

Introducción

INTRODUCCIÓN

1. EL DNA MITOCONDRIAL

1.1. Organización genómica

El DNA mitocondrial (mtDNA) es un material genético circular cerrado de doble cadena que se localiza en el interior de las mitocondrias celulares. Este genoma, de aproximadamente 16500 pares de bases (p.b.), codifica una pequeña fracción de las proteínas mitocondriales. Las proteínas restantes del mtDNA son codificadas por el DNA nuclear (nDNA).

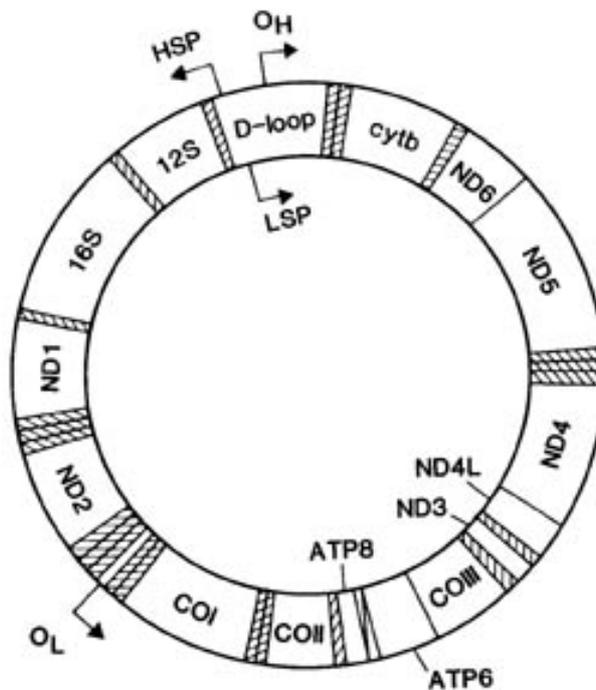
Las dos cadenas del mtDNA reciben el nombre de cadena L (ligera o *light*) y cadena H (pesada o *heavy*) atendiendo a su coeficiente de sedimentación. La mayor parte de las secuencias codificantes (28 genes) se encuentran en la cadena H. Estas secuencias se distribuyen en esta cadena de forma muy compactada, llegando incluso a solaparse. El mtDNA contiene información de 38 genes: 2rRNA (12S y 16S), 22tRNA y 13 genes estructurales, los cuales codifican diferentes subunidades de los complejos enzimáticos del sistema de fosforilación oxidativa:

- Tres subunidades de la citocromo c-oxidasa (COI, II y III) (Complejo IV).
- Una subunidad de la citocromo b (cyt b) óxido-reductasa (complejo III).
- Siete subunidades (ND-1, 2, 3, 4, 4L, 5 y 6) del complejo NADH deshidrogenasa (Complejo I).
- 2 subunidades (6 y 8) del complejo ATP sintetasa (Complejo V).

La región mayor no codificante, conocida como región control o *D-Loop*, ocupa 1122 pares de bases y se sitúa entre el gen que codifica para el RNA de transferencia de la prolina y el de la fenilalanina. En la región control se localizan el origen de replicación de la cadena H, los promotores para la transcripción de las dos hebras (P_L y P_H), dos lugares de unión de factores de transcripción (Tfam), tres bloques de secuencias conservadas asociados con el inicio de la replicación (CSBI, II y III) y las secuencias asociadas a la terminación de cadenas (TAS) (COSKUN *et al.* 2003;

FORAN *et al.* 1988). La región control destaca por su elevada tasa de mutación, aproximadamente 10 veces mayor que en las regiones codificantes (VIGILANT *et al.* 1991; GREENBERG *et al.* 1983) y por su elevada variabilidad intrapoblacional. Esta variabilidad se concentra en tres regiones o segmentos hipervariables: HVSI (posiciones 16024-16365), HVSII (posiciones 73-340) y HVSIII (posiciones 438-574) (LUTZ *et al.* 1998; VIGILANT *et al.* 1991). La más polimórfica es la HVSI (WAKELEY 1993) por lo que ha sido muy empleada por la Antropología, la Genética y la Medicina Forense (véanse apartados 1.3, 2.1 y 3.3.3 de este capítulo).

Figura F11. Organización genómica del mtDNA humano. OH y OL: Orígenes de replicación de las cadenas H y L respectivamente. LSP: Promotor de la cadena L. HSP: Promotor de la cadena H.



La secuencia completa del mtDNA humano –perteneciente a un individuo de origen europeo–, se publicó por primera vez en 1981 (ANDERSON *et al.* 1981). Desde entonces se emplea como la “secuencia consenso” (*CRS* o *Cambridge Reference Sequence*) con la que se comparan las diferencias encontradas en otros tipos mitocondriales. Un reanálisis reciente de esta secuencia ha revelado la existencia de 4 errores y de 7 polimorfismos poco comunes en la publicación original, ninguno de los cuales se encuentra en la HVSI (ANDREWS *et al.* 1999).

1.2. Características del mtDNA

Tres características convierten al mtDNA en una herramienta especialmente útil en la reconstrucción del pasado reciente de las poblaciones humanas:

- La poliplasmia.
- Su herencia materna.
- Su elevada tasa de cambio molecular.

1.2.1. Poliplasmia

Se designa así el elevado número de copias de mtDNA que existe en cada mitocondria y, por extensión, en la célula. En el interior de la mitocondria el DNA mitocondrial se encuentra unido a ciertas proteínas (*binding proteins*), formando un complejo denominado *nucleoide*. Una mitocondria puede contener entre 2 y 10 de esas estructuras, por lo que el número de copias de mtDNA por célula oscila entre 1000 y 10000 (MALYARCHUK *et al.* 2002, GILES *et al.* 1980). Tal particularidad hace muy recomendable el estudio de mtDNA en los casos en los que el material genético de partida es muy escaso o está muy degradado, como acostumbra a suceder en los restos antiguos.

1.2.2. Herencia matrilineal

El DNA mitocondrial se transmite de forma no mendeliana por línea materna. Aunque hombres y mujeres lo tienen, son únicamente las mujeres quienes lo transmiten a la descendencia. Esto se debe a que el citoplasma celular del cigoto, donde se localizan las mitocondrias, procede del óvulo. Los espermatozoides contienen un gran número de mitocondrias en la cola, pero raramente sobreviven a la primera división celular. De ahí se ha postulado que en el propio oocito hay un mecanismo que reconoce y elimina las mitocondrias de origen paterno (KANEDA *et al.* 1995; SUTOVSKY y SCHATTEN 1998; HOPKIN 1999). Otra explicación pudiera ser la diferencia en el número de moléculas de mtDNA de origen materno y paterno: mientras que un óvulo maduro puede llegar a contener entre 100000 y 200000 copias de mtDNA (MARCHINGTON *et al.* 1997), un espermatozoide sólo contiene unas 50 (ANKEL-SIMONS y CUMMINGS 1996). Por lo tanto, aún suponiendo que todas estas moléculas

penetraran en el oocito, su proporción respecto al mtDNA de origen materno sería ínfima.

En los últimos años se ha cuestionado, esporádicamente, la transmisión exclusivamente matrilineal del mtDNA. Algunos estudios de herencia del mtDNA en familias son contradictorios a este respecto. Mientras que PARSONS e IRWIN 2000, y TORRONI *et al.* 1998a, no hallan evidencias de herencia paterna en ninguno de los descendientes analizados, otros señalan evidencias de recombinación entre el DNA mitocondrial de origen materno y paterno (SCHWARTZ y VISSING 2002; HAGELBERG *et al.* 1999). Eyre-Walker Smith y John Maynard Smith sugieren que la elevada tasa de homoplasia en el mtDNA podría ser debida a la existencia de recombinación, más que a la presencia de “puntos calientes” o *hotspots* mutacionales (EYRE-WALKER *et al.* 1999a). Los autores establecieron una relación directa entre la distancia entre las posiciones del mtDNA y el desequilibrio de ligamiento (AWADAWA *et al.* 1999). Estos estudios han sido desmentidos y en algún caso se ha rectificado –como John Maynard Smith– (MACAULAY *et al.* 1999b; KIVISILD y VILLEMS 2000; JORDE y BAMSHAD 2000; KUMAR 2000). De hecho, ni el análisis de las bases de datos de mtDNA humano más extensas (INGMAN *et al.* 2000; ELSON *et al.* 2001) ni la aplicación de nuevos métodos para detectar la recombinación (WOROBEY 2001), consiguieron identificar ningún evento de recombinación potencial.

Erika Hagelberg (HAGELBERG 2003) insiste en que los eventos de recombinación proporcionan una respuesta satisfactoria a la homoplasia detectada en las reconstrucciones filogenéticas. Muchas de las posiciones catalogadas como sitios rápidos en las filogenias humanas actuales aparecen invariablemente en las tres secuencias completas de la HVSI de Neandertales publicadas (KRINGS *et al.* 1997, 2000; OVCHINNIKOV *et al.* 2000), lo que induce a pensar que se trata de mutaciones antiguas (GURVEN 2000). La recombinación explicaría que estas “posiciones antiguas” aparecieran hoy en diferentes líneas humanas no relacionadas (HAGELBERG 2003).

El debate continúa sin cerrarse ni superarse. De momento la evidencia de recombinación es bastante ocasional (HAGELBERG 2003; ELSON *et al.* 2001), pero si

se demostrara finalmente, la validez de las inferencias obtenidas de los estudios del mtDNA de las poblaciones humanas se vería seriamente comprometida.

1.2.3. Tasa de mutación

A pesar del importante papel que desempeña el genoma mitocondrial en la producción de energía celular, su tasa de evolución es mucho más rápida –hasta 10 veces– que la del genoma nuclear (BROWN 1980).

Ello se ha relacionado con las características de este material genético y con el entorno que lo rodea. El mtDNA no está protegido por proteínas tipo histonas como el nDNA y se ve continuamente expuesto a la acción de los radicales libres generados por el metabolismo oxidativo. Además, se ha demostrado que los sistemas genéticos de reparación del mtDNA son menos eficientes que los nucleares. Los tres factores bastarían, en principio, para explicar la elevada tasa de mutación del mtDNA, pero hay que añadirles, además, la elevada susceptibilidad de este material genético a los efectos de la deriva genética, mecanismo mediante el cual las mutaciones que experimenta tienen elevada probabilidad de fijación (WALLACE 1994).

El cálculo de la tasa de mutación del mtDNA puede realizarse por dos métodos: contando el número de mutaciones acumuladas entre dos o más especies o poblaciones relacionadas de las que se conoce la fecha de divergencia; o bien directamente a través del análisis de *pedigrees* familiares.

1.2.3.1. Métodos filogenéticos de estima de la tasa de mutación

El primero de los dos métodos recibe el nombre de *método filogenético* puesto que emplea reconstrucciones filogenéticas para calcular las diferencias acumuladas (divergencia) en las secuencias. Como método de datación es de tipo relativo pues se necesitan datos externos para calcular la tasa de mutación. Cuando se comparan especies suele recurrirse al registro paleontológico para estimar el momento de divergencia. Otro modo consiste en calcular la tasa de mutación a partir de la diversidad acumulada en una población que, en un momento conocido, ha colonizado una región

geográfica determinada y ha permanecido, desde entonces, en estricto aislamiento (STONEKING *et al.* 1986; TORRONI *et al.* 1994a).

1.2.3.2. Análisis de *pedigrees* familiares

Observando la evolución de un linaje a lo largo del tiempo –*i.e.* en diferentes generaciones–, puede estimarse la tasa de mutación de una región del genoma. Para ello se analiza el número de mutaciones nuevas que surgen en la descendencia de una familia determinada. La tasa de mutación se calcula, entonces, dividiendo el número de nuevos eventos mutacionales entre el número de generaciones analizadas.

Hay grandes diferencias en las tasas de mutación calculadas mediante cada uno de los dos métodos citados (PARSONS *et al.* 1997). En general, la tasa estimada a partir de los *pedigrees* familiares es muy superior a la proporcionada por los métodos filogenéticos. Un ejemplo: si empleáramos la tasa de mutación estimada a partir del primero de los dos métodos citados para calcular la edad del ancestro común más reciente (*Most Recent Common Ancestor* o *MRCA*) de la humanidad, llegaríamos a la conclusión de que éste había vivido hace no más de 6000 años (GIBBONS 1998).

Las discrepancias entre ambos tipos de estima de la tasa de mutación puede deberse a diversos factores no excluyentes. Uno es la existencia de tasas de mutación diferenciales en las posiciones polimórficas de la región control del DNA mitocondrial. La tasa de mutación es mayor en ciertas posiciones denominadas “puntos calientes” o *hot-spots* mutacionales (HEYER *et al.* 2001; STONEKING 2000; MEYER *et al.* 1999; EXCOFFIER y YANG 1999; HASEGAWA *et al.* 1993; WAKELEY 1993). Cabe también la posibilidad de que algunas de las elevadas tasas de mutación detectadas en los *pedigrees* familiares se deban a la recombinación del mtDNA de origen materno y paterno. Otras explicaciones alternativas son la aparición de mutaciones somáticas debido al proceso de envejecimiento o a la secuenciación de insertos mitocondriales en el DNA nuclear (NUMTs) (BALLARD y DEAN 2001).

