

OBJETIVOS

Objetivo General

El objetivo general de esta tesis doctoral, es valorar las posibilidades de los estudios de DNA antiguo a nivel poblacional, en la resolución de los problemas mencionados en los marcos teórico y de referencia, como la estimación de la tasa de evolución y el origen del acervo genético en Europa y sobretodo, en el desarrollo de métodos fiables para la detección de la contaminación con DNA moderno. Utilizando para ello, la secuenciación de un fragmento de la región de control del DNA mitocondrial y la tipificación del mtDNA mediante enzimas de restricción; en una población de la que se puede contar con muestras de individuos de dos etapas temporales bien diferenciadas.

Objetivos particulares

Los objetivos particulares comprenden:

1. Puesta a punto y valoración de distintos métodos de extracción de DNA antiguo, tanto publicados, como diseñados en el laboratorio.
2. Investigación bibliográfica y experimental, con el fin de mejorar las técnicas de extracción de DNA antiguo, especialmente en lo referente a la presencia de sustancias que inhiben las reacciones enzimáticas.
3. Extracción de DNA, amplificación y secuenciación de un segmento de mtDNA de al menos 15 individuos de una población catalana antigua.
4. La tipificación de los haplogrupos europeos del DNA mitocondrial de los individuos antiguos, de los cuales se tenga la secuencia del segmento de DNA mitocondrial mencionado en el objetivo anterior.
5. Valorar la utilidad de la correlación entre los polimorfismos de restricción y los polimorfismos de la secuencia en la autenticación del DNA extraído de restos antiguos.
6. Extracción de DNA, amplificación y secuenciación de un segmento de DNA mitocondrial de al menos 25 individuos de población catalana contemporánea.
7. La tipificación de los haplogrupos europeos del DNA mitocondrial de los individuos de la población contemporánea.
8. Comparar, con métodos estadísticos y filogenéticos, las poblaciones antigua y actual de Cataluña, entre sí y con otras poblaciones, principalmente europeas.
9. Evaluar mediante esta comparación la calidad y cantidad de información contenida en el segmento de mtDNA elegido para el análisis de la población antigua.

MATERIAL Y MÉTODOS

MATERIAL DE ESTUDIO

Poblaciones

Para este trabajo, se eligió como objeto de estudio la población catalana, de la comarca “del Vallès”, provincia de Barcelona, España. De esta población se extrajo DNA de individuos de dos épocas diferentes, se amplificó y secuenció un segmento de la región de control del mtDNA así como diferentes segmentos de las zonas codificantes, también de mtDNA, con el fin de tipificar las muestras para los haplogrupos mitocondriales europeos definidos por Torroni y col. (1996). Así mismo, se analizaron individuos de otras poblaciones antiguas, para utilizarlos como controles externos.

Población antigua

La población antigua procede de la necrópolis de la “Plaça Vella” de Terrassa, situada en el centro de la antigua villa medieval de la ciudad. Los restos humanos hallados en este yacimiento fueron excavados por el equipo de arqueología urbana del “Museu de Terrassa”, dirigido por el Sr. Antonio Moro, en colaboración con la Unitat d'Antropologia del Departament de Biologia Animal, Biologia Vegetal i Ecologia de la Universitat Autònoma de Barcelona. Las excavaciones se iniciaron en noviembre de 1993 y finalizaron en 1995.

Desde el punto de vista antropológico, la excavación proporcionó una población diversa, perteneciente a ambos sexos y a todas las edades y quizás de una posición social acomodada como se desprende de los datos procedentes del archivo histórico. Se recuperaron alrededor de 200 esqueletos en enterramientos individuales, unos 50 cráneos y algunos otros restos óseos aislados en una zona que se podría identificar como osario. Aproximadamente un centenar de esqueletos fueron depositados en la UAB lo que equivale a la mitad del material recuperado. Debemos resaltar sin embargo que desde el inicio del proyecto ha sido posible disponer del material dentario que ha sido preciso, además del material que el doctorando extrajo del yacimiento en condiciones controladas para evitar la contaminación (uso de mascarilla y guantes quirúrgicos, etc.).

En cuanto a la época a la que pertenecen los restos estudiados, no existen datos muy concretos. La necrópolis era totalmente desconocida por los historiadores de la

Terrassa medieval y se halló durante los trabajos de documentación histórica de la zona que debía convertirse en un aparcamiento subterráneo. Las excavaciones se efectuaron en la Plaça Vella de Terrassa durante campañas sucesivas, desde 1993 a 1995. La sorpresa del descubrimiento abrió diferentes hipótesis sobre los enterramientos, hipótesis que fueron cambiando a medida que avanzaban las excavaciones, aumentaba el número de individuos inhumados y se buscaba en los archivos. Según los datos actuales parece ser que este cementerio está relacionado con la “Confraria del Santíssim Sagrament” de la ciudad de Terrassa, autorizada el 12 de Septiembre de 1551. Cardús (1984) recoge datos de los libros de testamentos con disposiciones “dels confreres i confrereses, ... tots els quals ... disposen ésser soterrats a l'interior de l'esglèsia” aunque “la majoria, però, acostumen a disposar que volen ésser soterrats en la *tomba o carner de la Sanctíssima Confraria del Cors Preciós de Jesucrist en la Sglésia del Sanct Sperit* (sic) (Cardús, 1981). Datos recientes han permitido saber que el 18 de agosto de 1629 “el consell aprova profanar lo sementiri vell de dvant la plassa y fer mudar los cadavers al sementiri nou de la Iglesia parrochial del Sanct Sperit y Sant Pere de la present vila”. Estas disposiciones situarían la datación del cementerio entre los s. XVI y XVII, entre la fecha de autorización de la cofradía y la remodelación posterior del espacio donde se sitúa el cementerio. Es necesario comentar que la relación del cementerio con esta cofradía es una hipótesis que cobra fuerza, pero no se trata de una tesis definitiva (García et al. 1994); en todo caso los restos humanos tienen como mínimo esta antigüedad.

Población actual

Las muestras para la extracción de DNA actual, proceden de individuos de Terrassa, donantes del banco de sangre de la Mutua de Terrassa. Asimismo, se obtuvieron muestras de sangre de individuos de la ciudad de Barcelona, facilitadas por el banco de sangre de la ciudad. Por otra parte, también se analizaron muestras de bulbos capilares de individuos de diversas áreas del Vallès, estudiantes de la carrera de Biología de la UAB. Hasta donde fue posible se procuró que los individuos no estuviesen relacionados. No se investigó el sitio de nacimiento de los antecesores por vía materna, aunque se tienen registros de la muestra de la ciudad de Terrassa. Se debe considerar pues una muestra de población residente, no estrictamente autóctona.

Poblaciones control

Además de la población medieval de la Plaça Vella, se analizaron individuos procedentes de 11 colecciones arqueológicas:

- Iglesia de la Concepción (C). Población de los siglos XVI-XVII d.C. Sta. Cruz de Tenerife, Canarias, España.

- Ermita de la Soledad (ELS). S.~ XVII d.C. Huelva, España.
- Can Solà del Racó II (MCS). Población Medieval, ~S. IX d.C.

Matadepera, Barcelona, España.

- Can Bosch de Basea (CBB). Tardo-Romanos, S. V-VI d.C. Terrassa, España.

- Calle de la Era (CE). Población de los siglos IV-VI d.C. Murcia, España.
- La Molineta, Puerto de Mazarron (PM). S. IV-VI d.C. Murcia, España.
- Mas Rimbau/Mas Mallol (MR). Población Romana, S. III-VII d.C.

Tarragona, España.

- Can Solà del Racó I (MCS). Población Tardo – Romana, ~S. III-VII d.C.

Matadepera, Barcelona, España.

- Monte Albán (MA). Población precolombina. Oaxaca, México.
- s'Illot des Porros (IP). Población de la edad del Bronce, S. VI-II a.C.

Alcudia, Mallorca, España.

- Son Real (SR). Población de la edad del Bronce, S. VI-II a.C. Alcudia, Mallorca, España.

- Paja Sola (PS). Población Calcolítica aproximadamente del 2300 a. C. Cehegin, Murcia, España.

MÉTODO: DNA ANTIGUO

Muestreo

Los individuos analizados fueron elegidos al azar de entre la colección. Para la extracción del DNA, se utilizaron principalmente piezas dentales debido a que el DNA de los dientes suele presentar mejores condiciones que el recuperado de restos óseos (Pag. 47 y 64). Al seleccionar los dientes se evitó escoger dientes con patologías y fueron preferidos los que mostraban un desarrollo completo, salvo en algunas excepciones como en el caso de individuos infantiles. Las piezas fueron extraídas utilizando guantes quirúrgicos. En muy pocos casos fue necesario utilizar agujas de disección o pinzas para poder extraer el diente.

Una vez extraído, cada diente fue colocado en una bolsa nueva de poliuretano debidamente etiquetada indicando también el tipo de pieza. Para la clasificación se utilizaron 4 categorías: Incisivo, Canino, Premolar y Molar, sin distinguir si eran superiores o inferiores ni el número de pieza. En los casos en los que se utilizó hueso para la extracción del DNA, se buscaron piezas sin patologías evidentes y que presentaran un buen estado de conservación macroscópico. Al igual que los dientes, los huesos fueron guardados en una bolsa de plástico con la etiqueta correspondiente.

Algunos huesos y dientes fueron obtenidos por el autor de la tesis durante la excavación de la necrópolis de la Plaça Vella, utilizando guantes y mascarilla quirúrgica. Sin embargo, las muestras así obtenidas son muy pocas, por lo que este procedimiento no puede ser considerado como método general. No obstante, se indicará cuando una muestra haya sido obtenida *in situ*.

Control de la contaminación

Con el fin de eliminar y/o evitar la contaminación procedente tanto de fuentes naturales como de amplificaciones previas (*carryover*), durante el desarrollo experimental se siguieron diversos procedimientos que serán descritos a continuación.

Separación del trabajo: pre y post-PCR

La extracción del DNA y la preparación de la mezcla de reacción del PCR, se realizaron en una campana de seguridad biológica (Telstar), ubicada en una habitación separada y en la que nunca se habían manipulado ácidos nucleicos. Todo el material de vidrio y el material fungible utilizado en estos pasos fue de uso exclusivo y fue guardado en un compartimento separado. El equipo pre-PCR incluye una caja de esterilización de luz UV, un juego de pipetas, gradilla para tubos de microcentrífuga y una bata de uso exclusivo. Todos los reactivos fueron guardados en sitios separados, incluyendo los que requerían temperaturas de -20°C .

La reacción de amplificación se realizó en un termociclador situado en una habitación aislada y todos los procedimientos posteriores (post-PCR), tales como análisis electroforético, purificación del DNA amplificado y digestiones con enzimas de restricción, se realizaron en una tercera habitación. Para los procedimientos post-PCR se utilizó otro juego de pipetas y material estéril de uso exclusivo. También se utilizó una bata para estos pasos que nunca fue utilizada en los procedimientos pre-PCR. De igual forma, los guantes utilizados en los pasos post-PCR nunca fueron utilizados para manipular material pre-PCR. Todos los tubos con amplicones fueron guardados a -20°C en un compartimento separado.

También se realizó una separación temporal, dedicando días específicos para la extracción, la amplificación y el análisis post-PCR.

Material de un solo uso

En la medida de lo posible, tanto para la preparación de las soluciones como para llevar a cabo las reacciones, se utilizó material de un solo uso. Así por ejemplo, se utilizaron tubos de polipropileno de 15 y 50 ml (Corning) esterilizados mediante rayos gamma. También se utilizaron puntas de pipeta con filtro resistente a aerosoles y esterilizadas de fábrica (Aeroshield y ART). Se utilizaron rutinariamente guantes de látex de distintas marcas, que fueron cambiados con frecuencia y mascarillas quirúrgicas cuando fue necesario manipular muestras o reactivos fuera de la campana de seguridad biológica.

Limpieza del posible DNA contaminante

Limpieza de las áreas de trabajo

Periódicamente, la campana de seguridad biológica fue limpiada secuencialmente con hipoclorito de sodio al 10%, agua ultrapura (Milli-Q) y etanol al 70%, para finalmente irradiarla con luz ultravioleta durante toda la noche. Con este procedimiento (exceptuando la irradiación UV) fueron limpiadas otras superficies de trabajo y los rotores de las centrifugas utilizadas.

Limpieza de las piezas dentales

La extrema dureza que presenta el esmalte de las piezas dentales permite realizar un riguroso tratamiento con el fin de eliminar la posible contaminación superficial que pueda existir en ellas, debida a la manipulación por parte de los arqueólogos durante la excavación, o de los antropólogos durante el estudio morfológico.

El procedimiento es el siguiente:

- Sumergir la pieza en ácido acético al 30 % y agitar vigorosamente para eliminar todos los residuos de tierra. Este paso no es necesario si el diente ya esta limpio.
- Enjuagar con agua ultrapura (Milli-Q).
- Sumergir la pieza en hipoclorito de sodio al 10%, durante al menos 10 min., agitando de vez en cuando.
- Exponer el diente a luz ultravioleta (254nm), a 2 cm de la fuente de luz, durante al menos 10 min. por cara.

Es recomendable utilizar para la limpieza tubos estériles desechables de polipropileno de 15ml, aunque se puede utilizar cualquier otro recipiente de plástico desechable, siempre y cuando se utilice uno nuevo cada vez.

Para la exposición a UV, es recomendable poner la pieza dental dentro del tapón (invertido) del tubo utilizado para la esterilización, de esta forma una vez concluido el proceso puede introducirse el diente en el tubo sin necesidad de tocar la pieza. Alternativamente, en el caso de utilizar otro tipo de recipiente desechable, es recomendable taparlo con un trozo de papel de aluminio, procurando limpiarlo con las distintas soluciones con las que se limpia el diente. Posteriormente, este trozo de papel puede servir como soporte para la exposición UV y una vez concluida ésta, el papel de aluminio puede servir de envoltorio para transportar el diente hasta la campana de seguridad biológica.

Limpieza del material de dentista

Para la extracción del tejido interno de las piezas dentales, se utilizó un equipo de dentista que dispone de diferentes fresas y sierras circulares con borde diamantado, adquiridas en la casa DEVALSA (Barcelona). Las fresas intercambiables y la sierra circular también son rigurosamente limpiadas utilizando el mismo procedimiento que se utiliza para la limpieza de los dientes.

En cada proceso de extracción se esteriliza una fresa por diente, no obstante se utiliza una sola sierra circular, por lo tanto, es necesario limpiarla después de usarla con cada pieza dental. Para esta limpieza, se utiliza un procedimiento diferente: se sumerge la sierra en hipoclorito de sodio al 10%, accionando el rotor para que la sierra dé algunas vueltas. Después se sumerge la sierra en agua Milli-Q accionando nuevamente el rotor y finalmente se flamea la sierra utilizando una lámpara de alcohol. Alternativamente puede exponerse la sierra a luz UV durante algunos minutos. Después de la limpieza, la sierra se guarda en un estuche, previamente esterilizado con luz UV, hasta la siguiente utilización.

Limpieza de los huesos

Para la limpieza de los huesos, se utiliza una fresa de cabeza redonda de un tamaño relativamente grande (037) para remover la capa superficial del hueso del área de la que se extraerá la muestra. El tejido expuesto es el que será utilizado para la extracción del DNA.

