 523-524del	A2	021-580
33	129 183C 189 217 269 519	024-569	73 146 152 195 234 263 309+CC 315+C 499	B2	021-570
22	183C 189 217 519	024-569	73 263 309+C 315+C 499	B2	021-570
41	183C 189 223 256 298 325 327	024-560	73 249del 263 290-291del 309+C 315+C 489 523-524del	C1	021-555
39	189 223 298 325 327 362 519	024-560	73 195 249del 263 290-291del 315+C 489	C1	021-580
12	519	024-569	153 263 315+C 523-524del	H?	021-600
52	129 148 168 172 187 188G 189 223 230 278 293 311 320	024-560	93 95C 185 189 236 247 263 315+C 523-524del	L0a 1a	021-560
30	126 271 294 296 304 519	024-569	73 263 315+C	T2	021-580
96	051 129C 189 362	024-540	73 152 217 263 309+C 315+C 340 508	U2 e	021-600

HG = haplogroup

n.d. = non determined

Figure legends

Figure 1. Haplogroup patterns in America. Codes for populations are as follow: 1 = Chukchy, 2 = Eskimos (Starikovskaya et al. 1998); 3 = Inuit (collected from the HvrBase database (Burckhardt et al. 1999); 4 = Aleuts (Rubicz et al. 2003); 5 = Athapaskan (Shields et al. 1993); 6 = Haida (Ward et al. 1993); 7 = Apache (Budowle et al. 2002), 8 = Bella Coola (Ward et al. 1993); 9 = Navajo (Budowle et al. 2002); 10 = Sioux, 11 = Chippewa (Smith et al. 1999), 12 = Nuu-Chah-Nult (Ward et al. 1991); 13 = Cheyenne (Smith et al. 1999); 14 = Muskogean populations (Bolnick and Smith 2003); 15 = Cheyenne-Arapaho (Malhi et al. 2001); 16 = Yakima (Shields et al. 1993); 17 = Stillwell Cherokee (Malhi et al. 2001); 18 = Pima (Kittles et al. 1999); 19 = Mexico (Green et al. 2000); 20 = Quiche (Boles et al. 1995); 21 = Cuba (Boles et al. 1995); 22 = El Salvador (present study); 23 = Huetar (Santos et al. 1994b); 24 = Emberá (Kolman and Bermingham 1997); 25 = Kuna (Batista et al. 1995); 26 = Ngöbé (Kolman et al. 1995); 27 = Wounan (Kolman et al. 1995); 28 = Guahibo (Vona et al. 2005); 29 = Yanomamo from Venezuela (Williams et al. 2002); 30 = Gaviao (Ward et al. 1996); 31 = Yanomamo from Venezuela and Brazil (Merriwether et al. 2002); 32 = Colombia (Torres et al. 2006); 33 = Ecuador (general population), 34 = Cayapa (Rickards et al. 1999); 35 = Xavante (Ward et al. 1996); 36 = North Brazil (Santos et al. 1996); 37 = Brazil (Alves-Silva et al. 2000); 38 = Curiau (Ribeiro-Dos-Santos et al. 1996); 39 = Zoró (Ward et al. 1996); 40 = Ignaciano, 41 = Yuracare (Bert et al. 2004); 42 = Ayoreo (Dornelles et al. 2004); 43 = Araucarians (Horai et al. 1993); 44 = Pehuenche, 45 = Mapuche from Chile (Moraga et al. 2000); 46 = Coyas (Álvarez-Iglesias et al. 2007); 47 = Tacuarembó (Bonilla et al. 2004); 48 = Uruguay (Pagano et al. 2005); 49 = Mapuches from Argentina (Ginther et al. 1993); 50 = Yaghan (Moraga et al. 2005).



Artículo 10: Mitochondrial DNA haplogroup R0 variants show spatial geographical patterns in North Iberia

Álvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintáns B, Zarrabeitia M, Lareu MV, García O, Pérez-Jurado LA, Carracedo Á, Salas A

(Manuscrito en preparación)

Las variantes del haplogrupo R0 de ADN mitocondrial muestran un patrón geográfico espacial en el Norte de Iberia

ABSTRACT/RESUMEN:

El macro-haplogrupo R0 agrupa el linaje de ADN mitocondrial (ADNmt) más común en las poblaciones en Eurasia occidental, el haplogrupo H (~40%), y otros clados menos frecuentes. Los linajes de R0 están muy mal carecterizados en la región control, por lo que es necesario el análisis de polimorfismos de la región codificante para aumentar la resolución al inferir eventos demográficos en Europa occidental. En este trabajo genotipamos la región hipervariable I (HVS-I) y 71 SNPs de la región codificante de ADNmt, identificando así las principales ramas además de otras ramas menores de R0 en 518 muestras de tres regiones diferentes del norte de Iberia. Sólo se observan algunos clados en el norte de Iberia con frecuencias significativas.

De acuerdo con trabajos previos, los patrones geográfico-espaciales de H1, H3 y H5a presentan picos de frecuencia en la región Franco-Cantábrica, y patrones de frecuencia clinal decreciente hacia el Este y Sur de Europa. Además, en este trabajo caracterizamos un nuevo linaje autóctono del haplogrupo H, el H2a5 mediante la secuenciación de genomas completos. Este linaje tiene aproximadament 15.7±9 mil años, coincidiendo con el periodo de expansion poblacional que tuvo lugar después del último periodo glacial, que contrastanto con otros linajes del haplogrupo H que experimentaron una expansión fuera del refugio Cántabro-francés (e.g. H1 and H3), el H2a5 probablemente permaneció confinado a esta área hasta la actualidad.

Mitochondrial DNA haplogroup R0 variants show spatial geographical patterns in North Iberia

Álvarez-Iglesias V[1&], Mosquera-Miguel A[1&], Cerezo M[1], Quintáns B[2], Maite Zarrabeitia[3], Lareu MV[1], García O[4], Pérez-Jurado LA[5], Carracedo Á[1], Salas A[1*]

[1] *Unidade de Xenética, Instituto de Medicina Legal, Facultad de Medicina, 15782, Universidad de Santiago de Compostela, Galicia, Spain*

[2] *Fundación Pública Galega de Medicina Xenómica (FPGMX), CIBERER, Hospital Clínico Universitario, Universidad de Santiago de Compostela, 15706, Galicia, Spain*

[3] *Medicina Legal, Universidad de Cantabria, Santander, Spain*

[4] *Laboratorio de la Ertzaintza, Bilbao, Spain*

[5] *Unitat de Genètica, Universitat Pompeu Fabra, Barcelona, Spain*

[&] Both authors contributed equally to this study

[*]**Corresponding autor:** Unidade de Xenética Forense, Universidade de Santiago de Compostela, Galicia, Spain. Tel: +34–981–582327; Fax: +34–981–580336; E-mail: apimlase@usc.es

Keyword: mtDNA; coding region; HVS-I and HVS-II; SNP; haplotype; SNaPshot; phylogeny; forensics; population studies; haplogroup H

Abstract

R0 embraces the most common mitochondrial DNA (mtDNA) lineage in West Eurasian populations (~40%), namely haplogroup H, plus other lower frequency clades. R0 lineages are poorly defined in the control region, and therefore the analysis of haplogroup characteristic coding region polymorphisms are needed in order to gain resolution for the inference of demographic events in West Europe. Here we genotyped the first hypervariable region (HVS-I) as well as 71 mtDNA coding region SNPs identifying major and minor branches of R0 in 518 samples from three different regions in North Iberia. There are only few clades that were observed in North Iberia to significant frequencies.

In agreement with previous findings, the spatial geographical patterns of H1, H3 and H5a show frequency peaks in the Franco-Cantabrian region, declining from West towards the East and South Europe. In addition, we here characterized a new autochthonous clade of haplogroup H, named H2a5 by way of complete genome sequencing. Its coalescence age, 15.7±9 thousand years (kya), dates to the period immediately after the Last Glacial Maximum (LGM), but in contrast to other H lineages that experienced re-expansion outside the Franco-Cantabrian refuge (e.g. H1 and H3), H2a5 probably remained confined to this area till present days.

Introduction

Haplogroup R0, formerly known as pre-HV (Achilli et al. 2004), is defined by the absence of transitions A73G and G11719A. There are one main sub-branch of R0 defined by the lack of C14766T (haplogroup HV) and a minor branch known as R0a. HV embraces the most frequent haplogroup in Europe (~40%), namely, haplogroup H, which is defined by the lack of the characteristic transitions A2706G and C7028T. HV also contains some other less frequent clades, such as HV1, HV2, and specially HV0, where haplogroup V is nested. Most of the haplogroup H sub-lineages are most likely of Middle Eastern origin (as it is the case for the majority of the typical West European clades). Overall, R0 shows frequencies declining from West towards East and South (Achilli et al. 2004; Loogväli et al. 2004; Roostalu et al. 2007). By way of complete sequencing, Achilli et al. (2004) have recently identified numerous sub-branches of haplogroup H. For the first time, these authors demonstrated that, although haplogroup H overall in Europe is rather uniform, the sub-clades H1 and H3 showed frequency peaks centered in Iberia and surrounding areas. The phylogeographic distribution of these lineages and their coalescence ages (~11 kya) lead the authors to conclude that H1 and H3 represent a signal of late-glacial expansion of hunter-gatherers that repopulated Central and Northern Europe from about 15,000 years ago, after the LGM. These findings corroborated the patterns previously observed for haplogroup V (Torroni et al. 2001b), also showing a clear-cut clinal geographical distribution in Europe, with a peak in the Franco-Cantabrian area and coalescence ages ranging from 16.3±4.8 kya in West Europe to 8.5±2.3 kya in the East Europe. Therefore, these three lineages pointed to a recolonization period of Europe from western refugee locations after the LGM period.

The phylogenetic power provided by the dissection of the variability within haplogroup H demonstrated by this pioneering study of (Achilli et al. 2004) was further corroborated by another

contemporary study of Loogväli et al. (Loogväli et al. 2004). These authors added new sub-lineages to the phylogeny of haplogroup H in addition to those reported in the Achilli's et al. study. Among other findings, Loogväli et al. (2004) also showed that some sub-lineages (such as H1*, H1b, H1f, H2a, H3, H6a, H6b, and H8) have also distinct phylogeographic patterns in Europe. More recently, Roostalu et al. (2007) studied the population stratification of haplogroup H sub-lineages in West Eurasia, with special focus in Near Eastern and the Caucasus. Again, the authors demonstrated that most of the haplogroup H lineages of present-day Near Eastern-Caucasus area expanded after the LGM and presumably before the Holocene. The study by Brandstätter et al. (2007) also focused on the dissection of haplogroup H although in a more technical perspective. This is because the refined knowledge of the mtDNA phylogeny to the level of complete genomes opens the door to a wide spectrum of different applications, including medical and forensic genetics applications; see also (Quintáns et al. 2004; Álvarez-Iglesias et al. 2007; Álvarez-Iglesias et al. 2008).

The study from Abu-Amero et al. (2007) was also very useful in providing with new resolution at the level of complete genome sequencing for R0a.

On the other hand, due to the growing interest of laboratories in unraveling the sub-structure of haplogroup H, several conflicts have arisen in the nomenclature of R0 sub-clades. Although this is a common feature in mtDNA studies, there is also a continuous effort for resolving these conflicts in order to ease future population and phylogenetic studies; some examples are the study by Kong et al. (2006) on the East Asian phylogeny, Achilli et al. (2008) for the Native American tree, and Brandstätter et al. (2007) for the H haplogroup. Many problems still persist however within the R0 nomenclature. The recent publication of Roostalu et al. (2007) added new branches to the phylogeny, but for instance, the label H19 is used in parallel by Achilli

et al. (2007) for a different lineage. A more clear and recent example of nomenclature conflict is given in Behar et al. (2008). These authors refer to a new clade, R0a1, with three minor sub-clades (R0a1a, R0a1b, and R0a1c); however, they did notice the previous contribution by Abu-Amero et al. (2007) where new complete genomes and new sub-branches of R0a were already reported. Thus for instance, the R0a1 branch referred as novel in (Behar et al. 2008) matches a branch reported by (Abu-Amero et al. 2007) coined (preHV)1b (therefore using also the old nomenclature; see (Achilli et al. 2004)).

Finally, another goal of the present study is to contribute to enrich our knowledge on the mtDNA phylogeny by adding new complete genomes that characterize an autochthonous clade in the Basque Country, H2a5.

Material and Methods

Samples

We have collected samples from three different regions in North Iberia. A total of 282 healthy unrelated individuals were collected from Galicia (northwest Iberia) which is an independent sample from those Galicians reported in (Salas et al. 1998; Salas et al. 2000).

Three different locations were sampled in Cantabria (N=135; North-Central Iberia), including 39 healthy unrelated individuals from Valle del Pas, 45 from the Valle del Liébana, and 51 from Santander. It is important to note that several individuals from the Valle del Pas were previously reported for the HVS-I segment in (Maca-Meyer et al. 2003). For many purposes, these three locations will be lumped in a single group (Cantabrians).

A total of 101 individuals suffering autism were collected from Catalonia (northeast Iberia). It is highly probably that mtDNA lineages do not play a role as medium to high penetrance factors in autism

(which is likely to be a polygenic disease), and therefore this sample is here considered to represent (from a mtDNA point of view) a random sample from the region. Note however that some population admixture has occurred in this population, a fact that is reflected by the presence of non-Western European lineages in this sample (Supplementary Data S3).

Eight samples were also collected from the Basque Country. In order to give a geographical reference, it is important to note this region is bordering on Cantabria.

Amplification and minisequencing protocols

Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3) was used for designing the primers both for PCR amplification and minisequencing reaction. All of them have an annealing temperature around 60º C using. The sequence databases at the National Centre for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) were interrogated using BLAST in order to test the primers against possible repetitive sequences and sequence homologies in the nuclear genome. Each primer pair for PCR amplification and each single base extension primer were selected independently and AutoDimer (http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm) was used to test for potential hairpin structures and primer-dimer problems.

The SNPs were amplified in three independent PCR multiplex reactions. The size of the amplicons ranges from 66 and 195 bp. Same SNPs are located in the same amplicon, reducing substantially the number of amplicons needed in the whole multiplex design. We performed each multiplex reaction using 5 ng of DNA template and PCR master mix of QIAGEN Kit Multiplex PCR (Qiagen, Dusseldorf, Germany), amplification primers and their final concentrations are in Supplementary Data S1. Amplification was carried out in a 9700

Thermocycler (Applied Biosystems, Foster City, CA, USA). After a 95ºC pre-incubation step for 15 minutes, PCR was performed in a total of 30 cycles using the following conditions: 94ºC denaturation for 30 seconds, annealing at 58ºC for 90 seconds and extension at 72ºC during 90 seconds, followed by a 15 minutes of final extension at 72ºC and 4ºC until removed from thermocycler.

Previous to minisequening reaction, PCR products are treated with ExoSAP-IT (Amershan Biosciences, Uppsala, Sweden) to remove excess primers an un-incorporated dNTPs: 1 µl of PCR product was incubated together with 0.5 µl of ExoSAP-IT for 15 minutes at 37ºC followed by 15 minutes at 80ºC for enzyme inactivation. Minisequencing reaction is performed using SNaPshot™ Kit (AB). Minisequencing primers range in size from 21 to 86 bps, length primers were modified by the addition of non-homologous tails, poly(dGACT) added at the 5'-end (Supplementary Data S1). The minisiquencing reaction was carried out in a total volume of 10 µl comprising 3 µl of the SNaPshot™ Kit (AB), 1.5 µl PCR product, 1 µl of extension primers mix (final concentrations are between 0.06 and 0.6 µM) (Supplementary Data S1), and water-up to 10 µl. The reaction was performed in a 9700 Thermocycler (AB) following the recommendations of the manufacturer: 25 cycles of denaturation at 96 ºC for 10 seconds, annealing at 50 ºC for 5 seconds and extension at 60 ºC during 30 seconds. Un-incorporated ddNTPs are eliminated using SAP (Amershan Biosciences). The final volume (10 µl) was treated with 1 µl of SAP for 60 minutes at 37ºC followed by 15 minutes at 80ºC for enzyme inactivation.

The minisiquencing products (2 µl) were mixed with 9.5 µl of HiDi™ formamide (AB) and 0.3 µl of GeneScan –120 LIZ (AB) and capillary electrophoresis was undertaken on an ABI PRISM 3130xl Genetic Analyzer (AB) .Resulting data were analyzed using GeneMapper™ 3.7 Software (AB).

Automatic sequencing

The samples from Galicia, Cantabria, and Catalonia were sequenced for the HVS-I. PCR amplification was carried out in a 9700 Thermocycler (AB). The temperature profile for 32 cycles of amplification was 95º C for 10 sec, 60º C for 30 sec, and 72º C for 30 sec. Sequencing primers were previously described by (Wilson et al. 1995). PCR product purification and sequencing were performed as in (Salas et al. 1998). For those sequences containing a homopolymeric cytosine stretch from positions 16184 to 16193 (usually associated with length heteroplasmy), additional amplification and sequencing were performed using primers: L16209 (5'-CCC CAT GCT TAC AAG CAA GT-3') and H16164 (5'-TTT GAT GTG GAT TGG GTT T-3').

Complete genome sequencing

The eight samples from the Basque Country were sequenced for the complete genome. The primers used for PCR amplification and sequencing were those reported in (Torroni et al. 2001a). The PCR was performed in 10µl of reaction mix, containing 4 µl of *Taq* PCR Master Mix (Qiagen, Hilden, Germany), 0.5 µl 1µM of each primer, 1 µl sample template and 4 µl of water. This PCR were carried out in a termocycler GenAmp PCR System 9700 (Applied Biosystems) with one cycle of 95ºC for 15 minutes and then 35cycles of 94ºC for 30 seconds, 58ºC for 90 seconds and 72ºC for 90 seconds with a full extension cycle of 72ºC for 10 minutes. The PCR product was checked in agarose gels and purified using QIAquick^{R} Gel Extraction Kit (Qiagen).

Sequencing reaction was performed in 11.5 µl of reaction mixture, containing 2.5µl of sequencing buffer (5X), 0.5 µl of BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), 1 µl of the corresponding primer (final concentration was 1 µM), 3 µl of the purified PCR product and water up to 11.5 µl. Sequencing reaction was carried out in a termocycler GenAmp PCR System 9700 (Applied Biosystems)

with one cycle of 96ºC for 3 minutes and then 25cycles of 96ºC for 30 seconds, 50ºC for 15 seconds and 60ºC for 4 minutes or was carried out in a 9800 Fast Termal Cycler (Applied Biosystems) with one cycle of 96ºC for 1 minute then 25 cycles of 96ºC for 10 seconds, 50ºC for 5 seconds and 60ºC for 1 minute. To obtain 'clean' electropherograms, the sequencing product was doubled purified, first using Montage™ SEQ_{96} Sequencing Reaction Cleanup Kit (Millipore) according to manufacturer protocols, followed by purification with Sephadex™ G-10 (Amersham Biosciences, Uppsala, Sweden), the latter also according to manufacturer protocol. MtDNA automatic sequencing was carried out in a capillary electrophoresis ABI3730 (Applied Biosystems).

Monitoring genotyping errors

We have used the mtDNA tree as a reference to avoid as much as possible artefactual profiles and documentation errors in mtDNA sequences and in SNP genotypes (Bandelt et al. 2001; Bandelt et al. 2002; Yao et al. 2003; Bandelt et al. 2004a; Bandelt et al. 2004b; Bandelt et al. 2005b; Bandelt et al. 2005a; Salas et al. 2005a; Salas et al. 2005c; Salas et al. 2005b; Kong et al. 2006; Salas et al. 2006; Yao et al. 2006).