Se ha apuntado también la posibilidad de que exista cierta presión selectiva sobre el mtDNA. La selección podría estar operando sobre la región control –la más utilizada en las reconstrucciones filogenéticas– de forma directa. Alternativamente el

efecto de la selección sobre otras regiones del mtDNA afectaría también a esta misma región dado el carácter haploide del mtDNA. La selección –deletérea o adaptativa– sobre el mtDNA provoca una disminución del número de diferencias acumuladas en los linajes afectando por tanto a la tasa de mutación (PARSONS *et al.* 1997). En este sentido, el análisis de la variabilidad de las poblaciones antiguas podría ayudar en la calibración de las tasas de mutación comparando, por ejemplo, los linajes de una misma región geográfica antes y después de transcurrido el tiempo mínimo de aparición de nuevas mutaciones, estimado a partir de los *pedigrees* familiares.

La dificultad en la calibración de la tasa de mutación del mtDNA supone un grave problema, puesto que el empleo de tasas inadecuadas puede alterar sustancialmente el tiempo de datación de ciertas variantes mitocondriales, afectando a las conclusiones derivadas (RICHARDS *et al.* 1997). En palabras de Peter Forster (FORSTER 2004) podríamos decir que “*la tasa de mutación es el talón de aquiles de los estudios filogenéticos de mtDNA*”.

1.2.4. Heteroplasmia

El concepto de *heteroplasmia* hace referencia a la presencia de tipos diferentes de DNA mitocondrial en la misma mitocondria, célula o individuo. En contraposición, se denomina *homoplasmia* a la presencia de un único tipo de mtDNA.

La heteroplasmia en el mtDNA puede producirse por errores en su replicación o por el daño causado por radicales libres resultantes del metabolismo oxidativo. Cuando la heteroplasmia afecta a las moléculas de mtDNA del oocito puede transmitirse a la descendencia, la cual habitualmente manifiesta niveles variados de la misma debido a la segregación aleatoria de las mitocondrias durante la oogénesis.

En humanos la heteroplasmia del mtDNA suele estar relacionada con enfermedades mitocondriales (CHINNERY *et al.* 2000; CHINNERY y TURNBULL 1999; WALLACE 1999) y el porcentaje de heteroplasmia suele guardar una relación directa con la severidad de la enfermedad en el individuo (penetrancia). Inicialmente se consideró que la heteroplasmia en regiones neutras del mtDNA como la región control era muy rara (PARSONS e IRWIN 2000; BENDALL *et al.* 1996; IVANOV *et al.* 1996;

JAZIN *et al.* 1996; COMAS *et al.* 1995; MONNAT *et al.* 1986). Pero hoy se piensa que pudiera ser mayor a lo estimado anteriormente (CALLOWAY *et al.* 2000; HOWELL y SMEJKAL 2000; TULLY *et al.* 2000). Estas diferencias en la detección de la incidencia de heteroplasma pueden deberse a la escasa resolución de las técnicas de secuenciación, empleadas en los primeros citados estudios (BENDALL *et al.* 1996).

En la región control existe una elevada frecuencia de heteroplasma asociada a la longitud de los trectos de poli-Citosinas. En la HVSI la mutación 16189T→C, genera una heteroplasma de longitud en el tracto de poli-C entre las posiciones 16184 y 16193, que oscila entre 8 y 14 residuos (HOWELL y SMEJKAL 2000; MARCHINGTON *et al.* 1997; BENDALL y SYKES 1995). En la HVSII existen varios casos, algunos asociados a ciertos haplogrupos europeos. Este es el caso de la variabilidad asociada a la longitud del primer tracto de poli-Cs del motivo C₇-T-C₅ (posiciones 00303→00316). En el haplogrupo J se han descrito dos variantes (C₈-T-C₆ y C₇-T₂-C₆) con heteroplasma en la longitud del tracto citado. En algunos portadores del haplogrupo I europeo hay, también, heteroplasma asociada al tracto de poli-Citosinas en la misma región, a partir de la posición 00568. En la región control se ha descrito, asimismo, variabilidad en el número de repeticiones de CA a partir de la posición 00514 (HOWELL y SMEJKAL 2000).

La heteroplasma en las regiones no codificantes es común en las células somáticas, lo que probablemente guarde cierta relación con el proceso de envejecimiento (CALLOWAY *et al.* 2000). A su vez, la heteroplasma en las células germinales ha demostrado ser más común de lo que se afirmaba inicialmente (LAGERSTRÖM-FERMÉR *et al.* 2001; HOWELL *et al.* 1996; MARCHINGTON *et al.* 1997; CHEN 1995).

La heteroplasma es el primer nivel de manifestación del polimorfismo pues si surge una mutación en alguna molécula de mtDNA de un oocito, el individuo descendiente presentará inicialmente una mezcla de moléculas originales y mutantes. Para que el polimorfismo se establezca es necesario que la mutación pase a los descendientes, adquiera un estado homoplásmico a nivel individual y, finalmente, se fije en la población. La nueva mutación puede perderse en diversas fases del proceso. Sin embargo, hay evidencias de que la homoplasma puede llegar a alcanzarse en una o dos

generaciones. La segregación aleatoria de las mitocondrias durante la oogénesis no permite explicar la rapidez de este fenómeno por sí sola. Una posibilidad sugerida por varios autores es que, durante la oogénesis, se produzca una especie de cuello de botella en el que el número de moléculas de mtDNA se reduzca drásticamente (HAUSWIRTH y LAIPIS 1982; ASHLEY *et al.* 1989). En este caso, el mtDNA del oocito maduro y del embrión derivado del mismo descendería de un subconjunto del mtDNA total de la célula progenitora.

Aunque una mutación llegue a alcanzar el nivel de homoplasmia en una o dos generaciones, podría no fijarse en la población a causa de la deriva genética. El análisis de la variabilidad genética del pasado puede proporcionar una estima del número de líneas perdidas por deriva.

1.2.5. Neutralidad del mtDNA

Las inferencias realizadas a partir de los estudios de mtDNA parten del supuesto de que el mtDNA sigue un modelo de evolución neutral. Este tipo de modelo evolutivo postula que la mayor parte de las mutaciones son selectivamente neutras (KIMURA 1983) y su fijación se produce aleatoriamente por deriva genética. Las mutaciones neutras son mantenidas en una población mediante el equilibrio entre mutación (aporte de variantes) y deriva genética (eliminación de variantes), de manera que el polimorfismo en una población es el estado intermedio previo a la fijación.

Bajo condiciones neutras la cantidad de polimorfismo de una población es directamente proporcional al producto del tamaño efectivo de la población por la tasa de mutación:

$$\text{Polimorfismo neutro} = N_e \times \mu$$

Para comprobar si un marcador genético sigue un modelo de evolución neutral se suelen emplear los “tests de neutralidad”, que comparan el nivel de polimorfismo de dicho marcador con aquel esperado de acuerdo con dicho modelo. La aplicación de algunos tests de neutralidad sobre la variabilidad del mtDNA mostró que su evolución

se desviaba de dicha neutralidad, pues el nivel de polimorfismo actual era mucho menor que el que se esperaría bajo un modelo de evolución neutral (WHITTAM *et al.* 1986; EXCOFFIER 1990; MERRIWETHER *et al.* 1991). Existen dos explicaciones mutuamente no excluyentes de este resultado:

- Debe haber cierta presión selectiva sobre el mtDNA.
- La mayoría de las poblaciones humanas no están en equilibrio mutación-deriva, condición necesaria para la aplicación de los tests de neutralidad.

Las expansiones poblacionales de la Humanidad actual pueden haber desviado la población del equilibrio mutación-deriva; sin embargo, estudios recientes sugieren que la variabilidad y la distribución geográfica de los tipos mitocondriales observadas actualmente pueden ser explicadas mediante selección natural.

Tras aplicar los tests de neutralidad de Tajima (D) y de Fu y Li (D*, F*) (TAJIMA 1989; FU y LI 1993) a la distribución de las variantes mitocondriales – establecidas a partir de secuencias completas del mtDNA– de los diferentes haplogrupos mundiales, individualmente y agrupándolos por continentes, el equipo de Douglas Wallace concluyó que mientras que los haplogrupos africanos –en conjunto e individualmente– se ajustaban a un modelo de evolución neutral, los clusters europeos –M, N y sus derivados– y asiáticos –A, B, C, D, G, Z, Y, X– se desviaban significativamente del modelo (MISHMAR *et al.* 2003; RUIZ-PESINI *et al.* 2004). El análisis de las regiones codificantes reveló, además, que existía un alto número de sustituciones no sinónimas (ka) respecto a las sustituciones sinónimas (ks) en ciertos genes: ATP6, citocromo b y COI, y que sus niveles de variabilidad guardaban cierta relación con la procedencia geográfica individual. Así, los individuos actuales de zonas árticas y subárticas (Siberia y América) presentaban una mayor variabilidad asociada al gen de la ATP6. En aquellos procedentes de la zona templada –como Europa– dicha variabilidad se concentraba esencialmente en el gen del citocromo b, mientras que en los habitantes de las zonas tropicales y sub-tropicales (África) la variabilidad era mayor en la citocromo oxidasa I (COI). Se pudo comprobar, también, que muchas de las sustituciones no sinónimas alteraban aminoácidos conservados en otras especies. Según MISHMAR *et al.* 2003, estas sustituciones confirieron ventaja adaptativa a los individuos portadores, posibilitando la colonización de ambientes más fríos. Aunque se desconoce el efecto funcional de los cambios introducidos, éste pudiera estar

relacionado con la generación de calor a través del metabolismo de fosforilación oxidativa mitocondrial (OXPHOS). Puesto que la selección de regiones del mtDNA genera un “arrastre” de toda la molécula por ser ésta haploide y no recombinante, es de esperar que la selección natural de variantes de regiones codificantes hubiera afectado a las regiones tradicionalmente empleadas para la definición de los haplogrupos mitocondriales. Por lo tanto la distribución de éstos no respondería solamente a deriva genética sino también a selección natural.

Existen otros datos que sugieren cierta asociación entre algunos haplogrupos mitocondriales y ciertas regiones del genoma mitocondrial que pudieran haber modelado la frecuencia y distribución actual de dichos clusters. El haplogrupo T, por ejemplo, parece estar asociado con una reducida movilidad del esperma –concretamente con un defecto en el funcionamiento de los complejos I y IV–, y con el síndrome de Wolfram (RUIZ-PESINI *et al.* 2000). Este haplogrupo junto con el K y el H ha sido relacionado, también, con cierto efecto protector ante la enfermedad de Alzheimer (CHAGNON *et al.* 1999; CARRIERI *et al.* 2001; SHOFFNER *et al.* 1993). Por su parte, el haplogrupo J parece estar relacionado con la Neuropatía Hereditaria Óptica de Leber (LHON) (BROWN *et al.* 1997, 2002); y además con un aumento de la susceptibilidad a la esclerosis múltiple (KALMAN *et al.* 1999), que también afecta a algunas variantes del haplogrupo K. Estos dos últimos haplogrupos presentan cierto efecto protector frente a la enfermedad de Parkinson (VAN DER WALT *et al.* 2003). El haplogrupo J y ciertas variantes del haplogrupo M se han asociado a un incremento de la longevidad (IVANOVA *et al.* 1998; DE BENEDICTIS *et al.* 1999; ROSE *et al.* 2001; ROSS *et al.* 2001; NIEMI *et al.* 2005, 2003; BONAFE *et al.* 2002; TANAKA *et al.* 1998).

Si efectivamente la selección natural ha jugado un papel importante en la dispersión de las variantes mitocondriales, la tasa del reloj molecular mitocondrial no sería constante a lo largo de la historia de la humanidad, por lo que cabría replantearse la validez de las estimas realizadas con modelos de coalescencia.

1.2.6. Inserciones de fragmentos de mtDNA en el genoma nuclear

Algunos fragmentos del mtDNA pueden integrarse en el genoma nuclear, donde evolucionan como *pseudogenes* nucleares a un ritmo más lento que sus homólogos mitocondriales (COLLURA y STEWART 1995; ZISCHLER *et al.* 1995a; ZULLO *et al.* 1991). Estas integraciones –denominadas comúnmente numtDNA (*Nuclear mitochondrial DNA-like sequences*)– han sido descritas en diversos organismos, incluida la especie humana (COLLURA y STEWART 1995; ARCTANDER 1995; BLANCHARD y SCHMIDT 1996; CORRAL *et al.* 1989; KOCHER *et al.* 1989; LÓPEZ *et al.* 1994; SUNNUCKS y HALES 1996; NOMIYAMA *et al.* 1985; ZISCHLER *et al.* 1995a; HERRNSTADT *et al.* 1999). Se desconocen tanto el mecanismo mediante el cual se producen estas inserciones como las consecuencias de las mismas.

Análisis filogenéticos realizados con numtDNA humanas indican que, mientras que algunas de estas inserciones se realizaron antes de la formación del hombre moderno (COLLURA y STEWART 1995; FUKUDA *et al.* 1985; HU y THILLY 1994) hay otras mucho más recientes (WALLACE *et al.* 1997; ZISCHLER *et al.* 1995a).