Esterilización del material de vidrio

Todo el material de vidrio no desechable utilizado fue esterilizado con el siguiente procedimiento:

- Utilizando guantes limpios, se lava el material rigurosamente con detergente no iónico
 - Se enjuaga con agua corriente
 - Se le añade una cantidad de hipoclorito de sodio al 10%, se enrosca el tapón o se tapa el recipiente con un trozo de papel de aluminio y se deja durante al menos 30 minutos, agitando de vez en cuando procurando que el hipoclorito de sodio alcance también la superficie interna del papel de aluminio o el tapón
 - Se enjuaga con agua Milli-Q dos o tres veces
 - Se recubre el recipiente con otro trozo de papel de aluminio y se autoclava a la máxima temperatura permitida por el autoclave
 - Se repite el autoclavado

En el caso de las pipetas de vidrio de 5ml, se esterilizan introduciéndolas en una probeta de vidrio de 500ml con la que se sigue el procedimiento descrito, con la

diferencia de que antes del autoclavado se les pone un filtro de algodón a las pipetas. Las pipetas Pasteur se esterilizan de la misma forma poniéndoles también un filtro de algodón, pero no son reutilizadas.

En lo sucesivo cuando se haga mención de un recipiente esterilizado, será en referencia a que ha sido sometido a todo el proceso descrito.

Otras precauciones

Todos los tubos de microcentrífuga fueron autoclavados por lo menos dos veces. Para abrir los tubos de PCR se utilizó un abridor de tubos, con el fin de evitar salpicaduras.

Preparación de las soluciones

El agua ultrapura para la preparación de las soluciones fue preparada autoclavando dos veces agua del Milli-Q (Millipore) en un recipiente esterilizado. Los reactivos para preparar las soluciones fueron comprados y utilizados exclusivamente para la extracción del DNA antiguo. Fueron pesados utilizando mascarilla quirúrgica y la cantidad necesaria fue vaciada poco a poco en la balanza directamente de los recipientes para no introducir ningún objeto en ellos. Una vez pesados, los reactivos fueron transferidos a recipientes estériles y el agua ultrapura fue añadida mediante decantación para no introducir pipetas en el recipiente del agua.

Todas las soluciones *stock* autoclavables fueron preparadas en recipientes de vidrio esterilizados, filtradas con un filtro MILLEX-GS (Millipore) de 0.22 μ m y transferidas a otro recipiente de vidrio esterilizado; para después autoclavarlas dos veces a la máxima temperatura. Las soluciones *stock* no autoclavables fueron preparadas en tubos estériles de polipropileno de 50ml, filtradas y transferidas a un nuevo tubo estéril. La solución de Proteinasa K fue preparada de forma similar, pero utilizando tubos de microcentrífuga de 1.5ml estériles. Para medir el pH de todas las soluciones, se utilizaron soluciones paralelas, en las que se media el pH. Estas soluciones eran desechadas o utilizadas para fines distintos al de la extracción de DNA antiguo. De esta forma, en las soluciones para DNA antiguo nunca se introdujo el electrodo del potenciómetro. Las soluciones de trabajo fueron preparadas a partir de las soluciones *stock* en tubos de 50ml e irradiadas con luz UV (siempre sin Proteinasa K) durante al menos 30 minutos.

Los reactivos para el PCR fueron comprados en la presentación de menor tamaño disponible, con el fin de evitar la preparación de alícuotas y el riesgo de contaminación que conlleva este proceso. Cuando esto no fue posible (como en el caso de los cebadores) las alícuotas fueron preparadas con extremo cuidado, siendo la primera actividad del día después de limpiar la campana de seguridad biológica. Además, para este propósito se abría una nueva caja de puntas de pipeta.

Controles positivos y negativos

El uso de controles positivos fue evitado en la medida de lo posible con el fin de disminuir el riesgo de contaminación y cuando fue necesario, se utilizó una dilución de baja concentración de un extracto de DNA del autor de la tesis. Tal DNA fue caracterizado, analizando los polimorfismos que se estudiaron en este trabajo, ya que al estar directamente relacionado con todos los experimentos, el autor de la tesis es potencialmente la principal fuente de contaminación natural, por lo que es importante poder identificar su DNA.

La mayoría de las extracciones fueron llevadas a cabo en grupos de 5 muestras, más un control negativo de extracción (blanco de extracción). Para la amplificación de los segmentos del D-loop, los seis extractos resultantes (5 muestras + 1 blanco) fueron sujetos a amplificación, añadiendo en el proceso un control negativo de reactivos de PCR (blanco de PCR). Cuando fue necesario realizar una reamplificación mediante una segunda ronda de PCR, el producto de la primera reacción de PCR del blanco de extracción fue reamplificado junto con las muestras correspondientes, para obtener un control de reamplificación del extracto (blanco de reamplificación del extracto). Las muestras para los segmentos de las zonas codificantes se amplificaron de diez en diez aproximadamente, añadiendo uno o dos blancos de PCR.

La presencia de una banda del tamaño esperado (determinada mediante electroforesis en gel de agarosa) en alguno de los blancos utilizados, fue considerada como evidencia de contaminación. En el caso de detectar contaminación en el blanco de extracción, todas las muestras extraídas en el mismo grupo fueron descartadas y se sustituyeron los reactivos de extracción. Si la contaminación fue detectada en el blanco de PCR, todas las muestras amplificadas en ese grupo fueron descartadas, se reemplazaron los reactivos del PCR y se repitió la reacción utilizando nuevas alícuotas de los mismos extractos.

Puesta a punto de la técnica de extracción de DNA

Obtención del tejido

Una vez esterilizado, el diente se sujeta con unas pinzas (previamente expuestas a luz UV) y se corta transversalmente con la sierra circular por la línea amelo-cementaria. Después, utilizando una fresa redonda de tamaño pequeño (008) se procede a “raspar” las paredes internas de la cavidad pulpar, permitiendo que el polvo generado caiga directamente en el tubo de polipropileno (15ml) que será utilizado para la extracción. Se colectan alrededor de 0.05g de este polvo.

En el caso de muestras de hueso, se obtiene el tejido de la zona previamente limpiada, utilizando una fresa más grande (037), permitiendo también que el polvo caiga directamente dentro del tubo estéril. De este material es posible obtener más cantidad, por lo que se colecta alrededor de 1 gramo.

Extracción del DNA

Como se ha explicado en el Marco Teórico (Pag. 83) no existe un protocolo estándar para DNA antiguo, por lo que para las muestras analizadas en este trabajo fue necesario optimizar un protocolo. Uno de los problemas para esta optimización es que las variables que influyen en la preservación del DNA no son conocidas del todo, lo que complica más la estandarización de un método, incluso para las mismas muestras. La estrategia seguida consistió en probar una serie de métodos diferentes, para seleccionar uno de ellos basándonos tanto en la eficiencia como en criterios tales como los costos, el tiempo requerido y sobre todo el número de pasos necesarios en cada método, pues mientras menos pasos tenga un protocolo, existe menos riesgo de contaminación.

Se utilizaron los protocolos siguientes:

UAB-1

Este protocolo fue diseñado en la Unidad de Antropología a partir de un protocolo de extracción de DNA de sangre total, utilizado en el Instituto de Biología Fundamental de la UAB.

Soluciones stock

Tris-HCl 2 M pH 7.5

Tris-HCl 20 mM pH 8.1

EDTA 0.25 M Ph 8

SDS al 10%

CINa 5 M

Cl₂Mg 1 M
Nonidet P-40 al 0.1 %
Proteinasa K en concentración de 10mg/ml

Lisis celular. Añadir a la muestra la solución de lisis (Tris-HCl 20mM; Cl₂Mg 10mM) hasta alcanzar un volumen de 12-13ml. Agitar cuidadosamente hasta la disolución de la mezcla y dejar a 4°C durante 30 minutos, agitando de vez en cuando. Centrifugar durante 15 min. a 3000rpm a 8°C y desechar el sobrenadante. Añadir Nonidet P-40 al 0.1% hasta 13ml, mezclando suavemente. Centrifugar durante 15min a 3500rpm a 8°C y desechar el sobrenadante.

Digestión de proteínas. Agitar el sedimento golpeando con los dedos o con vórtice y añadir 5ml de solución A (CINa 0.1 M; EDTA 0.025 M; SDS al 0.5%). Agitar con vórtice hasta que se observe un aspecto homogéneo y añadir 20µl/ml de la solución de Proteinasa K. Sellar el tapón con parafilm e incubar a 56°C durante una noche con agitación.

Extracción del DNA. Añadir la mezcla de fenol/cloroformo/isoamílico (25:24:1) hasta un volumen de 14-15ml. Agitar vigorosamente y centrifugar durante 10 min. a 8°C a 3000 rpm. Extraer la fase superior con una pipeta y pasarla a otro tubo, desechar la fase inferior. Repetir la operación. Añadir la mezcla de cloroformo/isoamílico (24:1) hasta un volumen de 13-14ml. Agitar vigorosamente y centrifugar 10 min. a 8°C a 3000rpm y pasar la fase superior a otro tubo.

Precipitación. Añadir CINa 5M a razón de 400µl por cada 5ml de solución. Agitar cuidadosamente y añadir etanol absoluto (previamente enfriado a -20) hasta un volumen final de 14ml. Agitar cuidadosamente y centrifugar a 4°C, 30 min. a 4300rpm. Decantar el sobrenadante con cuidado y resuspender el sedimento en 1ml de etanol al 70% (enfriado a -20) y pasarlo a un tubo tipo eppendorf. Agregar más etanol al 70% hasta un volumen de 1.5ml. Centrifugar 5 o 10 min. a 12-13000rpm y decantar el sobrenadante. Secar el DNA en una estufa a 40-60°C. Resuspender en 50-80µl de agua ultrapura. Almacenar a -20°C.

Nota: Las mezclas de fenol/cloroformo/alcohol isoamílico y de cloroformo/alcohol isoamílico usadas en este protocolo y en el siguiente fueron preparadas en el laboratorio, tamponadas con Tris 20 mM a pH 8.1.

CINVESTAV, CINVESTAV/PK y CINVESTAV-UAB

El protocolo CINVESTAV fue diseñado en el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (IPN) de México. No ha sido publicado y fue amablemente cedido por Víctor Macías durante una estancia en el IPN del autor de la tesis. Este es el único protocolo en el que se han probado variantes; como se describe a continuación, en la variante CINVESTAV/PK se incorpora el uso de proteinasa K y en el protocolo CINVESTAV-UAB, se utiliza un tampón de digestión diferente del original.

Soluciones stock

Tris-HCl 20 mM pH 8.1

EDTA 0.25 M pH 8

SDS al 10%

Proteinasa K en concentración de 10mg/ml.

Digestión de proteínas. Agregar a la muestra 5ml de buffer de extracción (Tris-HCl 10 mM; EDTA 0.1 M; SDS al 0.5%). Opcionalmente añadir 20µl/ml de la solución de Proteinasa K (CINVESTAV/PK). Sellar el tapón con parafilm e incubar con agitación 1 hora a 37°C y después 2 horas a 50°C.

Extracción del DNA. Agregar 5ml de fenol saturado (pH 8) y agitar 15 min. a temperatura ambiente. Centrifugar 10 min. a 23°C a 3000rpm y recuperar la fase acuosa. Agregar 5ml de fenol/cloroformo/alcohol isoamílico y centrifugar 10 min. a 8°C a 3000rpm.

Precipitación. Recuperar la fase acuosa y agregar etanol absoluto hasta 12.5ml y 0.5ml de acetato de sodio 3 M. Guardar a -20°C una noche ó de 1 a 2 horas a -80°C. Centrifugar 35 min. a 4°C a 4300rpm. Lavar el sedimento con etanol al 70% y centrifugar 5-10 min. a 13000rpm. Resuspender en agua ultrapura durante una noche a 4°C. Almacenar a -20°C.

Nota: En el protocolo CINVESTAV-UAB se utiliza la solución A (ClNa 0.1 M; EDTA 0.025 M pH 8; SDS al 0.5%) para la digestión y también se añaden 20µl/ml de Proteinasa K.

UAB-2

Este protocolo es una modificación del protocolo de Hagelberg y col. (1989; Hagelberg y Clegg, 1991), similar al utilizado por C. Lalueza de la Universidad de Barcelona (1995). Originalmente está basado en el protocolo descrito por Maniatis y col. (1982, citado por Hagelberg et al. 1989).

Soluciones stock

Tris-HCl 1 M pH 8.0-8.5

EDTA 0.5 M pH 8

SDS al 10%

Proteinasa K en concentración de 10mg/ml.

Digestión. Añadir a la muestra 5ml de tampón de extracción (Tris-HCl 50 mM pH 8; EDTA 0.425 M pH 8; SDS al 0.5%) y 50µl de Proteinasa K. Sellar el tapón con parafilm e incubar toda la noche a 37°C.

Extracción. Añadir 5ml de fenol saturado pH 8 (Sigma P4557) y agitar vigorosamente. Centrifugar 5 min. a 8°C a 2000rpm y recuperar la fase inferior con pipeta Pasteur estéril. Añadir 5ml de fenol/cloroformo/alcohol isoamílico (Sigma P2069), centrifugar 5 min. a 8°C a 2000rpm y recuperar la fase superior con pipeta Pasteur estéril. Añadir 5ml de cloroformo (Sigma C2432) y centrifugar 5 min. a 8°C a 2000rpm.

Concentración. Transferir 2ml del sobrenadante a un Centricon 30 y centrifugar hasta obtener la máxima concentración; repetir la operación hasta terminar el sobrenadante (máximo 3 veces). Agregar 2ml de agua ultrapura, concentrar y repetir la operación. Invertir el Centricon, centrifugar y recuperar la solución retenida por el filtro para transferirlo a un tubo de 1.5ml. Almacenar a 4°C durante al menos 3 días.

Nota: Es conveniente sellar las tres partes del Centricon con parafilm, para evitar contaminaciones en la centrífuga. De acuerdo con el fabricante, el Centricon 30 no debe ser usado a más de 5000 g para no dañar la estructura del tubo. Tampoco debe centrifugarse una carga completa (2ml) más de 5 veces, pues de lo contrario podría abrirse el poro de la membrana o sufrir daños mayores. También debe tenerse especial cuidado con el cloroformo, ya que la membrana es sensible a este compuesto.

En este protocolo no es posible precipitar el DNA con etanol, ya que de acuerdo con Sambrook y col. (1989, p. E-10), cuando se utilizan tampones que contienen concentraciones de EDTA mayores de 10 mM o mayores de 1 mM de fosfato, estos compuestos coprecipitan con el DNA, evitando su correcta recuperación. La elevada concentración de EDTA en el tampón de extracción de este protocolo (0.425 M) es la responsable de que en el primer paso de la extracción se inviertan las fases, quedando la fase acuosa como inferior.

Mini-Extracción (UAB)

Este protocolo fue diseñado en la UAB, con el fin de minimizar los costos de extracción. Se lleva a cabo utilizando tubos de microcentrífuga de 1.5ml, con lo que el gasto en reactivos también es menor. Se ha utilizado exclusivamente para muestras dentales.

Soluciones stock

Tris-HCl 1 M pH 8 - 8.5

EDTA 0.5 M pH 8

SDS al 10%

NaCl 1 M

DTT en concentración de 0.5gr/ml

Proteinasa K en concentración de 20mg/ml

Digestión. Colectar la muestra en un tubo de 1.5ml y añadir 500µl de tampón de extracción (Tris-HCl 10 mM; EDTA 2 mM; SDS al 0.5%; NaCl 10 mM), 10µl de DTT y 12.5µl de Proteinasa K. Cerrar el tubo, sellar con parafilm e incubar toda la noche a 37 – 55°C.

Extracción. Añadir 550µl de fenol saturado pH 8 (Sigma P4557), agitar vigorosamente y centrifugar 7 min. a 9000rpm. Recuperar la fase acuosa y agregar 550µl de fenol/cloroformo/alcohol isoamílico (Sigma P2069), agitar vigorosamente y centrifugar 5 min. a 9000rpm. Recuperar la fase acuosa y agregar 550µl de cloroformo (Sigma C2432). Agitar y centrifugar 5 min. a 6000rpm.