When detecting some unexpected SNP pattern, we rechecked the genotyping by repeating the SNP genotyping using single plex minisequencing and automatic sequencing, as exercised in (Álvarez-Iglesias et al. 2007).

Genetic diversity estimates

DnaSP 4.10.3 software (Rozas et al. 2003) was use for computation of different diversity indices, including haplotype and nucleotide diversities and mean number of pairwise differences (Tajima 1983; Nei 1987; Tajima 1993). Departure from normal distribution of pairwise differences was checked using the Harpending`s r (raggedness)

index (Harpending 1994). Selective neutrality was tested using the Tajima (Tajima 1989) and Fu and Li tests (Fu and Li 1993). Principal component analysis was performed using Stata 9.1 (http://www.stata.com/).

Analysis of geographical patterns

A geographical description of allele frequency patterns was obtained using Surfer 8.0 (http://www.goldensoftware.com) by drawing maps of haplogroup frequencies for the most common lineages. We used the inverse-squared distance method. Haplogroup frequencies are presented in a regular grid covering part of Eurasia (including Europe), Middle East and the Arabian Peninsula. Only data points within the same landmass, either island or continent, were considered for interpolation.

In addition, we have carried out analysis of spatial autocorrelation using the Spatial Autocorrelation Analysis Program (SAAP; http://www.exetersoftware.com/cat/saap.html) in order to detect and evaluate statistically signals of gradients (clines), gradients irradiating from the center of a particular area (depressions) or isolation by distance models (see for instance (Barbujani 2000)).

Phylogeographic analysis and nomenclature

R0 and its different sublineages are the main focus of the present article, and therefore, only few studies are used here for population comparisons, in particular those presenting population data on a number of haplogroup R0 sub-lineages (Macaulay et al. 1999; Achilli et al. 2004; Loogväli et al. 2004; Quintáns et al. 2004; Achilli et al. 2005; Achilli et al. 2007; Roostalu et al. 2007). Nomenclature of lineages follows that proposed in the latter studies. The conflicts in nomenclature are clarified in the present study. We also describe in the text some new minor branch of haplogroup H.

MtDNA variation is referred to the revised Cambridge Reference Sequence (Andrews et al. 1999). A database containing more than 30,000 West Eurasian mtDNA profiles was used here for phylogeographic purposes. All the data in this database is available in the public literature on mtDNA European population studies.

Coalescence ages

Estimation of the time to the most recent common ancestor of each cluster and SDs were carried out according to (Saillard et al. 2000) and employing an evolutionary rate estimate of 1.26 x 0.08 x 10^{-8} base substitutions (other than a deletion or insertion) per nucleotide per year in the coding region (between 577 and 16023), corresponding to 5,140 years per substitution in the entire coding region (Mishmar et al. 2003).

Results

The R0 phylogeny and SNP selection

R0 differs from R* by lacking A73G and G11719A. R0 contains haplogroup HV which likewise embraces the most common haplogroup in Europe, H, but also haplogroup HV0a (where haplogroup V is nested) and some other minor branches such as HV1 and HV2. Within haplogroup H, there are more than 20 sub-haplogroups described to date; many of then can be further sub-divided in minor branches.

SNP selection in the present study considers the full set of SNPs reported in (Brandstätter et al. 2006) (with the exception of 14552 which is replace here by 3936; both leading to haplogroup H12) plus a selection of additional variants that define further sub-branches of R0 within Europe (see e.g. (Roostalu et al. 2007)). In addition, the analysis of the literature and complete genomes sequences available in GenBank has allowed us to infer new minor sub-lineages of R0; their diagnostic sites were also considered for genotyping (see below).

On the other hand, there are several inconsistencies in the nomenclature of haplogroup H and its sub-lineages (see also SNP selection section and Figure 2 in (Brandstätter et al. 2006) and compare with Figure S1 of (Roostalu et al. 2007)). The updated classification tree of haplogroup R0 and its sub-clades is shown in Figure 1 and 2. Nomenclature of new clades and changes with respect to previous nomenclature (e.g. Figure S1 of (Roostalu et al. 2007)) are as follow:

- T16209C, together with T16365C, is part of the motif of H1a1 (T16209C is present in all complete genomes from (Coble et al. 2004));
- T16263C, together with A9150G, is part of the motif of H1c1 (see for instance (Coble et al. 2004);
- According to (Brandstätter et al. 2007), 8764 is part of the motif of H1c2 coupled with C12858T;
- we keep the branches bearing T8473C and T8602C as considered in (Brandstätter et al. 2006); however, both carry transition G3010A, and therefore, the nomenclature would better be H1-8473 and H1-8602. Note however that these two sites are relatively unstable, and therefore, both can appear also outside H1. Position 709 is also highly recurrent, and therefore, we now record two different branches in the tree defined by this transition, namely, H-709 and H1-709;
- H2a5 is a new branch defined by position A1842G, T4592C, G13708A, C16291T and T16519C.
- variant A73G also characterizes a sub-clade of H3 supported by three complete genomes (AY495177, AY495180, AY738983 from (Coble et al. 2004) and (Achilli et al. 2004)), here named as H3d;

- there is a sub-branch of H4a1a carrying transition A73G (and 523-524del, H4a1a1; see also (Brandstätter et al. 2006)). The cluster H4a1a defined by (Behar et al. 2008) is here renamed as H4a1b in order to preserved the older nomenclature.
- 961G is part of the motif characterizing haplogroup H11, while H11a is defined by position 8898. Transition A16293G is not always present in H11 and H11a and therefore it is not considered here as part of the H11a motif;
- sub-clade H13a1 bears transition A4745G but also C13680T (note that there is only one out of 18 H13a1 complete genomes that lacks 13680, Genbank accession number: AM263187), and contains a minor new branch, H13a1a, defined by G7337A;
- 11377 is not part of the motif of haplogroup H14 as previously suggested by (Achilli et al. 2004);
- H16 embraces an small branch, H16a, defined by G8592A (data from (Coble et al. 2004));
- we keep the nomenclature of Roostalu et al. (2007) for H19, defined by transition 14869; while the H19 'Etruscan' lineage defined by G16145A A16227G is here renamed as H22;
- there is a new minor branch, H23, defined by transition 10211 (see (Coble et al. 2004) and (Herrnstadt et al. 2002)).
- We baptize new clades within HV0a. With respect to the nomenclature provided in (Palanichamy et al. 2004), we here define the following sub-clades: a) V1a1 is defined by A227G T485C and G709C (supported by data from (Finnilä et al. 2001)), b) V2a1 harbors T9088C and T14793C, and there is a sub-branch V2a1a defined by C5250T, c) V2b is characterized by C14770T T16519C, and V2b1 has in addition G15773A (see

data from (Coble et al. 2004)), d) haplogroup V3 bears variants T4550C A8347G A12810G T13500C G15346A T16519C (see AY495328 and AY495330 in (Coble et al. 2004)).

Apart from the mentioned SNPs, our assay also incorporates SNPs that are diagnostic of HV1 and HV2 (sister clades of H and HV0) and others covering several major branches of haplogroup R, namely, haplogroup U (A12308G) and JT (C15452A). The transition C12705T defining macro-haplogroup N was also included.

Table 1 summarizes the nomenclature used in different studies concerning R0 sub-clades and try to resolved the existent conflicts in order to facilitate future inter-population genetic studies. In order to present the last update of the R0 nomenclature, we have also included the new very recently described R0 clades from (Behar et al. 2008).Summarizing, we have genotyped SNPs that cover the whole known R0 phylogeny, which embraces the most prevalent European haplogroup (H).

The minisequencing high throughput SNP genotyping protocol

Analysis of mtDNA SNPs always supposes a challenge for those laboratories carrying out tedious RFLP analysis of several dozen coding region SNPs. In addition to the high cost and personal effort related to the genotyping of coding region SNPs, there is also an immanent risk for sample cross-over and contamination due to the number of PCR amplicons involved. In order to overcome this problem, the SNP genotyping carried out in the present study was based on a minisequencing technique that have demonstrated to be of great utility in many clinical, forensic, and population genetic application (Quintáns et al. 2004; Álvarez-Iglesias et al. 2007; Álvarez-Iglesias et al. 2008). The minisequencing reactions designed in the present study allow to genotype a total of 71 SNPs in three simple reactions. This method tries to avoid the hierarchical SNP genotyping approach which is the

standard procedure in Y-chromosome SNPs genotyping studies but also quite common in the mtDNA literature. The advantage of genotyping all the SNPs in all the samples (which tends to produce redundant data due to the haplotypic nature of the mtDNA variation) is to gather more information on mutation rates and phylogeny, but also it prevents e.g. artefactual recombination due to sample mixing.

Some phylogenetic inconsistencies were observed (and are indicated in Supplementary Table S3) in the reported dataset. It is noteworthy to mention some of them because some important diagnostic sites of the mtDNA phylogeny are involved. For instance, G11719A is present in three H* samples (then, all carry A2706G, C7028T, C14766T); two from Galicia (#C170 and #C203) and one from Catalonia (#AU87). Also interesting is the case of sample #AU74 from Catalonia. This sample belongs to HV (it lacks e.g. G11719A C14766T C16223T) and it carries A2706G and C7028T (it does not belong to haplogroup H). However, this sample carries G4580A which defines haplogroup V but it does not carries the necessary mutations C15904T and T16298C for this haplogroup. Sample #C25 belongs to H6a1 and carries the variant C15452T; note that the transversion C15452A defines haplogroup JT. To our knowledge there is only one record for C15452T, but since it occurs within haplogroup J (Uusimaa et al. 2004), it could just represent a documentation error from the necessary transversion at this position.

Diversity patterns of R0 in North Iberia

Several diversity indices were computed for the three North Iberian samples genotyped in the present study (Table 2). Overall, the Catalonian sample shows the highest values of sequence and nucleotide diversity (with no overlapping ranges with the other two North Iberian samples) and also for the average number of nucleotide differences. The

Cantabrian region shows the lowest values again for the three mentioned indices.

The patterns of variability within haplogroup H are quite different around Europe and Middle East. For instance, Galicia shows one of the lowest sequence diversity in Eurasia (Table 2) such as it was previously infered in an independent sample (Salas et al. 1998), and among the lowest values of nucleotide diversities (together with Cantabria).

Both the Tajima`s *D* and the Fu and Li's tests show significantly negative values in almost all the populations (Table 2) suggesting that all of them have experienced population expansions. The mismatch distributions (data not shown) corroborate this hypothesis as well as the raggedness *r* index (Table 2) indicating that the distributions approach quite well to the normality.

Phylogeographical patterns of R0 sub-lineages

Less than 10% of the lineages within haplogroup H could not be allocated to some of the H sub-branches. On average, ~42% of the lineages in the total sample belong to haplogroup H, but the Galician population has the highest frequency (~44%) and it slightly decreases to the ~39% in Cantabria and Catalonia. The maps of Figure 3 show the spatial frequency distribution of different sub-lineages of haplogroup R0. Some lineages get the highest frequencies in Iberia, such as H1, H3, and H5a or are only observed in this region H4; while other are virtually absent in Iberia but are significantly more prevalent in central Europe (e.g. H11).

Haplogroups H1, H3 and H5a display clinal patterns as determined by their spatial correlograms (Supplementary Data S5). The frequency of these three lineages have a peak in the the Franco-Cantabrian refuge area and decline towards West Europe.

The authoctonous nature of the H2a5 clade in the Basque Country

It was notice in a study by Pereira et al. (2004) by way of sequencing several small coding region fragments of the mtDNA, the presence of the coding region variant C4952T in 6% of their samples (4 or 5 mtDNAs) from the Basque Country. An scrutiny of more than 4,300 coding region segments (most of them available in GenBank) and in Google searches (*sensus* (Bandelt et al. 2006; Bandelt et al. 2008)) revealed that this variant was only reported twice in the literature, curiously in two medical studies, namely, one in (Schwartz and Vissing 2002) and another instance in (Pulkes et al. 2003). Therefore, the multiple occurrence of this transition in the Basque Country could point to a diagnostic site for an autochthonous lineage in this region. These features lead us to further investigate this mtDNAs by way of complete genome sequencing in the eight available Basques carrying the transition C4592T.

Six out of the eight complete genomes are identical while the other two show one private variant each. This analysis revealed a new sub-clade of haplogroup H5a2, all of the members share the following set of variants: A1842G C4592T G13708A C16291T (Figure 4). The coalescence age for this sub-lineage is 15.7±9 kya.

Discussion

A total of 518 samples from three main locations in North Iberia were sequenced for the HVS-I segment. 52% of them could be ascribed to R0. All these samples were further screened for a set of 71 coding region SNPs in order to sub-classify them into different R0 sub-clades; most of them sub-branches of haplogroup H.

The spatial frequency patterns of H1, H3 and H5a are statistically clinal in Europe, with frequency peaks in the Franco-Cantabrian region decreasing towards East Europe. This is compatible with patterns of

repopulation of Europe from this refugia after LGM period, as it was previously demonstrated by (Torroni et al. 2001b; Achilli et al. 2004). As indicated by the various diversity indices computed, Galicia and Cantabria show low values of diversity, especially for the overall haplogroup H.

We have also described a new autochthonous clade which seems to be autochthonous in the Basques, H2a5. This clade has been dated in 15.7±9 kya; this age fits with the period of population expansion that followed the LMG. However, this branch was exclusively found in this region and therefore it did probably not participate in the re-colonization of Europe from the Franco-Cantabrian refugee. Its low frequency in the Basque country (~6%) could also explain why members of H2a5 still remain undetected in other European locations.

Analysis of mtDNA variation based on exclusively on few RFLP markers and/or the HVS-I region have lead in the past to simplistic perceptions of Europe as a uniform population. The results presented in previous studies (Richards et al. 2002; Achilli et al. 2004; Loogväli et al. 2004; Brandstätter et al. 2007; Roostalu et al. 2007) and those showed here demonstrate that population stratification in European population can only be revealed when using high resolution genotyping. Analysis of complete genomes ideally provides with the maximum level of resolution; however, complete genome genotyping demands great economical and personal effort in population-scale projects. In fact, in most of the publications, complete genome sequencing is general restricted to the analysis of particular mtDNA lineages that present some interesting phylogenetic or population feature, such as the analysis of haplogroup H lineages in (Achilli et al. 2004; Loogväli et al. 2004), or as in the present study, the analysis of few members of haplogroup H2a5. The coding region genotyping strategy presented here represents a way to overcome the drawback of whole genome genotyping and allow at the same time to obtain high resolution information from

the mtDNA genome. There are other potential applications of the minisequencing design presented here. For instance, it performs well with typical forensic evidences that consist of low DNA samples because PCR primers were deliberately designed in order to render amplicons of small size. Other virtues of the minisequencing approach are described in previous studies (Álvarez-Iglesias et al. 2006; Montesino et al. 2006; (Crespillo et al. 2005). On the other hand, the present design covers an important part of the Eurasian mtDNA phylogeny and therefore this helps to increase significantly the discrimination power of the mtDNA test.

Acknowledgement

We would like to thank Francesc Calafell and Oscar Lao for their help with the Surfer software and the spatial representations. The 'Ramón y Cajal' Spanish programme from the Ministerio de Educación y Ciencia (RYC2005-3), the grant of the Xunta de Galicia (PGIDIT06PXIB208079PR), and the grant from the Fundación de Investigación Médica Mutua Madrileña, given to AS, partially supported this project.

References

Abu-Amero KK, González AM, Larruga JM, Bosley TM, Cabrera VM (2007) **Eurasian and African mitochondrial DNA influences in the Saudi Arabian population.** BMC Evol Biol 7: 32.

Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP et al. (2008) **The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies.** PLoS ONE 3(3): e1764.

Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A et al. (2005) **Saami and Berbers--an unexpected mitochondrial DNA link.** Am J Hum Genet 76(5): 883-886.

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A et al. (2004) **The molecular dissection of mtDNA haplogroup H confirms that the Franco-**

Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet 75(5): 910-918.

Achilli A, Olivieri A, Pala M, Metspalu E, Fornarino S et al. (2007) **Mitochondrial DNA variation of modern Tuscans supports the near eastern origin of Etruscans.** Am J Hum Genet 80(4): 759-768.

Álvarez-Iglesias V, Jaime JC, Carracedo Á, Salas A (2007) **Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups.** Forensic Sci Int: Genet 1: 44-55.

Álvarez-Iglesias V, Barros F, Carracedo Á, Salas A (2008) **Minisequencing mitochondrial DNA pathogenic mutations.** BMC Med Genet in press.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM et al. (1999) **Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA.** Nat Genet 23: 147.

Bandelt H-J, Salas A, Bravi CM (2004a) **Problems in FBI mtDNA database.** Science 305(5689): 1402-1404.

Bandelt H-J, Salas A, Lutz-Bonengel S (2004b) **Artificial recombination in forensic mtDNA population databases**. Int J Legal Med 118(5): 267-273.

Bandelt H-J, Salas A, Bravi CM (2006) **What is a 'novel' mtDNA mutation – and does 'novelty' really matter?** J Hum Genet 51(12): 1073-1082.

Bandelt H-J, Yao Y-G, Salas A (2008) **The search of 'novel' mtDNA mutations in hypertrophic cardiomyopathy: MITOMAPping as a risk factor.** Int J Cardiol: in press (doi:10.1016/j.ijcard.2007.1002.1049).

Bandelt H-J, Lahermo P, Richards M, Macaulay V (2001) **Detecting errors in mtDNA data by phylogenetic analysis.** Int J Legal Med 115(2): 64-69.

Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V (2002) **The fingerprint of phantom mutations in mitochondrial DNA data.** Am J Hum Genet 71(5): 1150-1160.

Bandelt H-J, Kong Q-P, Parson W, Salas A (2005a) **More evidence for non-maternal inheritance of mitochondrial DNA?** J Med Genet 42: 957-960.

Bandelt H-J, Achilli A, Kong Q-P, Salas A, Lutz-Bonengel S et al. (2005b) **Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies.** Biochem Biophys Res Commun 333(1): 122-130.

Barbujani G (2000) **Geographic patterns: how to identify them and why**. Hum Biol 72(1): 133-153.

Behar DM, Metspalu E, Kivisild T, Rosset S, Tzur S et al. (2008) **Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora.** PLoS ONE 3(4): e2062.

Bowen J, Richards T, Maravilla K (1998) **MR imaging and proton MR spectroscopy in A-to-G substitution at nucleotide position 3243 of leucine transfer RNA.** AJNR Am J Neuroradiol 19(2): 231-234.

Brandstätter A, Salas A, Niederstätter H, Gassner C, Carracedo Á et al. (2007) **Dissection of mitochondrial superhaplogroup H using coding region SNPs.** Electrophoresis(13): 2541-2550.

Brandstätter A, Zimmermann B, Wagner J, Göbel T, Röck A et al. (2008) **Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East.** BMC Evol Biol(submitted).

Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT et al. (2004) **Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians.** Int J Legal Med 118(3): 137-146.

Crespillo M, Paredes MR, Prieto L, Montesino M, Salas A et al. (2006) **Results of the 2003-2004 GEP-ISFG collaborative study on mitochondrial DNA: Focus on the mtDNA profile of a mixed semen-saliva stain.** Forensic Sci Int. 13;160 (2-3):157-67.

Finnilä S, Lehtonen MS, Majamaa K (2001) **Phylogenetic network for European mtDNA.** Am J Hum Genet 68: 1475-1484.

Fu YX, Li WH (1993) **Statistical tests of neutrality of mutations.** Genetics 133(3): 693-709.

Harpending HC (1994) **Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution.** Hum Biol 66(4): 591-600.