La existencia de secuencias numtDNA similares a sus homólogos mitocondriales dificulta la amplificación de secuencias exclusivamente mitocondriales a partir de un extracto de DNA total. Esto supone un gran problema tanto para las reconstrucciones filogenéticas a partir de mtDNA como para los análisis de DNA antiguo. En este último caso la dificultad no reside en la amplificación preferente de secuencias nucleares endógenas, puesto que, debido a la escasez y fragmentación del DNA antiguo, el DNA *template* de partida es mayoritariamente de origen mitocondrial. Sin embargo las secuencias numtDNA exógenas (de los investigadores, arqueólogos, del ambiente...) pueden convertirse en fuente de contaminación adicional de origen moderno. El problema aquí reside en la dificultad de identificar la fuente de origen, pues muchas de estas secuencias probablemente no están descritas. La putativa secuencia de dinosaurio de WOODWARD *et al.* 1994b pudiera ser un ejemplo de amplificación de secuencia numtDNA contaminantes (ZISCHLER *et al.* 1995b).

Los datos de inserciones recientes de mtDNA en el genoma nuclear que incluyen fragmentos de la región de control (ZISCHLER *et al.* 1995a: posiciones 59-16089) merecen especial atención por ser ésta la región más empleada en las reconstrucciones filogenéticas y en el análisis de DNA antiguo. Hasta el momento el número de diferencias encontrado entre estas secuencias y sus homólogas mitocondriales es muy superior a las que presentan estas últimas en la población actual (ZISCHLER *et al.* 1995a).

En consecuencia, el número de mutaciones puede orientar acerca del origen – mitocondrial endógeno o nuclear contaminante– de las secuencias obtenidas de restos antiguos de la especie humana. La comparación de los haplotipos antiguos obtenidos con secuencias de bases de datos de la humanidad actual y la asignación de haplogrupos mitocondriales pueden servir también como criterios diferenciadores.

1.3. El mtDNA y el estudio de las poblaciones humanas

Desde su desarrollo a finales de los años 80, el estudio de la variabilidad del DNA mitocondrial (mtDNA) de las poblaciones humanas se ha convertido en una poderosa herramienta. Con ella se ha rastreado y reconstruido la historia evolutiva de las poblaciones humanas.

1.3.1. Polimorfismos del mtDNA y haplotipos mitocondriales

La mayor parte de los polimorfismos del mtDNA son sustituciones puntuales, aunque también se han descrito deleciones e inserciones de una o varias pares de bases (CANN y WILSON 1983; HOWELL y SMEJKAL 2000). Si asumimos que el mtDNA se hereda en bloque, podemos considerar todo el genoma mitocondrial como un *locus* en el que la combinación de mutaciones, inserciones y deleciones constituiría un *alelo* o *haplotipo mitocondrial*. Idealmente, para una identificación correcta de todos los tipos mitocondriales debería secuenciarse todo el genoma mitocondrial. A efectos poblacionales ello no es práctico debido al elevado número de muestras y a la existencia de regiones del mtDNA con poca variabilidad. Los investigadores tienden, pues, a definir los haplotipos mitocondriales mediante el estudio de las regiones más variables – y por tanto más informativas– en un número elevado de individuos.

1.3.1.1. Polimorfismos de restricción enzimática y polimorfismos de secuencia

Clásicamente se han venido utilizando dos herramientas para el estudio de la variabilidad mitocondrial humana: el análisis de polimorfismos de restricción enzimática (“*Restriction Fragments Length Polymorphisms*” o “*RFLPs*”) –generalmente en todo el genoma mitocondrial–, y la secuenciación de ciertas regiones del mtDNA,– habitualmente las regiones hipervariables (HVS I y/o HVS II) de la Región Control–.

En el estudio de los polimorfismos de restricción enzimática las diferentes variantes mitocondriales o *haplotipos* se definen de acuerdo a la ausencia o presencia de “posiciones de restricción”, o lugares en los que se produce corte con ciertas endonucleasas o enzimas de restricción, en sitios concretos del genoma mitocondrial. Dependiendo del número de estos enzimas utilizados dichos análisis de restricción,

pueden ser de baja resolución (*Low resolution restriction analysis*) (GILES *et al.* 1980; MERRIWETHER *et al.* 1991), o de alta resolución (*High resolution restriction analysis*) (CANN y WILSON 1983, 1984; BALLINGER *et al.* 1991).

En la secuenciación de la totalidad o de ciertas regiones del genoma mitocondrial, por el contrario, es la combinación de cambios nucleotídicos o *mutaciones* en la secuencia de bases lo que define los haplotipos mitocondriales.

El uso del mtDNA como marcador molecular para la reconstrucción del pasado reciente de las poblaciones humanas fue encabezado por Westley Brown y Douglas Wallace a finales de los años 70. En los primeros estudios, de baja resolución, se estudiaron numerosas muestras mediante un único enzima de restricción (DENARO *et al.* 1981) o un número pequeño de éstas mediante varios enzimas (BROWN 1980). La tendencia en estudios posteriores fue a utilizar numerosos enzimas de restricción sobre un mayor número de muestras (CANN *et al.* 1987; TORRONI *et al.* 1996, 1994b; JONHSON *et al.* 1983; SANTACHIARA-BENERECETTI *et al.* 1988; SCOZZARI *et al.* 1988).

El desarrollo y generalización de la *PCR* y el avance de las técnicas de secuenciación posibilitaron una nueva aproximación al estudio molecular de las poblaciones humanas, consistente en la secuenciación de porciones de la molécula del mtDNA. La resolución proporcionada por los nuevos análisis de secuenciación era mucho mayor que la de los análisis de restricción más sofisticados y permitió realizar una disección más detallada de los tipos mitocondriales individuales. Inicialmente los análisis se centraron en la región control o D-loop mitocondrial, ampliándose posteriormente a otras regiones no codificantes del genoma mitocondrial.

1.3.2. Técnicas de estudio del mtDNA

En el estudio de la variabilidad genética del mtDNA de las poblaciones actuales pueden emplearse diversos métodos que repasaremos brevemente.

1.3.2.1. Medidas de diversidad

1.3.2.1.1. Índices de diversidad

Los índices de diversidad proporcionan una medida de la cantidad de variación de una población, siendo especialmente útiles para realizar comparaciones entre poblaciones *a grosso modo*. Los más empleados en el estudio de la variabilidad del mtDNA de las poblaciones actuales son: el número de posiciones variables (v), el número de haplotipos diferentes (k), la diversidad genética (o haplotípica) de Nei (H), la diversidad nucleotídica (π) y el número medio de diferencias por parejas (π_n).

1.3.2.1.2. Distribución de las diferencias por parejas o *mismatch distribution*

Otra forma habitual de representar la diversidad mitocondrial es la distribución de las diferencias por parejas, también conocida como “*mismatch distribution*” o distribución de las “*mean pairwise differences*”. Este tipo de representación es adecuado para aquellos datos en los que pueden computarse diferencias discretas entre los diferentes alelos, haplotipos en el caso del mtDNA. Las diferencias pueden ser sustituciones de bases, *RFLPs* o número de repeticiones de microsatélites.

En el caso de los haplotipos mitocondriales, las secuencias de una muestra determinada se comparan dos a dos y se calcula el número de diferencias. Por lo tanto, si existen n secuencias en la muestra será necesaria la realización de $n(n-1)/2$ de estas comparaciones. Finalmente se construye un histograma con la frecuencia observada para cada número de diferencias.

Además de describir la diversidad de una muestra, estas distribuciones permiten realizar inferencias acerca de la historia demográfica de las poblaciones. Se ha demostrado que los episodios de crecimiento demográfico, de declive o los cuellos de botella sufridos por una población dejan huellas características en la distribución de las diferencias por parejas. En general aquellas poblaciones con una distribución *gaussiana* de las diferencias por parejas del mtDNA han experimentado un período de crecimiento demográfico rápido a partir de una o unas pocas variantes. Por el contrario una distribución asimétrica de las diferencias por parejas es indicadora de una población

cuyo tamaño ha sido constante durante largo tiempo (ROGERS y HARPENDING 1992).

1.3.2.2. Los haplogrupos mitocondriales

Podríamos definir un haplogrupo o *cluster* mitocondrial como una agrupación de haplotipos que comparten ciertas sustituciones diagnósticas (definidas por enzimas de restricción o por secuenciación directa) y que presentan un origen común. Esto implica que los polimorfismos que definen cada haplogrupo se produjeron exclusivamente en las líneas antecesoras de todos los haplotipos que lo integran. El concepto de haplogrupo fue introducido por vez primera por A. Torroni, D. Wallace y colaboradores a inicios de los 90' (TORRONI *et al.* 1992).

La aplicación de este tipo de estudios en muestras de diferentes continentes por separado reveló la existencia de localizaciones geográficas de los haplogrupos mitocondriales, de manera que podían distinguirse clusters exclusivamente africanos, europeos y asiáticos respectivamente. Esta característica ha resultado ser de extrema utilidad para la inferencia de patrones migratorios dentro y fuera de cada continente.

La combinación de los datos de secuencia y de enzimas de restricción de regiones codificantes y no codificantes disponibles, así como el análisis de nuevos marcadores ha llevado a una disección cada vez más compleja de los diferentes haplogrupos mundiales, que queda reflejada en su compleja nomenclatura. La nomenclatura inicial propuesta por Wallace y Torroni (TORRONI *et al.* 1992) identifica los clusters principales designándolos con una letra mayúscula (Ej: A, B, C, D, H, I, L, U, V). Los clusters pueden dividirse, a su vez, en sub-clusters que se designan por su letra correspondiente seguida de un número (Ej: L2). Las subdivisiones al nivel siguiente se realizan alternando letras minúsculas y números (Ej: U5a1b). Las líneas que presentan los motivos (del inglés “*motif*”) propios de un haplogrupo pero no los de los diferentes sub-haplogrupos del mismo se designan con un asterisco (*) y habitualmente reciben el nombre de *paragrupos* en lugar de haplogrupos (Ej: J*). Cuando dos o más haplogrupos previamente definidos comparten una serie de motivos que sugieren un origen común, se tiende a agruparlos uniendo los nombres de los diferentes clados (Ej: HV, JT).

En la siguiente tabla (**Tabla II**) se muestran, respectivamente, los polimorfismos de secuencia y de restricción enzimática que caracterizan los diferentes haplogrupos mundiales.

Tabla II. Haplogrupos y subhaplogrupos mundiales, y polimorfismos de secuencia y de restricción enzimática que los caracterizan según RICHARDS *et al.* 2000; ACHILLI *et al.* 2004 y LÖÖGVÄLI *et al.* 2004 (subhaplogrupos H1-H15); TORRONI *et al.* 2001, 1998b (haplogrupos V y pre-V).

Haplogrupos			D-loop		Región codificante		
			HVRI (+16000)	HVRII	Mutaciones	RFLPs	
L	L1	L1a	148 172 187 188G 189 223 230 311 320	00073A		+3592HpaI	
		L1b	126 187 189 223 264 270 278 311	00073G		+3592HpaI	
		L1c	129 187 189 223 278 294 311 360	00073G		+3592HpaI	
		L1e	129 148 166 187 189 223 278 311	00073G		+3592HpaI	
		L1f	169 187 189 223 230 278 311 327	00073G		+3592HpaI	
		L2	223 278 390	00073G		+3592HpaI	
	L3	L3b	223	00073G			
		L3d	124 223 278 362	00073G		+10084TaqI	
		L3e	124 223	00073G		-8616MboI	
	M	M1	L3e	223	00073G		+2349MboI
					00073G		+10397AluI
					00073G		+10397AluI
		C		129 189 223 249 311	00073G		+10397AluI
				223 298 327	00073G		-13259HincII/ +13262AluI
		D		223 362	00073G		-5176AluI
				00073G		+10397AluI	
E			223 227 362	00073G		-7598HhaI	
				00073G		+10397AluI	
G			017 129 223	00073G		+4830HaeII/ +4831HhaI	
				00073G		+10397AluI	
N		Z		185 223 224 260 298	00073G		+10397AluI
				223	00073G		+10871MnII
				223	00073G		+10871MnII
N1		N1a		223	00073G		+10237HphI
			147A/G 172 223 248 355	00073G		+10871MnII	
				00073G		+10237HphI	
	N1b		145 176G 223	00073G		+10871MnII	
				00073G		+10237HphI	
	N1c		223 265	00073G		+10871MnII	
				00073G		+10237HphI	
	A	I		129 223 391	00073G		+10032AluI
				223 290 319	00073G		+663HaeIII
				223 292	00073G		-8994HaeIII
	W	I		189 223 278	00073G		+14465AccI
				CRS	00073G		
	X	R1		278 311	00073G		
				071	00073G		
	R	R2		189	00073G		9bp COII- tRNA ^{Lys}
				00073G		-	
F	F		304	00073G		12406HincII/- 12406HpaI	
				00073G		+7933MboI	
Y	Y		126 231 266	00073G		-8391HaeIII	
				00073G		+4216NlaIII	
JT	JT		126	00073G		+4216NlaIII	
				00073G		-13704BsrOI	
J	J		069 126	00073G		+4216NlaIII	
				00073G		-13704BsrOI	
J1	J1		069 126 261	00073G		+4216NlaIII	
				00073G		-13704BsrOI	
J1a	J1a		069 126 145 231 261	00073G		+4216NlaIII	
				00073G		-13704BsrOI	

Tabla II. (Continuación).