Concentración o precipitación. Transferir la fase acuosa a un Centricon 30, agregar agua ultrapura hasta 2ml y concentrar. Agregar 2ml más de agua y volver a concentrar. Repetir la operación, recuperar la solución retenida y almacenar a 4°C

durante al menos 3 días. Alternativamente puede precipitarse el DNA con etanol utilizando el procedimiento siguiente: añadir 40µl de NaCl 5 M y agitar con cuidado. Añadir etanol absoluto (previamente enfriado a -20°C) hasta 1.5ml, agitar y guardar a -20°C durante toda la noche o a -80°C durante al menos 1 hora. Centrifugar 25 min. a 13000rpm, decantar o extraer el sobrenadante con pipeta y agregar etanol al 70% (previamente enfriado a -20°C) hasta 1.5ml, centrifugar 10 min. a 13000rpm. Este paso puede repetirse dos veces. Decantar o extraer el sobrenadante con pipeta y secar el sedimento en estufa a 40-60°C. Resuspender el sedimento con agua ultrapura durante toda la noche (si es preciso agitar con vórtice). Almacenar a 4°C durante al menos 3 días.

Otros protocolos

Ocasionalmente fueron probados otros protocolos para analizar algún tipo de muestra en especial o como una aproximación a métodos basados en criterios distintos a los utilizados.

Persson (1992)

Este protocolo está basado en la afinidad que tiene la hidroxiapatita por el DNA. Se intenta recuperar el DNA que haya podido quedar unido al cristal después de liberarse del interior de la célula durante la autólisis post-mortem.

Soluciones stock.

Tris-HCl 1 M, pH 8

EDTA 0.2 M, pH 8

K₂HPO₄ 2.5 M, pH 7

NaCl 1 M.

Digestión. Agregar a la muestra 8ml de tampón de digestión (Tris-HCl 10 mM; EDTA 2 mM, NaCl 10 mM) y 4mg de colagenasa. Sellar el tapón con parafilm, agitar con vórtice e incubar toda la noche a 37°C con agitación.

Elución del DNA. Centrifugar a 3000rpm durante 10 min., descartar el sobrenadante y agregar 2ml de tampón de fosfato (K₂HPO₄ 2.5 M). Agitar con vórtice e incubar 15 min. a temperatura ambiente. Centrifugar a 3000rpm durante 10 min. y colectar el sobrenadante. Nota: se puede recuperar más DNA si se repite la operación.

Concentración y remoción de moléculas de bajo peso molecular. Diluir con agua el sobrenadante hasta 4ml. Transferirlo a un Centricon 30 (Millipore) y concentrar. Después de concentrar la solución hasta 50µl aproximadamente, diluirla con agua hasta 2ml y concentrarla de nuevo. Repetir esta operación dos veces para obtener una concentración final de fosfato de alrededor de 1mM (es muy importante para no inhibir la acción de la espermidina en el paso siguiente).

Purificación. Transferir la solución final de 50µl a un microtubo y precipitar el DNA añadiendo 0.08 volúmenes de espermidina trihidroclorada 1 M. Mezclar bien y

dejarla por lo menos 4 horas a 4°C. Centrifugar a 20,000g por 15 min. y descartar el sobrenadante. Después de la precipitación con espermidina, disolver el sedimento de DNA en una solución altamente salina (ej. tampón de PCR 10X) pero no en agua. Si se desea diluir el sedimento en agua, diluirlo en unos pocos microlitros de tampón y después diluirlo en agua hasta el volumen deseado.

Nota: si se diluye el DNA en tampón de PCR, cuando se prepara la mezcla de amplificación, hay que restar los microlitros de tampón que serán añadidos junto con el DNA.

Extracción de DNA de la Hidroxiapatita (EDH)

Este protocolo fue diseñado en la UAB con la idea de investigar si el DNA se conserva en las células nucleadas “momificadas” o libre unido a la hidroxiapatita. Nace de la observación de que en el protocolo de Persson, el tampón de digestión es desechado, mientras que en la mayoría de protocolos el DNA se extrae del tampón de digestión. Por otra parte, a excepción del protocolo de Hagelberg y col. (1989), todos los protocolos utilizan concentraciones de EDTA muy bajas que no son suficientes para descalcificar el hueso, por lo que asumen implícitamente que el DNA se encuentra todavía dentro de los restos celulares. El protocolo de Hagelberg utiliza una concentración de EDTA lo suficientemente alta como para permitir una descalcificación y en este sentido sería el más completo.

En el protocolo que se describe a continuación, basado en la cromatografía de DNA en columnas de hidroxiapatita (Sambrook et al. 1989, p. E-16 y E-30), se intenta separar el DNA conservado en diferentes sitios dentro del hueso para valorar cual es el de mejor calidad. Asimismo, este protocolo permite, en principio, separar el DNA de cadena sencilla (ssDNA) del DNA de doble cadena (dsDNA), con lo cual se pretende investigar la estabilidad del DNA monohebra en muestras antiguas. Se ha utilizado únicamente con muestras de hueso.

Soluciones stock

Fosfato de sodio 2 M pH 6.8
2-butanol (alcohol n-butílico)

Lavado. Agregar a la muestra 4ml de fosfato de sodio 0.01 M, agitar vigorosamente y desechar el sobrenadante por decantación. Repetir la operación 2 veces.

Elución del ssDNA. Añadir 3ml de fosfato de sodio 0.14 ó 0.16 M previamente calentado a 60°C. Incubar 5 min. a 60°C con agitación suave y centrifugar durante 5 min. a 1500rpm a 23°C. Recuperar el sobrenadante, repetir la operación dos veces pero utilizando sólo 2ml de tampón. Pasar a la extracción orgánica.

Elución del dsDNA. Añadir 3ml de fosfato de sodio 0.36 ó 0.40 M (a 60°C). Incubar 5 min. a 60°C con agitación suave y centrifugar durante 5 min. a 1500rpm a 23°C. Recuperar el sobrenadante, repetir la operación dos veces pero utilizando sólo 2ml de tampón. Pasar a la extracción orgánica.

Extracción orgánica. Dejar enfriar el sobrenadante a temperatura ambiente. Añadir un volumen de 2-butanol y agitar con vórtice. Centrifugar 5 min. a 1200g a 23°C, desechar la fase orgánica (superior). Repetir la extracción con 2-butanol hasta que la fase acuosa sea de 1.5-1.75ml (unas 3 veces). Añadir un volumen de cloroformo (Sigma C2432), agitar y centrifugar 5 min. a 1500rpm.

Concentración. Transferir la fase acuosa a un Centricon 30, añadir agua ultrapura hasta 2ml y concentrar. Añadir 2ml más de agua y concentrar. Repetir esta operación dos veces más para reducir la concentración final de fosfato a menos de 0.2 mM. Almacenar a 4°C.

Nota: después de la elución puede guardarse el sedimento para utilizarlo en una segunda extracción con protocolos basados en fenol-cloroformo.

Extracción mediante Formamida

Este protocolo está basado en el publicado por Panaccio y col. (1993) diseñado para extraer y amplificar DNA en un sólo procedimiento a partir de tejidos duros. La idea es de que a partir de un trozo de dentina o piel momificada se realice la extracción y amplificación directa. También se utilizó con una muestra de bulbo capilar.

Extracción. En un tubo de PCR (0.5ml) se introduce 1 mm³ de tejido o el equivalente de diente molido. Se agregan 18µl de formamida al 100% (desionizada; BRL) y se incuba durante 30 segundos a 95°C y 30 segundos a 72°C, repitiendo durante 30 ciclos. Después se realiza la amplificación en el mismo tubo, en las condiciones expuestas en la página 189.

Pääbo y col. (1988)

Este protocolo fue utilizado para la extracción de DNA de muestras antiguas de pelo y está basado en el publicado por Pääbo y col. (1988).

Soluciones stock

Tris-HCl 20 mM pH 8

NaCl 1 M

EDTA 0.25 M pH 8

SDS al 10%

DTT en concentración de 0.5gr/ml

Colagenasa en concentración de 20mg/ml

Proteinasa K en concentración de 20ml/ml

Digestión. Poner el pelo en un tubo de 15ml y añadir 5-8ml de tampón de extracción (Tris-HCl 10 mM pH 8; EDTA 2 mM; NaCl 10 mM). Agitar con vórtice y agregar 0.2ml de colagenasa. Cerrar el tubo, sellarlo con parafilm e incubar 3 horas a

37°C con agitación. Agregar 880µl de SDS, 160µl de DTT y 296µl de Proteinasa K. Incubar toda la noche a 37°C.

Extracción. Añadir fenol saturado pH 8 (Sigma P4557) hasta 10-15ml y centrifugar 5 min. a 2000rpm. Recuperar el sobrenadante, pasarlo a otro tubo y añadir fenol/cloroformo/alcohol isoamílico (Sigma P2069) hasta 10-15ml. Centrifugar 5 min. a 2000rpm, recuperar el sobrenadante, pasarlo a otro tubo y añadir cloroformo (Sigma C2432) hasta 10-15ml. Centrifugar y recuperar la fase acuosa.

Concentración. Poner el 20% del extracto en un Centricon 30 y concentrar. Agregar 2ml de agua y concentrar. Repetir la operación. Recuperar la solución retenida y almacenar a 4°C.

Inhibidores de la reacción en cadena de la polimerasa

Superar la acción inhibitoria

Durante el transcurso de este trabajo se notó una mejora en la amplificación cuando se aplicaba un paso previo al PCR consistente en almacenar los extractos de DNA a 4°C durante al menos 3 días (Montiel et al. 1997). A partir de este momento se aplicó este tratamiento de manera rutinaria. Sin embargo, un número considerable de muestras no fue sometido a este paso, lo que permitió contrastar los resultados obtenidos a partir de las muestras tratadas, en relación con los resultados obtenidos de las muestras en las que no se aplicó este tratamiento (ver Resultados).

El incremento en la eficiencia del PCR puede deberse a que las bajas temperaturas actúen sobre la sustancia inhibitoria, favoreciendo su precipitación y la formación de micelas, o bien favoreciendo su adsorción en las paredes del tubo de polipropileno (Montiel et al. 1997). Para profundizar en el esclarecimiento de estas posibilidades, se han realizado una serie de experimentos que pueden arrojar luz en cuanto a la naturaleza de las sustancias inhibitorias y podrían explicar el porqué del beneficio observado con el almacenamiento a 4°C. Estos experimentos no pretenden ser exhaustivos, ya que dilucidar la naturaleza de la sustancia inhibitoria no constituye el principal objetivo de esta tesis.

Caracterización de los inhibidores: espectrofotometría

Los análisis para caracterizar las sustancias inhibitorias parten de la hipótesis de que el frío promueve la precipitación o la adhesión a las paredes del tubo de las sustancias inhibitorias, reduciendo en ambos casos su concentración en la solución del DNA. Si esto es cierto, un tubo que haya contenido un extracto de DNA antiguo conservará las sustancias inhibitorias incluso después de ser vaciado, si ha sido mantenido a 4°C. Para corroborar esta hipótesis es necesario rellenar el tubo con agua ultrapura y agitar con vórtice para remover las posibles sustancias residuales y a partir de esta solución realizar algún tipo de análisis que permita determinar si efectivamente hay sustancias que permanecen en el tubo y las características de estas sustancias. Por otra parte, si el tubo no es sometido a este almacenamiento en nevera, sería de esperar que no contuviera residuos, o en su defecto, que contuviera una menor proporción, después de remover el extracto.

Dado que es de esperar que la concentración de los residuos recuperados sea mínima, se ha elegido como método de análisis la espectrofotometría. Comparando los espectros de emisión y de absorción de luz que presenten los residuos con los que presentan sustancias conocidas, es posible determinar las similitudes y eventualmente identificar el inhibidor. Como se ha visto en el marco teórico, algunos candidatos para explicar la inhibición podrían ser los compuestos porfirínicos, los ácidos húmicos o los productos de la reacción Maillard (Pag. 84).

Con relación a los compuestos porfirínicos, se ha utilizado una solución de Mesoporfirina IX dihidroclorada como modelo comparativo. También se han utilizado tres soluciones de distinta concentración de ácidos húmicos y ácidos fúlvicos. Los ácidos húmicos y fúlvicos provienen de tres diferentes suelos ácidos formados bajo diferentes condiciones climatológicas y litológicas y bajo diferente vegetación (Alegret et al. 1988). Estos ácidos fueron extraídos siguiendo el método recomendado por la "International Humic Substances Society" (1983). Desafortunadamente, no ha sido posible obtener productos Maillard para la comparación, por lo que no hemos podido hacer un análisis de compuestos de esta naturaleza. No obstante, existe en la bibliografía cierta información útil para este fin.

Espectros de absorción de luz

Las mediciones de absorción de luz fueron hechas con un espectrofotómetro de barrido modelo 8452A de Hewlett Packard (EE.UU.), en el rango de 190 a 820nm. Se analizaron las siguientes muestras:

- Dos muestras de 20µl de extracto de DNA antiguo diluidas en 2ml de agua.
- Distintas soluciones de los residuos obtenidas después de extraer todo el extracto de DNA antiguo de los tubos, añadiendo 300µl de agua en cada uno y agitando con vórtice.
 - Tres soluciones de distinta concentración de ácidos fúlvicos y húmicos, preparadas diluyendo 0.005, 0.01 y 0.15 mg de ácido por mililitro de Tris-HCl (20 mM pH 8).
 - Una solución de Mesoporfirina de 5.4mg en 10ml de medio básico (Carbonato de sodio 1.68×10^{-5} M).

Espectros de emisión de luz

Para las mediciones de emisión, se utilizó un espectrofotómetro de fluorescencia modelo F-4500 de Hitachi (Japón), excitando a diferentes longitudes de onda para encontrar los picos de máxima emisión.

De acuerdo con la hipótesis de trabajo, después de mantener los tubos con los extractos de DNA a 4°C durante 3 días, las sustancias inhibitorias permanecerán en el tubo aún después de retirar la solución. Con la idea de contrastar esta hipótesis se analizaron tanto las soluciones de los residuos de los tubos después del almacenamiento a 4°C, como las soluciones de los residuos de los tubos que no fueron sometidos al frío.

Con fines comparativos también fueron analizadas las soluciones de ácidos húmicos y fúlvicos, así como la solución de Mesoporfirina.

Todas las mediciones fueron hechas en condiciones ambientales de temperatura ($22 \pm 2^\circ\text{C}$) y oxígeno.