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM et al. (2002) **Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences from the major African, Asian, and European haplogroups.** Am J Hum Genet 70: 1152-1171.

Kong Q-P, Bandelt H-J, Sun C, Yao Y-G, Salas A et al. (2006) **Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations.** Hum Mol Genet 15(13): 2076-2086.

Loogväli E-L, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T et al. (2004) **Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia.** Mol Biol Evol 21(11): 2012-2021.

Maca-Meyer N, Sánchez-Velasco P, Flores C, Larruga JM, González AM et al. (2003) **Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain).** Ann Hum Genet 67(Pt 4): 329-339.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F et al. (1999) **The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs.** Am J Hum Genet 64(1): 232-249.

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG et al. (2003) **Natural selection shaped regional mtDNA variation in humans.** Proc Natl Acad Sci U S A 100(1): 171-176.

Nei N (1987) **Molecular evolutionary genetics**. New York. Columbi University Press.

Palanichamy Mg, Sun C, Agrawal S, Bandelt H-J, Kong Q-P et al. (2004) **Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia.** Am J Hum Genet 75(6): 966-978.

Pereira L, Richards M, Alonso A, Albarrán C, Garcia O et al. (2004) **Subdividing mtDNA haplogroup H based on coding-region polymorphisms—a study in Iberia.** Int Congress Series 1261: 416-418.

Pulkes T, Liolitsa D, Nelson IP, Hanna MG (2003) **Classical mitochondrial phenotypes without mtDNA mutations: the possible role of nuclear genes.** Neurology 61(8): 1144-1147.

Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV et al. (2004) **Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing.** Forensic Sci Int 140(2-3): 251-257.

Richards M, Macaulay V, Torroni A, Bandelt H-J (2002) **In search of geographical patterns in European mitochondrial DNA.** Am J Hum Genet 71(5): 1168-1174.

Roostalu U, Kutuev I, Loogväli E-L, Metspalu E, Tambets K et al. (2007) **Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective.** Mol Biol Evol 24(2): 436-448.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) **DnaSP, DNA polymorphism analyses by the coalescent and other methods.** Bioinformatics 19(18): 2496-2497.

Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) **mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion.** Am J Hum Genet 67(3): 718-726.

Salas A, Bandelt HJ, Macaulay V, Richards MB (2006) **Phylogeographic investigations: The role of trees in forensic genetics.** Forensic Sci Int in press.

Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo Á (1998) **mtDNA analysis of the Galician population: a genetic edge of European variation.** Eur J Hum Genet 6(4): 365-375.

Salas A, Lareu MV, Sánchez-Diz P, Calafell F, Carracedo A (2000) **mtDNA hypervariable region II (HVII) sequences in human evolution**

studies: impact of mutation rate heterogeneity. Progress in Forensic Genetics 8: 329-331.

Salas A, Carracedo Á, Macaulay V, Richards M, Bandelt H-J (2005a) **A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics.** Biochem Biophys Res Commun 335(3): 891-899.

Salas A, Yao Y-G, Macaulay V, Vega A, Carracedo Á et al. (2005b) **A critical reassessment of the role of mitochondria in tumorigenesis.** PLoS Med 2(11): e296.

Salas A, Prieto L, Montesino M, Albarrán C, Arroyo E et al. (2005c) **Mitochondrial DNA error prophylaxis: assessing the causes of errors in the GEP'02-03 proficiency testing trial.** Forensic Sci Int 148(2-3): 191-198.

Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. N Engl J Med 347(8): 576-580.

Tajima F (1983) **Evolutionary relationship of DNA sequences in finite populations.** Genetics 105(2): 437-460.

Tajima F (1989) **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** Genetics 123(November): 585–589.

Tajima F (1993) **Unbiased estimation of evolutionary distance between nucleotide sequences.** Mol Biol Evol 10(3): 677-688.

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D et al. (2001a) **Do the four clades of the mtDNA haplogroup L2 evolve at different rates?** Am J Hum Genet 69(6): 1348-1356.

Torroni A, Bandelt H-J, Macaulay V, Richards M, Cruciani F et al. (2001b) **A signal, from human mtDNA, of postglacial recolonization in Europe.** Am J Hum Genet 69(4): 844-852.

Uusimaa J, Finnilä S, Remes AM, Rantala H, Vainionpää L et al. (2004) **Molecular epidemiology of childhood mitochondrial encephalomyopathies in a Finnish population: sequence analysis of entire mtDNA of 17 children reveals heteroplasmic mutations in tRNAArg, tRNAGlu, and tRNALeu(UUR) genes.** Pediatrics 114(2): 443-450.

Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B (1995) **Validation of mitochondrial DNA sequencing for forensic casework analysis.** Int J Legal Med 108(2): 68-74.

Yao Y-G, Macaulay V, Kivisild T, Zhang Y-P, Bandelt H-J (2003) **To trust or not to trust an idiosyncratic mitochondrial data set.** Am J Hum Genet 72(5): 1341-1346; author reply 1346-1349.

Yao YG, Salas A, Bravi CM, Bandelt HJ (2006) **A reappraisal of complete mtDNA variation in East Asian families with hearing impairment**. Hum Genet 119(5): 505-515.

Table 1. Nomenclature of haplogroup R0 clades. The SNP/motif column only list diagnostic variants for the tip branches of each (sub)-clade; the full phylogenetic path for each sub-clade is indicated in Figure 1 and 2.

SNP/Motif	[1] *1	[2] *1	[3] *1	[4] *1	[5] *1	[6]	[7]	[8]
7028	H	H	H	H	H	H	H	H
2706								
709				H-709				H-709
3010	H1	H1	H1	H1	H1		H1	H1
16519								
73/16162		H1a	H1a	H1a	H1a			H1a
6365				H1a1	H1a1			H1a1
16209								
8271T				H1a2	H1a2			H1a2
16189/16356		H1b	H1b	H1b	H1b		*2	H1b
3796								
16362								
477				H1c	H1c			H1c
9150				H1c1	H1c1			H1c1
16263								
12858				H1c2	H1c2			H1c2
8764								
456					H1d			H1d
4452/7309/9066/16189		H1f		H1f	H1f			H1f
16093								
5460							H1e	H1e
8602/14212				H1-8602				H1g
267/14053							H1o	H1o
13470							H1p	H1p
8473				H1-8473				H1-8473
709								H1-709
1438		H2	H2	H2	H2	H2		H2
4769	H2				H2a			H2a
951		H2a		H2a	H2a1			H2a1
16354		H2a1		H2a1				
750		H2b		H2b	H2a2			H2a2
10810/16274			H2c	H2c	H2a3			H2a3
11140/16519					H2a4			H2a4
1842/4592/13708/16291/16519								H2a5
8598/16311					H2b			H2b
6776	H3	H3	H3	H3	H3			H3
13404			H3a	H3a	H3a			H3a
2581				H3b	H3b			H3b
12957				H3c	H3c			H3c
73								H3d
3992	H4	H4	H4	H4	H4		H4	H4
5004/9123								
4024/14365/14582					H4a		H4a	H4a
8269		H4a		H4a	H4a1		H4a1	H4a1
10044				H4a1	H4a1a			H4a1a
073				H4a1a				H4a1a1
195/5773/13889							H4a1a	H4a1b
10166					H4b			H4b
456/16304		H5	H5	H5	H5			H5
4336	H5	H5a	H5a	H5a	H5a			H5a
15833		H5a1		H5a1	H5a1			H5a1
239/16362/16482		H6	H6	H6	H6		H6	H6
3915	H6	H6a	H6a	H6a	H6a		H6a	H6a
9380		H6a1		H6a1	H6a1		H6a1	H6a1
4727								
11253							H6a1a	H6a1a
7325/16311							H6a1a1	H6a1a1

10589							H6a1b	H6a1b
6218/7859/16284/16519							H6a1b1	H6a1b1
16300		H6b	H6b	H6b	H6b			H6b
4793	H7	H7	H7	H7	H7			H7
13101C/16288/16362		H8	H8	H8	H8			H8
709								
146								
195								
6869/9804*3		H9						
3591/4310/13020			H9	H9	H9			H9
16168								
14470A		H10	H10	H10	H10			H10
8448/13759/16311		H11a	H11	H11	H11			H11
195								
961G				H11a	H11a			
16293								
8898			H11					H11a
3936/14552			H12	H12	H12			H12
16287								
195								H12
16519							H13	
14872			H13	H13	H13			H13
2259					H13a		H13a	H13a
4745				H13a	H13a1			H13a1
13680								
7337			H13a	H13a1				H13a1a
709					H13a2		H13a2	H13a2
1008							H13a2a	H13a2a
183/11151							H13a2a1	H13a2a1
5899.1C/13762G/16311							H13a2b	H13a2b
10217			H14		H14		H14	H14
7645				H14				
11377								
16256/16352					H14a		H14a	H14a
146/7864/12870							H14a1	H14a1
55/57/6253			H15	H15	H15			H15
10394				H16	H16			H16
8592								H16a
3915/6296				H17	H17			H17
13708					H18			H18
14869					H19			H19
16218/16328A					H20		H20	H20
249d/292/16362							H20a	H20a
8994					H21			H21
16145/16227								H22*4
10211								H23
9620/16519							H25	H25
195/8638/14241/16192							H30	H30
14766	HV	*6	HV*	HV	*6	HV	*6	HV
8014T/15218/16067			HV1	HV1			HV1	HV1
8277/15927/16355							HV1a	HV1a
4227							HV1a1	HV1a1
9554							HV1a1a	HV1a1a
150/4257							HV1a1a1	HV1a1a1
2626/4739/16274							HV1b	HV1b
182/13933/15930/16172							HV1c	HV1c
9336/16217								HV2
16298	pre-V		Pre-V					HV0*5
72								
15904			Pre*V2	pre*V2				HV0a*5
195/198/8520							HV0b	HV0b
4580	V		V	V				V
4639/8869								V1
5263								V1a
227/485/709								V1a1
13105								V2

12438								V2a
9088/14793								V2a1
5250								V2a1a
14770/16519								V2b
15773								V2b1
4550/8347/12810/13500/15346/16519								V3
2442/3847/13188/16126/16362			pre-HV)1	(pre-HV)1		(preHV)1	R0a	R0a*5
64							R0a1	
2355/15674						(preHV)1b		R0a1
60+T								R0a1b
8701						(preHV)1b1		R0a1b1
601T							R0a1	R0a1a
11719/12040/14544*6							R0a1a	R0a1a1
5237/15924/16305T							R0a1b	R0a1a2
58/13474/16304							R0a1c	R0a1a3
827						(preHV)1a		R0a2*7
8292/11716/16355						(preHV)1a1		R0a2a*7
57+T/2355/15674								R0a3*7
73/11719	*6	*6	pre-HV	*6	*6	pre-HV	R0*5	R0*5

Note: Variation around the homopolymeric tracts (e.g. around 16189 and 310) and the dinucleotide marker 515-524 is not considered here

*[1][1] Quintáns et al. (2004), [2] Loogväli et al. (Bowen et al. 1998; Loogväli et al. 2004), [3] Achilli et al. (Achilli et al. 2004), [4] Brandstätter et al. (Brandstätter et al. 2007) [5] Roostalu et al. (Roostalu et al. 2007), [6] Abu-Amero et al. (2007), [7] Behar et al. (2008), and [8] Present study.

*[2]Note that Behar et al. (2008) used only control region information for H1b profiles probably assuming that 16189 and 16356 are part of the diagnostic motif (see their Supplementary Data Table S1).

*[3]As reported in (Achilli et al. 2004), this branch is actually asub-clade of H6

*[4] H22 was previously designated by (Achilli et al. 2007) as H19 (see text)

*[5] Nomenclature according to (Achilli et al. 2007)

*[6] These motifs are not explicitly mentioned in these studies (although it is universally accepted) and therefore we keep empty this cells in order to be consistent with the rest of the Table.

*[7] We here modify the nomenclature by (Abu-Amero et al. 2007) in order to be consistent with previous nomenclature and the new branches of R0 reported in (Behar et al. 2008)

Table 2. Summary statistics of HVS-I sequences in the North Iberian populations analyzed in the present article and other European populations.

HG	Population	*N*	*k*	*S*	N_{mut}	*H* ± SE	π ± SE	*M*	$V_O(M)$	*r*	*D*	*FL*
All the sample												
	Galicia[*1]	282	150 (0.53)	93	102	0.952 ± 0.010	0.0138 ± 0.001	3.76	5.06	0.012	-2.328**	-4.727**
	Catalonia[*1]	101	79 (0.78)	71	73	0.984 ± 0.007	0.0166 ± 0.001	4.59	6.29	0.014	-2.187**	-3.557**
	Cantabria[*1]	135	61 (0.45)	60	62	0.971 ± 0.007	0.0135 ± 0.001	3.72	3.85	0.018	-2.099*	-2.596*
HG-H												
	Galicia[*1]	124	51 (0.41)	49	50	0.800 ± 0.038	0.006 ± 0.001	1.73	2.08	0.035	-2.528***	-4.447**
	Catalonia[*1]	44	30 (0.68)	33	33	0.937 ± 0.030	0.009 ± 0.001	2.48	1.93	0.043	-2.300**	-3.836**
	Cantabria[*1]	52	26 (0.50)	25	26	0.875 ± 0.042	0.006 ± 0.001	1.78	1.33	0.067	-2.251**	-2.480**
	Volga-Ural[*2]	50	18 (0.36)	17	18	0.819 ± 0.049	0.006 ± 0.001	1.61	1.39	0.050	-1.884*	-1.966
	Finnland[*2]	31	16 (0.52)	15	16	0.908 ± 0.035	0.009 ± 0.001	2.42	1.53	0.092	-1.338	-1.083
	Estonia[*2]	50	31 (0.62)	30	31	0.936 ± 0.026	0.009 ± 0.001	2.54	2.31	0.035	-2.114*	-3.113*
	Slovakia[*2]	50	30 (0.60)	31	30	0.939 ± 0.027	0.009 ± 0.001	2.49	2.23	0.045	-2.090*	-2.455*
	France[*2]	50	19 (0.38)	17	19	0.762 ± 0.063	0.005 ± 0.001	1.31	1.33	0.097	-2.187**	-2.569*
	Balkans[*2]	50	31 (0.62)	30	31	0.953 ± 0.018	0.009 ± 0.001	2.52	1.77	0.053	-2.120*	-2.852*
	Turkey[*2]	50	31 (0.62)	27	31	0.914 ± 0.032	0.008 ± 0.001	2.13	1.59	0.055	-2.311**	-3.113*
	Near East[*2]	50	36 (0.72)	30	36	0.943 ± 0-023	0.009 ± 0.001	2.56	2.08	0.040	-2.301**	-4.097**
	Asia[*2]	48	29 (0.60)	26	29	0.947 ± 0.019	0.010 ± 0.001	2.89	2.37	0.029	-1.962*	-2.261
	Eastern Slavs[2]	50	30 (0.60)	31	30	0.944 ± 0.023	0.009 ± 0.001	2.35	1.67	0.057	-2.162*	-3.280*
	Arabian Peninsula[*3]	52	29 (0.56)	30	30	0.947 ± 0.017	0.008 ± 0.001	2.32	1.34	0.074	-2.153*	-3.050*
	Armenia[*3]	54	27 (0.50)	33	33	0.914 ± 0.031	0.009 ± 0.001	2.53	2.35	0.030	-2.158*	-1.685
	Daghestan[*3]	60	26 (0.43)	33	33	0.859 ± 0.042	0.008 ± 0.001	2.17	2.28	0.023	-2.268**	-2.323
	Georgia[*3]	30	15 (0.50)	16	16	0.874 ± 0.050	0.008 ± 0.001	2.11	2.12	0.031	-1.617	-0.682
	Jordan[*3]	33	18 (0.55)	25	25	0.847 ± 0.062	0.008 ± 0.001	2.24	2.30	0.024	-2.227**	-2.586*
	Karatchaians-Balkanians[*3]	50	21 (0.42)	23	23	0.943 ± 0.017	0.012 ± 0.001	3.23	2.00	0.059	-1.202	0.411
	Lebanon[*3]	34	20 (0.59)	23	23	0.907 ± 0.041	0.008 ± 0.001	2.09	1.88	0.061	-2.171*	-3.548**
	Northwest Caucasus[*3]	69	35 (0.51)	38	38	0.895 ± 0.034	0.009 ± 0.001	2.42	2.70	0.026	-2.256**	-2.953*
	Ossetia[*3]	45	22 (0.49)	26	27	0.883 ± 0.002	0.009 ± 0.001	2.58	2.84	0.029	-1.950*	-2.445
	Syria[*3]	28	19 (0.68)	23	23	0.966 ± 0.019	0.009 ± 0.001	2.38	1.38	0.098	-2.139*	-2.667*
	Turkey[*3]	90	46 (0.51)	44	46	0.898 ± 0.029	0.008 ± 0.001	2.24	2.10	0.037	-2.408**	-2.957*
	Austria[*4]	964	116 (0.12)	75	81	0.683 ± 0.017	0.005 ± 0.001	1.15	1.07	0.041	-2.468***	-5.322
	Germany[*4]	28	20 (0.71)	20	20	0.952 ± 0.030	0.010 ± 0.001	2.73	1.88	0.042	-1.657	-1.116
	Hungary[*4]	55	15 (0.27)	22	22	0.677 ± 0.070	0.006 ± 0.001	1.64	2.61	0.073	-2.059*	-2.160
	Macedonia[*4]	88	30 (0.34)	28	29	0.892 ± 0.025	0.007 ± 0.001	2.01	1.84	0.058	-2.000*	-1.707
	Romania[*4]	100	29 (0.29)	29	29	0.917 ± 0.017	0.009 ± 0.001	2.48	2.04	0.034	-1.690	-2.160

N = sample size, *k* = number of different haplotypes (divided by *N* in brackets); *S* = Number of polymorphic (segregating) sites; N_{mut} = total number of mutations; *H* = haplotype diversity and standard error; π = nucleotide diversity and standard error; *M* = average number of nucleotide differences; $V_O(M)$= observed variance of *M*; *r* = Harpending's (raggedness) index; *D* = Tajima's test of selective neutrality; *FL* = Fu and Li's D* statistics; Statistical significance: *, *P*-value < 0.05; **, *P*-value < 0.02. [*1] Present study; [*2] (Loogväli et al. 2004); [*3] (Roostalu et al. 2007); [*4] (Brandstätter et al. 2008)

Figure 1. Phylogeny of haplogroup the main branches of R0. The detailed phylogeny of haplogroup H is shown in Figure 2. In bold are those variants included in one of the three multiplexes designed.