Haplogrupos	D-loop		Región codificante	
	HVRI (+16000)	HVRII	Mutaciones	RFLPs
J1b	069 126 145 222 261	00073G		+4216 <i>Nla</i> III -13704 <i>Bst</i> OI
	J1b1 069 126 145 172 222 261	00073G		+4216 <i>Nla</i> III -13704 <i>Bst</i> OI
J2	069 126 193	00073G		+4216 <i>Nla</i> III -13704 <i>Bst</i> OI +4216 <i>Nla</i> III
T	126 294	00073G		+13366 <i>Bam</i> HI/ -13367 <i>Ava</i> II/ +13367 <i>Mbo</i> I +15606 <i>Alu</i> I -15925 <i>Msp</i> I +4216 <i>Nla</i> III -12629 <i>Ava</i> II +13366 <i>Bam</i> HI/
T1	126 163 186 189 294	00073G		-13367 <i>Ava</i> II/ +13367 <i>Mbo</i> I +15606 <i>Alu</i> I -15925 <i>Msp</i> I +4216 <i>Nla</i> III +13366 <i>Bam</i> HI/
T2	126 294 304	00073G		-13367 <i>Ava</i> II/ +13367 <i>Mbo</i> I +15606 <i>Alu</i> I -15925 <i>Msp</i> I +4216 <i>Nla</i> III +13366 <i>Bam</i> HI/
T3	126 292 294	00073G		-13367 <i>Ava</i> II/ +13367 <i>Mbo</i> I +15606 <i>Alu</i> I -15925 <i>Msp</i> I +4216 <i>Nla</i> III +13366 <i>Bam</i> HI/
T4	126 294 324	00073G		-13367 <i>Ava</i> II/ +13367 <i>Mbo</i> I +15606 <i>Alu</i> I -15925 <i>Msp</i> I +4216 <i>Nla</i> III +13366 <i>Bam</i> HI/
T5	126 153 294	00073G		-13367 <i>Ava</i> II/ +13367 <i>Mbo</i> I +15606 <i>Alu</i> I -15925 <i>Msp</i> I
U	CRS	00073G		+12308 <i>Hinf</i> I -4990 <i>Alu</i> I +12308 <i>Hinf</i> I
U1	249	00073G		-13103 <i>Hinf</i> I/ +13104 <i>Mbo</i> I +14068 <i>Taq</i> I -4990 <i>Alu</i> I +12308 <i>Hinf</i> I
U1a	189 249	00073G		-13103 <i>Hinf</i> I/ +13104 <i>Mbo</i> I +14068 <i>Taq</i> I -4990 <i>Alu</i> I +12308 <i>Hinf</i> I
U1b	249 327	00073G		-13103 <i>Hinf</i> I/ +13104 <i>Mbo</i> I +14068 <i>Taq</i> I +12308 <i>Hinf</i> I
U2	051 129C	00073G		+15907 <i>Rsa</i> I
U3	343	00073G		+12308 <i>Hinf</i> I +4643 <i>Rsa</i> I
U4	356	00073G		+11329 <i>Alu</i> I +12308 <i>Hinf</i> I
U5	270	00073G		+12308 <i>Hinf</i> I

Tabla II. (Continuación).

Haplogrupos		D-loop		Región codificante	
		HVRI	HVRII	Mutaciones	RFLPs
	U5a		192 270	00073G	+12308HinfI
	U5a1		192 256 270	00073G	+12308HinfI
		U5a1a	256 270 399	00073G	+12308HinfI
	U5b		189 270	00073G	+12308HinfI
	U5b1		144 189 270	00073G	+12308HinfI
	U6		172 219	00073G	+12308HinfI
	U6a		172 219 278	00073G	+12308HinfI
	U6a1		172 189 219 278	00073G	+12308HinfI
	U6b		172 219 311	00073G	+12308HinfI
	U7		318T	00073G	+12308HinfI
	K		224 311	00073G	-9052HaeII/ -9053HhaI
pre-HV			CRS	00073A	+12308HinfI +11718HaeIII
	pre-HV(1)		126 362	00073A	11719, 2442, 3847, 13188 +11718HaeIII
HV			CRS	00073A	11719, 14766 +11718HaeIII -14766MseI
	HV1		067	00073A	8014T, 15218 +11718HaeIII -14766MseI -7025AluI
	H		CRS	00073A	2706, 7028 +11718HaeIII -14766MseI
	H1			00073A	3010
	H1a		162	00073A	3010
	H1b		189, 356	00073A	3010, 3796
	H1f		189	00073A	3010, 4452, 7309, 9066
	H2			00073A	1438, 4769
	H2a			00073A	1438, 4769, 951
		H2a1	354	00073A	1438, 4769, 951
	H2b			00073A	1438, 4769, 750
	H2c		274	00073A	1438, 4769, 10810
	H3			00073A	6776
	H3a		239G	00073A	6776, 152, 13404
	H4			00073A	3392, 4024, 5004, 9123, 14365, 14582
	H4a			00073A	3392, 4024, 5004, 9123, 14365, 14582, 8269
	H5		304	00073A	456
	H5a		304	00073A	456, 4336
		H5a1		00073A	15833
	H6		362	00073A	239, 16482
	H6a		362	00073A	239, 16482, 3915
		H6a1	362	00073A	239, 16482, 3915, 9380
	H6b		300	00073A	239, 16482
	H7			00073A	4793
	H8		288, 362	00073A	13101C
	H9		168	00073A	3591, 4310, 13020
	H10			00073A	14470A
	H11		311	00073A	961G, 8448, 13759
	H12		287	00073A	3936, 14552
	H13			00073A	2259, 4745, 13680, 14872
	H14			00073A	7645, 11377
	H15			00073A	55, 57, 6253
Pre-V			298	00073A	72, 15904
	V		298	00073A	72, 15904, 4580

1.3.2.3. *Relaciones filogenéticas entre los haplotipos mitocondriales. Median Networks.*

La clasificación de los haplogrupos mitocondriales según las posiciones diagnósticas de la región control no es 100% exacta. Esto se debe esencialmente a dos fenómenos: la retromutación y la homoplasia.

El avance en los estudios de secuenciación de la región hipervariable I (HVSI) ha revelado que la tasa de mutación es diferencial a lo largo del segmento. Mientras que algunas posiciones de la HVSI son extremadamente variables entre los linajes, otras son relativamente constantes (WAKELEY 1993; PIERCY *et al.* 1993). Un determinado cambio nucleotídico puede ser frecuente en una población por dos causas: porque se trate de un cambio ancestral que ha sido heredado por los descendientes de ese linaje; o porque se haya producido de manera independiente en una línea no relacionada (homoplasia). Cualquier análisis filogenético que otorgue el mismo peso a todas las posiciones –como el de máxima parsimonia o de diferencias en parejas de nucleótidos–, tenderá a agrupar aquellas secuencias que compartan posiciones, creando grupos polifiléticos. Cualquier interpretación evolutiva derivada de estas agrupaciones será, pues, errónea. De manera análoga, si una misma posición ha mutado varias veces en la evolución puede resultar difícil, si no imposible, determinar el estatus ancestral del haplotipo. Las reconstrucciones filogenéticas convencionales, en estos casos, separarían variantes estrechamente relacionadas.

Debido a lo anterior, a medida que se fue acumulando un mayor número de datos de la variabilidad del mtDNA de las poblaciones humanas, los métodos “clásicos” de reconstrucción filogenética se mostraron cada vez menos efectivos y fue necesario el desarrollo de nuevas herramientas específicas para el análisis poblacional de este peculiar material genético. PENNY *et al.* 1995 desarrollaron un método que les permitía explorar el espacio de los árboles más parsimoniosos generados a partir de un conjunto de datos. Sin embargo la aproximación de mayor éxito fue la de BANDELT *et al.* 1995. Este método de reconstrucción filogenética –denominado “Método de las redes medias” o “*Median Networks*”– permite resumir los árboles más parsimoniosos de un grupo de datos en forma de una única red, e informa, además, acerca de la frecuencia de cada haplotipo observado. En este tipo de reconstrucción se identifican las sustituciones

homoplásicas y se presentan en forma de puntos reticulares, lo que facilita su interpretación evolutiva.

Cuando el número de haplotipos diferentes es muy elevado, las reticulaciones pueden aumentar hasta hacer el árbol ininteligible. En este caso se suelen eliminar los haplotipos menos frecuentes, partiendo de la asunción que son los que se han generado más recientemente (RICHARDS *et al.* 1998). Otra estrategia válida consiste en identificar las posiciones causantes de las reticulaciones y otorgarles un peso bajo, considerando que se trata de sitios hipermutables. Un método alternativo para la construcción de networks con pocas reticulaciones es el *Median Joining* (BANDELT *et al.* 1999), que consiste en la introducción de los nodos ancestrales más probables en la red de secuencias observadas.

1.3.2.3.1. Topología de los *networks*

La topología de los *Median Networks* o *Median Joining Networks* puede orientar acerca de la historia demográfica de una población determinada. Por ejemplo, las filogenias en forma de estrella –con un haplotipo central del que surgen varios haplotipos separados de éste por un único paso mutacional–, son signatures de una expansión poblacional. Este patrón puede explicarse de la siguiente manera: una determinada variante mitocondrial experimenta condiciones reproductivas favorables durante un largo período de tiempo y aumenta de frecuencia. Tras muchas generaciones algunos de los descendientes portadores de este haplotipo adquieren nuevas mutaciones. En algunos casos la variante ancestral se extingue por deriva genética. Si alguno de los descendientes sufre el mismo proceso que la variante inicial, en la filogenia en forma de estrella aparecen subgrupos o sub-clusters (FORSTER 2004).

1.3.2.3.2. Datación de los *clusters*

Para la datación de los diferentes *clusters* se estima el tiempo del ancestro común más reciente (TMRCA, *Time of the Most Recent Common Ancestor*) de cuantas líneas lo integran. Para ello se consideran aquellos haplotipos relacionados, se calcula su diversidad y se estima el tiempo que ésta ha tardado en generarse con una tasa de mutación determinada. La estima no se corresponde exactamente con el momento en

que surgieron las mutaciones que caracterizan a dichos clusters, pues desde ese momento hasta la fijación de dichas mutaciones en las líneas descendientes puede haber transcurrido mucho tiempo. Por otra parte, la estima del TMRCA se basa en aquellos haplotipos derivados del linaje original. Muchos de los descendientes de este MRCA pueden haberse extinguido por procesos de deriva genética por lo que el tiempo estimado puede ser menor al real.

Entre las medidas de diversidad más empleadas para calcular el TMRCA está el estadístico ρ (Rho) (FORSTER *et al.* 1996). La datación mediante este estadístico requiere de la reconstrucción filogenética entre los haplotipos que conforman un cluster determinado, habitualmente en forma de *Median networks* o *Median joining networks*. El estadístico ρ representa el número medio de cambios mutacionales entre el haplotipo original (*root*) y los haplotipos individuales de la muestra. Estos cambios se estiman a partir del Network puesto que éste tiene en cuenta las reversiones y paralelismos en los sitios hipermutables. Este estadístico se relaciona directamente con el tiempo mediante la ecuación: $\rho = \mu t$ (FORSTER *et al.* 1996). Puede calcularse un error estándar de cada fecha en base al número de haplotipos muestreados y a la subestructura de las ramas (SAILLARD *et al.* 2000).

Las principales fuentes de error de este tipo de estimas del TMRCA son la incertidumbre en los parámetros, esencialmente en la tasa de mutación y el tiempo de generación, y en la historia demográfica de la población. Para datar los clusters mediante el método descrito es necesario asumir que:

- a. la filogenia presenta una forma de estrella.
- b. desde el MRCA cada haplotipo ha experimentado una historia demográfica independiente del resto.

Estas condiciones únicamente son aplicables a aquellas poblaciones que han experimentado un crecimiento demográfico rápido a partir de unas pocas variantes fundadoras. Habitualmente estos métodos proporcionan una infraestima del tiempo del MRCA (JOBBLING *et al.* 2003).

Existe una tendencia a interpretar las dataciones de los clusters según la información del registro prehistórico. Sin embargo conviene recalcar que el TMRCA

marca el momento de separación de determinados linajes relacionados, no de las poblaciones en las que se encuentran actualmente. Habitualmente el TMRCA de un conjunto de haplotipos es mucho más antiguo que la población o poblaciones en las que éstos se encuentran hoy. Su estima sólo se aproxima a la fecha de origen de una población si ésta desciende de unos pocos haplotipos porque ha experimentado un fuerte cuello de botella o efecto fundador.

1.3.2.4. Análisis filogeográficos

Mediante la integración de la información genética y su localización geográfica se ha intentado evaluar el impacto de determinados procesos históricos sobre la diversidad genética actual.

1.3.2.4.1. Mapas de frecuencias

Existen diversos métodos para representar la variabilidad geográfica de las frecuencias de diferentes alelos, haplotipos o haplogrupos. La forma más sencilla consiste en dibujar las frecuencias sobre un mapa. Esta aproximación se emplea comúnmente para indicar la frecuencia de los haplogrupos mitocondriales en las poblaciones muestreadas y ha sido aplicada en la presente tesis doctoral.