Preparación de las reacciones de amplificación

La preparación de la reacción de amplificación es uno de los pasos más críticos en el estudio del DNA antiguo, pues es necesario manipular diversos reactivos y los propios extractos de DNA, con el consiguiente riesgo de contaminación tanto con DNA de fuentes externas como con DNA de las mismas muestras. Como se expone en el Marco Teórico (Pag. 117), para minimizar este riesgo es recomendable premezclar los reactivos para la amplificación, preparando cantidad suficiente para varias muestras. A esta mezcla se le denomina “mezcla maestra”. Preparar esta solución tiene la ventaja que reduce el número de pipeteos que se debe hacer en relación con el número que se haría preparando la solución para cada muestra por separado. Esto representa un menor riesgo de contaminación y a la vez el ahorro de soluciones que normalmente se desperdician debido a que se impregnan en las paredes externas de las puntas. El procedimiento es el siguiente:

- Encender la campana por lo menos media hora antes de su utilización con el fin de purgar el aire que contiene.
- Utilizando guantes, colocar todo el material necesario como puntas, pipetas, gradilla y reactivos en el interior de la campana.
- Poner una alícuota de aceite de parafina ligera (Serva) en un microtubo para exposición a luz UV (254nm) a 2 cm de la fuente de luz y en posición vertical.
- Colocar en una gradilla tubos estériles de 1.5ml a los que se les corta la tapa, con el fin de utilizarlos como soporte tanto para los tubos de PCR de 0.5ml como para los tubos de los reactivos y el tubo dónde se preparará la mezcla maestra.
- Cambiarse los guantes.
- En un tubo estéril de 1.5ml se prepara el mezcla maestra, añadiendo la cantidad previamente calculada de cada reactivo. El cálculo se realiza teniendo en cuenta el número de muestras, el blanco y cantidad para una muestra más, con el fin de compensar la pérdida que ocasiona el pipeteo. Se añade el agua ultrapura y después los reactivos siguientes: tampón de PCR, Cl_2Mg , dNTP's, cebadores y BSA. Para reducir el riesgo de contaminación, el aire de la pipeta es expulsado antes de colocarle la punta, de esta forma se evita transferir el aire del interior de la pipeta a las puntas o dentro de los reactivos.
- Después de cerrar el tubo, se agita unos segundos con vórtice a baja velocidad y se irradia con luz UV (254nm) colocando el tubo a 2 cm de la fuente de luz y en posición vertical durante 8 min. y 45 seg. Durante este tiempo se retira de la campana el material y los reactivos que ya no sean necesarios.
- Cambiarse los guantes.
- Se añade la *Taq* DNA polimerasa. Debido a que la *Taq* viene disuelta en glicerina, es necesario agitar el tubo para que la enzima no quede en el fondo del tubo. El problema es que la *Taq* puede ser desnaturalizada si la agitación es muy intensa. Por este motivo se agita con vórtice a una velocidad muy baja y durante no más de 12 segundos.
- Se reparte la mezcla maestra en cada tubo y los tubos se colocan en un bloque isotérmico enfriado previamente a $-20^{\circ}C$, con el fin de evitar que se produzcan reacciones inespecíficas cuando se añade el DNA. El bloque utilizado cuenta con unos adaptadores para colocar tubos de 0.5ml que al ser removibles pueden ser esterilizados antes de ser usados.

-
- Se añade en cada tubo el aceite de parafina ligera.
 - Se retira la gradilla y el material que no sea necesario.
 - Se añade el DNA de cada muestra y se traslada el bloque a la habitación del termociclador, se espera a que el bloque del termociclador alcance la temperatura de 94°C y entonces se colocan los tubos para iniciar la reacción, en las condiciones específicas para cada par de cebadores.

Secuencias de la región de control

Para la amplificación de los segmentos de la región de control, se utilizaron cuatro oligonucleótidos con los que, mediante distintas combinaciones, es posible variar el tamaño del segmento objetivo. Los cebadores fueron sintetizados y purificados mediante HPLC ó HPSS por la casa MWG (Alemania). Las secuencias de los *primers* se exponen a continuación. En la nomenclatura utilizada, una L significa que la secuencia corresponde a la cadena ligera, una H hace referencia a la cadena pesada y el número indica la posición del nucleótido en 3'; siguiendo la numeración de la secuencia de referencia de Cambridge (Anderson et al. 1981):

L-15996: 5'-CTCCACCATTAGCACCCAAAGC-3'
L-16209: 5'-CCCATGCTTACAAGCAAGT-3'
H-16255: 5'-CCTAGTGGGTGAGGGGTGGC-3'
H-16401: 5'-TGATTTCACGGAGGATGGTG-3'

Los cebadores L-15996 y H-16401 son los mismos que los utilizados por Vigilant y col. (1991); el *primer* L-16209 tiene un nucleótido más con respecto al utilizado por Horai y col. (1989) y la secuencia del *primer* H-16255 nos fue cedida por Paolo Francalacci. Las posibles combinaciones y el tamaño del fragmento objetivo, incluyendo *primers*, son los siguientes:

L-15996 – H-16401 → 446 pb
L-15996 – H-16255 → 300 pb
L- 16209 – H-16401 → 231 pb
L- 16209 – H-16255 → 85 pb

En una primera etapa exploratoria, se usaron todas las combinaciones, no obstante, no fue posible amplificar los fragmentos de 300 y 446pb, por lo que se decidió no utilizarlos rutinariamente sino únicamente en aquellas muestras en las que se obtuviera una buena amplificación con el segmento de 231pb. En esta primera etapa, el segmento de 85pb fue amplificado con bastante éxito, sin embargo, los fragmentos obtenidos presentaron dificultades inusuales en la secuenciación. Además, con este fragmento tan pequeño, pronto se reveló la dificultad de controlar la contaminación, resultado que está en acuerdo con la bibliografía ya comentada (Pag. 119). Por estos motivos, este fragmento tampoco fue amplificado en las muestras que constituyen el análisis formal.

Protocolo de amplificación

Durante la primera etapa exploratoria, se definieron también las condiciones de PCR para el segmento del D-loop, en cuanto a composición de la mezcla de reacción y el perfil de temperatura de los ciclos. Estas condiciones se han mantenido constantes para todas las muestras. La reacción estándar para la amplificación del segmento de 231 pb contenía 2 U de *Taq* DNA polimerasa, tampón de reacción 1X, 50 picomoles de cada

primer, 200 μ M de cada dNTP, Cl_2Mg 2 mM, 160 mg/ml de BSA y 1 μ l de extracto de DNA en un volumen final de 50 μ l cubiertos con 33 μ l de aceite de parafina ligera (Serva). En la Tabla MM-1 puede verse la composición de los distintos tampones de reacción 10X utilizados, de acuerdo con la casa comercial.

Enzima	Tampón 10X
Promega Taq	Tris-HCl 100 mM pH 9.0 a 25°C; KCl 500 mM; 1% Triton X-100
Promega Tth	Tris-HCl 100 mM pH 9.0 a 25°C; KCl 500 mM; 1% Triton X-100
Pharmacia Taq	Tris-HCl 100 mM pH 9.0 a temperatura ambiente; KCl 500 mM; Cl_2Mg 15mM
Finnzymes DyNAzyme	Tris-HCl 100 mM pH 8.8 a 25°C; KCl 500 mM; Cl_2Mg 15 mM; 1% Triton X-100
Bioline Taq	Tris-HCl 670 mM pH 8.8 a 25°C; $(NH_4)_2SO_4$ 160 mM; 0.1% Tween-20
Boehringer Taq	Tris-HCl 100 mM pH 8.3 a 25°C; KCl 500 mM
Roche Expand H-F	Tris-HCl 100 mM; 500 mM KCl (pH 8.3 a 20°C)

Tabla MM-1 Composición, a concentración 10X, de los tampones de PCR utilizados.

El programa del termociclador consistió de un paso inicial de desnaturalización de 8 min. a 94°C, seguido de 40 ciclos de 45 seg. a 94°C, 1 min. y 30 seg. a 60°C y 1 min. a 72°C. Finalmente se realizó un paso de extensión de 5 min. a 72°C.

Los productos de la amplificación fueron analizados mediante electroforesis en gel de agarosa (2%) y las muestras que presentaron una banda bien definida del tamaño esperado (Fig. MM-1), fueron reamplificadas. Las muestras en las que no se observó banda alguna no fueron utilizadas.



Figura MM-1 Electroforesis de la amplificación de 5 extractos (carriles 1-5) de DNA antiguo, el respectivo blanco de extracción (KE) y el blanco de PCR (K-); M, marcador de peso molecular (múltiplos de 123pb).

Muestras en Formamida

En el caso de la amplificación de las muestras extraídas con el protocolo basado en formamida, se utilizó la misma concentración de *primers*, dNTP's y BSA que para el resto de muestras, sin embargo, se agregaron 2 unidades de *Tth* DNA polimerasa (Promega), Cl_2Mg 1.75 mM y el tampón de reacción correspondiente en un volumen final de 100 μ l. Se utilizaron los 18 μ l de la muestra, ya que como se expone en el método de extracción, la amplificación se realizó en el mismo tubo. El perfil de los ciclos consistió de 5 minutos a 85°C seguidos de 30 ciclos de 30 segundos a 85°C; 30 segundos a 46°C y 30 segundos a 60°C.

Reamplificación

La reamplificación representa un proceso aún más delicado que la amplificación en lo referente al riesgo de contaminación, sobretodo si se tiene en cuenta que es un paso que está en la frontera entre los procedimientos pre-PCR y post-PCR. Para minimizar este riesgo, se decidió no purificar los productos de la primera reacción, utilizando directamente un microlitro para la segunda reacción. Para evitar el conflicto entre los pasos pre y post-PCR, se llevó a cabo el siguiente procedimiento:

- En la campana de pre-PCR se prepara la mezcla maestra siguiendo exactamente el mismo procedimiento que se sigue para un PCR normal.
- Una vez que los tubos situados en el bloque isotérmico contenían la mezcla de reacción y el aceite (y habiendo guardado todo el material y equipo) se trasladan las muestras a otra campana, en la que utilizando puntas y pipeta diferentes dedicadas para este fin, se añade el microlitro de la muestra a reamplificar. Como ya se ha mencionado, el producto de la amplificación del blanco de extracción también fue reamplificado junto con las muestras correspondientes.
- Los productos de la reamplificación se analizaron en gel de agarosa (2%) y las muestras que presentaron una banda intensa del tamaño esperado (Fig. MM-2) fueron purificadas para ser secuenciadas.

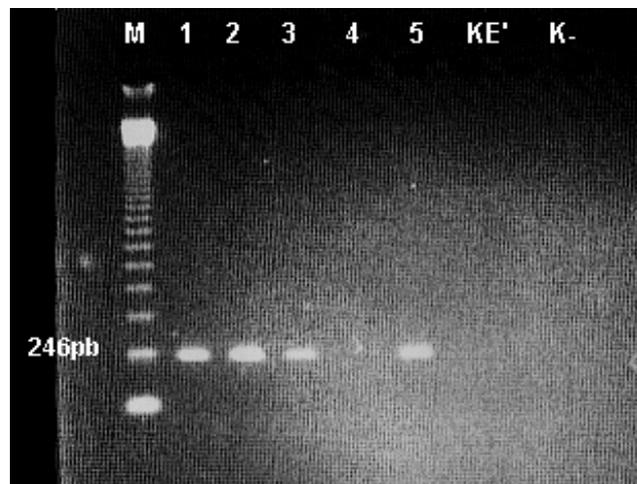


Figura MM-2 Electroforesis de la reamplificación de 5 muestras (carriles 1-5), el blanco de reamplificación del extracto (KE') y el blanco de PCR (K-); M, marcador de peso molecular (múltiplos de 123pb).

Purificación con gel de agarosa

La purificación de las muestras que fueron sometidas a la reacción de secuenciación, se realizó mediante gel de agarosa y extracción con fenol-cloroformo, siguiendo el procedimiento que se describe a continuación:

- Cargar 45µl de cada muestra en los pocillos de un gel de agarosa (2%) y realizar la electroforesis durante 1 hora y 30 minutos a 5 voltios/cm.
- Cubrir la superficie del transiluminador con un vidrio, cubierto a su vez, con un film de plástico
- Trasladar el gel al transiluminador y antes de exponerlo a luz UV cortar los carriles del gel. Después, con exposición a luz UV, cortar la banda de cada carril evitando exponer cada banda a la luz durante más de 10-12 seg. para no dañar el DNA (Orrego y King, 1990). Utilizar una hoja de bisturí estéril para cada muestra
- Colocar la banda en un microtubo de 0.5ml, al que previamente se le ha introducido lana de vidrio lavada (Panreac).
- Cortar el fondo del tubo con un bisturí estéril, para crear un orificio de aproximadamente un milímetro de diámetro. El tubo de 0.5ml se introduce dentro de otro tubo de 1.5ml
- Centrifugar el conjunto durante 10 min. a 13,000 rpm, para que la fuerza centrífuga haga pasar la banda de gel a través de la lana de vidrio. De esta forma la agarosa es retenida pero el tampón conteniendo el DNA pasa al tubo inferior. Dado que el rotor de la microcentrífuga utilizada es de ángulo fijo, cierta cantidad de material se acumula en una de las paredes, por lo que es necesario girar los tubos para que en la siguiente centrifugación, el material acumulado tenga la oportunidad de ser filtrado.
- Centrifugar 5 min. a 13,000 rpm.
- Desechar el tubo que contiene los restos de agarosa y la lana de vidrio y conservar el tubo de 1.5ml que contiene el DNA.
- Agregar fenol saturado pH 8 (Gibco-BRL) a razón de una vez el volumen y centrifugar 8 min. a 10,000 rpm.
- Recuperar la fase acuosa, pasarla a otro tubo y repetir la operación.
- Recuperar la fase acuosa y agregar un volumen de cloroformo (Merck) y centrifugar 5 min. a 10,000 rpm.
- Recuperar la fase acuosa, pasarla a otro tubo y agregar 2.5 volúmenes de etanol absoluto.
- Guardar a -80°C durante al menos 1 hora.
- Centrifugar 25 min. a 13,000 rpm. Decantar cuidadosamente y secar en estufa.
- Resuspender en agua ultrapura durante una noche a 4°C.
- Agitar intensamente con vórtice, centrifugar brevemente y almacenar a -20°C.

Secuenciación

La secuenciación fue llevada a cabo en el Departamento de Genética y Microbiología de la UAB, en cuyos laboratorios no se manipula DNA humano. Se utilizó un secuenciador automático Alf-express (Pharmacia) usando el *primer* H-16401. Las secuencias que mostraban más de dos sitios ambiguos, fueron descartadas. De esta forma se descarta la posibilidad de que las secuencias provengan de muestras con una muy baja cantidad de moléculas endógenas iniciales (Pag.122). Las secuencias fueron alineadas manualmente usando la secuencia de referencia de Cambridge (CRS, Anderson et al. 1981)

Inferencia del haplogrupo mitocondrial

Debido a que el enfoque pretendido en este trabajo es el análisis poblacional, se consideró de mayor importancia la obtención de secuencias de un mayor número de individuos diferentes, antes que la repetición de la secuenciación de las muestras obtenidas. Una vez obtenido un número considerable de secuencias se pensó en su autenticación. Para este fin, existían dos posibilidades, por un lado, intentar repetir el proceso con diferentes muestras de los mismos individuos para obtener secuencias independientes, o buscar algún método alternativo. Dada la relativa baja eficiencia del proceso de obtención de las secuencias, se optó por buscar algún otro método de autenticación.

Cómo se ha explicado en el marco de referencia, en el mtDNA existe una correlación entre determinadas substituciones de la secuencia de la región de control y ciertas substituciones en las regiones codificantes que permiten su clasificación dentro de un determinado haplogrupo. Así, la opción elegida fue la de caracterizar las muestras en función de los haplogrupos europeos definidos por Torroni y colaboradores (1996), para contrastar el resultado con el haplogrupo inferido en base a las secuencias. Para inferir el haplogrupo se utilizaron los datos de Macaulay y col. (1999a) y el criterio de Francalacci y col. (1999) (ver Marco de Referencia, Tabla MR-2).

Análisis de los polimorfismos de restricción

Sitios polimórficos

Para el análisis de restricción fueron seleccionados los sitios polimórficos: 1715 *Dde* I, 4577 *Nla* III, 7025 *Alu* I, 8994 *Hae* III, 9052 *Hae* II, 10394 *Dde* I, 12308 *Hinf* I, 13704 *Bst* 0I y 15606 *Alu* I. El número indica la posición de la diana de restricción siguiendo la numeración de Anderson y col. (1981). Estos sitios permiten definir 9 haplogrupos europeos (Torroni et al. 1996), de acuerdo con la Tabla MM-2.