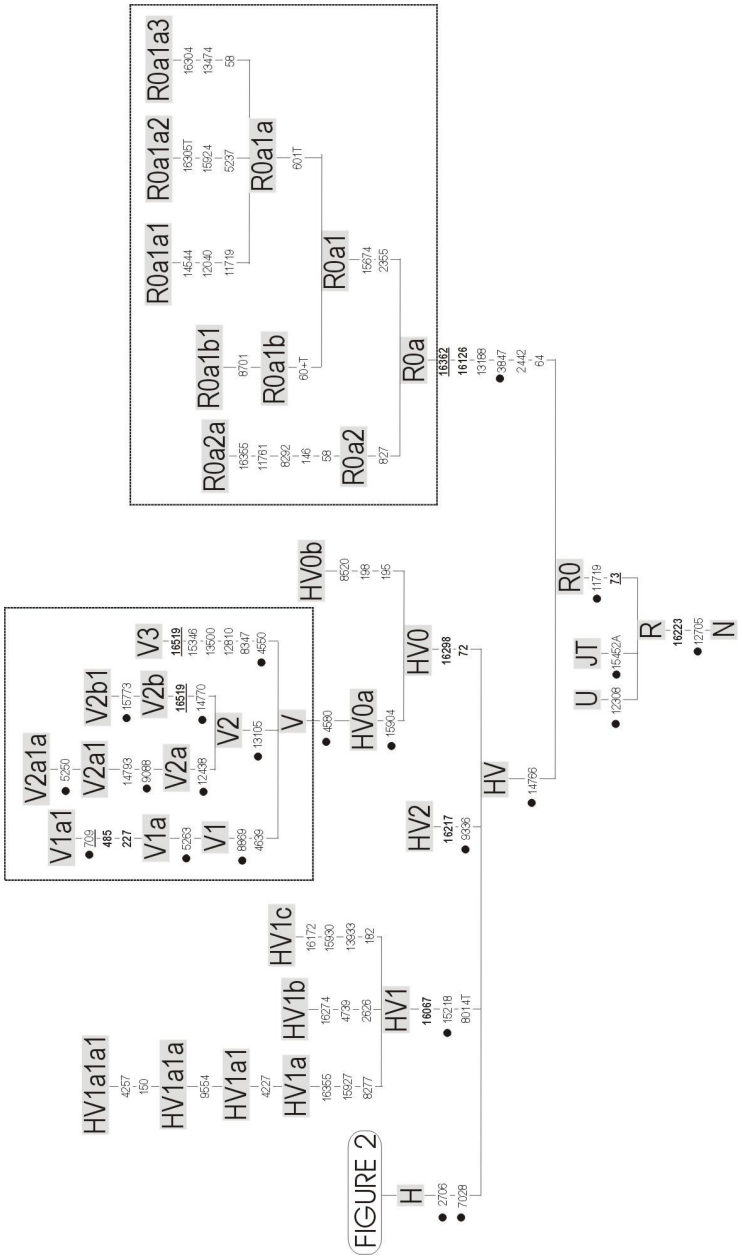


Figure 2. Phylogeny of haplogroup H.

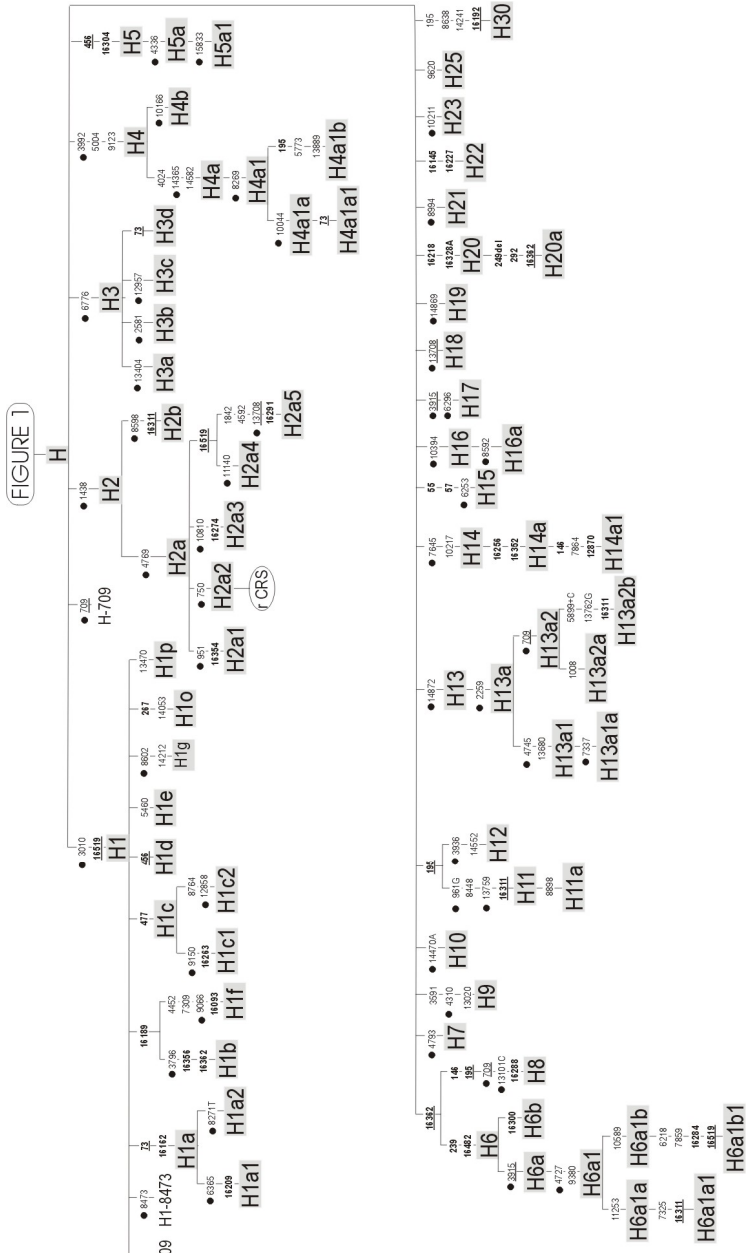


Figure 3. Maps of haplogroup frequencies for haplogroups H*, H1, H2a, H3, H4, H5a, H6a, H7, H8, H11. Dots in the map corresponding to H* indicate the location of the populations used.

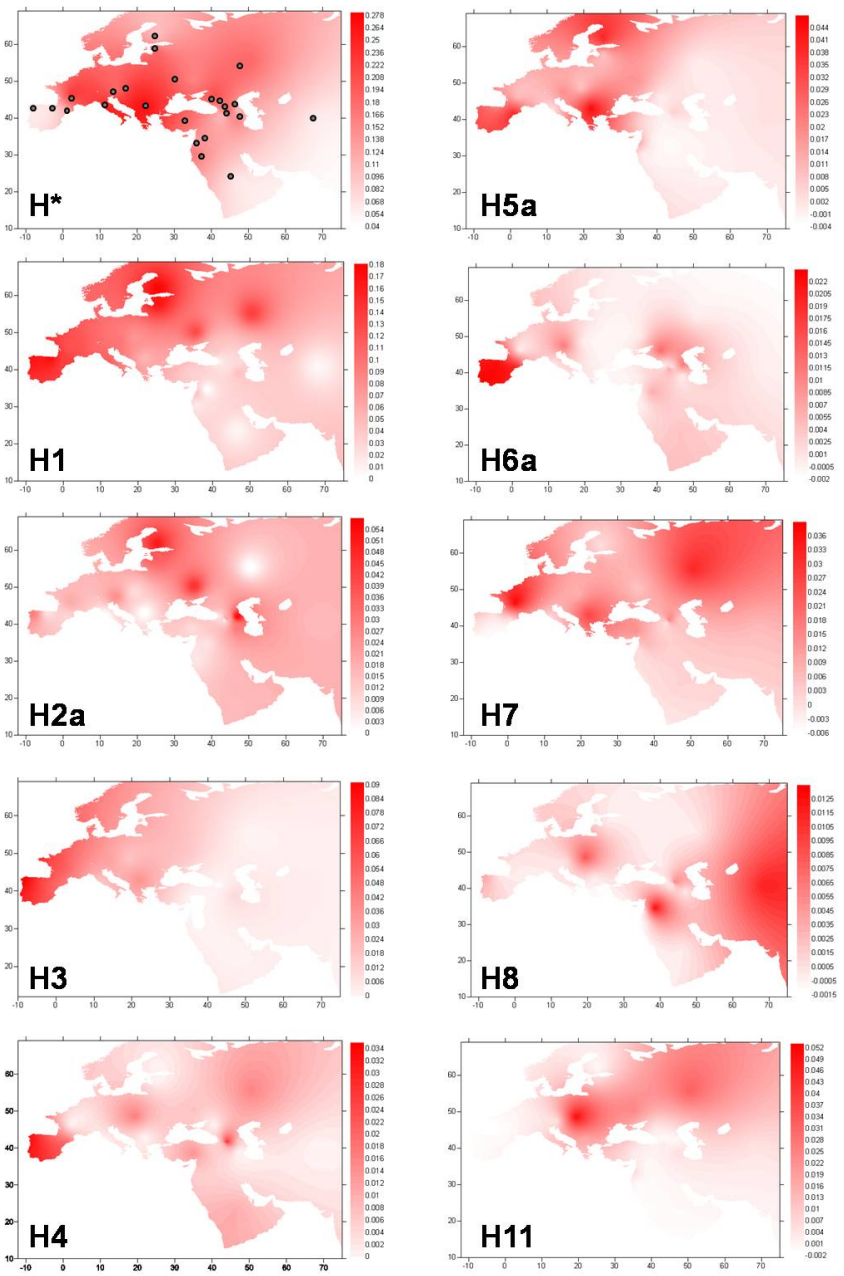
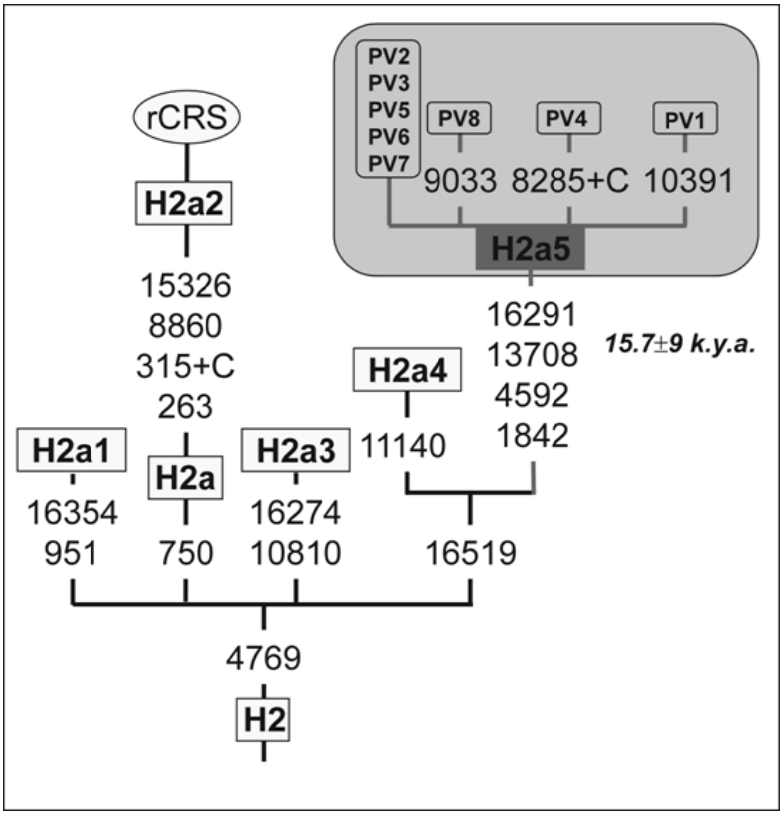


Figure 4. Phylogeny of haplogroup H2a5



Supplementary data S1: Amplification and minisequencing primers

Table 1. Amplification primers for multiplex 1. F.C. = final concentration

SNP	Primer forward and reverse	Size (bp)	F.C. (µM)
709/750	GGCTCACATCACCCCATAAA CGTTTTGAGCTGCATTGCT	161	0.2
2581/2706	GCCTGCCCAGTGACACAT GCTCCATAGGGTCTTCTCGT	195	0.2
3010	CAATAACTTGACCAACGGAACA CGGTCTGAACTCAGATCACGTA	179	0.4
3796/3847	TCAACATTACTAATAAGTGGCTCCTTT GGTTCGGTTGGTCTCTGCTA	135	0.2
4550/4580	CAACCCGTCATCTACTCTACCAT CTTCTGTGGAACGAGGGTTTATT	148	0.2
6253/6296/6365	TGACTCTTACCTCCCTCTCTCC GATGAAATTGATGGCCCCTA	189	0.2
6776	GCTTCCTAGGGTTTATCGTGTG GAGTGTGGCGAGTCAGCTAAA	140	0.4
7337	GGCTCATTCATTTCTCTAACAGC TCCAGGTTTATGGAGGGTTC	110	0.4
10810	GCTAAAACTAATCGTCCCAACA AATTAGGCTGTGGGTGGTTG	97	0.2
12858/12957	CAACACAGCAGCCATTCAAG GAGGCCTAGTAGTGGGGTGA	157	0.2
13708/13759	AACGAAAATAACCCCACCCTA GTTGTTTGGAAGGGGGATG	113	0.2
14365/14470	CCACCCCATCATACTCTTTCA TAGGGGGAATGATGGTTGTC	159	0.3
14766/14770	TCAACTACAAGAACACCAATGACC GGAGGTCGATGATGAGTGG	82	0.2
15218	ACTATCCGCCATCCCATACA GGGCAAGATGAAGTGAAAGG	110	0.4

Table 2. Amplification primers for multiplex 2. F.C. = final concentration

SNP	Primer Forward and Reverse	Size (bp)	F.C. (µM)
951/961G	TCACACGATTAACCCAAGTCA ACTCAGGTGAGTTTTAGCTTTATTG	87	0.2
3915/3936/3992	TAGCAGAGACCAACCGAACC GAAGATTGTAGTGGTGAGGGTGT	158	0.6
4310/4336	AGCATTCCCCCTCAAACCTA TTTTGGATTCTCAGGGATGG	127	0.4
4727/4745/4769/4793	TCCTTCTAATAGCTATCCTCTTCAACA TGGGTAACCTCTGGGACTCA	154	0.4
7028	CACCGTAGGTGGCCTGACTGGC GTGTAGCCTGAGAATAGGGG	168	0.4
7645	ACATGCAGCGCAAGTAGGTC AAAATGATTATGAGGGCGTGA	90	0.2
8269/8271	TAGGGCCCGTATTTACCCTAT AAGAGGTGTTGGTTCTCTTAATCTTT	110	0.2
8473/8592/8598/8602	CCCAACTAAAAATATTAAACACAAACT GGAGGTGGGGATCAATAGAG	193	0.2
9066/9088/9150	CCTACTCATGCACCTAATTGGA GGCTTACTAGAAGTGTGAAAACGTA	155	0.3
10044	CCGTTAACTTCCAATTAACTAGTTTTG AAGGCTAGGAGGGTGTTGATT	91	0.6
10394	CCATGAGCCCTACAAACAACT TGAGTCGAAATCATTCGTTTTG	159	0.3
13404	TATGTGCTCCGGGTCCATC TGGTGAGGGAGGTTGAAGTG	104	0.2

Table 3. Amplification primers for multiplex 31. F.C. = final concentration

SNP	Primer Forward and Reverse	Size(bp)	F.C. (µM)
1438	AACTTAAGGGTCGAAGGTGGA AGGGCCCTGTTCAACTAAGC	66	0.1
2259	TCAAGCTCAACACCCACTACC TGCGGAGGAGAATGTTTTCA	131	0.2
5250/5263	ATTCCATCCACCCTCCTCTC GGTGGGGATGATGAGGCTAT	111	0.6
8869	GGACTCCTGCCTCACTCATTT AAGTGGGCTAGGGCATTTTT	128	0.4
8994	AATGCCCTAGCCCACTTCTT AGGTGGCCTGCAGTAATGTT	140	0.1
9336	GCCATGTGATTTCACTTCCA GTGGCCTTGGTATGTGCTTT	117	0.2
10166/10211	ACTACCACAACTCAACGGCTACA AGGGGTAAAAGGAGGGCAAT	145	0.2
11140	CATTCACAGCCACAGAACTAATCAT GTTCTGGCTGGTTGCCTCAT	99	0.2
11719	CAGCCATTCTCATCCAAACC GCGTTCGTAGTTTGAGTTTGC	113	0.3
12308	CTGCTAACTCATGCCCCCATG ATTACTTTTATTTGGAGTTGCACCAAGATT	106	0.3
12438	CCACCCTAACCCTGACTTCC GTGGATGCGACAATGGATTT	106	0.1
12705	TGTAGCATTGTTCGTTACATGG AGTTGGAATAGGTTGTTAGCGG	147	0.2
13101/13105	CAGCCCTACTCCACTCAAGC TGGGCTATTTTCTGCTAGGG	83	0.2
14869/14872	CAACATCTCCGCATGATGAA AGGCGTCTGGTGAGTAGTGC	104	0.2
15452A	AGACGCCCTCGGCTTACTT GTCGCCTAGGAGGTCTGGTG	78	0.2
15773	CCGCAGACCTCCTCATTCTA CGGATGCTACTTGTCCAATG	81	0.1
15833/15904	CCCTTTTACCATCATTGGACA AAAGGTTTTCATCTCCGGTTT	162	0.2

Table 4. Minisequencing primers for multiplex 1. F.C. = final concentration

SNP	Minisequencing primer	Size (pb)	Base change	Chain	F.C. (μM)
709	$(gact)_3$TTACACATGCAAGCATCCCC	32	G-A	L	0.2
750	CTCTAAATCACCACGATCAAAAGG	24	A-G	L	0.2
2581	TGATTATGCTACCTTTGCACGGT	23	A-G	H	0.2
2706	$(gact)_2$gAGGGTCTTCTCGTCTTGCTGTGT	32	A-G	H	0.2
3010	$(gact)_3$gAACCTTTAATAGCGGCTGCACCAT	37	G-A	H	0.2
3796	$(gact)_2$gaCTAATAAGTGGCTCCTTTAACCTCTCC	37	A-G	L	0.1
3847	$(gact)_5$ATTACTCCTGCCATCATGACCC	42	T-C	L	0.2
4550	$(gact)_{13}$gaGCGCTAAGCTCGCACTGATT	74	T-C	L	0.3
4580	$(gact)_3$gaTTACCTGAGTAGGCCTAGAAATAAACAT	42	G-A	L	0.2
6253	$(gact)_6$gaTGTTCCTGCTCCGGCCTCCACT	48	T-C	H	0.1
6296	$(gact)_6$AACAGGTTGAACAGTCTACCCTCC	48	C-T	L	0.2
6365	$(gact)_7$gaGATGGCCCCTAAGATAGAGGAGAC	54	T-C	H	0.2
6776	$(gact)_6$acCGTGTGTCTACGTCTATTCCTACTGTAAATAT	58	T-C	H	0.3
7337	$(gact)_{10}$TGATTTGAGAAGCCTTCGCTTC	62	G-A	L	0.3
10810	$(gact)_5$gacCAACAATTATATTACTACCATTGACATGACT	54	T-C	L	0.2
12858	$(gact)_8$gacGCAGCCATTCAAGCAATCCTATA	58	C-T	L	0.3
12957	$(gact)_9$CAACAAATAGCCCTTCTAAACGCTAA	62	T-C	L	0.3
13708	$(gact)_{13}$ gaCTACTAAACCCCATTAAAGGCCTG	78	G-A	L	0.3
13759	$(gact)_{10}$TTCTCATTACTAACAACATTTCCCCC	66	G-A	L	0.1
14365	$(gact)_{11}$ gaGTTAGCGATGGAGGTAGGATTGGT	70	C-T	H	0.3
14470A	$(gact)_{13}$gCCTCAATAGCCATCGCTGTAGTATA	78	T-A	L	0.4
14770	$(gact)_{11}$GAATGAGTGGTTAATTAATTTTATTAGGGG	74	C-T	H	0.3
14766	$(gact)_{11}$ gCAATGACCCCAATACGCAAAA	66	T-C	L	0.3
15218	$(gact)_{11}$TCCTCAGATTCATTGAACTAGGTCTG	70	A-G	H	0.3