Otra forma consiste en unir en el mapa los puntos con igual frecuencia génica mediante unas líneas denominadas *líneas isogénicas*.

La desventaja principal de los mapas de frecuencia es que sintetizan la información de un único alelo por mapa. Sin embargo CAVALLI-SFORZA *et al.* 1993 desarrollaron un método que permitía representar la información de varios alelos en un mismo mapa. La información de dichos alelos se condensa previamente en un número reducido de variables mediante el método de Componentes Principales. Cada componente principal se calcula como la función lineal optimizada de todas las frecuencias génicas disponibles y explica un porcentaje total de la variación genética observada. Si se representan en un mapa las frecuencias génicas utilizadas para la construcción de cada componente principal se obtiene el *mapa sintético* de cada una de ellas. La presencia en estos mapas de un gradiente de frecuencias en forma de anillos

concéntricos en torno a un área concreta se interpreta como signo de una expansión demográfica con origen geográfico en dicha región (CAVALLI-SFORZA *et al.* 1993). El gradiente indica la existencia de diferencias en las frecuencias génicas entre el foco de expansión y las zonas próximas. Ello requiere que las poblaciones inmigrante y receptora presenten cierto grado de diferenciación genética en el momento de la expansión. Para que ello ocurra además de éste deben cumplirse otros requisitos, como un cierto grado de mezcla entre la población que se expande y las circundantes. Si la mezcla es demasiado elevada o si la población receptora experimenta un incremento demográfico paralelo –por ejemplo, tras absorber las innovaciones culturales responsables de la expansión de sus vecinos–, el gradiente tendrá escaso alcance.

Una de las principales críticas a este método es que los movimientos migratorios que hayan podido dejar una huella genética similar en las frecuencias se integran en un mismo mapa, siendo imposible determinar, finalmente, el número real de migraciones que han tenido lugar.

1.3.2.4.2. Análisis fundador

El análisis fundador (RICHARDS *et al.* 2000) permite identificar aquellas variantes mitocondriales que llegan a nuevos territorios y datar la migración. Se compara la diversidad molecular en las poblaciones original y receptora y se identifican los haplotipos fundadores que abandonaron la primera para colonizar la segunda. El tiempo de migración se estima midiendo la diversidad acumulada en estos haplotipos fundadores en la población receptora. Esta diversidad se determina sustrayendo la diversidad de la población original de la diversidad de la población receptora. Se identifican las líneas introducidas en la población original por migración posterior desde la población receptora (*back-migrants*) y se excluyen del análisis.

Habitualmente una población contiene diversos fundadores que han llegado en diferentes épocas desde otra región geográfica. Cuando el ambiente en la nueva región es propicio, muchos de estos fundadores experimentan un incremento rápido en su frecuencia seguido de una diversificación de linajes que formará el patrón de estrella característico. Parte de este patrón puede perderse, sin embargo, por deriva genética si la

migración es muy antigua, enmascarando el incremento demográfico inicial (RICHARDS *et al.* 2000).

Puesto que este tipo de análisis se realiza con la variabilidad observada en las poblaciones actuales y no en las poblaciones originales fuente de la migración, su aplicación hace necesaria una serie de supuestos:

- Las poblaciones actuales son descendientes de las poblaciones del pasado y se encuentran localizadas en la misma región geográfica.
- La diversidad genética de las poblaciones actuales es muy similar a la de las poblaciones del pasado.

Este método ha sido aplicado, entre otros, para determinar el origen, tiempo e impacto de las diferentes migraciones que entraron en Europa y modelaron su composición genética (RICHARDS *et al.* 2000).

2. EL ORIGEN DEL ACERVO GENÉTICO DE LAS POBLACIONES EUROPEAS

Desde la genética de poblaciones se han propuesto diversas hipótesis sobre el origen y procedencia de las poblaciones humanas basadas en el análisis de su composición genética actual.

Muchos autores otorgan a las expansiones geográficas acontecidas durante la prehistoria un papel primordial en la constitución genética de las poblaciones actuales. Estos movimientos, de carácter individual o local, han constituido la mayor fuente de cambio genético y cultural entre grupos, siendo su efecto final la homogenización genética de las poblaciones (CAVALLI SFORZA *et al*, 1993, 1994).

Entre las expansiones poblacionales que han podido modelar el acervo genético de la población europea se encuentran: la primera expansión del hombre moderno fuera de África, las expansiones del Paleolítico Superior, la difusión del Neolítico en Europa y las expansiones post-Neolíticas. El impacto de todas ellas sobre la composición genética de las poblaciones europeas actuales ha sido evaluado a través del análisis de diversos marcadores: marcadores “clásicos”, DNA mitocondrial y, más recientemente, Cromosoma Y. A continuación se comentan algunos de los estudios más importantes y sus principales conclusiones.

2.1. La primera expansión fuera de África

Los primeros pobladores de Europa fueron los descendientes de *Homo erectus* que hace 1.9 millones entraron en Europa por el Cáucaso (Dmanisi) y quizá algo después por Gibraltar (Orce) y entre 1 y 2 millones de años se dispersaron ampliamente por Eurasia (KAIFU *et al.* 2005). Hace unos 350.000 años, poblaciones arcaicas de *Homo sapiens* provenientes de África ocuparon Europa, originando la línea pre-neandertalense (Atapuerca, Sima de los Huesos) y después neandertalense en sentido estricto. Esta línea se extinguió hace 28.000 años aproximadamente (BRAUER 1997).

Los primeros estudios del mtDNA revelaron una mayor diversidad en las poblaciones africanas actuales respecto al resto de poblaciones mundiales analizadas (JOHNSON *et al.* 1983; CANN *et al.* 1987; HORAI Y HAYASAKA 1990; HORAI 1995; VIGILANT *et al.* 1991). La reconstrucción filogenética de estos polimorfismos mediante el método de la máxima parsimonia mostró, además, que las líneas más antiguas se localizaban en África. La datación del árbol filético situaba el origen de las poblaciones humanas en el continente negro hace aproximadamente 180.000 años (CANN *et al.* 1987; VIGILANT *et al.* 1991).

Tales resultados fueron interpretados como evidencia de un origen africano único y reciente del hombre anatómicamente moderno (HAM). Tras la publicación de estos trabajos se abrió un amplio debate entre defensores y detractores de esta teoría. La principal crítica radicaba en los métodos de reconstrucción filogenética empleados (SAITOU y OMOTO 1987; SPUHLER 1988; EXCOFFIER y LANGANEY 1989; KRÜGER y VOGEL 1989; HEDGES *et al.* 1991; MADDISON 1991; TEMPLETON 1991, 1993, 1994; GEE 1992; GOLDMAN y BARTON 1992; PESOLE *et al.* 1992; AYALA 1995). A pesar de la crítica inicial, estudios posteriores en mtDNA y en otros marcadores –genes nucleares y Cromosoma Y– han corroborado las principales conclusiones derivadas de los primeros (HARDING *et al.* 1997, 2000; CLARK *et al.* 1998; HEY y HARRIS 1999; THOMSON *et al.* 2000; INGMAN *et al.* 2000; RICHARDS y MACAULAY 2001; QUINTANA-MURCI *et al.* 1999; TAKAHATA *et al.* 2001; FORSTER 2004).

Los análisis de mtDNA muestran que hay dos líneas casi exclusivamente africanas: los haplogrupos L1 y L2. El haplogrupo L1 es el más antiguo (130.000 años) y contiene la raíz de la filogenia mundial de mtDNA (FORSTER 2004). Algunas líneas de este clado son hoy exclusivas de ciertas regiones geográficas. Así las líneas L1d y L1k aparecen en grupos bosquimanos de África del Sur, el clado L1c se localiza en África Central y del Oeste y los clados L1e y L1f en ciertas poblaciones de África del Este (WATSON *et al.* 1997). Los haplogrupos más ampliamente distribuidos y predominantes en África son L2 y L3. Surgieron en África, entre hace 80.000 y 60.000 años, y se expandieron por todo el continente desplazando a las líneas L1 preexistentes. Fuera de África los linajes del mtDNA se engloban en dos grandes super-clusters M y N, derivados del haplogrupo L3, que reúnen el resto de haplogrupos mundiales (**Figura F12**). Ello ha llevado a muchos autores a hipotetizar que la expansión desde África estuvo formada esencialmente por portadores de líneas L3 los cuales se habrían diferenciado después en M y N, dentro o fuera del continente africano. Los haplogrupos M y N se expandieron por toda Europa y Asia diversificándose y dando lugar a variantes específicas de cada continente. M dio lugar a la mayoría de haplogrupos asiáticos: C, D, E, G, Z y Q, y N a los europeos y algún asiático: A, I, W, X, Y, P, B, F, H, U y JT. Hace aproximadamente 25.000 años una o varias migraciones desde Liberia, a través del estrecho de Bering, habría llevado los haplogrupos A, B, C y D al continente americano.

El estudio de otro marcador haploide, el Cromosoma Y, ha proporcionado un panorama muy similar. En la filogenia reconstruida a partir de marcadores binarios de la porción no recombinante de este cromosoma se distinguen dos líneas predominantemente africanas (A y B). El resto de variantes, tanto en África como fuera de ella, comparten dos sustituciones características (M168 y P9) que se han relacionado con las líneas que abandonaron África y poblaron Eurasia. Estas líneas se subdividen en tres: C, DE y F, siendo el último grupo el más ampliamente distribuido fuera de África (Y CHROMOSOME CONSORTIUM 2002; JOBLING *et al.* 2003).

Las fechas para la separación entre las líneas estrictamente africanas y aquellas distribuidas por el resto del mundo obtenidas a partir de las filogenias de estos dos marcadores teniendo en cuenta los intervalos de confianza, son consistentes y permiten situar dicha separación hace entre 40.000 y 60.000 años.

En la vertiente opuesta a la teoría de un origen único africano reciente “*Out of Africa*” se encuentra la Teoría Multiregional (THORNE y WOLPOFF 1992) según la cual existió un flujo genético importante entre los humanos anatómicamente modernos y las poblaciones descendientes directas de *Homo erectus*, los Neandertales en el caso del continente Europeo. De ser cierta esta última hipótesis cabría buscar los orígenes de la diversidad genética europea mucho más atrás en el tiempo, probablemente en el stock ancestral de *Homo erectus*.

El avance de las técnicas de extracción y amplificación de DNA fósil ha permitido la recuperación de secuencias de mtDNA de varios individuos Neandertales de diferente procedencia (KRINGS *et al.* 1997, 2000; OVCHINNIKOV *et al.* 2000; SERRE *et al.* 2004). El número y naturaleza de los cambios nucleotídicos y el análisis de las diferencias por parejas (MPD) muestran que la variabilidad genética de dichos especímenes está fuera de la variabilidad genética humana actual, excluyendo por tanto la posibilidad de una contribución genética importante de los Neandertales al *stock* genético actual (más información en el apartado 3.3.1. de este mismo capítulo).

2.2. La difusión del Neolítico en Europa

2.2.1. Neolítico y neolitización

J. Lubbock en 1856 acuñó por primera vez el término “Neolítico” (véase LUBBOCK 1987). En ésta su definición original, se refiere exclusivamente al cambio tecnológico consistente en la aparición de un nuevo conjunto útil de piedra pulimentada (*neos-lithos*, nueva piedra), que vendría a sustituir al de piedra tallada característica del Paleolítico. Como resultado de la investigación arqueológica esta acepción inicial se ha visto completada por otro conjunto de transformaciones de aparición más o menos simultánea: cambios económicos como la transformación de la economía de subsistencia en una economía de producción resultado de la domesticación de plantas y animales y el consecuente desarrollo de la agricultura y la ganadería; cambios sociales, como la aparición de poblados resultado de la sedentarización de los últimos grupos cazadores-recolectores o el desarrollo de intercambios, y cambios culturales y/o artesanales como la aparición de la cerámica. El concepto de Neolítico, tal y como hoy lo entendemos, define un período arqueológico de profundas transformaciones de carácter tecnológico, social, cultural y económico clave para el desarrollo de las sociedades humanas modernas.

Con el término “neolitización” se designa el conjunto de procesos que operaron en la transición del modo de vida cazador recolector al productor de alimentos. Los mecanismos que propiciaron esta transición, la naturaleza de sus artífices, y el mecanismo de su adopción prácticamente universal como método de subsistencia continúan todavía sin resolver y han sido objeto de un continuo debate en el último siglo.

2.2.2. Origen y difusión del Neolítico en Europa

Existe consenso entre los diferentes investigadores acerca de que la economía agro-pastoral (Neolítico) se originó hace aproximadamente 10000 años en ciertas regiones de Oriente Próximo, lo que se ha denominado el creciente o la media luna fértil y que se corresponde con los territorios actuales de Siria e Israel.