Haplogrupos	Sitios polimórficos									
	1715 <i>Dde</i> I	4577 <i>Nla</i> III	7025 <i>Alu</i> I	8994 <i>Hae</i> III	9052 <i>Hae</i> II	10394 <i>Dde</i> I	10397 <i>Alu</i> I	12308 <i>Hinf</i> I	13704 <i>Bst</i> 0I	15606 <i>Alu</i> I
X	–	+	+	+	+	–	–	–	+	–
I	–	+	+	+	+	+	–	–	+	–
V	+	–	+	+	+	–	–	–	+	–
H	+	+	–	+	+	–	–	–	+	–
W	+	+	+	–	+	–	–	–	+	–
K	+	+	+	+	–	+	–	+	+	–
U	+	+	+	+	+	–	–	+	+	–
J	+	+	+	+	+	+	–	–	–	–
T	+	+	+	+	+	–	–	–	+	+
M	+	+	+	+	+	+	+	–	+	–

Tabla MM-2 Definición de los haplogrupos sobre la base de la presencia (+) o ausencia (–) de los sitios de restricción analizados.

Cebadores utilizados

Para analizar estos sitios se amplificaron 9 segmentos de las regiones codificantes, utilizando los siguientes *primers*:

1715 <i>Dde</i> I	L1637: 5'-ACAAAGCACCCAACTTACAC-3' H1764: 5'-ACTATATCTATTGCGCCAGG-3'
4577 <i>Nla</i> III	L4538: 5'-CACTCATCACAGCGCTAAGC-3' H4621: 5'-TGGCAGCTTCTGTGGAAC-3'
7025 <i>Alu</i> I	L6968: 5'-CCGTAGGTGGCCTGACTGGC-3' H7052: 5'-TGATGGCAAATACAGCTCCT-3'
8994 <i>Hae</i> III	L8927: 5'-TTCTTACCACAAGGCACACC-3' H9014: 5'-AGGTGGCCTGCAGTAATGT-3'

9052 Hae II	L8997: 5'-AACCAATAGCCCTGGCC-3' H9121: 5'-GCGATTTCTAGGATAGTCAG-3'
10394 Dde I	L10366: 5'-ATCCTAGCCCTAAGTCTGGC-3' H10420: 5'-GTCGAAATCATTGTTTTGT-3'
12308 Ninf I	L12237: 5'-CACAGAAGCTGCTAACTCATGC-3' H12309: 5'-ATTACTTTTATTTGGAGTTGCACCAAGATT-3'
13704 Bst 01	L13627: 5'-TCGAATAATTCTTCTCACCC-3' H13725: 5'-TAGTAATGAGAAATCCTGCG-3'
15606 Alu I	L15503: 5'-AGACCTCCTAGGCGACC-3' H15620: 5'-GATGGATAGTAATAGGGCAAG-3'

Las secuencias de los cebadores L4538, H4621, L8927, H9014, L12237 y H12309 fueron tomadas de Izagirre (1998). Las secuencias de los *primers* L6968, H7052, L10366 y H10420 fueron cedidas por Paolo Francalacci. Los cebadores restantes fueron diseñados utilizando el programa informático PrimerDesign 1.12b (Escrito por J. Napiwotzki y A. Becker, 1995, <http://www.chemie.uni-marburg.de/~becker>) siguiendo los criterios de composición, tamaño y temperatura de hibridación, expuestos en el Marco Teórico (Pag. 112). También se utilizó el programa PRIMER DESIGNER Version 1.01 (Scientific & Educational Software, USA, 1990) para visualizar los posibles dímeros de *primers*, lo que facilitó la elección final de los cebadores.

Las muestras que no fueron definidas analizando estos sitios, fueron escrutadas para el sitio polimórfico 10397 *AluI*, que define el haplogrupo M, utilizando el fragmento amplificado con los *primers* H10420 y L10366.

Condiciones de amplificación

La composición de la mezcla de reacción para la amplificación de los segmentos de las regiones codificantes fue la misma que se utilizó para el fragmento de la región de control. También el perfil de los ciclos para el PCR fue el mismo, con excepción de la temperatura de hibridación, ya que se utilizó la adecuada para cada pareja de cebadores; en todo caso, siempre se utilizó una temperatura de hibridación relativamente elevada. En la Tabla MM-3 se muestran las temperaturas de hibridación y el tamaño del fragmento objetivo correspondientes a cada juego de *primers*. Todas las muestras analizadas fueron sometidas a PCR para los 9 segmentos. La mezcla maestra se preparó según el procedimiento expuesto anteriormente. Las reamplificaciones fueron evitadas en la medida de lo posible.

Nombre del Fragmento	Sitio polimórfico	Cebadores	T° de hibridación	Tamaño del segmento ^a	Fragmentos de Restricción ^a	
					No polimórficos	Polimórficos y no polimórficos
X	1715 Dde I	L1637 - H1764	48°C	166	20/30/116	20/30/48/68
V	4577 Nla III	L4538 - H4621	51°C	120	---	62/58
H	7025 Alu I	L6968 - H7052	56°C	123	108/15	78/30/15
W	8994 Hae III	L8927 - H9014	51°C	125	119/6	88/31/6
K	9052 Hae II	L8997 - H9121	49°C	160	---	76/84
E	10394 Dde I	L10366 - H10420	56°C	93	10/83	10/38/45
U	12308 Hinf I	L12237 - H12309	53°C	123	---	93-96/30-27
J	13704 Bst 0I	L13627 - H13725	47°C	137	---	98/39
T	15606 Alu I	L15503 - 15620	49°C	154	---	121/33

Tabla MM-3 Temperaturas de hibridación de los cebadores utilizados, el tamaño del fragmento objetivo y el tamaño de los fragmentos de restricción generados, tanto por los cortes no polimórficos como por los polimórficos. El nombre del fragmento fue designado con el nombre del haplogrupo para el cual resulta de mayor relevancia (el fragmento X también es relevante para el I), con excepción del fragmento E, que es relevante para la determinación de distintos haplogrupos. ^a en pares de bases.

Reacciones de restricción

Cada fragmento amplificado fue sometido a restricción con la enzima correspondiente, utilizando de 5 a 10µl de muestra amplificada según la concentración del producto, determinada mediante electroforesis en gel de agarosa. Los productos de la amplificación no fueron purificados. La mezcla de reacción se preparó en un tubo estéril de 1.5ml añadiendo agua, el tampón de reacción requerido y el enzima, en cantidades suficientes para las muestras a procesar. En tubos de 0.5ml se añadió la muestra y la cantidad correspondiente de la mezcla de reacción. La incubación se llevó a cabo en un baño María termostático durante 1-3 horas. Algunas muestras fueron incubadas durante 10 horas con el fin de mejorar la digestión. Excepto para la enzima *Bst* 0I que fue incubada a 60°C, todas las digestiones se realizaron a 37°C. En la Tabla MM-3 se observa el tamaño de los fragmentos producidos por el corte de cada enzima en el fragmento correspondiente.

Las digestiones parciales fueron consideradas como restricciones positivas, ya que representan un fenómeno común en muestras de DNA antiguo (Izagirre, 1998; Kolman y Tuross, 2000).

Criterios de autenticidad de resultados

Para garantizar la autenticidad de los resultados no basta con la ausencia de amplificación en los diferentes blancos utilizados, ya que con este criterio no es posible descartar el efecto *carrier* (Pag. 91). Por este motivo es necesario utilizar otros medios de verificación que permitan obtener un cierto nivel de confianza, en el sentido que el DNA recuperado pertenezca en verdad al individuo en estudio y no a múltiples eventos de contaminación o a artefactos producidos por errores de la *Taq* DNA polimerasa. A continuación se exponen los criterios seguidos en este trabajo.

Extracciones múltiples

Como se ha expuesto en el marco teórico, uno de los criterios más recomendados es la extracción de DNA de diferentes muestras de un mismo individuo para corroborar que los resultados no sean producto de amplificaciones fortuitas. En la medida de lo posible en este trabajo se ha intentado realizar más de una extracción de los individuos analizados. En algunos casos esto no ha sido posible debido a falta de muestra.

Poblaciones “control”

Una medida que hasta ahora no ha sido recomendada expresamente en la bibliografía, es la utilización de poblaciones control. En este estudio se han utilizado individuos de poblaciones diferentes a la población analizada, con la idea de descartar la posibilidad de una contaminación generalizada a nivel de laboratorio. En caso de existir una contaminación de este tipo, los individuos procedentes de distintas poblaciones presentarían una alta incidencia de los mismos polimorfismos. En este sentido lo ideal sería la utilización de poblaciones con polimorfismos bien definidos y diferentes, como las de origen asiático o africano. Desafortunadamente en el laboratorio no contábamos con poblaciones de tales características y únicamente ha sido posible utilizar individuos de otras poblaciones españolas, de distintas épocas y de distintas localizaciones geográficas. No obstante, también se analizaron algunas muestras de individuos de una población precolombina.

Correspondencia secuencia – haplogrupo

La correlación establecida entre los polimorfismos de restricción y las sustituciones de la región de control (Pag. 135) puede ser muy útil en la autenticación de los resultados. De hecho, este criterio ya ha sido previamente utilizado en poblaciones antiguas americanas (Stone y Stoneking, 1998). Si una o dos sustituciones específicas son encontradas en una secuencia de la región de control, los correspondientes polimorfismos de restricción que definen el haplogrupo correlacionado deberían también

ser encontrados en la misma muestra, siempre y cuando la secuencia no sea un artefacto y la muestra esté libre de contaminación. Es muy poco probable que se encuentre esta correspondencia debido a errores de la *Taq* DNA polimerasa, ya que haría falta que la enzima se equivocara en dos o más sitios específicos de forma independiente. Más adelante se calcula esta probabilidad tomando como base la tasa de error de la *Taq*.

El segmento amplificado de la región de control para ser secuenciado, comprende 231pb. No obstante, la secuencia finalmente obtenida consta de un número de pares de bases menor, pues es necesario descontar 40pb correspondientes a las secuencias de los *primers* y algunos nucleótidos más, debido a que durante la secuenciación es normal perder definición en los primeros y en los últimos pares de bases leídos. Así, el segmento efectivo queda reducido aproximadamente a 160pb, con lo que algunas substituciones relevantes para la definición de un haplogrupo en base a la secuencia, no se encuentran dentro del segmento analizado, por lo que no puede ser diferenciado.

Cálculo de la probabilidad de error

Para valorar de forma cuantitativa la fiabilidad del criterio de autenticación basado en la correspondencia secuencia – haplogrupo, hemos calculado la probabilidad de encontrar esta correspondencia en una muestra debido a errores de la *Taq* ocurridos durante la amplificación. En primer término es necesario estimar la probabilidad de que habiendo un nucleótido distinto de N en la posición X de la secuencia original, obtengamos por error el nucleótido N en la secuenciación. Debemos considerar también la probabilidad de que se produzcan dos o más errores de forma independiente en los sitios relevantes para la correspondencia secuencia – haplogrupo. Tomando como ejemplo el haplogrupo T, realizaremos este cálculo basándonos en la tasa de error de la *Taq* DNA polimerasa, estimada por Eckert y Kunkel (1990) en 10^{-5} .

Dos de las características del mtDNA del haplogrupo T, son la presencia de una Timina en el nucleótido 16294 y la de un sitio de restricción para *Alu* I en la posición 15606, debido a la substitución 15607G. Así, para caracterizar por error un DNA mitocondrial como haplogrupo T, sería necesario que la *Taq* produjera artificialmente estos dos nucleótidos de forma independiente. La probabilidad de que en la posición 16294 no exista una T cuando obtenemos en la secuenciación una T puede expresarse como:

$$P\{X_R \neq T \mid X_L = T\} \tag{1}$$

donde X_R y X_L representan, respectivamente, el nucleótido real y el nucleótido leído en la posición X . Sabiendo que $P\{X_R \neq T \mid X_L = T\} + P\{X_R = T \mid X_L = T\} = 1$

por el teorema de Bayes

$$P\{X_R \neq T \mid X_L = T\} = \frac{P\{X_R \neq T \mid X_L = T\}}{P\{X_L = T\}} = \frac{P\{X_L = T \mid X_R \neq T\} P\{X_R \neq T\}}{P\{X_L = T\}}$$

$$= \frac{P\{X_L = T | X_R \neq T\} \cdot P\{X_R \neq T\}}{P\{X_L = T | X_R \neq T\} \cdot P\{X_R \neq T\} + P\{X_L = T | X_R = T\} \cdot P\{X_R = T\}} \quad (2)$$

donde $P\{X_L = T | X_R \neq T\}$ es la probabilidad de leer una T en la posición X cuando en la secuencia original existe un nucleótido diferente de T. Esta probabilidad es un subconjunto de $P\{X_L = N | X_R \neq N\}$, que es la probabilidad de leer un nucleótido N en la posición X cuando en realidad existe cualquier otro nucleótido diferente de N, es decir, la tasa de error de la *Taq*. De acuerdo con Eckert y Kunkel

$$P\{X_L = N | X_R \neq N\} = 10^{-5}$$

por lo tanto,

$$P\{X_L = T | X_R \neq T\} = \frac{1}{3} 10^{-5} \quad (3a)$$

$$P\{X_L \neq T | X_R = T\} = \frac{3}{4} 10^{-5} \quad (3b)$$

$$P\{X_L = T | X_R = T\} = 1 - \frac{3}{4} 10^{-5} \quad (3c)$$

ya que si un nucleótido es diferente de T, podría ser cualquiera de los otros 3 nucleótidos.

La probabilidad de que exista una T en una posición determinada ó $P\{X_R = T\}$, puede ser estimada a partir de la frecuencia de T que presenta la población en la posición en cuestión. De acuerdo a la base de datos de Handt y col. (1998), en población europea la frecuencia de T en la posición 16294 es de $143/1219 = 0.1173$, por lo que $P\{16294_R = T\} = 0.1173$ y $P\{16294_R \neq T\} = 1 - 0.1173$.