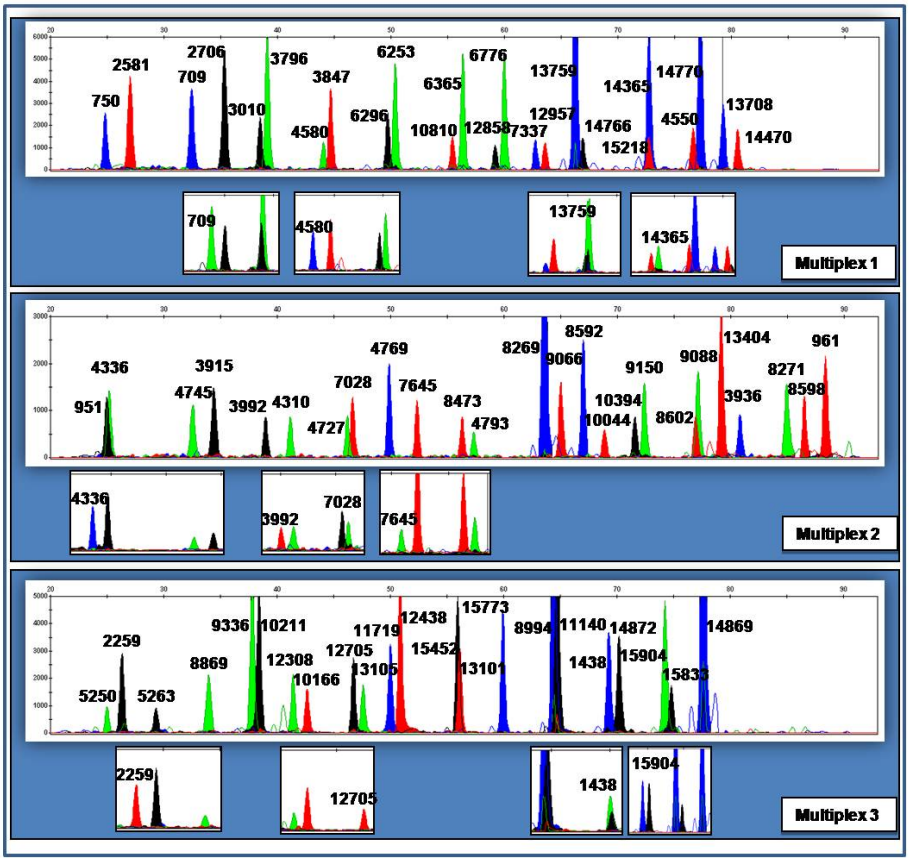
Table 5. Minisequencing primers for multiplex 2. F.C. = final concentration

SNP	Minisequencing primer	Size (pb)	Base change	Chain	F.C. (μM)
951	CTTTATTGGGGAGGGGGTGAT	21	G-A	H	0.2
961G	(gact)$_{15}$CGTAAAGAGTGTTTTAGATCACCCCC	86	T-G	L	0.06
3915	(gact)$_{2}$GAAGCCTGAGACTAGTTCGGACTC	32	G-A	H	0.2
3936	(gact)$_{14}$gTGCGGCGTATTCGATGTTGAA	78	C-T	H	0.6
3992	(gact)$_{3}$gaCCCTATTCTTCATAGCCGAATACA	38	C-T	L	0.3
4310	TCTGATAAAAGAGTTACTTTGATAGAGTAAATAATAGG	38	A-G	L	0.4
4336	AGGGATGGGTTCGATTCTCAT	21	T-C	H	0.2
4727	(gact)$_{5}$gaTACTCTCCGGACAATGAACCAT	44	A-G	L	0.2
4745	gactAATGAACCATAACCAATACTACCAATCA	32	A-G	L	0.2
4769	(gact)$_{4}$gaACCAATCAATACTCATCATTAATAATCATAAT	50	A-G	L	0.2
4793	(gact)$_{5}$gATAATCATAATAGCTATAGCAATAAAACTAGGAAT	56	A-G	L	0.4
7028	(gact)$_{4}$gacTACACGACACGTACTACGTTGTAGC	44	C-T	L	0.2
7645	(gact)$_{6}$GCTACTTCCCCTATCATAGGAGAGCT	50	T-C	L	0.2
8269	(gact)$_{8}$gacTGAAATAGGGCCCGTATTTACCCTATA	62	G-A	L	0.06
8271T	(gact)$_{15}$GGGCCCGTATTTACCCTATAGC	82	A-T	L	0.3
8473	(gact)$_{6}$gaAAAAATATTAAACACAAACTACCACCTACC	56	T-C	L	0.2
8592	(gact)$_{11}$gacGCCTACCCGCCGCAGTACT	66	G-A	L	0.1
8598	(gact)$_{15}$gacACCCGCCGCAGTACTGATCAT	84	T-C	L	0.4
8602	(gact)$_{13}$ CGCCGCAGTACTGATCATTCTA	74	T-C	L	0.4
9066	(gact)$_{8}$gaGTGTAGAGGGAAGGTTAATGGTTGATAT	62	A-G	H	0.3
9088	(gact)$_{10}$AGTAGAATTAGAATTGTGAAGATGATAAGTGTAG	74	T-C	H	0.2
9150	(gact)$_{12}$gCCTAGAAATCGCTGTCGCCTT	70	A-G	L	0.2
10044	(gact)$_{9}$TTAAGGCGAAGTTTATTACTCTTTTTTGAA	66	A-G	H	0.6
10394	(gact)$_{9}$gaCTGGCCTATGAGTGACTACAAAAAGGATTAGA	70	C-T	L	0.3
13404	(gact)$_{11}$gaATCATCCACAACCTTAACAATGAACAAGATAT	78	T-C	L	0.3

Table 6. Minisequencing primers for multiplex 3. F.C. = final concentration

SNP	Minisequencing primer	Size (pb)	Base change	Chain	F.C. (µM)
1438	$(gact)_{10}$gaGTCGAAGGTGGATTTAGCAGTAAACT	68	A-G	L	0.06
2259	TCCCAAACATATAACTGAACTCCTCA	26	C-T	L	0.2
5250	TTCTTCGATAATGGCCCATTTGGGCA	26	T-C	H	0.2
5263	$(gact)_2$gacGGCTTTTTGCCCAAATGGG	30	C-T	L	0.2
8869	$(gact)_2$gTTATGAGCGGGCACAGTGATT	30	A-G	L	0.6
8994	$(gact)_9$gacGCCTACTCATTCAACCAATAGCCCT	64	G-A	L	0.06
9336	$(gact)_3$gacTTCCACTCCATAACGCTCCTC	36	A-G	L	0.15
10166	$(gact)_4$gaGGCTACATAGAAAAATCCACCCC	41	T-C	L	0.3
10211	$(gact)_4$TCCCCCGCCCGCGTCCCTTT	36	C-T	L	0.15
11140	$(gact)_7$gacCATATTTTATATCTTCTTCGAAACCACACTTAT	64	C-T	L	0.2
11719	$(gact)_7$gaATTCTCATAATCGCCCACGG	50	G-A	L	0.4
12308	$(C)_{24}$ATTGGTCTTAGGCCCCA	41	A-G	L	0.4
12438	$(gact)_6$gCTAACAAAAAAAACTCATACCCCCA	50	T-C	L	0.15
12705	$(gact)_3$gaAACATTAATCAGTTCTTCAAATATCTACTCAT	46	C-T	L	0.2
13101C	$(gact)_7$gaGAAGCGGATGAGTAAGAAGATTCC	54	A-C	H	0.2
13105	$(gact)_4$gaCACTCAAGCACTATAGTTGTAGCAGGA	45	A-G	L	0.2
14869	$(gact)_{13}$gacGCTCACTCCTTGGCGCCTGCCT	77	G-A	L	0.2
14872	$(gact)_{12}$CTCCTTGGCGCCTGCCTGAT	68	C-T	L	0.4
15452A	$(gact)_8$gGCCCTCGGCTTACTTCTCTTC	54	C-A	L	0.4
15773	$(gact)_9$gacCCTGAATCGGAGGACAACCA	59	G-A	L	0.1
15833	$(gact)_9$gacGACAAGTAGCATCCGTACTATACTTCACAACAATC	74	C-T	L	0.2
15904	$(gact)_{10}$TTTCATCTCCGGTTTACAAGACTGGTGTATTA	72	C-T	H	0.1

Supplementary data S2: SNaPshot electropherogram. The three large pictures show an haplogroup V profile. The small pictures indicate some peak (SNP) variants that would allow the identification of other R0 branches.



Supplementary Table S3: HVS-I and coding region SNP variation for the Iberian samples analyzed in the present study.

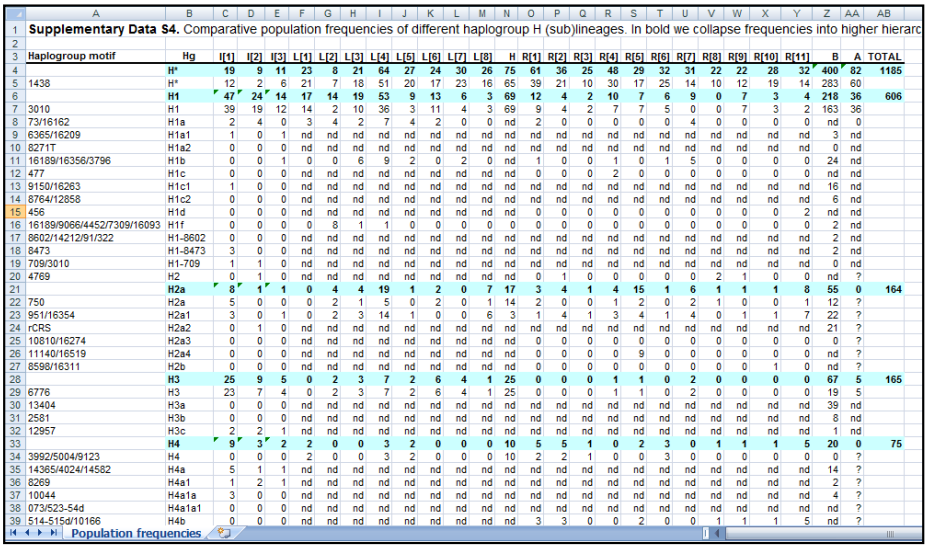
Supplementary Table S2. HVS-I and coding region SNP profiles for the Galician, Cantabrian and Catalonian population.

SNP genotyping as described in the present study

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY
3	R0/nonR0	Sample ID	Iberian region	Sub-region	HVS-I	Sequence range	Hg	709	750	951	961	1438	2259	2581	2706	3010	3796	3847	3915	3936	3992	4310	4336	4550	4580	4727	4745	4769	4793	5250	5263	6253	6296	6365	6776	7028	7337	7645	8269	8271	8473	8592	8598	8602	8869	8994	9066	9088	9150	9336	10044
4	-	rCRS	-	-	-		-	G	A	G	G	A	C	A	A	G	A	T	G	C	C	A	T	T	G	A	A	A	A	T	C	T	C	T	T	C	G	T	G	T	T	G	T	T	A	G	A	T	A	A	A
5	-	Multiples	-	-	-		-	1	1	2	2	3	3	1	1	1	1	1	2	2	2	2	2	1	1	2	2	2	2	3	3	1	1	1	1	2	1	2	2	2	2	2	2	2	3	3	2	2	2	3	2
6	R0	C155	Galicia	Lugo	291	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	R0	C170	Galicia	Mariña Lucense	093 129 179 316 319	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	R0	C203	Galicia	Mariña Lucense	093 129 316	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	R0	C235	Galicia	Noroeste	209	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	R0	C270	Galicia	Rías Baixas	rCRS	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	R0	C324	Galicia	Noroeste	311	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	R0	C325	Galicia	Noroeste	311	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	R0	C348	Galicia	Santiago	rCRS	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	R0	C401	Galicia	-	293 519	16024-16569	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	R0	C45	Galicia	Ourense	399	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	R0	C8	Galicia	Lugo	rCRS	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	R0	C86	Galicia	Ourense	218	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	R0	A371	Cantabria	Santander	rCRS	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19	R0	A360	Cantabria	Santander	93	16024-16400	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	R0	AU141	Catalonia	-	rCRS	16024-16569	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	R0	AU156	Catalonia	-	176 519	16024-16569	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	R0	AU58	Catalonia	-	335	16024-16569	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
23	R0	AU61	Catalonia	-	311	16024-16569	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
24	R0	AU64	Catalonia	-	184 248	16024-16569	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	R0	AU87	Catalonia	-	129 316 519	16024-16569	H*	-	G	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	R0	C100	Galicia	Ourense	093 261 269 270 294	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27	R0	C123	Galicia	Lugo	rCRS	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
28	R0	C15	Galicia	Lugo	129	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
29	R0	C154	Galicia	Lugo	140	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
30	R0	C163	Galicia	Lugo	189	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
31	R0	C176	Galicia	Mariña Lucense	111 311	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
32	R0	C178	Galicia	Mariña Lucense	234	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
33	R0	C181	Galicia	Lugo	rCRS	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
34	R0	C199	Galicia	Lugo	176 183C 189	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
35	R0	C212	Galicia	Lugo	075 183C 189 356	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
36	R0	C22	Galicia	Lugo	rCRS	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
37	R0	C231	Galicia	Santiago	305T 320A	16024-16400	H1	-	G	-	-	G	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Extracto de la tabla S3, que debido a su gran tamaño sólo puede estar disponible en su versión electrónica.

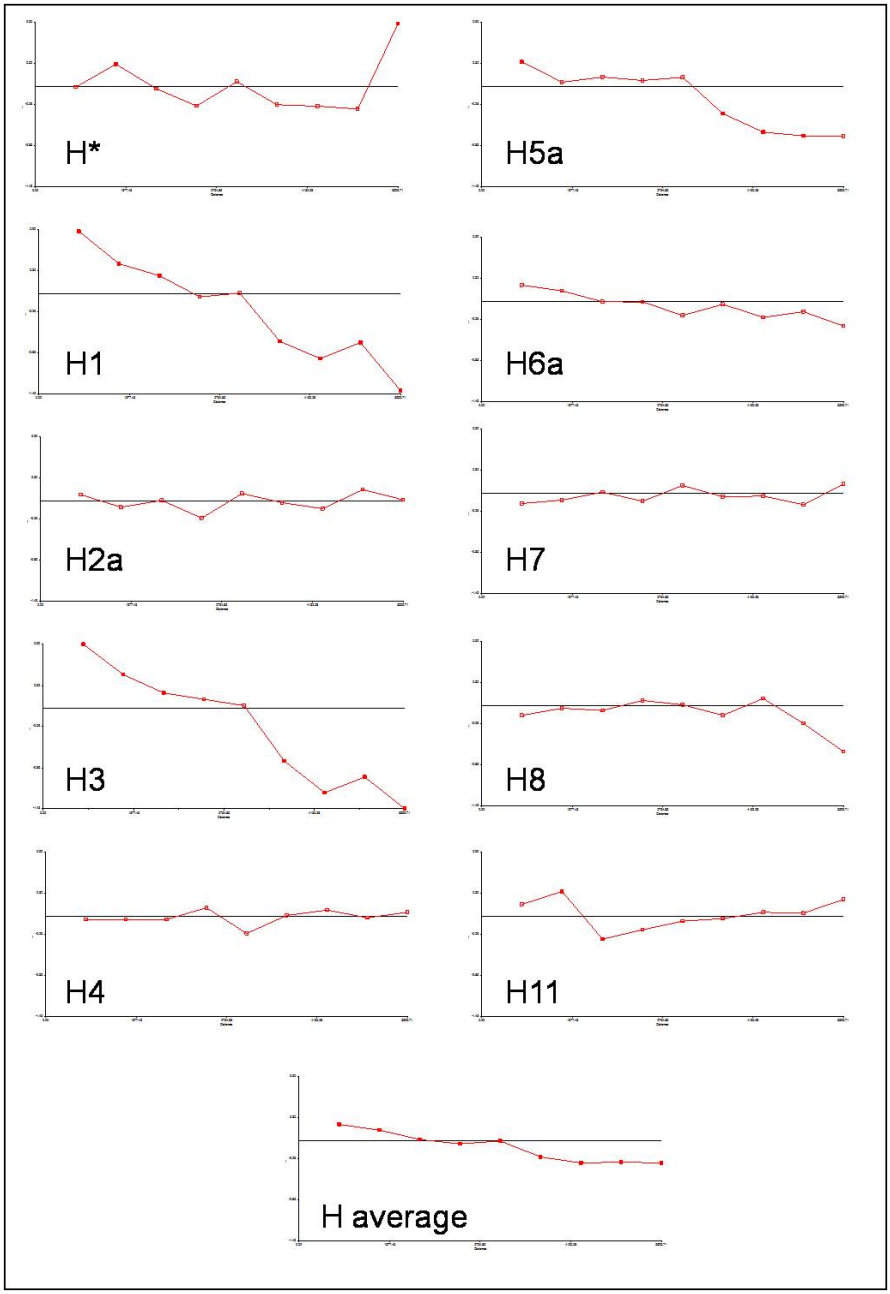
Supplementary Data S4. Comparative population frequencies of different haplogroup H (sub)lineages. In bold we collapse frequencies into higher hierarchical phylogenetic clades as a function of the SNPs genotyped in the referred studies, such that only these 'bolded' categories are fully comparable between the different studies considered. This is because haplogroup categories are not fully comparable among populations when the samples have undetermined (nd) SNP; for instance, H* embraces different lineages in our study because we genotyped to a higher resolution than in previous attempts (where different lineages were already collapsed into H*). For nomenclature we follow the scheme of Figure2.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	**Supplementary Data S4.** Comparative population frequencies of different haplogroup H (sub)lineages. In bold we collapse frequencies into higher hierarc																											
2																												
3	**Haplogroup motif**	**Hg**	**I[1]**	**I[2]**	**I[3]**	**L[1]**	**L[2]**	**L[3]**	**L[4]**	**L[5]**	**L[6]**	**L[7]**	**L[8]**	**H**	**R[1]**	**R[2]**	**R[3]**	**R[4]**	**R[5]**	**R[6]**	**R[7]**	**R[8]**	**R[9]**	**R[10]**	**R[11]**	**B**	**A**	**TOTAL**
4		**H***	**19**	**9**	**11**	**23**	**8**	**21**	**64**	**27**	**24**	**30**	**26**	**75**	**61**	**36**	**25**	**48**	**29**	**32**	**31**	**22**	**22**	**28**	**32**	**400**	**82**	**1185**
5	1438	H*	12	2	6	21	7	18	51	20	17	23	16	65	39	21	10	30	17	25	14	10	12	19	14	283	60	
6		**H1**	**47**	**24**	**14**	**17**	**14**	**19**	**53**	**9**	**13**	**6**	**3**	**69**	**12**	**4**	**2**	**10**	**7**	**6**	**9**	**0**	**7**	**3**	**4**	**218**	**36**	**606**
7	3010	H1	39	19	12	14	2	10	36	3	11	4	3	69	9	4	2	7	7	5	0	0	7	3	2	163	36	
8	73/16162	H1a	2	4	0	3	4	2	7	4	2	0	0	nd	2	0	0	0	0	0	4	0	0	0	0	nd	0	
9	6365/16209	H1a1	1	0	1	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	3	nd	
10	8271T	H1a2	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	0	nd	
11	16189/16356/3796	H1b	0	0	1	0	0	6	9	2	0	2	0	nd	1	0	0	1	0	1	5	0	0	0	0	24	nd	
12	477	H1c	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	0	0	0	2	0	0	0	0	0	0	0	nd	nd	
13	9150/16263	H1c1	1	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	16	nd	
14	8764/12858	H1c2	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	6	nd	
15	456	H1d	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	0	0	0	0	0	0	0	0	0	0	2	nd	nd	
16	16189/9066/4452/7309/16093	H1f	0	0	0	0	8	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	nd	
17	8602/14212/91/322	H1-8602	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	2	nd	
18	8473	H1-8473	3	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	2	nd	
19	709/3010	H1-709	1	1	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	0	nd	
20	4769	H2	0	1	0	nd	nd	nd	nd	nd	nd	nd	nd	0	1	0	0	0	0	0	2	1	0	0	nd	?	
21		**H2a**	**8**	**1**	**1**	**0**	**4**	**4**	**19**	**1**	**2**	**0**	**7**	**17**	**3**	**4**	**1**	**4**	**15**	**1**	**6**	**1**	**1**	**1**	**8**	**55**	**0**	**164**
22	750	H2a	5	0	0	0	2	1	5	0	2	0	1	14	2	0	0	1	2	0	2	1	0	0	1	12	?	
23	951/16354	H2a1	3	0	1	0	2	3	14	1	0	0	6	3	1	4	1	3	4	1	4	0	1	1	1	7	22	?
24	rCRS	H2a2	0	1	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	21	?
25	10810/16274	H2a3	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	0	0	0	0	0	0	0	0	0	0	0	0	0	?
26	11140/16519	H2a4	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	0	0	0	0	9	0	0	0	0	0	0	nd	?	
27	8598/16311	H2b	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	0	0	0	0	0	0	0	0	0	1	0	nd	?	
28		**H3**	**25**	**9**	**5**	**0**	**2**	**3**	**7**	**2**	**6**	**4**	**1**	**25**	**0**	**0**	**0**	**1**	**1**	**0**	**2**	**0**	**0**	**0**	**0**	**67**	**5**	**165**
29	6776	H3	23	7	4	0	2	3	7	2	6	4	1	25	0	0	0	1	1	0	2	0	0	0	0	19	5	
30	13404	H3a	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	39	nd	
31	2581	H3b	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	8	nd	
32	12957	H3c	2	2	1	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	1	nd	
33		**H4**	**9**	**3**	**2**	**2**	**0**	**0**	**3**	**2**	**0**	**0**	**0**	**10**	**5**	**5**	**1**	**0**	**2**	**3**	**0**	**1**	**1**	**1**	**5**	**20**	**0**	**75**
34	3992/5004/9123	H4	0	0	0	2	0	0	3	2	0	0	0	10	2	2	1	0	0	3	0	0	0	0	0	0	?	
35	14365/4024/14582	H4a	5	1	1	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	14	?	
36	8269	H4a1	1	2	1	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	2	?	
37	10044	H4a1a	3	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	4	?	
38	073/523-54d	H4a1a1	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	?	
39	514-515d/10166	H4b	0	0	0	nd	nd	nd	nd	nd	nd	nd	nd	3	3	0	0	2	0	0	1	1	1	5	nd	?	