Desde un punto de vista clásico se considera que la difusión del Neolítico en Europa incluye todos los elementos novedosos: domesticación de plantas y animales (agricultura y ganadería), herramientas de piedra pulida, construcción de viviendas permanentes y cerámica, conjunto que recibe el nombre de *paquete neolítico*. Existen tres grandes complejos cerámicos asociados a la difusión del Neolítico: el complejo de cerámica pintada de los Balcanes-Anatolia, la cultura cerámica impresa cardial y la cultura de cerámica lineal (LBK). Desde Oriente Próximo el *paquete neolítico* se habría expandido siguiendo básicamente dos rutas, una a través de la costa Mediterránea hacia Italia, Francia y finalmente la Península Ibérica asociada a la cerámica cardial y otra hacia el centro y norte de Europa asociada al complejo cerámico LBK. La velocidad de difusión del Neolítico en Europa ha sido determinada datando por radiocarbono los depósitos con cerámica y evidencias de domesticación de los yacimientos neolíticos más tempranos. La proyección de estas fechas sobre un mapa evidencia un gradiente sureste-noreste que sugiere que la transición al modo de vida Neolítico de Europa habría tardado 2500 años en completarse (CLARK 1965; PARZINGER 1993; MÜLLER 1994; AMMERMAN y CAVALLI-SFORZA 1984). (**Figura FI2**).

Ciertos hallazgos recientes contradicen la difusión conjunta de todos los elementos que conforman el *paquete neolítico*. Se ha documentado por ejemplo la existencia de asentamientos permanentes y la domesticación de especies animales y vegetales anteriormente a las fechas proporcionadas por las dataciones de radiocarbono anteriormente citadas (PRICE 2000).

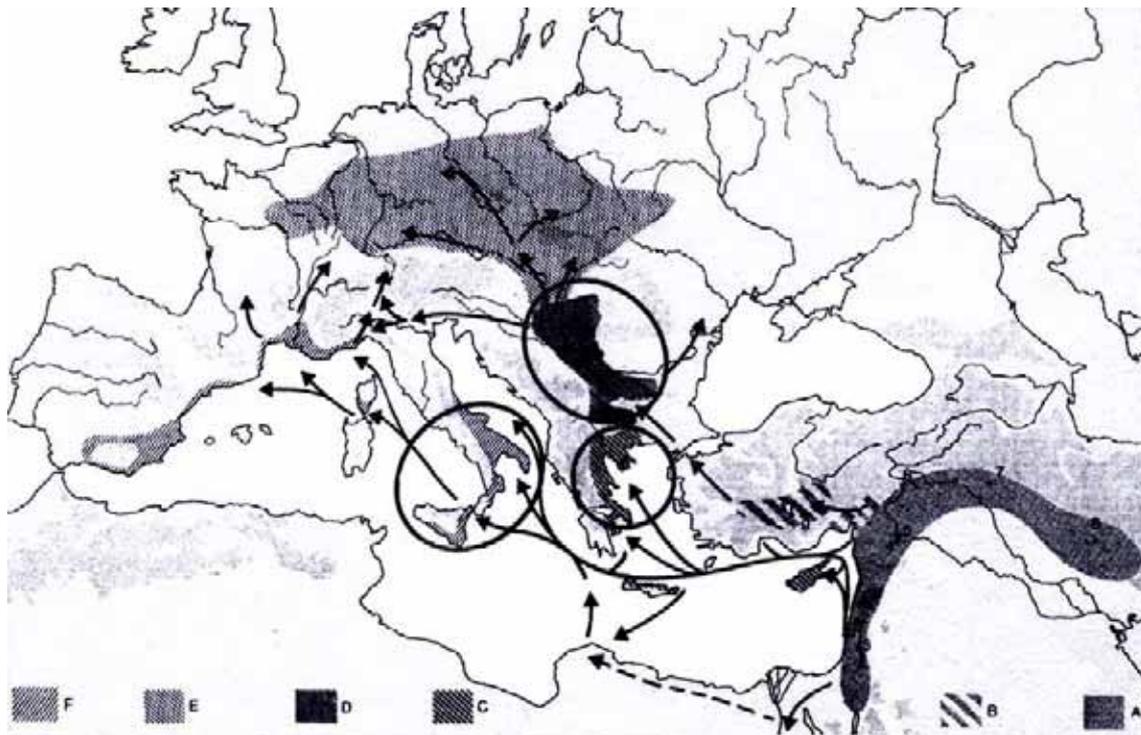


Figura FI2. Rutas de difusión del Neolítico desde Próximo Oriente a Europa. Tomado de BUDJA 2001.

2.2.3. Naturaleza de la expansión del Neolítico en Europa. Implicaciones genéticas.

Si bien se conocen el foco de origen del Neolítico y sus rutas de dispersión en Europa la naturaleza de dicha dispersión a día de hoy sigue siendo una cuestión muy controvertida. En el debate sobre el mecanismo de transición neolítica se barajan básicamente dos puntos de vista: el modelo *difusionista* vs. el *indigenista* o de *aculturación*.

El modelo *difusionista* mantiene que la transición neolítica en Europa se produjo como consecuencia de una sustitución poblacional de los grupos cazadores recolectores por parte de una comunidad agrícola en expansión procedente de Próximo Oriente (CHILDE 1925; PIGGOTT 1965; CASE 1969; LICHARDUS y LICHARDUS-ITTEN 1985; AURENCHE y CAUVIN 1989; CAUVIN 1994; AMMERMAN y CAVALLI-SFORZA 1984; CAVALLI-SFORZA *et al.* 1994; CAVALLI-SFORZA y MINCH 1997).

Según el modelo *indigenista* la transición neolítica tuvo lugar mediante la adopción de este sistema productivo por parte de los grupos locales sin recambio

poblacional (DENNELL 1983; BARKER 1985; TILLEY 1994; THOMAS 1988, 1996; WHITTLE 1996; PLUCIENNIK 1998).

Entre ambos extremos se encuentran modelos intermedios que postulan cierto grado de mezcla entre los inmigrantes y los grupos locales de cazadores recolectores, propiciada por diferentes mecanismos como la difusión démica (AMMERMAN y CAVALLI-SFORZA 1984), dominancia de élite (RENFREW 1987), la colonización pionera (ARNAUD 1982; ZILHAO 1993; RENFREW 1996; RENFREW y BOYLE 2000) o la movilidad de fronteras (ZVELEBIL 2001) entre otros.

La consideración de uno u otro modelo tiene diferentes implicaciones sobre el origen de la composición genética de las poblaciones europeas actuales. De haberse tratado de una expansión principalmente démica la variabilidad genética de las poblaciones actuales procedería esencialmente de Oriente Próximo y no se remontaría más allá de los orígenes del Neolítico (10000A.P.). Se esperaría por lo tanto encontrar una gran similitud genética entre las poblaciones Europeas y de Oriente Próximo. Sin embargo si el fenómeno fue más cultural que démico, la expansión del Neolítico no habría modificado la composición genética de Europa, que tendría su origen más atrás en el tiempo, en las expansiones del Paleolítico o antes si se tiene en consideración la Hipótesis Multiregional. En este caso la distinción entre las poblaciones de Europa y Oriente Próximo dependerá del grado de diferenciación genética acumulada individualmente. Un modelo intermedio en el que los grupos de agricultores se mezclaran con los grupos de cazadores-recolectores locales generaría una clina o gradiente de frecuencias con un máximo en Próximo Oriente que disminuiría hacia el noroeste.

Los estudios de la composición genética de las poblaciones europeas y de Próximo Oriente actuales han tratado de determinar el impacto de las expansiones neolíticas sobre el acervo genético europeo. Existe cierto consenso general respecto a que ninguna de las dos hipótesis extremas, difusión estrictamente poblacional o cultural, es realista. El debate generado en torno a estos estudios afecta al grado relativo de contribución genética de la población autóctona e inmigrante al acervo genético europeo actual.

2.2.3.1. Estudios de frecuencias de marcadores clásicos

El estudio de las frecuencias de ciertos marcadores en las poblaciones europeas reveló la existencia de diferentes gradientes o clinas interpretadas como firmas de expansiones poblacionales del pasado, siendo la expansión del Neolítico desde Próximo Oriente la más evidente (MENOZZI *et al.* 1978; CAVALLI-SFORZA *et al.* 1993; CAVALLI-SFORZA *et al.* 1994; PIAZZA *et al.* 1995; SEMINO *et al.* 1996).

En el primero de estos estudios (MENOZZI *et al.* 1978) se construyeron componentes principales y se trazaron mapas sintéticos a partir de las frecuencias de 38 alelos de 10 loci diferentes, dos del Sistema de Histocompatibilidad HLA (HLA-A y HLA-B) y el resto de diversos grupos sanguíneos, enzimas plasmáticas y proteínas eritrocitarias (ABO, Rh, MN, Lewis, Duffy, Haptoglobina, Fosfatasa ácida y Fosfoglucomutasa). Posteriormente, los mismos autores ampliaron este estudio al análisis de 95 genes, confirmando y completando los resultados obtenidos anteriormente (CAVALLI-SFORZA *et al.* 1994).

El mapa sintético de la primera componente principal, que resumía cerca del 30% de la varianza total, mostraba un gradiente Sudeste-Noroeste con foco en Próximo Oriente que prácticamente se superponía con el mapa de la expansión de la agricultura en Europa reconstruido a partir de fechas de radiocarbono (AMMERMAN y CAVALLI-SFORZA 1984) (**Figura FI3. Mapas A y B**). Ello fue interpretado en favor de un modelo démico de expansión del Neolítico en Europa, con un profundo impacto en la constitución genética de las poblaciones. Un resultado similar fue obtenido en la primera componente principal del estudio de PIAZZA *et al.* 1995, en el que fueron analizados 35 genes (**Figura FI3 C**). En la **Tabla I2** se resumen los porcentajes de variabilidad total explicados por cada una de las diferentes componentes principales de estos estudios.

Los resultados fueron confirmados por el estudio de SOKAL *et al.* 1991. Estos autores emplearon un *análisis de autocorrelación espacial* para estimar el grado de significación estadística entre las distancias genéticas observadas y unas distancias especialmente diseñadas para representar la difusión de la agricultura manteniendo constantes las distancias geográficas.

Tabla 12. Porcentajes de variabilidad explicados por las diferentes componentes principales en tres estudios de frecuencias de marcadores clásicos.

Componentes principales	Porcentaje de variabilidad explicado por las componentes principales		
	Menozzi <i>et al.</i> 1978 38 genes (HLA y no HLA)	Cavalli-Sforza <i>et al.</i> 1994 95 genes (HLA y no HLA)	Piazza <i>et al.</i> 1995. 35 genes (HLA y no HLA)
Primera	27,0	28,1	26,0
Segunda	18,4	22,2	20,6
Tercera	11,0	10,6	14,3
Cuarta		7,0	
Quinta		5,3	
Sexta		3,2	
Total 3 primeras	56,4	60,9	60,9

RENDINE *et al.* 1986 tras simular un patrón de múltiples expansiones démicas en Europa, demostraron que la impronta genética dejada por las migraciones tempranas era detectable mediante un análisis de componentes principales. Las clinas establecidas por estas migraciones prehistóricas eran muy resistentes a movimientos migratorios continuos posteriores. La simulación puso sin embargo de manifiesto la dificultad a la hora de separar algunas de estas posteriores expansiones demográficas menores. (CAVALLI-SFORZA *et al.* 1994).