Substituyendo los términos en la ecuación (2) tenemos que habiendo un nucleótido distinto de T en la posición 16294, la probabilidad de obtener en la secuenciación un nucleótido T por error de la *Taq* es

$$P\{16294_R \neq T | 16294_L = T\} = 2.5 \times 10^{-4}.$$

Ahora bien, en el ejemplo del haplogrupo T también sería necesario que ocurriese un error en la posición 15607, en el que el nucleótido original fuese substituido por una Guanina dando origen a la diana de restricción para *Alu I*, para considerar la muestra como perteneciente al haplogrupo T. La probabilidad de que se produzcan dos errores de forma independiente en dos posiciones X e Y, puede expresarse como

$$P\{(Y_R \neq G | X_R \neq T) | (Y_L = G | X_L = T)\} \quad (4)$$

aplicando el teorema de Bayes,

$$= \frac{P\{(Y_R \neq G | X_R \neq T) | (Y_L = G | X_L = T)\}}{P\{Y_L = G | X_L = T\}} =$$

$$= \frac{P\{(Y_L = G | X_L = T) | (Y_R \neq G | X_R \neq T)\} \cdot P\{Y_R \neq G | X_R \neq T\}}{P\{Y_L = G | X_L = T\}} \quad (5)$$

El primer término del numerador en (5) es la probabilidad de leer G y T cuando en realidad existen nucleótidos distintos de G y T en las posiciones respectivas. Dado que los errores se producirían de forma independiente, pues la amplificación de los segmentos se realizan en reacciones distintas de PCR,

$$\begin{aligned} & P\{(Y_L = G | X_L = T) | (Y_R \neq G | X_R \neq T)\} = \\ & = P\{Y_L = G | (Y_R \neq G | X_R \neq T)\} \cdot P\{X_L = T | (X_R \neq T | Y_R \neq G)\} = \\ & = P\{Y_L = G | Y_R \neq G\} \cdot P\{X_L = T | X_R \neq T\}. \end{aligned}$$

Asumiendo que la correspondencia secuencia – haplogrupo se cumple en cerca del 100% de los casos tenemos que

$$\Phi(Y_R \neq G | X_R \neq T) = \Phi(Y_R \neq G) = \Phi(X_R \neq T)$$

por lo que

$$P\{Y_L = G | Y_R \neq G\} \cdot P\{X_L = T | X_R \neq T\} = (P\{X_L = T | X_R \neq T\})^2 \quad (5a)$$

Asimismo, el segundo término del numerador en (5), que es la probabilidad de que en la posición Y no exista una G y que en la posición X no exista una T sería

$$P\{Y_R \neq G | X_R \neq T\} = P\{X_R \neq T\} \quad (5b)$$

Finalmente, el término en el denominador de (5)

$$\begin{aligned} & P\{Y_L = G | X_L = T\} = \\ & = P\{(Y_L = G | X_L = T) | X_R \neq T\} \cdot P\{X_R \neq T\} + \\ & + P\{(Y_L = G | X_L = T) | X_R = T\} \cdot P\{X_R = T\} = \\ & = P\{(Y_L = G | X_L = T) | (X_R \neq T | Y_R \neq G)\} \cdot P\{X_R \neq T\} + \\ & + P\{(Y_L = G | X_L = T) | (X_R = T | Y_R = G)\} \cdot P\{X_R = T\} \quad (5c) \end{aligned}$$

donde el término

$$P\{(Y_L = G | X_L = T) | (X_R = T | Y_R = G)\}$$

que es la probabilidad de leer G y T cuando en realidad existen G y T en las posiciones respectivas, es igual a

$$P\{Y_L = G | (X_R = T | Y_R = G)\} \cdot P\{X_L = T | (X_R = T | Y_R = G)\}$$

debido a que el acierto (o error) en la lectura de un sitio es independiente del acierto (o el error) en la lectura del otro. Y como asumimos que la correspondencia secuencia-haplogrupo es casi del 100%,

$$= P\{Y_L = G | Y_R = G\} \cdot P\{X_L = T | X_R = T\} = (P\{X_L = T | X_R = T\})^2$$

y como de acuerdo con (5a)

$$P\{(Y_L = G | X_L = T) | (X_R \neq T | Y_R \neq G)\} = (P\{X_L = T | X_R \neq T\})^2$$

(5c) puede escribirse

$$(P\{X_L = T | X_R \neq T\})^2 \cdot P\{X_R \neq T\} + (P\{X_L = T | X_R = T\})^2 \cdot P\{X_R = T\}.$$

Substituyendo (5a), (5b) y (5c) en (5), tenemos que

$$\begin{aligned} & P\{(Y_R \neq G | X_R \neq T) | (Y_L = G | X_L = T)\} = \\ & = \frac{(P\{X_L = T | X_R \neq T\})^2 \cdot P\{X_R \neq T\}}{(P\{X_L = T | X_R \neq T\})^2 \cdot P\{X_R \neq T\} + (P\{X_L = T | X_R = T\})^2 \cdot P\{X_R = T\}} \end{aligned}$$

De acuerdo a (3a) y (3c) y asumiendo que, por la correspondencia secuencia – haplogrupo, la frecuencia de T en la posición 16294 (0.1173) es igual a la frecuencia de G en la posición 15607,

$$\begin{aligned} & P\{(15607_R \neq G | 16294_R \neq T) | (15607_L = G | 16294_L = T)\} = \\ & = \frac{\left(\frac{1}{3}10^{-5}\right)^2 \cdot (1 - 0.1173)}{\left(\frac{1}{3}10^{-5}\right)^2 \cdot (1 - 0.1173) + \left(1 - \frac{3}{4}10^{-5}\right)^2 \cdot 0.1173} = 8.36 \times 10^{-11} \end{aligned}$$

lo que demuestra que la probabilidad de caracterizar erróneamente una muestra como un determinado haplogrupo utilizando ampliificaciones independientes, es en realidad ínfima.

Cabe hacer notar que no sólo se utilizan dos ampliificaciones, ya que se amplifican los 9 segmentos descritos anteriormente para cada muestra y los resultados deben ser completamente concordantes, con lo que la posibilidad de error es todavía más pequeña. Así por ejemplo, en el caso del haplogrupo T, las muestras también deben carecer de una diana de restricción para *Dde I* en el sitio 10394 debido a la presencia de una C en la posición 10400.

Variabilidad a nivel de la secuencia de la región de control

En una muestra aleatoria, libre de sesgo por consanguinidad, de una población normal, debe existir cierta variabilidad entre las secuencias de los individuos. Como se ha visto en el Marco de Referencia, las poblaciones europeas presentan una similitud inusual en cuanto a la variabilidad de la secuencia de la región de control. En este sentido un buen indicador de la fiabilidad de los resultados obtenidos, es comparar la variabilidad obtenida en la población de estudio con la variabilidad de otras poblaciones europeas. Por el momento no se han publicado datos de la variabilidad de la secuencia del D-loop en poblaciones antiguas europeas, no obstante, afortunadamente existe un gran número de poblaciones europeas actuales analizadas por diferentes investigadores que pueden servir de referencia. A nivel teórico es de esperar que la variabilidad producida por eventos de contaminación en relación al número de muestras analizadas, sea mucho más reducida que la variabilidad de una población natural.

Distribución de los haplogrupos

Las poblaciones europeas analizadas hasta el momento presentan algunas diferencias en la distribución de los haplogrupos, sin embargo, presentan también ciertas similitudes, sobre todo en relación con la frecuencia de algunos haplogrupos, que pueden ser “comunes” como el H, o “raros” como el X. Aunque es de esperar que la distribución de una muestra de una población europea no sea muy similar a la de otras poblaciones europeas, ciertamente debe compartir características básicas, como una elevada frecuencia de los haplogrupos comunes y una escasa presencia de los haplogrupos raros, sobretodo si se tiene en cuenta el origen común de las poblaciones europeas como lo indican los datos del mtDNA.

Desde luego se debe tener en cuenta que el tamaño de la población disponible, así como el tamaño de la muestra, son de fundamental importancia para la aplicación de este criterio, ya que con poblaciones grandes y un número de muestras importante es posible minimizar los efectos de factores como las relaciones familiares.

MÉTODO: DNA ACTUAL

Muestreo

Las muestras de sangre procedentes del banco de sangre de Terrasa fueron colectadas y enviadas en tubos de muestra sanguínea sin anticoagulantes, por lo que fueron procesadas inmediatamente. Las muestras de sangre procedentes del banco de sangre de Barcelona contenían heparina.

Las muestras de pelo fueron obtenidas de estudiantes de la UAB, extrayéndolas utilizando pinzas estériles, tanto de la cabeza como de las cejas. Los pelos fueron colocados en tubos de polipropileno de 1.5ml y procesados en el mismo día en que fueron colectados. Los pelos largos fueron cortados con unas tijeras hasta un centímetro aproximadamente, a partir de la raíz.

Extracción de DNA

Muestras sanguíneas

Para la extracción del DNA de las muestras de sangre se utilizó el siguiente protocolo basado en el que se utiliza en el Instituto de Biología Fundamental de la UAB:

Soluciones stock

Tris-HCl 2 M pH 7.5

Tris-HCl 20 mM pH 8.1

EDTA 0.25 M pH 8

SDS al 10%

ClNa 5 M

Cl₂Mg 1 M

Nonidet P-40 al 0.1 %

Proteinasa K en concentración de 10mg/ml

Limpieza. Lavar 5ml de sangre total con suero fisiológico (NaCl 0,9%), centrifugando 15 min. a 3000rpm; desechar el sobrenadante. Repetir 2 o 3 veces si el sobrenadante se ve muy sucio. Congelar el sedimento celular a -20°C hasta su procesamiento.

Lisis de los hematíes. Añadir al sedimento celular la solución de lisis de eritrocitos (Tris-HCl 20 mM; Cl₂Mg 10 mM) hasta alcanzar un volumen de 12-13ml. Agitar cuidadosamente hasta la disolución de la mezcla. Enfriar a 4°C durante 30 minutos, agitando de vez en cuando. Centrifugar durante 15 min. a 3000rpm a 8°C y desechar el sobrenadante.

Lisis de leucocitos. Añadir al sedimento Nonidet P-40 (0,1%) hasta 14ml y mezclar suavemente. Centrifugar durante 15 min. a 3500rpm a 8°C, desechar el sobrenadante.

Digestión de proteínas. Agitar el sedimento golpeando con los dedos o con vórtice. Añadir 5ml de solución A (ClNa 0.1 M; EDTA 0.025 M; SDS al 0.5%) y agitar nuevamente con vórtice hasta que se observe un aspecto homogéneo. Añadir 20 µl/ml de proteinasa K e incubar a 56°C durante una noche.

Extracción del DNA. Añadir la mezcla de fenol/cloroformo/isoamílico (25:24:1) hasta un volumen de 14-15ml. Agitar vigorosamente y centrifugar durante 10 minutos a 8°C, a 3000rpm. Recuperar la fase superior y repetir la operación. Añadir la mezcla de cloroformo/isoamílico (24:1) hasta un volumen de 13-14ml. Agitar vigorosamente y centrifugar 10 min. a 8°C a 3000rpm; recuperar la fase superior. Centrifugar 5 min. a 8°C, a 3000rpm y recuperar la fase superior. Añadir ClNa 5 M a razón de 400µl por cada 5ml de solución. Agitar cuidadosamente y añadir etanol absoluto (previamente enfriado a -20°C) hasta un volumen final de 14-15ml. Agitar cuidadosamente y centrifugar a 4°C durante 30 min. a 4300rpm. Decantar el sobrenadante con cuidado y resuspender el

sedimento en 1ml de etanol al 70% (enfriado a -20°C) y pasarlo a un tubo de 1.5ml. Agregar más etanol al 70% hasta un volumen de 1.5ml. Centrifugar 5 o 10 min. a 12-13000rpm y decantar el sobrenadante. Secar el DNA en una estufa a $40-60^{\circ}\text{C}$. Resuspender en 50-80 μl de agua destilada y almacenar a -20°C .

Muestras capilares

En el tubo de 1.5ml que contiene el bulbo capilar, se agregan 0.4 ml de tampón de PCR (Tris-HCl 10 mM pH 9.0; KCl 50 mM; Cl_2Mg 2.5 mM) y 1 μl de proteinasa K (de una solución de 10 mg/ml). Se incuba a 56°C durante una hora y después durante 10 minutos a 95°C para inactivar la enzima. Se centrifuga brevemente y se guarda la muestra a -20°C hasta su utilización.

Análisis del DNA

Amplificación

La amplificación del DNA extraído de sangre se realizó con la misma composición de tampón y perfil de ciclos que los utilizados para el DNA antiguo, tanto para el segmento del D-loop como para las zonas codificantes, salvo que se utilizaron de 1 a 3 µl de extracto de DNA. Sin embargo, para la amplificación del D-loop se utilizaron los cebadores L-15996 y H-16401 que permiten amplificar un segmento de 446 pb.

La amplificación de las muestras de DNA procedente de bulbos capilares, requirió una modificación para compensar el hecho de que el DNA queda disuelto en tampón de PCR, teniendo esto en cuenta cuando se prepara la “mezcla maestra”. Por otra parte, dado que la concentración del DNA en los extractos de bulbos capilares es muy baja, es necesario utilizar de 25 a 35µl de muestra. No obstante, las concentraciones finales del tampón de reacción, fueron las mismas que las utilizadas para las muestras de los extractos de DNA de sangre.

Secuenciación y análisis de restricción

Sólo algunas de las amplificaciones del segmento de 446 pb del D-loop de las muestras de DNA de sangre fueron secuenciadas, utilizando el *primer* H-16401.

Todas las muestras de sangre y de pelo fueron sometidas al análisis con enzimas de restricción. La estrategia seguida para las muestras de DNA moderno varió respecto a la de las muestras de DNA antiguo. Primero se definió para todas las muestras el estatus del sitio 7025 *Alu* I, con el fin de identificar las muestras que pertenecían al haplogrupo H, ya que es el de mayor frecuencia en población europea. Después se definió para todas las muestras el estatus del sitio 10394 *Dde* I, con el fin de separar las muestras en dos conjuntos de haplogrupos, ya que los haplogrupos X, V, H, W, U y T, carecen de esta diana de restricción, en tanto que los haplogrupos I, J y K, si la presentan (ver Tabla MM-2). De esta forma también se constató el resultado de las muestras del haplogrupo H, pues el haplogrupo H presenta la carencia de las dos dianas. Posteriormente, las muestras fueron discriminadas siguiendo el orden de las frecuencias de los haplogrupos, empezando por los más frecuentes de cada grupo. Así por ejemplo, dentro de las muestras que habían sido agrupadas como posibles I, J ó K, se determinó el estatus del sitio 13704 *Bst* 0I, pues el haplogrupo J carece de esta diana. Una vez identificadas las muestras del haplogrupo J, se procedió a identificar entre las muestras restantes, las que pertenecían al K, que carecen del sitio de restricción 9052 *Hae* II. Finalmente, se corroboró que las muestras que no eran ni J ni K pertenecieran al haplogrupo I analizando el sitio 1715 *Dde* I cuya ausencia, en combinación con la presencia de la diana 10394 *Dde* I, es la principal característica de este haplogrupo. De esta forma, los haplogrupos comunes fueron definidos por un mínimo de dos sitios de restricción, en

tanto que para los raros se incrementó el número de sitios analizados. En las muestras en las que se detectó algún resultado contradictorio se repitió el análisis y se analizaron todos los sitios.

Se analizaron los mismos sitios polimórficos descritos para el DNA antiguo, utilizando los mismos métodos.

ANÁLISIS ESTADÍSTICO

Valoración metodológica e inhibidores del PCR

Extracción y amplificación del DNA antiguo

Los resultados en cuanto a contaminación y eficiencia en la extracción y amplificación del DNA antiguo fueron analizados para evaluar las diferencias de acuerdo a distintas variables, utilizando para ello los paquetes estadísticos SPSS v.9.0 y SAK v.2.40 (escrito por G. E. Dallal). Para el SPSS, cada extracto fue introducido como un caso y las variables fueron factores como el protocolo de extracción, la polimerasa utilizada, la procedencia de la muestra etc. La variable dependiente fue el resultado de la amplificación, introducida de forma bivariada con los valores 0 y 1, para indicar si la amplificación fue negativa o positiva, respectivamente. Se utilizaron los procedimientos “tablas de contingencia” y “regresión logística”. El programa SAK está concebido para el análisis de datos resumidos, por lo que se utilizó el total de casos positivos y negativos para cada variable a comparar. Se utilizaron las opciones de tablas de contingencia y prueba exacta de Fisher para tablas 2x2.

Inhibidores del PCR

La diferencia de medias entre las intensidades de fluorescencia fue analizada mediante la prueba de t del programa SAK, utilizando las medias y las desviaciones estándar de cada uno de los grupos comparados. También se utilizó, para la comparación de medias, la prueba no paramétrica de Mann-Whitney mediante el programa SPSS.

ANÁLISIS ESTADÍSTICO Y FILOGENÉTICO

Haplogrupos

Tanto para la comparación poblacional como para la reconstrucción filogenética se utilizó la frecuencia de los haplogrupos de cada población. Los haplogrupos fueron considerados alelos de un mismo locus, bajo el supuesto de que no existe recombinación significativa en el modo hereditario del DNA mitocondrial y que los sitios que definen los haplogrupos son suficientemente estables, es decir, que presentan una tasa de mutación relativamente baja. La frecuencia se calculó dividiendo el número de individuos de cada haplogrupo entre el número total de individuos analizados.