Extracto de la tabla S4, que debido a su gran tamaño sólo puede estar disponible en su versión electrónica.

Supplementary Data S5. Autocorrelograms for haplogroups H*, H1, H2a, H3, H4, H5a, H6a, H7, H8, H11 and H average.



Artículo 11: Minisequencing of coding region mitochondrial DNA SNPs: forensic validation

Mosquera-Miguel A, Álvarez-Iglesias V, Cerezo M, Fondevila M, Lareu MV, Carracedo Á, Salas A

(Manuscrito en preparación)

Minisecuenciación de SNPs de la región codificante de ADN mitocondrial: validación forense

ABSTRACT/RESUMEN:

La necesidad de protocolos eficientes de genotipado para tipar SNPs (*Single Nucleotide Polymorphisms*) de la región codificante del ADN mitocondrial (ADNmt) es un campo que ha despertado un interés creciente en la genética forense. El uso de la minisecuenciación se ha vuelto un método muy popular en la literatura forense, pero su empleo está, todavía, restringido a unos pocos laboratorios. Esto es debido, en parte, a la escasez de estudios que muestren la eficiencia real y la reproducibilidad de esta técnica cuando se aplica en la resolución de muestras forenses complejas. En este trabajo probamos un ensayo de minisecuenciación que incluye 71 SNPs de la región codificante de ADNmt que definen las ramas y las sub-ramas conocidas de la filogenia de R0. Las muestras utilizadas para validar el multiplex incluyen huesos y dientes degradados, cabellos sin bulbo y diluciones seriadas de ADN. El tamaño de los amplicones de PCR de este multiplex que va desde 66 a 195 pb (media=135; SD = 35) unido a la eficiencia de la técnica de minisecuenciación, permiten que este multiplex funcione bien en las muestras degradadas y/o con baja cantidad de ADN probadas. No observamos inconsistencias filogenéticas en los haplotipos genotipados que puedan indicar posibles artefactos de genotipado a causa de contaminación o a amplificaciones espúreas como por ejemplo debidas a NUMTs.

Minisequencing of coding region mitochondrial DNA SNPs: forensic validation

Mosquera-Miguel A[1&], Álvarez-Iglesias V[1&], Cerezo M[1], Fondevila M[1], Lareu MV[1], Carracedo Á[1,2], Salas A[1,2*]

[1]*Unidade de Xenética, Instituto de Medicina Legal, Facultad de Medicina, 15782, Universidad de Santiago de Compostela,*

[2]*Grupo de Medicina Xenómica, Hospital Clínico Universitario, 15706 Santiago de Compostela, Galicia, Spain*

[&] Both authors contributed equally to this study

* Corresponding author: Dr. Antonio Salas, Unidad de Genética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, 15782, Galicia, Spain. Tel: +34–981–582327; Fax: +34–981–580336; E-mail: apimlase@usc.es

Keywords: mtDNA; coding region; HVS-I and HVS-II; SNP; haplotype; SNaPshot; phylogeny; forensics; population studies; haplogroup H

Abstract

The need of efficient genotyping protocols for mitochondrial DNA (mtDNA) testing coding region SNPs (Single Nucleotide Polymorphisms) is a field of growing interest in forensic genetics. The use of minisequencing is becoming a popular method in the forensic literature, but its use is still restricted to few laboratories. In part, this is due to the lack of studies showing the real efficiency and the reproducibility of this technique when applied to real and complex forensic samples. Here we have tested a minisequencing assay that includes 71 coding region SNPs defining known branches and sub-branches of the R0 phylogeny. The samples used for validating the multiplex assay included degraded bones and teeth, as well as hair shafts serial dilutions. The fact that the PCR amplicons of this multiplex assay ranges in size from 66 to 195 bp (mean = 135; SD = 35) coupled with the efficiency of the minisequencing procedure allow this multiplex to perform well with all the samples tested either degraded or/and those containing low DNA amounts. We did not observe phylogenetic inconsistencies in the 71-SNP mtDNA haplotypes genotyped that could indicate potential genotyping artifacts due for instance to contamination or spurious amplification of i.e. NUMTs.

Introduction

Genotyping of coding region mtDNA SNPs is a topic of growing interest among forensic geneticists. The common practice of sequencing the first (and some times the second) hypervariable segment (HVS-I) is most of the times insufficient due to its limited power of discrimination in forensic casework. However, for highly degraded samples (and/or those containing low amounts of DNA) the analysis of mtDNA is often the only choice basically because the high copy number of mtDNA molecules per cell in comparison to the autosomal DNA contents.

Sequencing strategies based on the amplification of small amplicons are needed when dealing with highly degraded samples, but these sequencing strategies are time demanding, costly, and require intense lab work. A variety of screening and genotyping strategies have been proposed in the past for analyzing the variation at the coding region of the mtDNA molecule (see for instance (Barros et al. 1997; Salas et al. 2001)). During these last few years, minisequencing genotyping has gaining adepts among forensic geneticists because this technique has several advantages: a) it can be a good complement to the common forensic practice of sequencing the mtDNA control region because a solid design can substantially increase the discrimination power of the mtDNA test or can simply be useful for pure screening purposes, b) it allows to genotype a selection of SNPs in multiplex assays such that the interrogated SNPs can be (and are generally) located at distant parts of the mtDNA molecule, c) it is rapid and cost-effective, and d) using multiplex genotyping prevents artifactual recombinations (Bandelt et al. 2004a; Bandelt et al. 2004b; Bandelt et al. 2005; Salas et al. 2005b) which can easily occur when genotyping one SNP at a time.

Several multiplex assays have been published during these last few years for genotyping coding region SNPs. For instance, Quintáns et al. (2004) described a multiplex protocol that target SNPs defining main

West European branches. A similar multiplex assay was also independently developed by (Vallone et al. 2004). Recently, Brandstätter et al. (2006) have published an SNaPshot design allowing to genotype a total of 45 coding region SNPs for resolving sub-lineages of haplogroup H. This SNP selection is of special interest for forensic geneticist interested in genotyping samples of West European ancestry, because haplogroup H can constitute more than 40% of all the haplotypes in a typical European population. In addition, this haplogroup is very poor in terms of variation at the control region segment and therefore, coding region information helps to substantially increase the discrimination power of the mtDNA test. Other multiplexes have been designed to provide with a reasonable resolution of the East Asian and Native American phylogeny (Álvarez-Iglesias et al. 2007); another one developed by (Umetsu et al. 2005) focused also in the East Asian phylogeny. Recently, have show the ability of minisequencing for genotyping historic samples from Andaman Islanders (Endicott et al. 2006).

When proposing a new genotyping multiplex protocol, one of the most demanded requirements among forensic analysts is to amplify PCR amplicons of small size; this obviously will better guarantee a good performance when dealing with complex forensic samples. These requirements are only reached by some previous attempts, but covering different parts of the phylogeny (Álvarez-Iglesias et al. 2007) or with a more broad phylogenetic coverage (e.g. West European branches (Quintáns et al. 2004)). We here give details on our validation study on typical casework forensic samples of a 71-SNP multiplex minisequencing assay for targeting the most prevalent branch of the West European phylogeny.

Material and Methods

Forensic samples

We have analyzed three different hair shafts of 1 cm long collected from three unrelated donors. DNA extraction was carried out using the Bio Robot EZ1 robot (Qiagen; Hilden, Germany) using the manufacture protocols. Two bone and two teeth samples were extracted using a phenol-chlorophorm protocol from the Instituto de Medicina Legal of Santiago de Compostela.

Serial DNA dilution samples

Four different DNA samples were selected for serial dilution experiments. The samples were initially quantified using the Applied Biosystems' Quantifiler human DNA quantification kit (Applied Biosystems, Foster City, US), and also the IPC (internal PCR control) PCR inhibitor detecting feature.

The reactions were run on an Applied Biosystems 7300 real time PCR device following the manufacturer's specifications. This quantification technology consists on a short amplicon PCR reaction associated with the liberation of a quenched fluorescent molecule. The accumulation curve of fluorescence signal mimics that of the amplification product. The output data that the system uses to calculate the DNA concentration on the sample is the PCR cycle on which the fluorescence signal exceeds a detection threshold. Depending on the initial concentration of the sample this level will be achieved in an earlier or a later cycle. There is a parallel reaction on each tube, called IPC (internal PCR control), consistent on a self amplifying primer. The cycle on which the IPC surpasses the threshold is, so, independent of the initial DNA concentration of the sample, the only situation on which the IPC could yield a erroneous result (a high threshold surpassing cycle) would be in case the reaction is somehow inhibited; for instance, the presence of inhibitors that frequently shows-up in

highly degraded forensic samples. In order to avoid artifact related to the contamination/inhibition of the quantification procedure, we have used good sample DNA as template for the dilution experiments.

Finally, all the samples were diluted as follows: 1000, 500, 200, 100, 50, 25, and 10 pg/µl, and genotyped for the whole set of SNPs considered in the present report.

SNP selection and SNaPshot design

Details concerning SNP selection and SNaPshot design are given in Álvarez-Iglesias. et al. (author's unpublished data). The design consists of three different multiplex reactions that allow the genotyping of 71 SNPs that are diagnostic of different haplogroup R0 branches (Table 1).

Positive and negative controls were used during the whole extraction, amplification and genotyping process. Negative controls are mock extractions run with the same reagents as the sample extractions, but with not sample added.

Automatic sequencing

All the samples were also sequenced in the forward and the reverse directions for the HVS-I segment (the serial dilutions were only sequenced for the 500 pg/µl dilution). Note however that the forensic samples were highly degraded. An indirect indication of the level of the DNA degradation was that the standard protocols based on PCR of large amplicons such as those described in (Álvarez-Iglesias et al. 2007) did not perform well. Most of the samples were therefore amplified and sequenced using smaller amplicons in at least two independent and overlapping reactions (author's unpublished data).

Control for genotyping errors

We have used the mtDNA tree as a reference to check phylogenetic inconsistencies and avoid as much as possible artifactual

results and documentation errors (Bandelt et al. 2001; Bandelt et al. 2002; Yao et al. 2003; Bandelt et al. 2005; Salas et al. 2005a; Salas et al. 2005b; Yao et al. 2006).

Results and Discussion

A total of 71 coding region SNPs covering the main branch of the West European phylogeny were genotyped in a set of typical casework forensic samples. The average amplicon size of these multiplexes is very low (135 bp), which allow increasing its efficiency and applicability in real complex forensic casework samples. Thus, the success rate of the multiplex assay was ~100% for a control well-preserved DNA sample (good quality DNA and DNA concentration > 1-2 ng). Figure 1 shows two minisequencing profiles (Figure 1A = haplogroup R*; Figure 1B = haplogroup H1) that correspond with samples #Femur_1 and #Tooth_1 (Table S1), respectively.

The minisequencing assay performs very well for the forensic samples analyzed, especially if we take into account that all of them could not be properly sequenced using standard protocols or yielded positive results after carrying out various repetitive and tedious sequencing analyses (see M&M). Moreover, only small sequencing amplicons succeed for the several sequence reactions. Such a procedure is not feasible for routine casework in laboratories where hundred of samples have to be processed.

Reagent blank extraction controls associated with the casework extracts were also investigated. We did not observe false positives due to contamination.

Series dilution experiments showed that the average genotyping success (measured by the number of SNPs genotyped unambiguously) was high for those samples containing 1000 to 25 pg (average genotyping success = ~95%). Dilutions containing just 10 picograms experienced a clear lost of electrophoretic signal so that some SNPs

could not be genotyped unambiguously. However, the lowest percentage of success rate for the whole set of samples used was ~45%.

We observed a moderate correlation (Pearson's correlation coefficient; $r^2 = 0.47$) between the multiplex design and the genotyping success. This correlation however does not seems to depend on the average amplicon size or the number of SNPs contained by each multiplex since these values are similar in the three designs.

SNP typing allowed the allocation of most of the samples to specific haplogroup backgrounds to a high level of phylogenetic resolution (see Table S1).

Final remarks

All casework samples presented a highly degraded status as inferred by the intense sequence effort that was carried out in order to just sequence the HVS-I segment. We did not observed mtDNA mixtures ('haplotype/haplogroup' mixtures) that could indicate for instance sample mix-up (Montesino et al. 2007; Prieto et al. 2007).

The 71 SNP assays (Álvarez-Iglesias et al.; in preparation) genotyped in the present study show that the minisequencing procedure performs well with low DNA and degraded samples, apart from improving the discrimination power of the mtDNA test (in this particular case, the SNPs targeted are specifically designed for analyzing samples of a likely West European origin. As mentioned in previous studies (Álvarez-Iglesias et al. 2007), this kind of assays is cheaper than many other techniques and can be executed in few hours for a large amount of SNPs and samples.

To sum-up, we have genotyped a multiplex assay capable of typing 71 SNPs spanning the mtDNA molecule in highly degraded samples and in serial dilutions. SNaPshot minisequence performs very well with typical forensic DNA evidences containing low amount or degraded DNA. In serial samples, we observed that 10 pg can be enough

to obtain ~45% of the SNPs; this information could be relevant to discriminate between different DNA biological sources. As recommended in previous studies (Álvarez-Iglesias et al. 2007), it is mandatory to interpret the SNP-haplotype profiles under the view of the mtDNA phylogeny, in order to detect potential artifacts. When observing unexpected SNP patterns, it is worth to carry out the affected SNPs in singleplex PCR products and, in case the same pattern still persists, it would be recommended to carry out sequencing analysis in order to rule out potential artifacts (Álvarez-Iglesias et al. 2007), such as involuntary amplification of NUMTs.

Acknowledgements

The 'Ramón y Cajal' Spanish programme from the Ministerio de Educación y Ciencia (RYC2005-3), the grant from the Xunta de Galicia (PGIDIT06PXIB208079PR), and a grant from the Fundación de Investigación Médica Mutua Madrileña awarded to AS supported this project.

References

Álvarez-Iglesias V, Jaime JC, Carracedo Á, Salas A (2007) **Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups.** Forensic Sci Int: Genet 1:44-55

Bandelt H-J, Kong Q-P, Parson W, Salas A (2005) **More evidence for non-maternal inheritance of mitochondrial DNA?** J Med Genet 42:957-960

Bandelt H-J, Lahermo P, Richards M, Macaulay V (2001) **Detecting errors in mtDNA data by phylogenetic analysis.** Int J Legal Med 115:64-69

Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V (2002) **The fingerprint of phantom mutations in mitochondrial DNA data.** Am J Hum Genet 71:1150-1160

Bandelt H-J, Salas A, Bravi CM (2004a) **Problems in FBI mtDNA database.** Science 305:1402-1404

Bandelt H-J, Salas A, Lutz-Bonengel S (2004b) **Artificial recombination in forensic mtDNA population databases.** Int J Legal Med 118:267-273

Barros F, Lareu MV, Salas A, Carracedo A (1997) **Rapid and enhanced detection of mitochondrial DNA variation using single-strand conformation analysis of superposed restriction enzyme fragments from polymerase chain reaction-amplified products.** Electrophoresis 18:52-54

Brandstätter A, Salas A, Niederstätter H, Gassner C, Carracedo Á, Parson W (2007) **Dissection of mitochondrial superhaplogroup H using coding region SNPs**. Electrophoresis 13:2541-2550

Endicott P, Metspalu M, Stringer C, Macaulay V, Cooper A, Sánchez JJ (2006) **Multiplexed SNP typing of ancient DNA clarifies the origin of andaman mtDNA haplogroups amongst south Asian tribal populations.** PLoS ONE 1:e81

Montesino M, Salas A, Crespillo M, Albarran C, Alonso A, Alvarez-Iglesias V, Cano JA, et al. (2007) **Analysis of body fluid mixtures by mtDNA sequencing: An inter-laboratory study of the GEP-ISFG working group.** Forensic Sci Int 168:42-56

Prieto L, Alonso A, Alves C, Crespillo M, Montesino M, Picornell A, Brehm A, et al. (2007) **2006 GEP-ISFG Q1 collaborative exercise on mtDNA. Reflections about interpretation, artefacts, and DNA mixtures.** Forensic Sci Int Genet *in press*

Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo Á (2004) **Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing.** Forensic Sci Int 140:251-257

Salas A, Carracedo Á, Macaulay V, Richards M, Bandelt H-J (2005a) **A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics.** Biochem Biophys Res Commun 335:891-899

Salas A, Rasmussen EM, Lareu MV, Morling N, Carracedo A (2001) **Fluorescent SSCP of overlapping fragments (FSSCP-OF): a highly**

sensitive method for the screening of mitochondrial DNA variation. Forensic Sci Int 124:97-103

Salas A, Yao Y-G, Macaulay V, Vega A, Carracedo Á, Bandelt H-J (2005b) **A critical reassessment of the role of mitochondria in tumorigenesis.** PLoS Med 2:e296

Umetsu K, Tanaka M, Yuasa I, Adachi N, Miyoshi A, Kashimura S, Park KS, Wei YH, Watanabe G, Osawa M (2005) **Multiplex amplified product-length polymorphism analysis of 36 mitochondrial single-nucleotide polymorphisms for haplogrouping of East Asian populations.** Electrophoresis 26:91-98

Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ (2004) **A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome.** Int J Legal Med 118:147-157

Yao Y-G, Macaulay V, Kivisild T, Zhang Y-P, Bandelt H-J (2003) **To trust or not to trust an idiosyncratic mitochondrial data set.** Am J Hum Genet 72:1341-1346; author reply 1346-1349

Yao YG, Salas A, Bravi CM, Bandelt HJ (2006) **A reappraisal of complete mtDNA variation in East Asian families with hearing impairment.** Hum Genet 119:505-515

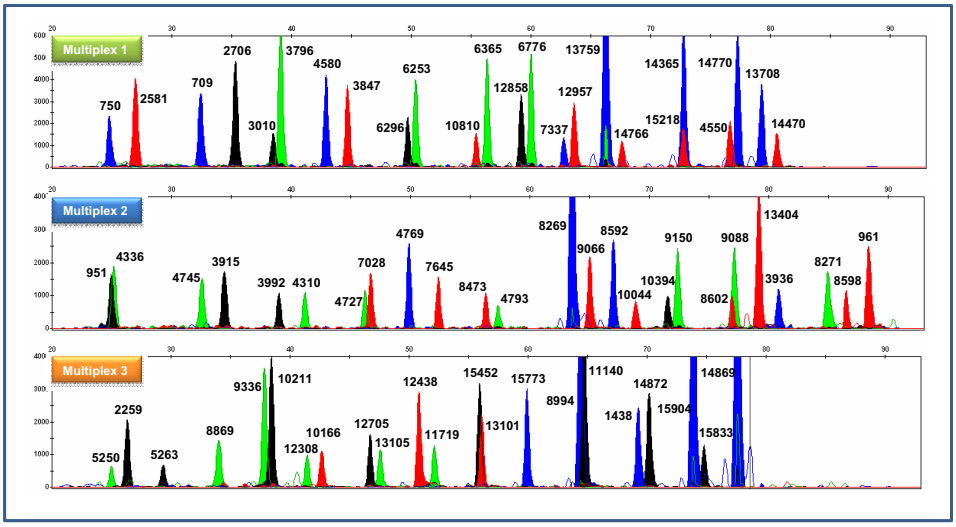
Table 1. List of SNPs genotyped in the present study

Marker	Allele	Multiplex	Marker	Allele	Multiplex	Marker	Allele	Multiplex
709	G-A	1	**951**	G-A	2	**1438**	A-G	3
750	A-G	1	**961G**	T-G	2	**2259**	C-T	3
2581	A-G	1	**3915**	G-A	2	**5250**	T-C	3
2706	A-G	1	**3936**	C-T	2	**5263**	C-T	3
3010	G-A	1	**3992**	C-T	2	**8869**	A-G	3
3796	A-G	1	**4310**	A-G	2	**8994**	G-A	3
3847	T-C	1	**4336**	T-C	2	**9336**	A-G	3
4550	T-C	1	**4727**	A-G	2	**10166**	T-C	3
4580	G-A	1	**4745**	A-G	2	**10211**	C-T	3
6253	T-C	1	**4769**	A-G	2	**11140**	C-T	3
6296	C-T	1	**4793**	A-G	2	**11719**	G-A	3
6365	T-C	1	**7028**	C-T	2	**12308**	A-G	3
6776	T-C	1	**7645**	T-C	2	**12438**	T-C	3
7337	G-A	1	**8269**	G-A	2	**12705**	C-T	3
10810	T-C	1	**8271T**	A-T	2	**13101C**	A-C	3
12858	C-T	1	**8473**	T-C	2	**13105**	A-G	3
12957	T-C	1	**8592**	G-A	2	**14869**	G-A	3
13708	G-A	1	**8598**	T-C	2	**14872**	C-T	3
13759	G-A	1	**8602**	T-C	2	**15452A**	C-A	3
14365	C-T	1	**9066**	A-G	2	**15773**	G-A	3
14470A	T-A	1	**9088**	T-C	2	**15833**	C-T	3
14766	C-T	1	**9150**	A-G	2	**15904**	C-T	3
14770	C-T	1	**10044**	A-G	2			
15218	A-G	1	**10394**	C-T	2			
			13404	T-C	2			

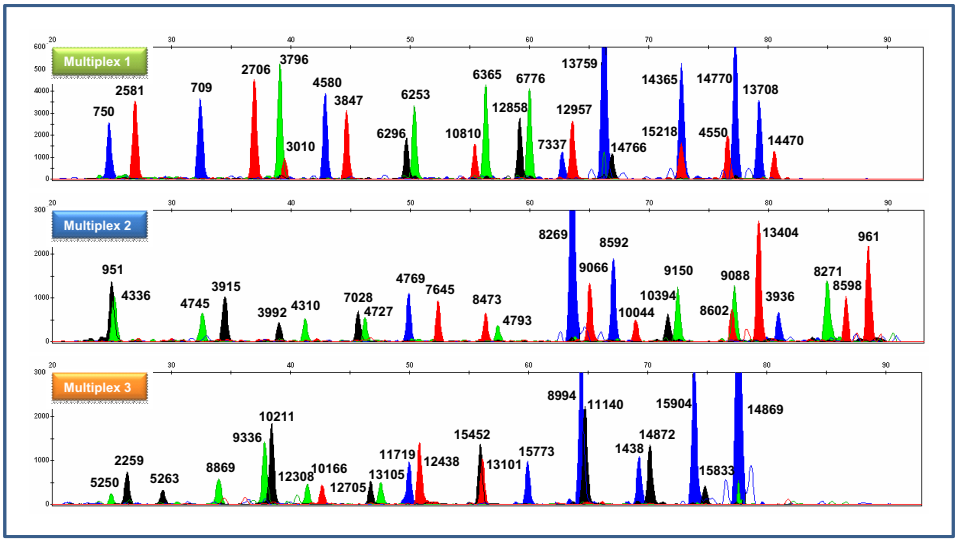
Legend to figures

Figure 1. Haplogroup R* (A) and H1 (B) SNP profiles for samples Femur_1 and Tooth_1, respectively (see Table S1)

(A)



(B)



CONCLUSIONES

De interés general y/o metodológicas:

- La técnica de minisecuenciación implementada en SNaPshot es adecuada para el genotipado simultáneo de SNPs en multiplexes. Entre sus múltiples ventajas respecto a otras técnicas clásicas (e.j. RFLP) están su bajo coste, la rapidez del genotipado, y la reproducibilidad de los resultados.
- Las técnicas de genotipado multiplex previenen errores que pudieran surgir debido a mezcla de muestras (*sample mix-up*) y contaminación. Estos errores surgen fundamentalmente cuando una misma muestra se analiza de manera independiente (vía PCR) para distintos marcadores moleculares (e.j. HVR-I y HVR-II o diferentes SNPs).
- Los errores de genotipado en ADNmt son relativamente fáciles de detectar cuando se examina la variabilidad mitocondrial (e.j. *artefactual recombinants*) debido a que los patrones que se generan de manera artificial difieren de la variabilidad conocida en las poblaciones humanas. La detección de errores consiste básicamente en comparar los resultados de genotipado (secuenciación o SNPs) con el árbol de clasificación mitocondrial. Se pueden dar varias circunstancias: a) falta variabilidad que debería estar presente dadas las características del perfil, b) existe mezcla de linajes (por contaminación o mezcla de muestras), c) aparecen mutaciones fantasma (especialmente detectables cuando las variantes son raras o diagnósticas de haplogrupos concretos), y d) existen errores de documentación.

De interés forense:

- Aunque el ADNmt no es el marcador más apropiado para la detección de mezclas de distintos fluídos biológicos, su análisis puede ser muy útil en aquellos contextos en los que no se

pueden analizar marcadores autosómicos o de cromosoma Y debido al estado de la muestra o al bajo número de copias de ADN en la muestra.

- En ocasiones es difícil interpretar los patrones de mezcla debido a que la propia química de secuenciación no presentará de manera homogénea la carga mitocondrial de cada componente de la muestra (i.e. heteroplasmías). Tal y como se demuestra en estudios de mezclas controladas, no se puede predecir *a priori* el aporte concreto de mitocondrias de cada uno de los componentes/donantes.
- La interpretación filogenética de los perfiles de mezcla es por lo tanto la única herramienta que podría permitir separar los distintos perfiles que pudieran haber contribuído a la mezcla.
- En aquellas ocasiones en las que la amplificación de la región control (HVS-I/II) no ha tenido éxito en una determinada muestra debido a su estado de degradación severa o su escaso contenido de ADN, el análisis de SNPs mitocondriales se presenta como una técnica alternativa de gran utilidad debido a que la probabilidad de conseguir genotipar una muestra (amplificación y minisecuenciación) es superior a la probabilidad de obtener resultados de secuenciación.
- El poder de discrimación de la región control es frecuentemente escaso, especialmente en poblaciones Europeas y Nativo Americanas. El análisis de SNPs codificantes puede aumentar este poder, especialmente genotipando variantes con una alta tasa de mutación o aquellos que determinan sub-linajes mitocondriales concretos.

De interés poblacional:

- La población Afro-Colombiana está fundamentalmente compuesta por linajes mitocondriales de ancestralidad sub-Sahariana. Sin embargo presenta también un porcentaje elevado de linajes Nativo Americanos y algún linaje europeo. La población 'Mestiza' está fundamentalmente compuesta por linajes mitocondriales Nativo Americanos. La población de 'Mulatos' está fundamentalmente compuesta por linajes mitocondriales de ancestralidad Africana.
- La variabilidad de los linajes mitocondriales Nativos Americanos observada en la población colombiana es muy alta cuando ésta se valora a través de diferentes índices de diversidad génica o a través de criterios filogeográficos. No existen evidencias de deriva génica en dichos linajes indicando que el proceso de 'mestizaje' fue gradual y permitiendo que un porcentaje importante de la variabilidad hoy persista en la población 'mestiza'.
- Tal y como cabría esperar, la mayor parte de la variabilidad Nativo Americana en Colombia es afín a Centro América (Panamá, Costa Rica, etc) y al Norte de Suramérica, especialmente a grupos indígenas de países vecinos de Venezuela, Ecuador, y Brazil. Existe algún linaje sin embargo que paradójicamente ha sido detectado únicamente en poblaciones indígenas que actualmente residen en regiones ubicadas al norte del continente americano.
- La variabilidad genética de los linajes mitocondriales Africanos en Colombia es genéticamente muy diversa, reflejando el gran tamaño efectivo de población de esclavos que fueron forzosamente llevados a Colombia en tiempos del tráfico de esclavos trans-Atlántico. Tal y como cabría esperar, la

procedencia de la mayor parte de los linajes Afro-Colombianos de ancestralidad Africana es Senegambia y en general la costa oeste y centro-oeste de África. También el sur-oeste (Angola, Cabinda) y el sureste de Africa (Mozambique) han contribuído substancialmente a esta variabilidad.

- La ancestralidad basada en la autodefinicición de la etnicidad (*seft-reported etnicity*) no siempre es indicativa de la ancestralidad genética. Además, muestras independientes de poblaciones mestizadas parecen presentar distintos patrones de frecuencias de haplogrupos. Esto puede tener importantes consecuencias cuando se generan bases de datos con fines forenses y se considera tales grupos pseudos-étnicos como genéticamente homogéneos; este es, de hecho, el problema que caracteriza bases de datos importantes tales como las de la SWGDAM. Este problema se puede agravar en países en donde el concepto de 'raza' está fuertemente arraigado en la sociedad, por lo que las bases de datos de grupos pseudos-étnicos cobran la apariencia de grupos genéticamente homogéneos.
- El uso de los polimorfismos genéticos para examinar la historia de la población de América es muy complicado por la posibilidad de mezcla reciente. Esta mezcla podría tener un mayor efecto en los marcadores del cromosoma Y, dada la evidencia de una mezcla preferencial en las Américas entre el hombre inmigrante y la mujer nativa americana.
- Para el ADN mitocondrial, la población de Córdoba presenta una alta prevalencia de linajes Nativo Americanos (representados por los haplogrupos A2, B2, C1, y D1), así como trazas de linajes sub-Saharianos.
- La variabilidad de los linajes mitocondriales Nativo Americanos observada en la población de Córdoba es muy alta cuando ésta

se valora a través de diferentes índices de diversidad génica o a través de criterios filogeográficos. No existen evidencias de deriva génica en dichos linajes, indicando que el proceso de mestizaje con la población europea fue gradual y permitiendo que un porcentaje importante de la variabilidad hoy persista en la población.

- Tal y como cabría esperar, la mayor parte de la variabilidad mitocondrial Nativo Americana en Córdoba es afín a poblaciones de Sudamérica, especialmente a grupos indígenas Argentinos (Mapuches, Coyas, etc.) y de países vecinos como Chile y Brasil.La procedencia africana más probable de los escasos linajes Afro-Americanos observados en Córdoba es la costa centro-oeste y el suroeste Atlántico (Angola).

- La mayoría de los haplotipos de ADNmt encontrados son típicamente nativos americanos, mayoritariamente pertenecientes al haplogrupo A2. La contribución de linajes europeos de ADNmt en la muestra estudiada es bajo (aproximadamente 2%), lo que contrasta con el porcentaje observado para los linajes de cromosoma Y.

- Coincidiendo con la documentación histórica, la contribución de linajes Afro-Americanos de ADNmt en la muestra estudiada en la población de El Salvador es muy baja, prueba de ello es que sólo se detectó una muestra perteneciente al haplogrupo de origen claramente sub-Sahariano L01a1.

- No se detectó presencia de deriva génica intensa (e.j. cuellos de botella) en la población de El Salvador.

- Los resultados obtenidos en el estudio de las muestras de El Salvador han permitido confirmar la hipótesis de que los salvadoreños han preservado el componente matrilineal

indígena original de los primeros habitantes de la región, sin detectarse una influencia importante demográfica de otras poblaciones nativas americanas del Norte, Centro, Suramérica, así como de otras poblaciones no americanas.

- Se observa una clara asimetría en el aporte de ancestralidad Europeo en el componente masculino (cromosoma Y) y femenino (mtDNA). Este resultado confirma lo que cabría esperar de acuerdo a los antecedentes históricos. El conocimiento de la estructura genómica de la población de El Salvador es importante no sólo para la casuística forense, sino también para estudios relacionados con la investigación médica de enfermedades raras y complejas; estas últimas a través por ejemplo de estudios de asociación basados en poblaciones (e.j. caso-control) o basados en métodos tales como '*admixture mapping*'.

- La técnica de minisecuenciación en genética de poblaciones aplicada al genotipado de SNPs codificantes tiene una finalidad doble: a) aumentar el poder de discriminación del test mitocondrial optimizando su poder informativo una vez la región control (que en si mismo también contiene posiciones diagnósticas de utilidad filogenética y filogeográfica) ha sido secuenciada, y b) clasificar los linajes mitocondriales en sus haplogrupos correspondientes.

- Existe una gran confusión en la nomenclatura de los haplogrupos del ADN mitocondrial en la literatura. Un ejemplo claro de dicha confusión se puede encontrar en R0 y sus sub-linajes, entre los cuales cabe destacar el haplogrupo H. El esfuerzo de estadarización es necesario entonces para poder permitir estudios futuros inter-poblacionales.

- Existen tres linajes que muestran picos de frecuencias en la región Cántabra y patrones de frecuencia clinal decreciente hacia el Este y Sur de Europa (hacia Oriente Medio), H1, H3 y H5a. Estos resultados corroboran los obtenidos previamente por otros autores. Los patrones de frecuencias de estos haplogrupos responden a los eventos de expansión poblacional y recolonización de Europa desde zonas de refugio después del último periodo glacial.

- Existen diferencias de frecuencias para linajes mitocondriales en el Norte de Iberia. Estas diferencias pueden ser relevantes en estudios forenses y clínicos. En los primeros debido al hecho de que la estratificaicón poblacional podría afectar de manera significativa a la evaluación de la prueba forense. En los segundos, debido al hecho de que los estudios clínicos basados en poblaciones (e.j. caso-control en enfermedades complejas) son altamente sensibles a la estratificación poblacional, siendo este un factor principal de error tipo I en la asociación.

- A través del estudio de genomas completos se ha descrito un sub-linaje nuevo en la población Vasca, denominado H2a5. Este linaje tiene aproximadamente 15,000 años, coincidiendo por lo tanto con el periodo de expansión poblacional después del último periodo glacial que caracterizó de manera especial al refugio Cántabro-Francés. Curiosamente, este linaje tan sólo se ha detectado en el País Vasco, con lo que se le puede atribuir un carácter autóctono en esta región.

De interés clínico

- Existe una alta prevalencia de la mutación m.11778G>A en los pacientes Españoles de LEBER analizados.

- El haplogrupo J está sobrerrepresentado en la muestra de pacientes de Leber analizados. Resultados similares se han observado recientemente en otras poblaciones de ancestralidad europea y asiática. Esta asociación podría indicar que existe una predisposición genética asociada a la presencia de dicho haplogrupo.

- No se puede descartar la posibilidad de que en algunas ocasiones se presenten mutaciones altamente deteléreas en individuos sanos. El hecho de que dichas mutaciones aparezcan por debajo de un umbral determinado en las células de los individuos sanos explicaría por que dichas mutaciones no tienen consecuencias sobre el fenotipo del individuo y además porqué estas mutaciones se mantienen en las poblaciones con una prevalencia relativamente alta. También explicaría porqué dichas mutaciones parecen estar asociadas a determinados efectos fundadores en las poblaciones; este podría ser el caso de la mutación m.14484T>C en pacientes de Leber en la población Gallega.

- Mitomap se utiliza como el *gold-standard* para la búsqueda y el almacenamiento de mutaciones patogénicas en genoma mitocondrial. Sin embargo, existen muchos artefactos relacionados con esta práctica recurrente en los laboratorios de diagnóstico molecular y consejo genético. Por ejemplo, la literatura está actualmente plagada de falsos posivos, ya que cualquier mutación 'novel' que aparezca en pacientes se interpreta frecuentemente como mutación patogénica. Por lo tanto, es importante que el genetista haga el esfuerzo de hacer búsquedas más exhaustivas en otros recursos existentes en la Web y en

otras bases de datos, incluyendo la literatura que aun no ha sido registrada en i.g. Mitomap.

- El diseño de multiplexes de minisecuenciación es eficaz en los laboratorios de diagnóstico molecular ya que esta técnica reduce el tiempo de genotipado y el coste significativamente. Sin embargo, cuando no se detecta la mutación causal, se debería genotipar el genoma mitocondrial completo de forma sistemática. Toda la variabilidad detectada en un paciente debe además interpretarse a la luz de la filogenia para evitar en la medida de lo posible resultados espúreos de asociación positiva.

BIBLIOGRAFÍA

Bibliografía:

- **Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD.** (1996) Biología molecular de La Célula. 3ª edición. Barcelona: Ediciones Omega.
- **Ávarez-Iglesias V, Jaime JC, Carracedo Á, Salas A.** (2007) Coding region mitochondrial DNA SNPs: Targeting East Asian and Native American haplogroups. *FSI:Genetics* 1(1): 44-55.
- **Álvarez-Iglesias V, Barros F, Carracedo A, Salas A.** (2008) Minisequencing mitochondrial DNA pathogenic mutations. *BMC Medical Genetics* 9:26.
- **Andreu AL, Dimauro S** (2003) Current classification of mitochondrial disorders. *J Neurol* 250:1403-1406.
- **Andreu AL, Gonzalo-Sanz R** (2004) Las enfermedades mitocondriales: una clasificación para el siglo XXI. *Neurología* 19(1):15-22.
- **Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N** (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23 (2):147.
- **Bandelt HJ, Lahermo M, Richards M, Macauly V** (2001) Detecting errors in mtDNA data by phylogenetic analysis. *Int J Legal Med* 115: 64-69.
- **Bandelt HJ, Salas A, Lutz-Bonengel S** (2004a) Artificial recombination in forensic mtDNA population databases. *Int J Legal Med* 118(5):267-73.
- **Bandelt HJ, Salas A, Bravi CM** (2004b) Problems in FBI mtDNA database. *Science* 305: 1402-1404.