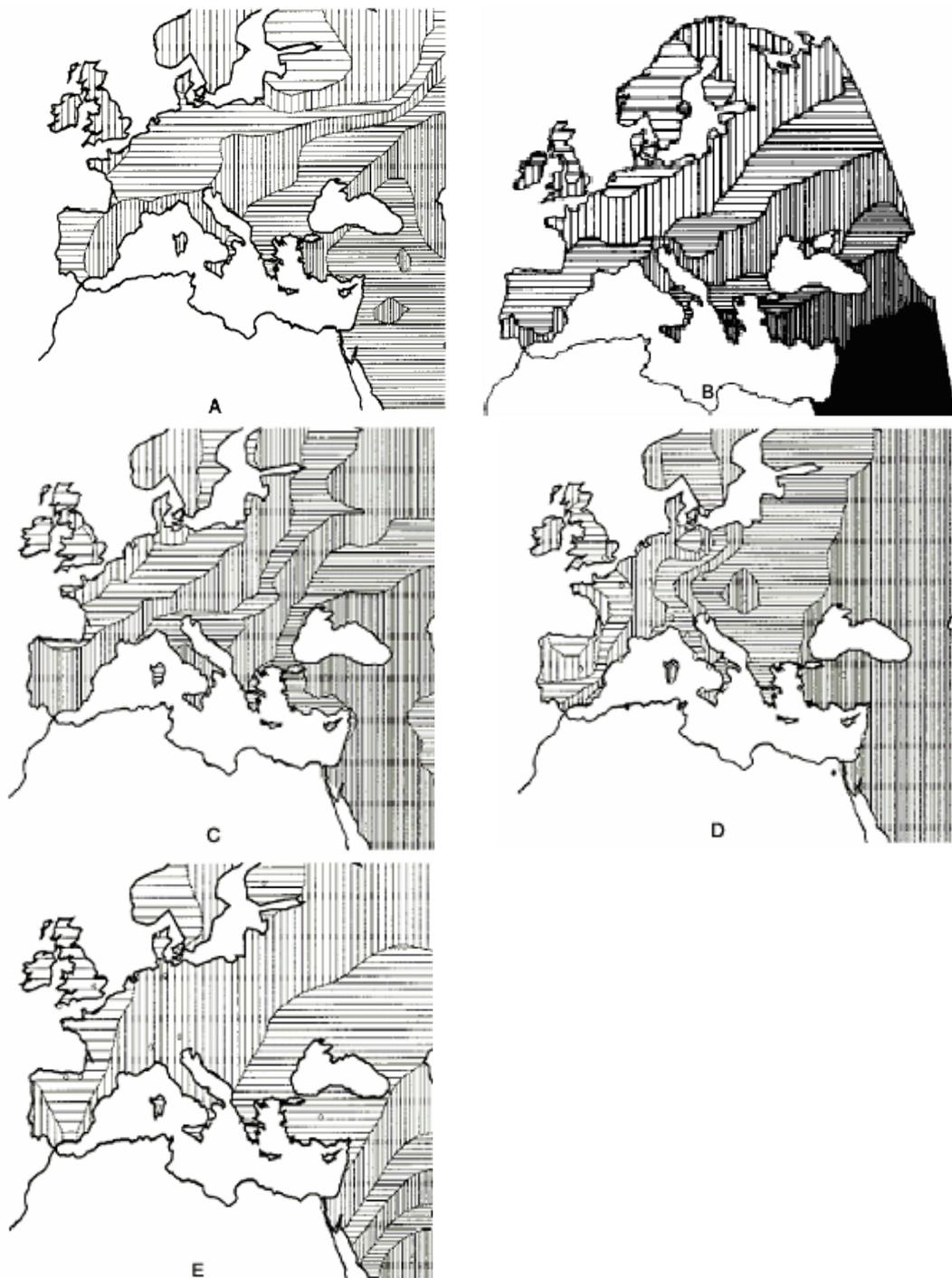


Figura FI3. Representación espacial del primer componente principal de los estudios de B: CAVALLI-SFORZA *et al.* 1994 (95 marcadores autosómicos); C: PIAZZA *et al.* 1995 (91 marcadores autosómicos); D: SEMINO *et al.* 1996 (2 STRs cromosoma Y). El mapa E fue reconstruido por CAVALLI-SFORZA y MINCH 1997 a partir de los datos de mtDNA de RICHARDS *et al.* 1996. A: Difusión de la agricultura en el Neolítico basado en fechas de radiocarbono de 106 yacimientos arqueológicos. Modificada de CAVALLI-SFORZA y MINCH 1997.

2.2.3.2. Estudios de DNA mitocondrial

Los patrones de diversidad mitocondrial europea contradicen los estudios de marcadores clásicos y proponen un origen Paleolítico para la diversidad genética de las poblaciones europeas actuales.

El primero de estos estudios (RICHARDS *et al.* 1996) abrió un intenso debate entre los defensores de la teoría difusionista y los de la teoría de aculturación. El análisis de la variabilidad de la HVSI del mtDNA de 821 individuos de 14 poblaciones europeas reveló la existencia de cinco clusters o haplogrupos mitocondriales principales que fueron nombrados del 1 al 5 y de algunos subgrupos dentro de éstos (2A, 2B y 3A). La estima del TMRCA de cada haplogrupo proporcionó fechas anteriores a la difusión del Neolítico en todos los casos, menos en dos subgrupos del cluster 2A distribuidos en el oeste (2A-W) y centro (2A-C) de Europa. Este resultado fue interpretado como un origen predominantemente Paleolítico de la diversidad mitocondrial europea, con una contribución Neolítica minoritaria (12%) restringida a las variantes dentro del sub-cluster 2A. Según estos autores la difusión de la agricultura fue un fenómeno predominantemente cultural con muy poca difusión démica.

El mismo equipo investigador en posteriores estudios procedió a incrementar los tamaños muestrales de las poblaciones europea (RICHARDS *et al.* 1998), y de Oriente Próximo (MACAULAY *et al.* 1999a; RICHARDS *et al.* 2000). Introdujeron también una nomenclatura cladística para designar los clusters aunando la información acumulada de los análisis de restricción enzimática (TORRONI *et al.* 1994b, 1996; CALAFELL *et al.* 1996; COMAS *et al.* 1996) y de HVSI (CÔRTE-REAL *et al.* 1996; RICHARDS *et al.* 1996) e incorporando nueva información de la HVSII para su correcta clasificación. Esta información permitió una disección aún más detallada de los diferentes haplogrupos mitocondriales europeos (RICHARDS *et al.* 1998).

En el trabajo de RICHARDS *et al.* 2000 se empleó una nueva aproximación metodológica para el análisis de la variabilidad mitocondrial de la población europea. Mediante el análisis fundador (ver apartado 1.3.2.4.2. del presente capítulo) se detectaron aquellas variantes originarias de Próximo Oriente que habían migrado a Europa y posteriormente se estimó el tiempo aproximado en que estas migraciones

habían tenido lugar. Con este método se pudieron identificar las variantes asociadas a cada una de las expansiones poblacionales que supuestamente han modelado el acervo genético de las poblaciones europeas actuales.

Los resultados de estos estudios corroboraron la principal hipótesis del trabajo de RICHARDS *et al.* 1996, según la cual la mayor parte de la variabilidad mitocondrial europea actual tiene sus orígenes en las expansiones desde Próximo Oriente durante el Paleolítico. Según RICHARDS *et al.* 2000 el haplogrupo U fue introducido en Europa al inicio del Paleolítico Superior (*Early Upper Palaeolithic* o EUP) por las expansiones del hombre anatómicamente moderno. Los haplotipos fundadores de los haplogrupos HV*, U1 y posiblemente también de U2 y U4 llegaron después, durante el Paleolítico Superior Medio (MUP). A finales del Paleolítico Superior (*Late Upper Palaeolithic* o LUP) penetraron en Europa desde Próximo Oriente algunos sub-clusters del haplogrupo H, como el H-CRS, H-16304 y H-16362,16482. Únicamente aquellas variantes situadas en la base de los haplogrupos J, T1, U3 y algunos sub-clusters de los haplogrupos H y W podrían haber sido introducidos en Europa durante las expansiones del Neolítico (**Figura F14**).

Según estos estudios, teniendo en cuenta la frecuencia de los diferentes haplotipos fundadores en la actualidad en Europa y Próximo Oriente, el porcentaje de variabilidad genética actual asociada a las expansiones neolíticas no sería superior al 23%. El porcentaje asociado a estas expansiones en las poblaciones mediterráneas sería de un 9-12% mientras que en centro Europa éste sería sensiblemente mayor, entre un 15 y un 22% aproximadamente.

Tabla 13. Estimaciones del TMRCA de los diferentes haplogrupos y sub-haplogrupos Europeos realizadas en tres estudios.1: RICHARDS *et al.* 1996; 2: RICHARDS *et al.* 1998; 3: RICHARDS *et al.* 2000. Los haplogrupos se nombran con letras mayúsculas de acuerdo con TORRONI *et al.* 1992. Entre paréntesis se señala la nomenclatura empleada por RICHARDS *et al.* 1996.

Haplogrupo	Europa					Próximo Oriente		
	Filogenia	TMRCA			Filogenia	TMRCA		
		1	2	3		1	3	
HV	Estrella			24300-29000	Estrella		20700-22800	
H (1)	Estrella	23500	20500±2500	23200-28400	Estrella	52500	19200-21400	
V (1)	Estrella		12500±3000	9500-43900	Estrella		11100-16900	
HV1	Estrella			11300-24800	Estrella		22200-58300	
Pre-HV	Estrella			18600-31800	Estrella		15400-41600	
U	Estrella			50400-58300	Estrella		53600-58900	
K (4)	No estrella	17500	15500±3000	15500-25500	No estrella		12900-18300	
U1a	Estrella			17000-33100	Estrella		20500-49900	
U1b	Estrella			14000-40800	Estrella		2400-56200	
U2	No estrella			14000-42300	Estrella		23600-48000	
U3 (1)	Estrella			16300-26600	Estrella		11900-26800	
U4 (1)	Estrella		25000	16300-35500	No estrella		16100-24700	
U5 (5)	No estrella	36500	52500±6500	46000-75000	No estrella		45100-52800	
	U5a		40000					
	U5a1		30000					
	U5a1a		20000					
U6 (6)								
U7	No estrella			23900-53600	No estrella		11900-45400	
JT (2)		35500				57500		
J (2A)	No estrella	23500	28000±400	42400-53700	No estrella	25000	22000-27400	
	J1a (2A-C)	6000	6500±3000					
	J1b1 (2A-W)	12500						
T (2B)	No estrella	35500	46500±6000	41900-52900	No estrella	42000	33100-40200	
	T1	Estrella		16700-28400	Estrella		6100-12800	
3		50500				65000		
I (3A)	No estrella	34000	35000±8500	32300-58400	No estrella	21000	27200-40500	
X (3B)	Estrella		24000±6500	13700-26600	Estrella		17000-30000	
W (3C)	Estrella		18500±6500	18000-38400	Estrella		17100-28400	
N1b	Estrella			8900-24900	Estrella		21100-59300	

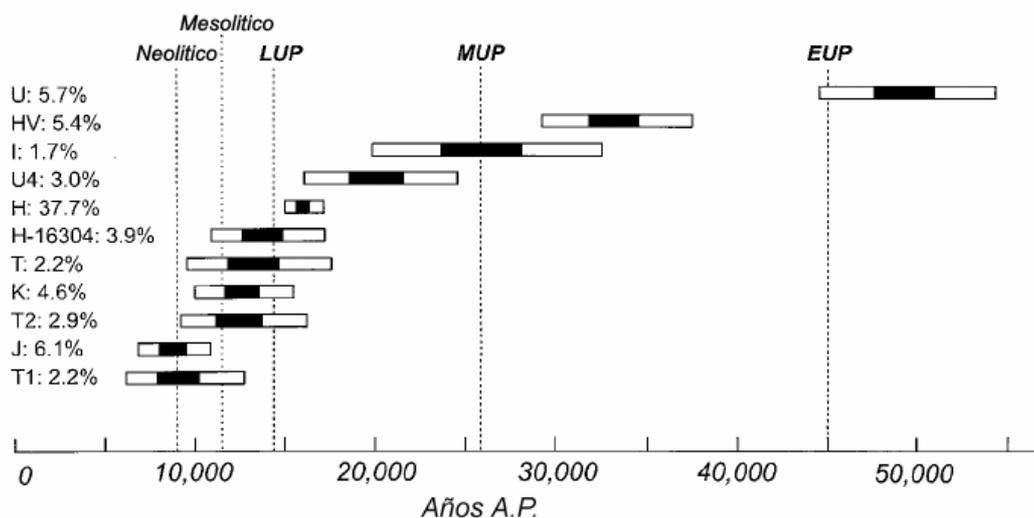
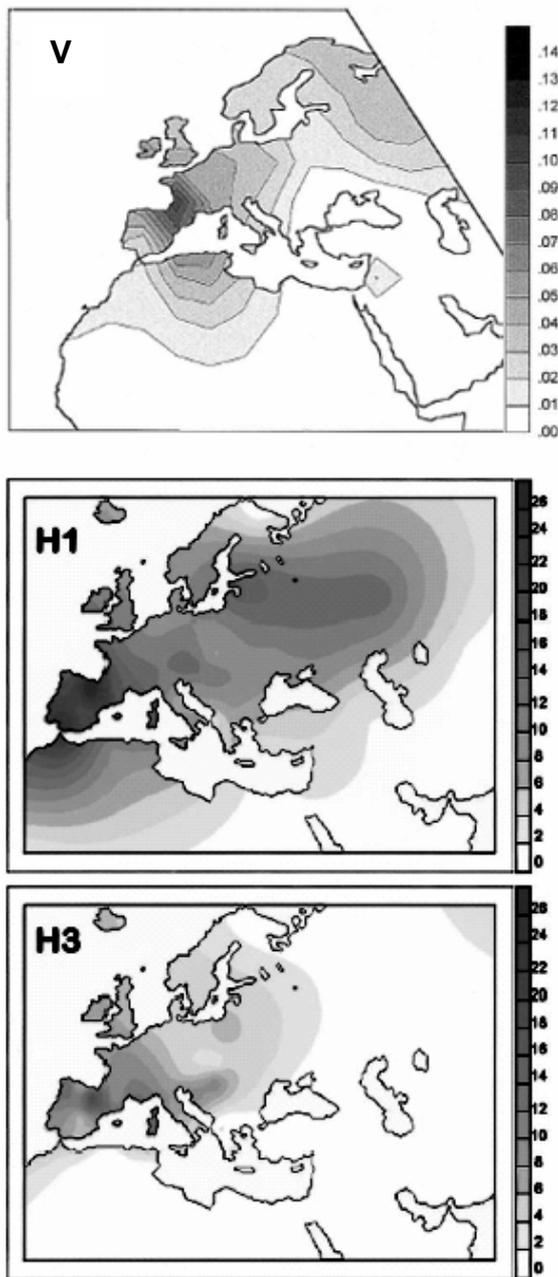


Figura F14. Porcentaje de fundadores principales de cada haplogrupo y sub-haplogrupo introducidos durante diferentes periodos. EUP: Paleolítico Superior Inicial; MUP: Paleolítico Superior Medio; LUP: Paleolítico Superior Final. Tomada de RICHARDS *et al.* 2000.