Comparación poblacional

Para la comparación de los datos obtenidos de las poblaciones antigua (PV) y actual (CAT) de Cataluña se utilizó la distribución de los haplogrupos publicada para las siguientes muestras poblacionales: 56 individuos de la isla mediterránea de Córcega (COR), Francia y 96 individuos de la isla de la Cerdeña, Italia, incluyendo 51 de Galura (GALU), un área del norte de la isla relacionada histórica y lingüísticamente con Córcega y 45 de Barbagia (BAR), la zona montañosa central de la isla, aislada cultural y genéticamente (Francalacci et al. 1999).

Asimismo, se incluyeron 48 individuos no relacionados de diversas poblaciones del sur de la región Toscana (TOS) de Italia (Torrioni et al. 1996) y 86 sujetos no relacionados de Escandinavia, incluyendo 49 finlandeses (FIN) de distintas regiones de Finlandia y 37 suecos (SUE) estudiantes de medicina de la Universidad de Gotenburgo (Torrioni et al. 1996). Las abuelas maternas de los individuos de Córcega, Galura, Barbagia y Toscana nacieron en la misma zona que los individuos muestreados. La información resumida de las poblaciones utilizadas se presenta en la Tabla MM-4.

Localidad	Clave	N	Descripción	ANT	Referencia
Galicia ^a	GALI	92	Diversos puntos de Galicia	–	Salas et al. 1998
País Vasco ^a	VAS	45	Provincia de Guipúzcoa	+	Bertranpetit et al. 1995
Terrasa	P.VELLA	24	Población medieval Catalana	^b	Esta tesis
Terrasa/Barcelona	CAT	90	Población actual Catalana	–	Esta tesis
Barbagia	BAR	45	Centro de Cerdeña	+	Francalacci et al. 1999
Galura	GALU	51	Norte de Cerdeña	+	Francalacci et al. 1999
Córcega	COR	56	Corte y áreas aledañas	+	Francalacci et al. 1999
Toscana	TOS	48	Zona sur de la región	+	Torrioni et al. 1996
Finlandia	FIN	49	De diversas zonas	–	Torrioni et al. 1996
Suecia	SUE	37	Estudiantes Univ. Gotenburgo	–	Torrioni et al. 1996

Tabla MM-4 Poblaciones utilizadas para el análisis estadístico y filogenético mediante la frecuencia de los haplogrupos; ANT, indica si se conocen los antecedentes familiares por línea materna hasta la segunda generación (abuelas); ^a la información de los haplogrupos fue inferida a partir de las secuencias; ^b no se conoce la ascendencia de los individuos pero se asume que la muestra no está influida significativamente por la inmigración.

Además, se infirió la distribución de los haplogrupos de las siguientes muestras: 92 individuos no relacionados autóctonos de Galicia (GALI), España (Salas et al. 1998) y 45 individuos no relacionados de descendencia materna vasca de la provincia de Guipúzcoa, País Vasco (VAS), España (Bertranpetit et al. 1995). El haplogrupo fue definido a partir de las secuencias originales mediante el criterio de Francalacci y col. (1999) basado en la correspondencia secuencia – haplogrupo. Esta inferencia es tentativa y en ningún caso pretende ser definitiva (Tablas MM-5 y MM-6).

Haplotipo	N	Haplogrupo	Haplotipo	N	Haplogrupo
HAP 1	1	V	HAP 15	1	T
HAP 2	1	V	HAP 16	1	H
HAP 3	1	H	HAP 17	1	H
HAP 4	1	H	HAP 18	2	H
HAP 5	1	T	HAP 19	1	U5a
HAP 6	2	U?	HAP 20	4	U5a
HAP 7	1	U?	HAP 21	1	U5a
HAP 8	1	U1?	HAP 22	1	H
HAP 9	1	X	HAP 23	1	U5a
HAP 10	1	K	HAP 24	1	X
HAP 11	3	V	HAP 25	1	H
HAP 12	3	H	HAP 26	1	J1a1.1
HAP 13	2	H	HAP 27	1	J2a
HAP 14	9	H	TOTAL	45	

Tabla MM-5 Haplogrupo inferido de cada secuencia diferente (haplotipo) encontrada en los 45 individuos de población vasca del estudio de Bertranpetit y col. (1995). N= número de individuos con el mismo haplotipo.

Ind.	Hap.	Ind.	Hap.	Ind.	Hap.	Ind.	Hap.
GAL2	H	GAL48	J1a2.1	GAL121	? (JT?)	GAL43C	H
GAL4	H	GAL51	H	GAL124	J2	GAL46	H
GAL7	H	GAL60	V	GAL128	H	GAL141	H
GAL53	H	GAL111	V	GAL129	H	GAL143	H
GAL12	U6	GAL62	H	GAL133	I	GAL159	H
GAL13	U5a1	GAL64	H	GAL137	J	GAL160	H
GAL51C	U5a1	GAL76	J	GAL138	H	GAL117	H
GAL16	U5a	GAL126	J	GAL142	H	GAL134	H
GAL17	K	GAL78	H	GAL42	H	GAL136	H
GAL18	H	GAL81	X	GAL156	T2a	GAL140	H
GAL21	H	GAL85	T2a	GAL163	H	GAL6	H
GAL22	W	GAL86	H	GAL61	U?	GAL19	H
GAL23	H	GAL34	H	GAL104	U?	GAL20	H
GAL24	H	GAL102	H	GAL38	U?	GAL49	H
GAL27	U?	GAL105	H	GAL59	H	GAL57	H
GAL28	H	GAL108	H	GAL135	H	GAL63	H
GAL29	K	GAL109	H	GAL48C	V	GAL79	H
GAL77	K	GAL47	H	GAL42C	W	GAL87	H
GAL30	J	GAL113	U6	GAL40C	? (I,X,W)	GAL97	H
GAL40	J	GAL116	H	GAL44	U5a	GAL107	H
GAL103	J	GAL118	H	GAL39	H	GAL106	H
GAL31	H	GAL119	H	GAL37	U4	GAL112	H
GAL32	K	GAL120	H	GAL45C	H	GAL115	H

Tabla MM-6 Haplogrupo inferido de la secuencia de cada individuo de la muestra de 92 individuos de Galicia del estudio de Salas y col. (1998).

Diversidad genética

Para todas las poblaciones se calculó la diversidad genética, mediante la fórmula de Nei (1987, p. 180), utilizando el programa Arlequín v.1.1 (Schneider, et al. 1997). Este índice se define como la probabilidad de que dos haplotipos elegidos al azar en una muestra sean diferentes. Para nuestros datos implica la probabilidad de que dos individuos elegidos al azar presenten diferente haplogrupo mitocondrial.

Prueba exacta de diferenciación poblacional

Esta prueba es análoga a la prueba exacta de Fisher que se utiliza para tablas de contingencia 2x2, pero extendida para tablas de contingencia $r \times k$ (poblaciones x alelos). La hipótesis nula postula que la distribución alélica es idéntica entre las poblaciones. El

valor insesgado de P , fue calculado según el método de Raymond y Rousset (1995), basado en una cadena de Markov, utilizando el programa Arlequín v.1.1 para la comparación de todas las parejas posibles de poblaciones y el programa STRUC del paquete estadístico GENEPOP v.3.2 (Raymond y Rousset, 2000) para el análisis de grupos de más de 2 poblaciones. En ambos casos se utilizaron 5000 pasos de desmemorización y 200,000 iteraciones en la cadena de Markov (200 tandas de 1000 iteraciones en el caso del programa STRUC).

Reconstrucción filogenética

Para la reconstrucción filogenética, se calculó a partir de la frecuencia de los haplogrupos la distancia genética de la medida de la cuerda (Cavalli-Sforza y Edwards, 1967) y la distancia genética de Reynolds y col. (1983). Las matrices de distancias fueron utilizadas para construir árboles filogenéticos sin raíz utilizando el método de Fitch y Margoliash (1967) y con raíz utilizando el método UPGMA (Sneath y Sokal, 1973, p. 230).

Por otra parte, también se construyó un árbol sin raíz utilizando un método restringido de máxima verosimilitud (Felsenstein, 1981), que no requiere de la estimación previa de una distancia genética.

Todos los cálculos y reconstrucciones se realizaron con el paquete informático Phylip v.3.57c (Felsenstein, 1995). Los programas y opciones utilizadas se especifican para cada uno de los árboles obtenidos en la sección de resultados.

Secuencias

Al igual que con los datos de haplogrupos, se utilizó la información de las secuencias para realizar una comparación poblacional utilizando distintos índices de diversidad. En general, se consideró que cada secuencia diferente de mtDNA representa un haplotipo. También se realizó la reconstrucción filogenética, tanto intra como interpoblacional.

Poblaciones analizadas

De la población Catalana actual se obtuvo un segmento de 391pb, de la posición 16001 a la 16391 de 25 individuos. Para este segmento únicamente fue posible utilizar una población para la comparación, se trata de una muestra de 100 individuos no relacionados de raza blanca de 6 diferentes regiones de Inglaterra y País de Gales (Piercy et al. 1993).

Reduciendo el análisis a un segmento de 360pb (16024-16383) se utilizaron, además de la anterior, las siguientes muestras poblacionales: 45 individuos de población Vasca (Bertranpetit et al. 1995); 48 individuos Toscanos (Torroni et al. 1996); 92 individuos de Galicia (Salas et al. 1998); una submuestra de 27 individuos de Galura (Francalacci et al. 1999) de los utilizados para el análisis de la frecuencia de haplogrupos; 69 secuencias de Cerdeña (Italia), obtenidas de mujeres muestreadas en hospitales de maternidad de Cagliari, Nuoro, Oristano, Ozieri y Sassari (Di Rienzo y Wilson, 1991); 47 individuos de Finlandia de lengua finlandesa (Sajantila et al. 1995); 45 individuos de villas rurales de la península de Anatolia, Turquía (Comas et al. 1996); una muestra heterogénea de 22 individuos asiáticos y una muestra de 22 individuos Hadza, de Tanzania (Vigilant et al. 1991).

Como ya se ha mencionado, la población de Galura está situada en la isla de la Cerdeña (ver Pag. 212), no obstante para este trabajo se considera esta muestra como una

unidad distinta de la muestra de población Sarda de DiRienzo y Wilson (1991), ya que esta última es una muestra heterogénea de toda la isla.

Con el fin de investigar la posible influencia de poblaciones del norte de Africa en la península ibérica y concretamente en la población Catalana, se incluyó también una muestra de 85 beréberes Mozabites de Ghardia, norte de Argelia (Côte-Real et al. 1996), para lo cual fue necesario reducir el análisis a un fragmento de 302pb (16069-16370). De la población antigua de Cataluña (Plaça Vella), se obtuvo un segmento de 153pb (16210-16362), así que el análisis se realizó utilizando este segmento de todas las poblaciones anteriormente descritas.

Así, para el segmento de 391pb se analizaron dos poblaciones, para el de 360 fueron 11, para el de 302 se utilizaron 12 y para el de 153pb se analizaron un total de 13 poblaciones. La información de todas las poblaciones se resume en la Tabla MM-7.

Localidad	Clave	N	Descripción	ANT	Referencia
Argelia	ARG	85	Bereberes Mozabites de Ghardia	-	Côte-Real et al. 1996
Gran Bretaña	BRI	100	6 regiones de Inglaterra y País de Gales	-	Piercy et al. 1993
País Vasco	VAS	45	Provincia de Guipúzcoa	+	Bertranpetit et al. 1995
Galicia	GALI	92	Diversos puntos de Galicia	-	Salas et al. 1998
Terrasa/Barcelona	CAT	25	Población actual Catalana	-	Esta tesis
Terrasa	P.VELLA	24	Población medieval Catalana	^a	Esta tesis
Galura	GALU	27	Norte de Cerdeña	+	Francalacci et al. 1999
Cerdeña	SAR	69	De Cagliari, Nuoro, Oristano, Ozieri y Sassari	+	DiRienzo et al. 1991
Toscana	TOS	48	Zona sur de la región	+	Torroni et al. 1996
Finlandia	FIN	47	De lengua Finlandesa	-	Sajantila et al. 1995
Anatolia, Turquía	TRK	45	Villas rurales de la península de Anatolia	^b	Comas et al. 1996
Asia	ASI	22	Diversas localidades	-	Vigilant et al. 1991
Tanzania	HAD	17	Población Hadza	-	Vigilant et al. 1991

Tabla MM-7 Poblaciones utilizadas para el análisis estadístico y filogenético basado en las secuencias de la región de control; N, tamaño de muestra; ANT, indica si se conocen los antecedentes familiares por línea materna hasta la segunda generación (abuelas); ^a no se conoce la ascendencia de los individuos pero se asume que la muestra no está influida significativamente por la inmigración; ^b se conoce el lugar de nacimiento de las madres. La población de Galura, es una submuestra de la utilizada en el análisis de la distribución de haplogrupos.

Índices de diversidad

Con el fin de comparar las poblaciones en cuanto a su grado de diversidad, se calcularon diversos estimadores. Los índices fueron obtenidos para los distintos segmentos de todas las poblaciones, con el fin de observar las variaciones debidas al tamaño y la zona del segmento analizado.

Haplotipos diferentes y sitios variables

El porcentaje de haplotipos distintos se obtuvo en relación al número de individuos analizados y el porcentaje de sitios variables en relación a la longitud, en número de nucleótidos, del segmento analizado.

Diversidad genética

Se calculó el índice de diversidad genética de Nei (1987) definido como la probabilidad de que sean diferentes dos haplotipos de una muestra elegidos al azar. El cálculo se realizó utilizando el programa Arlequin 1.1.

Diversidad nucleotídica

Definido como la probabilidad de que sean diferentes dos nucleótidos homólogos elegidos al azar, es equivalente al índice de diversidad genética pero al nivel de nucleótido. Fue calculado para los distintos segmentos de todas las poblaciones dividiendo la media de las diferencias por parejas en la población entre el número de nucleótidos analizados (Tajima, 1983; Nei, 1987 p. 257). El número medio de las diferencias por parejas en la población fue ponderado con el método de Tamura-Nei (Tamura y Nei, 1993), utilizando la corrección de distribución gamma con parámetro $\alpha=0.47$ (ver Wakeley, 1993). Los cálculos se realizaron con el programa Arlequin 1.1.

Historia demográfica

Como se ha visto en el Marco de Referencia, la estructura genética de las poblaciones puede utilizarse para inferir su historia demográfica. Se utilizaron los dos procedimientos descritos a continuación.

Distribución de las diferencias por parejas

De acuerdo con Rogers y Harpending (1992), los episodios de crecimiento (o declive) poblacional dejan huellas características en la distribución de las diferencias entre pares de individuos. En esta distribución un episodio de crecimiento genera una ola que viaja hacia la derecha, recorriendo una unidad del eje horizontal cada $1/2u$ generaciones, donde u es la tasa de mutación. Las distribuciones empíricas pueden ajustarse a un modelo de expansión súbita desarrollado por Rogers y Harpending (1992). Se asume que una población inicial en equilibrio con una diferencia por parejas esperada $\theta = \theta_0$ aumenta o disminuye rápidamente a un nuevo tamaño en el que $\theta = \theta_1$ y que el inicio de este crecimiento ocurrió t unidades de tiempo mutacional antes del presente.

La distribución de las diferencias observadas por parejas de individuos se obtuvo mediante los programas Arlequin 1.1. y DnaSP 3.14 (Rozas y Rozas, 1999). Para evaluar la robustez de las distribuciones empíricas se utilizó un método de *bootstrap* (Efron, 1982). Muestreando los sitios con reemplazamiento se generaron 1000 conjuntos de datos, cada uno con el mismo número de secuencias que la muestra original. Después se obtuvo la distribución de las diferencias por parejas de las 1000 muestras y se derivaron los errores estándar. Este procedimiento se llevó a cabo con un programa escrito para tal fin por Mauro Santos del Departamento de Genética de la UAB.