- **Bandelt HJ, Kong QP, Parson W, Salas A.** (2005) More evidence for non-maternal inheritance of mitochondrial DNA? *J Med Genet* 42(12):957-60.
- **Bandelt HJ, Salas A, Bravi CM** (2006) What is a "novel" mtDNA mutation- and does "novelty" really matter? *J Hum Genet* 51: 1073-1082.
- **Bandelt HJ, Yao YG, Salas A** (2007) The search of 'novel' mtDNA mutations in hypertrophic cardiomyopathy: MITOMAPping as a risk factor. *Int J Cardiol.* (In press).
- **Bannwarth S, Procaccio V, Rouzier C, Fragaki K, Poole J, Chabrol B, Desnuelle C, Pouget J, Azulay JP, Attarian S, Pellissier JF, Gargus JJ, Abdenur JE, Mozaffar T, Calvas P, Labauge P, Pages M, Wallace DC, Lambert JC, Paquis-Flucklinger V** (2008) Rapid identification of mitochondrial DNA (mtDNA) mutations in neuromuscular disorders by using surveyor strategy. *Mitochondrion* 8: 136-145.
- **Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M, Lincoln PJ, Mayr W, Morling N, Olaisen B, Schneider PM, Tully G, Wilson M** (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Int J Legal Med* 113(4):193-6.
- **Barros F, Carracedo A, Lareu MV, Rodríguez MS** (1997) Rapid and enhaced detection of mitochondrial DNA variation using single-strand conformation analysis of superposed restriction enzyme fragments from polymerase chain reaction-amplified products. *Electrophoresis* 18: 1041-45.
- **Bendall KE, Macaulay VA, Baker JR, Sykes BC** (1996) Heteroplasmic point mutations in the human mtDNA control region. *Am J Hum Genet* 59(6):1276-87.

- **Bodenteich A, Mitchell LG, Polymeropoulos MH, Merril CR** (1992) Dinucleotide repeat in the human mitochondrial D-loop. *Hum Mol Genet* 1(2):140.
- **Brandon MC, Lott MT, Nguyen KC, Spolim S, Navathe SB, Baldi P, Wallace DC** (2005) MITOMAP: a human mitochondrial genome database--2004 update. *Nucleic Acids Res.* 33 (Database issue):D611-3.
- **Brandstätter A, Parsons TJ, Parson W** (2003)Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med* 117(5):291-8.
- **Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, Parsons TJ** (2004) Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the stablishment of a forensic database. *Int J Legal Med* 118: 294-306.
- **Brandstätter A, Salas A, Niederstätter H, Gassner C, Carracedo Á, Parson W** (2007)Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis* 13:2541-2550.
- **Brown WM, George M Jr, Wilson AC** (1979) Rapid evolution of animal mitochondrial DNA. *Proc natl Acad Sci USA* 76: 1967-71.
- **Budowle B, Allard MW, Wilson WR, Chakraborty R** (2003) Forensics and mitochondrial DNA: Applications, debates and foundations. *Annu Rev Genomics Hum Genet* 4: 119-141.
- **Calloway CD, Reynolds RL, Herrin GL Jr, Anderson WW** (2000) The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet* 66(4):1384-97.
- **Cann RL, Stoneking M, Wilson AC** (1987) Mitochondrial DNA and human evolution. *Nature* 325: 31-36.

- **Carothers AM, Urlaub G, Mucha J, Grunberger D, Chasin LA** (1989) Point mutation analysis in a mammalian gene: rapid preparation of total RNA, PCR amplification of cDNA, and Taq sequencing by a novel method. *Biotechniques* 7: 494-496, 498-499.
- **Carracedo A, D'Aloja E, Dupuy B, Jangblad A, Karjalainen M, Lambert C, Parson W, Pfeiffer H, Pfitzinger H, Sabatier M, Syndercombe Court D, Vide C** (1998) Reproducibility of mtDNA analysis between laboratories: a report of the European DNA Profiling Group (EDNAP). *Forensic Sci Int* 97(2-3):165-70.
- **Carracedo A, Bar W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M** (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int* 110(2):79-85.
- **Cassandrini D, Calevo MG, Tessa A, Manfredi G, Fattori F, Meschini MC, Carrozzo R, Tonoli E, Pedemonte M, Minetti C, Zara F, Santorelli FM, Bruno C** (2006) A new method for analysis of mitochondrial DNA point mutations and assess levels of heteroplasmy. *Biochem Biophys Res Commun* 7; 342(2):387-93.
- **Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP** (1996) Accessing genetic information with high-density DNA arrays. *Science* 274(5287): 610-4.
- **Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC** (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 57(1):133-49.
- **Crespillo M, Paredes MR, Prieto L, Montesino M, Salas A, Albarran C, Alvarez-Iglesias V, Amorin A, Berniell-Lee G,**

Brehm A, Carril JC, Corach D, Cuevas N, Di Lonardo AM, Doutremepuich C, Espinheira RM, Espinoza M, Gomez F, Gonzalez A, Hernandez A, Hidalgo M, Jimenez M, Leite FP, Lopez AM, Lopez-Soto M, Lorente JA, Pagano S, Palacio AM, Pestano JJ, Pinheiro MF, Raimondi E, Ramon MM, Tovar F, Vidal-Rioja L, Vide MC, Whittle MR, Yunis JJ, Garcia-Hirschfel J (2006) Results of the 2003-2004 GEP-ISFG collaborative study on mitochondrial DNA: focus on the mtDNA profile of a mixed semen-saliva stain. *Forensic Sci Int.* 13;160 (2-3):157-67.

- **Crews S, Ojala D, Posakony J, Nishiguchi J, Attardi G** (1979) Nucleotide sequence of a region of human mitochondrial DNA containing the precisely identified origin of replication. *Nature* 277(5693):192-8.
- **Denaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli-Sforza LL, Wallace DC** (1980) Ethnic variation in Hpa I endonuclease patterns of human mitochondrial DNA. *Proc Natl Acad Sci USA* 78: 5768-5772.
- **Dimauro S, Schon EA** (2001) Mitochondrial DNA mutations in human disease. *Am J Med Genet* 106:18-26.
- **Dimauro S** (2004) Mitochondrial diseases. *Biochim Biophys Acta* 1658(1-2):80-88.
- **Dimauro S, Davidzon G** (2005) Mitochondrial DNA and disease. *Annals of Medicine* 37(3):222-32.
- **DiMauro S** (2007) Mitochondrial DNA medicine. *Biosci Rep* 27(1-3):5-9.
- **Endicott P, Metspalu M, Stringer C, Macaulay V, Cooper A, Sanchez JJ** (2006) Multiplexed SNP typing of ancient DNA clarifies the origin of andaman mtDNA haplogroups amongst south Asian tribal populations. *PLoS ONE* 20;1:e81.

- **Filippini S, Blanco A, Fernandez-Marmiesse A, Alvarez-Iglesias V, Ruiz-Ponte C, Carracedo A, Vega A** (2007) Multiplex SNaPshot for detection of BRCA1/2 common mutations in Spanish and Spanish related breast/ovarian cancer families. *BMC Med Genet* 8:40.
- **Giles RE, Blanc H, Cann HM, Wallace DC** (1980) Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA* 77:6715-19.
- **Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K** (1994) Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 6(2):130-5.
- **Grignani P, Peloso G, Achilli A, Turchi C, Tagliabracci A, Alu M, Beduschi G, Ricci U, Giunti L, Robino C, Gino S, Previdere C** (2006) Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification. *Int J Legal Med* 120(3):151-6.
- **Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB, Bowden DK** (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc R Soc London B Biol Sci* 266:485-92.
- **Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB, Bowden DK** (2000) Evidence for mitochondrial DNA recombination in a human population of island Melanesia: correction. *Proc R Soc London B Biol Sci* 267:1595-96.
- **Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N** (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 70(5):1152-71.

- **Holt IJ, Harding AE, Morgan-Hughes JA** (1988) Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* 331: 717-719.
- **Hopgood R, Sullivan K, Gill P** (1992) Strategies for automated sequencing of human mitochondrial DNA directly from PCR products. *Biotechniques* 13: 82-92.
- **Horai S, Hayasaka K** (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am J Hum Genet* 46(4):828-42.
- **Howell N, Kubacka I, Mackey DA** (1996) How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet* 59(3):501-9.
- **Ingman M, Kaessmann H, Paabo S, Gyllensten U** (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813):708-13. Erratum in: *Nature* (2001) 410(6828):611.
- **Ingman M, Gyllensten U** (2001) Analysis of the complete human mtDNA genome: methodolody and inferences for human evolution. *J Hered* 92:454-461.
- **Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M, Gyllensten U** (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nat Genet* 18(2):109-10.
- **Jeffreys AJ, Wilson V, Thein SL** (1985) Individual-specific 'fingerprints' of human DNA. *Nature* 4-10; 316(6023):76-9.
- **Kmiec M, Woloszynska M, Janska H** (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Curr Genet* 50: 149-159.
- **Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP** (2003) Phylogeny of east Asian mitochondrial DNA lineages inferred from

complete sequences. *Am J Hum Genet* 73(3):671-6; 75(1):157 (erratum).

- **Kong QP, Yao YG, Sun C, Zhu CL, Zhong L, Wang CY, Cai WW, Xu XM, Xu AL, Zhang YP** (2004) Phylogeographic analysis of mitochondrial DNA haplogroup F2 in China reveals T12338C in the initiation codon of the ND5 gene not to be pathogenic. *J Hum Genet* 49(8):414-23
- **Kong QP, Bandelt HJ, Sun C, Yao YG, Salas A, Achilli A, Wang CY, Zhong L, Zhu CL, Wu SF, Torroni A, Zhang YP** (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet* 15(13):2076-86.
- **Lee HY, Yoo JE, Park MJ, Chung U, Kim CY, Shin KJ** (2006) East Asian mtDNA haplogroup determination in Koreans: haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis. *Electrophoresis* 27(22):4408-18.
- **Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM** (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13.
- **Maxam AM and Gilbert W** (1977) New method for sequencing DNA. *Proc Natl Acad Sci USA* 74:560-564.
- **Margulis, L** (1976) Genetic and evolutionary consequences of symbiosis. *Exp Parasitol* 39: 277-349
- **McKenzie M, Liolitsa D, Hanna MG** (2004) Mitochondrial disease: mutations and mechanisms. *Neurochemical Research* 29 (3): 589-600.

- **McFarland R, Taylor RW, Elson JL, Lightowlers RN,Turnbull DM, Howell N** (2004a) Proving Pathogenicity: when evolution is not enough. *Am J Med Genet* 131A: 107-108
- **McFarland R, Elson JL, Taylor RW, Howell N, Turnbull DM** (2004b) Assigning pathogenicity to mitochondrial tRNA mutations: when "definitely maybe" is not good enough. *Trends Genet* 20 (12): 591-596.
- **McFarland R, Taylor RW, Turnbull DM** (2007) Mitochondrial disease- Its impact, etiology and pathology. *Current Topics in Developmental Biology* 77:113-155.
- **Miller KW, Dawson JL, Hagelberg E** (1996) A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. *Int J Legal Med*. 109(3):107-13
- **Mitchell AL, Elson JL, Howell N, Taylor RW, Tumbull DM** (2006) Sequence variation in mitochondrial complex I genes: mutation or polymorphism? *J Med Genet* 43(2):175-179.
- **Monnat RJ Jr, Reay DT** (1986) Nucleotide sequence identity of mitochondrial DNA from different human tissues. *Gene* 43(3):205-211.
- **Montesino M, Salas A, Crespillo M, Albarran C, Alonso A, Álvarez-Iglesias V, Cano JA, Carvalho M, Corach D, Cruz C, Di Lonardo A, Espinheira R, Farfan MJ, Filippini S, Garcia-Hirschfeld J, Hernandez A, Lima G, Lopez-Cubria CM, Lopez-Soto M, Pagano S, Paredes M, Pinheiro MF, Rodriguez-Monge AM, Sala A, Sonora S, Sumita DR, Vide MC, Whittle MR, Zurita A, Prieto L** (2007) Analysis of body fluid mixtures by mtDNA sequencing: An inter-laboratory study of the GEP-ISFG working group. *Forensic Sci Int* 168: 42-56.

- **Orita M, Iwahana H, Kanazawa K, Sekiya T** (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA* 86: 2766-2770.
- **Pakendorf B, Stoneking M** (2005) Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6: 165-183.
- **Parson W, Brandstatter A, Alonso A, Brandt N, Brinkmann B, Carracedo A, Corach D, Froment O, Furac I, Grzybowski T, Hedberg K, Keyser-Tracqui C, Kupiec T, Lutz-Bonengel S, Mevag B, Ploski R, Schmitter H, Schneider P, Syndercombe-Court D, Sorensen E, Thew H, Tully G, Scheithauer R** (2004) The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. *Forensic Sci Int* 139(2-3):215-26.
- **Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM** (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* 15(4):363-8.
- **Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A** (2004) Tiping of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int* 130(2-3): 97-111.
- **Salas A, Rasmussen EM, Lareu MV, Morling N, Carracedo A** (2001) Fluorescent SSCP of overlapping fragments (FSSCP-OF): a highly sensitive method for the screening of mitochondrial DNA variation. *Forensic Sci Int* 124(2-3):97-103.
- **Salas A, Carracedo A, Macaulay V, Richards M, Bandelt HJ** (2005a) A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. *Biochem Biophys Res Commun* 335(3):891-9.

- **Salas A, Prieto L, Montesino M, Albarran C, Arroyo E, Paredes-Herrera MR, Di Lonardo AM, Doutremepuich C, Fernandez-Fernandez I, de la Vega AG, Alves C, Lopez CM, Lopez-Soto M, Lorente JA, Picornell A, Espinheira RM, Hernandez A, Palacio AM, Espinoza M, Yunis JJ, Perez-Lezaun A, Pestano JJ, Carril JC, Corach D, Vide MC, Alvarez-Iglesias V, Pinheiro MF, Whittle MR, Brehm A, Gomez J** (2005b) Mitochondrial DNA error prophylaxis: assessing the causes of errors in the GEP'02-03 proficiency testing trial. *Forensic Sci Int* 148(2-3):191-8.
- **Salas A, Bandelt HJ, Macauly V, Richards M** (2006) Phylogeographic investigations: The role of trees in forensic genetics. *Forensic Sci Int* 168(1):1-13.
- **Sanger F, Nicklen S, Coulson AR** (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5468.
- **Schapira AH** (2006) Mitochondrial disease. *Lancet* 368(9529):70-82.
- **Sigurdsson S, Hedman M, Sistonen P, Sajantila A, Syvänen AC** (2006) A microarray system for genotyping 150 single nucleotide polymorphisms in the coding region of human mitochondrial DNA. *Genomics* 87(4):534-42.
- **Sigurgardottir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P** (2000) The mutation rate in the human mtDNA control region. *Am J Hum Genet* 66(5):1599-609.
- **Singer-Sam J, Tanguay RL, Riggs AD** (1998) Use of chelex to improve the PCR signal from a small number of cells. *Amplification* 3:11.
- **Sobrino B, Brión M, Carracedo A** (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 154:181-194.

- **Stoneking M, Sherry ST, Redd AJ, Vigilant L** (1992) New approaches to dating suggest a recent age for the human mtDNA ancestor. *Philos Trans R Soc Lond B Biol Sci* 29;337 (1280):167-75.
- **Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G** (1999) Ubiquitin tag for sperm mitochondria. *Nature* 402(6760):371-2.
- **Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G** (2000) Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod* 63(2):582-90.
- **Taylor RW, Turnbull DM** (2005) Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 6(5):389-402.
- **Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC** (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130(1):153-62.
- **Tully G, Sullivan KM, Nixon P, Stones RE, Gill P** (1996) Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. *Genomics* 34(1):107-13.
- **Tully G, Bar W, Brinkmann B, Carracedo A, Gill P, Morling N, Parson W, Schneider P** (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Sci Int* 124: 83-91.
- **Tully G, Barritt SM, Bender K, Brignon E, Capelli C, Dimo-Simonin N, Eichmann C, Ernst CM, Lambert C, Lareu MV, Ludes B, Mevag B, Parson W, Pfeiffer H, Salas A, Schneider**

PM, Staalstrom E (2004) Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts. *Forensic Sci Int* 10;140(1):1-11.

- **Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ** (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. *Int J Legal Med* 118(3):147-57.
- **Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC** (1989) Mitochondria DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci USA* 86: 9350-9354.
- **Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza A, Elsas LJ, Nikoslainen EK** (1988) Mitochondrial DNA mutation associated with Leber´s hereditary optic neuropathy. *Science* 242: 1427- 1430.
- **Wallace DC, Brown MD, Lott MT** (1999) Mitochondrial DNA variation in human evolution and disease. *Gene* 238(1):211-230.
- **Wang X** (2001) The expanding role of mitochondria in apoptosis. *Genes Dev* 15: 2922-2933.
- **Ward RH, Franzier BL, Dew-Jager, Pääbo S** (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc. *Natl Acad Sci USA* 88:8720-8724.
- **Wiesbauer M, Meierhofer D, Mayr JA, Sperl W, Paulweber B, Kofler B** (2006) Multiplex primer extension analysis for rapid detection of major European mitochondrial haplogroups. *Electrophoresis* 27(19):3864-8.
- **Wyman AR, White R** (1980) A highly polymorphic locus in human DNA. *Proc Natl Acad Sci USA* 77(11):6754-8.

- **Yao YG, Salas A, Bravi CM, Bandelt HJ** (2006) A reappraisal of complete mtDNA variation in East Asian families with hearing impairment. *Hum Genet* 119(5):505-15.
- **Zsurka G, Kraytsberg Y, Kudina T, Kornblum C, Elger CE, Khrapko K, Kunz WS** (2005) Recombination of mitochondrial DNA in skeletal muscle of individuals with multiple mitochondrial DNA heteroplasmy. *Nat Genet* 37(8):873-7.
- **Zsurka G, Hampel KG, Kudina T, Kornblum C, Kraytsberg Y, Elger CE, Khrapko K, Kunz WS** (2007) Inheritance of mitochondrial DNA recombinants in double-heteroplasmic families: potential implications for phylogenetic analysis. *Am J Hum Genet* 80(2):298-305

Bibliografía adicional

A lo largo de este documento de tesis se han utilizado otros recursos bibliográficos:

Acosta A (2007) *Estudio de la variabilidad a nivel molecular del DNA autosómico, mitocondrial y del cromosoma Y en una muestra poblacional del sur occidente de Colombia.* Tesis doctoral, Universidade de Santiago de Compostela

Lovo S (2006) *Diversidad genética de las poblaciones de El Salvador. Estudio de marcadores del ADN mitocondrial, ADN autosómico y del cromosoma Y*. Tesis doctoral, Universidade de Santiago de Compostela.

Quintáns B (2004) *Nuevos retos en Genética Forense: SNPs de ADN mitocondrial y cromosoma Y, nuevos microsatélites de cromosoma Y e influencia de la subestructurapoblacional.* Tesis doctoral, Universidade de Santiago de Compostela.

Salas A (1999) *Estudio de la variabilidad molecular del ADN mitocondrial humano*. Tesis doctoral, Universidade de Santiago de Compostela.