Los parámetros del modelo de expansión súbita de Rogers y Harpending (1992) fueron estimados con los programas Arlequin 1.1 y DnaSP 3.14 mediante el método de los momentos (Rogers, 1995):

asumiendo que $\theta_1 \rightarrow \infty$

$$\hat{\theta}_0 = \sqrt{v - m}$$

$$\tau = m - \hat{\theta}_0$$

donde m y v son, respectivamente, la media y la varianza observadas de las diferencias por parejas. Estas estimaciones no pueden realizarse si la varianza es menor que la media, pero para fines prácticos puede considerarse que $\hat{\theta}_0 = 0$ si $v < m$ (por lo que $\tau = m$) y que $\tau = 0$ si $m < \hat{\theta}_0$ para ajustar los datos observados al modelo (Rogers, 1995).

La distribución de las frecuencias esperadas, según el modelo de expansión súbita, para los distintos tamaños de los segmentos analizados se obtuvo con el programa DnaSP 3.14. La bondad de ajuste se evaluó mediante una prueba de χ^2 con la hipótesis nula de que la distribución observada y la esperada son iguales (es lo que se esperara cuando existe un buen ajuste).

Por otra parte, de acuerdo con Slatkin y Hudson (1991) una población que ha estado creciendo exponencialmente durante un período de tiempo largo, muestra una distribución de las diferencias por parejas muy similar a la distribución de Poisson. Por este motivo, la distribución observada de las distintas poblaciones también se ajustó a la distribución de Poisson tomando la media de las diferencias por parejas como el parámetro λ de la distribución. El ajuste a esta distribución también fue evaluado con la prueba de χ^2 .

Para todas las poblaciones se calculó el índice de irregularidad (r) de Harpending (1994), con la fórmula:

$$r = \sum_{i=1}^{d+1} (x_i - x_{i-1})^2$$

donde d es el número máximo de diferencias observadas entre haplotipos y x_i es la frecuencia relativa observada del i -ésimo número de diferencias ($i = 0, 1, 2, \dots$). Este índice

es una medida de la irregularidad de la curva de la distribución de las diferencias por parejas y alcanza valores mayores en las poblaciones en equilibrio que en las poblaciones que han sufrido una expansión súbita.

Debido a los problemas que presenta la estimación de la tasa de evolución del mtDNA, (ver Marco Teórico), no es posible calcular adecuadamente el tiempo desde que ocurrió la expansión. Sin embargo, independientemente de si la tasa utilizada es errónea o no, el cálculo de este tiempo en las distintas poblaciones resulta útil para fines comparativos. Con esta idea se realizó este cálculo utilizando la tasa obtenida para el segmento hipervariable I (378 bases; posiciones 16024-16401) por Horai y col. (1995) de 2.575×10^{-6} por sitio por generación, asumiendo un tiempo generacional de 25 años. El tiempo en años desde la expansión, asumiendo generaciones de 25 años, fue obtenido con la siguiente ecuación:

$$t = \frac{25t}{2\mu l}$$

donde l es la longitud en nucleótidos del segmento analizado y μ es la tasa de mutación por sitio por generación (ver Rogers y Harpending, 1992).

Inferencia de máxima verosimilitud

La historia demográfica de máxima verosimilitud dentro de varias clases de modelos de expansión poblacional dadas la diferencia media por parejas y el número de sitios variables de la muestra, fue inferida con el método de Weiss y von Haeseler (1998).

Con este método se investiga una clase de modelo de expansión en el que una población en equilibrio comienza a crecer o a decrecer exponencialmente desde un cierto tiempo t_0 en el pasado hasta el tamaño actual de la población. Este modelo está definido por tres parámetros: N_0 es el parámetro de la población en equilibrio en el pasado (proporcional al tamaño que tenía la población); t_0 es el tiempo en el que el tamaño

poblacional comenzó a cambiar (medido en unidades de $1/\mu$); y λ define la razón entre el tamaño actual y el tamaño inicial de la población. Si $\lambda > 1$ la población se ha expandido, si $\lambda < 1$ la población ha reducido su tamaño. Si λ se iguala a 1, se obtiene el modelo básico de tamaño constante como un caso especial del modelo de expansión; en cuyo caso el parámetro λ queda sin especificar. Así, con este método es posible contrastar tres clases de historia poblacional.

Si se asume que se conoce el proceso de mutación (en este caso se asume el modelo de Tamura-Nei), la evolución de una muestra de secuencias estará completamente caracterizada por los tres parámetros λ , μ y ν . Para estimar estos tres parámetros se utiliza un método de verosimilitud, condicionando la verosimilitud del conjunto de parámetros a la diferencia media entre parejas de individuos k y al número de posiciones variables s que presenta la muestra. De esta forma se hace un uso implícito de la relación entre la media de las diferencias por parejas y el número de posiciones variables.

El conjunto de parámetros que maximice la verosimilitud $L(\lambda, \mu, \nu | k, s)$ define la historia poblacional más probable dentro de la clase de modelo analizado (crecimiento, reducción o equilibrio). Para jerarquizar la plausibilidad de un conjunto de parámetros (λ, μ, ν) se utiliza la razón de verosimilitud $L(\lambda, \mu, \nu | k, s) / L_A$, donde L_A es el valor máximo verosímil dentro de la clase de historia poblacional considerada. Si se analizan distintas clases de historias poblacionales se jerarquizan los conjuntos de parámetros con la razón $L(\lambda, \mu, \nu | k, s) / L_A^{pop}$ donde L_A^{pop} es el valor máximo verosímil de todas las clases de historias analizadas y corresponde a la historia demográfica más probable de la población en estudio.

Para determinar el valor de verosimilitud de cada conjunto de parámetros se utilizan simulaciones por ordenador. En cada simulación se genera una secuencia ancestral en la raíz de una genealogía que evoluciona bajo el modelo de Tamura-Nei (con heterogeneidad de la tasa de mutación). La construcción de la secuencia ancestral se basa en una composición de bases (porcentaje de cada base) determinada empíricamente a partir de un gran número de secuencias homólogas del tamaño del fragmento analizado provenientes de distintas poblaciones a nivel mundial (Weiss y von Haeseler, 1998).

El valor aproximado de la verosimilitud $L(_, _, _ | k, s)$ para un conjunto de datos real con media de diferencias por parejas k y s posiciones variables se basa en $j = 1, \dots, B$ simulaciones utilizando en cada una de ellas los mismos parámetros especificados $(_, _, _)$. Para cada simulación j se computa la media de las diferencias por parejas k_j y el número de posiciones variables s_j . El valor de verosimilitud del conjunto de parámetros se aproxima por

$$L(\theta, \tau, \rho | k, s) \approx \frac{1}{B} \sum_{j=1}^B I_{\delta}(j)$$

donde $I_{\delta}(j)$ es el indicador de la variable y B es el número de simulaciones del conjunto específico de parámetros. El indicador de la variable

$$I_{\delta}(j) = \begin{cases} 1, & \text{si } |k_j - k| \leq \delta \text{ y } s_j = s \\ 0, & \text{en cualquier otro caso} \end{cases}$$

donde δ es un número pequeño positivo que define un intervalo con k como centro. Así, el indicador de la variable es igual a 1 para simulaciones razonablemente cercanas a nuestra muestra según los estadísticos k y s observados en ella y el valor de verosimilitud corresponde a la fracción de simulaciones cuyo indicador es igual a 1. La elección de δ es un compromiso entre la eficiencia de las simulaciones y la precisión de la aproximación. De acuerdo con Weiss y von Haeseler (1998), un valor de $\delta = 0.2$ resulta útil cuando $B = 25000$.

Los valores de verosimilitud se determinan en un panel de distintas combinaciones de parámetros para obtener el valor aproximado máximo verosímil. Se analiza un panel por cada clase de historia poblacional y el conjunto de parámetros que muestra el valor más alto se considera como la historia poblacional más probable. Dado que el análisis requiere de un tiempo computacional que puede resultar excesivo, únicamente hemos realizado un análisis exhaustivo para la población actual de Cataluña mediante $B = 25000$ simulaciones. Otras poblaciones fueron analizadas, con carácter preliminar, mediante un número de simulaciones reducido para cada conjunto de parámetros del panel. Utilizando $B = 1000$, se efectúa un sondeo para observar el espacio probable de máxima verosimilitud para cada clase de historia poblacional. Las

simulaciones se realizaron con el programa IPHULA (facilitado amablemente por G. Weiss) utilizando las opciones predefinidas.

Análisis filogenético intrapoblacional

Las relaciones filogenéticas de los haplotipos dentro de las poblaciones fueron analizadas mediante dos procedimientos distintos:

Reconstrucción filogenética a partir de distancias

La distancia genética entre los haplotipos se obtuvo con el método de Tamura y Nei (1993), ya que es un modelo explícito de evolución molecular desarrollado a partir de datos de la región de control del mtDNA; más aún, Weiss y von Haeseler (1998) han demostrado que es el máximo verosímil para el segmento hipervariable I. Se utilizó la corrección de la distribución gamma de la tasa de mutación, utilizando el parámetro $\alpha=0.47$, pues es el valor estimado por Wakeley (1993) para el segmento hipervariable I. Los árboles filogenéticos fueron construidos sin raíz a partir de las matrices de distancias utilizando el método de Neighbor-Joining (Saitou y Nei, 1987). El cálculo de las distancias y la construcción de los árboles se realizó con el programa MEGA 1.02 (Kumar et al. 1993). Para el dibujo de los árboles se utilizó el programa TreeView (Page, 1996).

Redes medias reducidas

Las relaciones entre los haplotipos también fueron examinadas con el método de las redes medias reducidas de Bandelt y col. (1995). Las redes fueron calculadas y construidas con el programa Network 2.0b (Röhl, 1998), utilizando los mismos pesos que los utilizados por Richards y col. (1998) para los sitios inestables.

Análisis filogenético inter poblacional

Distancias genéticas

La distancia genética inter poblacional se obtuvo mediante la transformación de Reynolds y col. (1983) del estadístico F_{ST} (Slatkin, 1991) calculado para cada pareja de poblaciones¹.

De acuerdo con Slatkin (1991)

$$F_{ST} = \frac{f_0 - \bar{f}}{1 - \bar{f}} = \frac{\bar{t} - \bar{t}_0}{\bar{t}}.$$

donde f_0 es la probabilidad de identidad de dos genes (en nuestro caso secuencias) obtenidos de la misma población; \bar{f} es la probabilidad de identidad de dos secuencias obtenidas al azar de la colección de poblaciones consideradas; \bar{t}_0 es el promedio del tiempo de coalescencia de dos secuencias extraídas de la misma población y \bar{t} es el promedio del tiempo de coalescencia de dos secuencias extraídas de la colección de poblaciones. Como se conocen las secuencias, las t 's pueden estimarse directamente utilizando una distancia genética. En nuestro caso, la distancia entre los haplotipos fue estimada con el método de Tamura y Nei (1993) utilizando la corrección de la distribución gamma con el parámetro $a=0.47$ (Wakeley, 1993).

Con la transformación de Reynolds y col. (1983) se obtiene la distancia genética entre las poblaciones,

$$D = -\ln(1 - F_{ST}).$$

Todos los cálculos se realizaron con el programa Arlequin 1.1.

¹ Con la transformación de Slatkin (1995) se obtuvieron idénticos resultados que no son presentados

A partir de las matrices de distancias se construyeron los árboles filogenéticos sin raíz con el método de Neighbor-Joining (Saitou y Nei, 1987), utilizando el programa NEIGHBOR del paquete informático Phylip 3.5c.

Relaciones entre los haplotipos de las poblaciones

Otra aproximación para analizar las relaciones entre las poblaciones consiste en reunir los haplotipos de las distintas poblaciones y reconstruir un árbol o una red con esta combinación de haplotipos (Côrte-Real et al. 1996; Rando et al. 1998).

Aunque este análisis puede realizarse incluyendo indistintamente todos los haplotipos de las poblaciones consideradas, utilizar todos los haplotipos puede producir árboles o redes poco claras si se analizan muchas poblaciones debido al elevado número de haplotipos resultante. En nuestro caso hemos optado por separar los haplotipos de cada población de acuerdo al haplogrupo al que pertenecen, para poder analizar todas las secuencias de cada haplogrupo de todas las poblaciones estudiadas. Este método tiene la ventaja de que se evita en gran medida la confusión generada por sitios de tasa de mutación elevada que producen mutaciones paralelas entre los haplogrupos evitando su diferenciación (ver Resultados del análisis intrapoblacional). Por otra parte, las relaciones de las secuencias dentro de los distintos haplogrupos europeos están lo suficientemente estudiadas (Richards et al. 1998; Francalacci et al. 1999; Macaulay et al. 1999a) como para proveer un marco de referencia adecuado a la hora de interpretar los resultados de un análisis por haplogrupos.

Las secuencias fueron clasificadas dentro de los distintos haplogrupos, de acuerdo a la caracterización mediante enzimas de restricción cuando se conocía, y mediante las substituciones diagnósticas en las secuencias cuando no se conocía esta caracterización. La validez de la clasificación por medio de las substituciones puede contrastarse *a posteriori* observando la posición de los haplotipos dentro de la filogenia obtenida.

Los árboles filogenéticos se construyeron con el método de Neighbor-Joining utilizando la distancia de Tamura-Nei con la corrección de distribución gamma de la tasa de mutación. El parámetro a de la distribución fue calculado para cada haplogrupo con un método de máxima verosimilitud utilizando el programa PUZZLE © (1995-99, K. Strimmer y A. von Haeseler; Versión 4.0.2); excepto para el haplogrupo V, ya que el bajo número de secuencias diferentes encontradas en el conjunto de poblaciones no permitió realizar una estimación fiable. Para este haplogrupo se utilizó un valor de $a=0.47$.

Para la reconstrucción sólo se utilizaron haplotipos únicos, pero la información de su frecuencia fue incluida en el árbol, representada por un número entre paréntesis. En el nombre de los haplotipos se incluyó el haplogrupo (y subgrupo) al que pertenecen de acuerdo a la designación previa. Cuando se encontraron haplotipos idénticos en dos o más poblaciones, se les dio un nombre específico compuesto por una letra representativa de cada población, precedido por un asterisco. El código de una letra para cada población utilizado es el siguiente: A, ARG; B, BRI; C, CAT; F, FIN; I, GALI; U, GALU; S, SAR; T, TOS; K, TRK y V, VAS. Así por ejemplo, el haplotipo *BCSK-H (6) corresponde a una secuencia del haplogrupo H encontrada en las poblaciones BRI, CAT, SAR y TRK con una frecuencia total observada de 6. La contribución de cada población a la frecuencia total, se indica en el pie de figura del árbol correspondiente. Las muestras de Asia (ASI) y de la población africana Hadza (HAD), no fueron incluidas en este análisis.

Para el análisis del haplogrupo V se incluyeron las secuencias que presentaban la substitución 16298C de la muestra de 115 individuos de población Saami de Finlandia, Noruega y Suecia del estudio de Sajantila y col. (1995). Esta población presenta el porcentaje más alto de esta substitución (característica del haplogrupo V) y es una de las poblaciones en las que pudo haber surgido originalmente el haplogrupo V (Torrioni et al. 1998b). Se utilizaron las letras FSA, KSA (Finlandia), NSA (Noruega) y SSA (Suecia) para identificar los haplotipos de la población Saami. En el código de una letra, se utilizó la letra M para identificar esta población.