Como corolario a estos trabajos, el equipo encabezado por Antonio Torroni, tras analizar la frecuencia y distribución actual de ciertos haplogrupos europeos como el V y de ciertos subgrupos del haplogrupo H como el H1 y H3, concluyó que estos clusters, presentes ya en Europa a finales del Paleolítico Superior, sufrieron nuevas expansiones desde la franja Cantábrica hacia el norte tras la última glaciación hace aproximadamente 11000 años (TORRONI *et al.* 1998b; TORRONI *et al.* 2001; ACHILLI *et al.* 2004). En la **Figura F15** se muestran los mapas de frecuencias de estos haplogrupos y sub-haplogrupos.

Figura F15. Distribución espacial de las frecuencias en Europa del haplogrupo V y de los subhaplogrupos H1 y H3. Tomado de TORRONI *et al.* 2001 (haplogrupo V) y ACHILLI *et al.* 2004 (subhaplogrupos H1 y H3).

Las críticas realizadas por los defensores de la hipótesis démica de expansión del Neolítico hacia estos estudios hacen referencia a la validez de la región control del mtDNA como marcador para la reconstrucción del pasado de las poblaciones humanas, y al análisis basado en la comparación de haplotipos (CAVALLI-SFORZA y MINCH 1997). Estos últimos autores, sin embargo, han reconocido en posteriores trabajos que la contribución de los inmigrantes neolíticos al acervo mitocondrial europeo fue de un 26% (SEMINO *et al.* 2000), valor muy similar al deducido a partir de los estudios de mtDNA.

2.2.3.3. Estudios de Cromosoma Y

Los estudios de cromosoma Y proporcionan evidencias contradictorias frente a una contribución predominantemente paleolítica o neolítica en el acervo genético europeo actual.

Por una parte la distribución de frecuencias de algunos polimorfismos de la región no recombinante del cromosoma Y (NRY) muestra un gradiente sureste-noroeste muy similar al mapa sintético de la primera componente principal de los estudios realizados con marcadores clásicos, que ha sido interpretada como la huella genética de la expansión neolítica (SEMINO *et al.* 1996, 2000; ROSSER *et al.* 2000). Las variantes asociadas a dicha expansión, tal y como sucedió con el mtDNA, se han identificado y en algunos casos cuantificado y datado.

La variante “neolítica” del estudio de SEMINO *et al.* 1996, la delección 12f2, identifica el haplogrupo J de ROSSER *et al.* 2000, que presenta una distribución similar. SEMINO *et al.* 2000 realizaron una disección más detallada de la filogenia del cromosoma Y, estudiando 22 marcadores binarios que permitieron identificar 6 haplogrupos diferentes. Entre éstos, el F*, E3b, G y J2 mostraban un gradiente de frecuencias compatible con la expansión del Neolítico. El porcentaje de estas líneas en la actualidad sirvió a estos autores para realizar una estima de la contribución Neolítica al acervo genético Europeo de un 22%.

Además de identificar las variantes actuales de origen supuestamente neolítico, los tres estudios citados distinguen también aquellos haplotipos remanentes de un sustrato paleolítico anterior. SEMINO *et al.* 1996 señaló ya el haplotipo 15 del marcador 49f como posible signatura paleolítica en el *pool* genético europeo actual. Este haplotipo se corresponde con el haplogrupo R* de ROSSER *et al.* 2000, de distribución geográfica similar. SEMINO *et al.* 2000 relacionó además los Y-haplotipos R e I respectivamente con las culturas Auriñaciense (35.000-40.000 años) y Gravetiense (22.000 años) del Paleolítico Superior. La distribución geográfica de frecuencias de otras variantes con máximo en centro-norte Europa (E, R1a) y al este del báltico (N3) ha sido interpretada como evidencia de las expansiones post-glaciales desde los refugios de la Península Ibérica y los Balcanes.

En contraposición a estos estudios la distribución de las diferencias por parejas (*mismatch distribution*) de 22 SNPs del cromosoma Y no muestra el patrón acampanado característico que se esperaría bajo un modelo de expansión démica (PEREIRA *et al.* 2001).

2.2.3.4. Estudios de simulación

Las estimas de los porcentajes de contribución de genes neolíticos y paleolíticos al acervo genético europeo llevadas a cabo mediante simulaciones informáticas han proporcionado también resultados totalmente opuestos.

CHIKHI *et al.* 2002, tras reanalizar los datos de SEMINO *et al.* 2000, sugieren que la contribución de genes neolíticos en Europa fue muy considerable (entre un 50 y un 60%) y superior a lo estimado hasta el momento (22% según SEMINO *et al.* 2000 y RICHARDS *et al.* 2000). En estos últimos estudios la medida de la aportación neolítica se basa en la cuantificación de las variantes cuya distribución de frecuencias sigue un patrón compatible con la dispersión del Neolítico en Europa. En el trabajo de CHIKHI *et al.* 2002 la composición genética de las poblaciones ancestrales “neolítica” y “paleolítica” se estima, sin embargo, a partir de la de sus poblaciones descendientes; Turquía, Líbano y Siria en el primer caso y País Vasco o Cerdeña en el segundo. Se modeliza, además, la variabilidad genética acumulada en ambas poblaciones por efecto del aislamiento y la deriva genética.

De forma similar CURRAT y EXCOFFIER (2005) simulan el patrón de diversidad esperada en un marcador como los SNPs de acuerdo con la expansión neolítica, teniendo en cuenta diferentes grados de mezcla con la población paleolítica. De estas simulaciones se concluye que incluso una contribución pequeña de genes del Paleolítico habría sido suficiente para borrar la huella genética de la difusión de los primeros agricultores desde Próximo Oriente.

Tanto CHIKHI *et al.* 2002 como CURRAT y EXCOFFIER 2005 coinciden en la importancia de la selección de los haplotipos para estimar correctamente la historia demográfica de una región. La elección de aquellas variantes más frecuentes en

PEREIRA *et al.* 2001 y SEMINO *et al.* 2000, podría explicar las diferencias entre sus respectivos estudios y el de CHIKHI *et al.* 2002.

2.2.4. Consideraciones acerca de las conclusiones derivadas de los estudios de variabilidad genética en poblaciones europeas

Independientemente de los resultados obtenidos, todos los estudios anteriormente mencionados parten de un conjunto de supuestos que no tienen por qué ser necesariamente ciertos:

- Se piensa, en todos los casos, que Próximo Oriente es el foco del que parten la mayoría de expansiones prehistóricas, incluidas las del Paleolítico y la del Neolítico.
- Se supone, también, que las poblaciones europea y de Próximo Oriente estaban lo suficientemente diferenciadas genéticamente cuando se produjeron los diferentes eventos migratorios.
- Se considera que la variabilidad genética de Europa y de Próximo Oriente es representativa de la existente en estas áreas cuando se produjeron las migraciones paleolíticas o neolíticas.

Llegados a este punto, los estudios de DNA antiguo disponen del potencial para testar la validez de algunos de estos supuestos.

2.3. Expansiones post-neolíticas

Los estudios genéticos de diferentes marcadores sitúan el origen de la estructura genética de las poblaciones europeas bien en las expansiones del Paleolítico, o bien en la difusión de los agricultores desde Próximo Oriente a inicios del Neolítico. Sin embargo, muchas expansiones pudieron haber dejado huella en el panorama genético europeo, algunas prehistóricas no documentadas arqueológicamente y otras bien documentadas posteriores al neolítico.

Entre estas últimas citemos, por ejemplo, la dispersión del megalitismo durante el Calcolítico (6000-5000 A.P). Destaca también la expansión del campaniforme en la Península Ibérica, Reino Unido, Francia y Centro Europa y de otras culturas cerámicas en el Norte de Europa (5500-4500 A.P.).

Por otra parte, en las estepas del este de Europa se desarrolló, hace unos 5000 años, un segundo tipo de economía ganadera, conocida como el “pastoreo nómada”, que se caracterizaba por la domesticación del caballo. Esta cultura dominó toda la estepa (3500 A.P.) y se extendió luego hacia el sur, India, península de Anatolia y Mesopotamia. Marija Gimbutas relaciona la expansión de esta cultura hacia el oeste con la difusión de la lengua Indoeuropea (GIMBUTAS 1970).

Durante la edad de los metales hubo, asimismo, diversos eventos de difusión cultural que pudieron estar asociados a movimientos poblacionales paralelos. Uno de los más destacados es sin duda el que llevó la tradición de los campos de urnas (*urnenfelder*) desde Centro Europa al este, al sur y en menor medida al norte y al oeste a finales de la Edad de Bronce (3200 años A.P.). Ya en la Edad de Hierro las culturas celtas de Hallstat y la Tene se extendieron desde su foco de origen en centro Europa (3000 A.P.) hasta ocupar todo el sureste y suroeste del continente (400-200 A.P).

Finalmente, entre las colonizaciones históricas más relevantes cabe citar las de las colonias griegas, las guerras púnicas –que introdujeron contingentes norteafricanos en Iberia–, el imperio romano, las invasiones germánicas que se sucedieron a su caída, los movimientos poblacionales durante la expansión del Islam y las Cruzadas. Añádase el sesgo que provocaron mortandades medievales en la estructura genética europea.

Ciertos autores consideran nulo el impacto estas últimas expansiones en la constitución genética de las poblaciones europeas, mientras que otros les atribuyen un efecto minoritario, como es el caso de ciertos estudios de frecuencias alélicas de marcadores genéticos clásicos (MENOZZI *et al* 1978; CAVALLI-SFORZA *et al* 1993, 1994; PIAZZA *et al* 1995). Según CAVALLI-SFORZA *et al.* 1994 estas expansiones no han contribuido sustancialmente al acervo genético europeo debido a que la densidad poblacional existente en Europa ha sido muy alta desde el Neolítico.

3. DNA ANTIGUO

3.1. Definición de DNA antiguo

Según Herrmann y Hummel 1994, el DNA antiguo (aDNA: ancient DNA) es “cualquier parte o traza de DNA procedente de un organismo vivo o de partes de él, así como DNA extracorpóreo de un organismo vivo o de una parte de él, que sirva como material de partida en los estudios de biología evolutiva, antropología histórica o ciencia forense”.

3.2. Historia

3.2.1. Los primeros pasos de una nueva disciplina

El primer éxito en la recuperación de secuencias antiguas de DNA consistió en la clonación molecular del DNA obtenido de piel modificada del *Equus quagga*, un équido de África del Sur extinguido hace unos 140 años (HIGUCHI *et al.* 1984). Tras comparar las secuencias del *quagga* con las de otros équidos, estos autores descubrieron que mostraban un estrecho parentesco con las cebras. Un año después Svante Pääbo publicaba secuencias de DNA obtenidas de pieles de momias, de 2500 años de antigüedad, empleando el mismo procedimiento experimental (PÄÄBO 1985a y b). Sin embargo, la escasez del material genético conservado en los restos antiguos, su elevada fragmentación y alteración molecular hizo inviable la replicación de estos experimentos, condición necesaria para probar su autenticidad (PÄÄBO *et al.* 1989; HANDT *et al.* 1994b).

El desarrollo de la reacción en cadena de la polimerasa (*PCR*) en 1985 (SAIKI *et al.* 1985; MULLIS y FALOONA 1987) supuso una auténtica revolución dentro del campo del DNA antiguo, al posibilitar la obtención de millones de copias a partir de cantidades ínfimas y muy fragmentadas de DNA original. Se empleó por primera vez en un extracto de aDNA poco después de su desarrollo. Svante Pääbo y Alan Wilson (PÄÄBO y WILSON 1988) replicaron el trabajo de HIGUCHI *et al.* 1984 empleando la *PCR*. Hallaron que dos de las sustituciones detectadas en el mtDNA del *quagga* no eran

correctas. En el experimento de Higuchi, las posiciones dañadas en la hebra de DNA original, probablemente habían sido reparadas por enzimas bacterianas provocando los errores de la secuencia final.

Estos resultados tan prometedores impulsaron a otros equipos de investigación a estudiar restos antiguos. DORAN *et al.* 1986 demostraron la existencia de DNA en un cerebro humano, de 8000 años de antigüedad, excelentemente preservado procedente de un depósito de turba del fondo de un estanque. Por su parte ROLLO *et al.* 1988, tras analizar diversas muestras de maíz precolombino (1500 años A.P.), obtuvieron dos secuencias cortas correspondientes a una subunidad del gen de la citocromo c oxidasa y a una fracción repetitiva de heterocromatina (H2a).

En un excelente trabajo publicado en 1989, Svante Pääbo profundizó en el conocimiento de las características físico-químicas del DNA antiguo así como en los procesos que contribuyen a su degradación. Tras estudiar doce especímenes de diferente procedencia y antigüedad mediante diversas técnicas, Pääbo concluyó que su material genético se encontraba modificado de forma similar en todos ellos, siendo la fragmentación y las modificaciones oxidativas las manifestaciones más abundantes. Al final de este trabajo señaló una de las principales dificultades de los estudios de aDNA desde el desarrollo de la *PCR*: la contaminación con DNA exógeno, a la vez que recomendó ciertas precauciones para minimizar su efecto.

3.2.2. Hitos del DNA prehistórico

Desde la publicación de estos trabajos se ha conseguido aislar DNA de restos de una gran variedad de especies vegetales y animales, incluido el hombre, así como de diversos tejidos. Podríamos destacar las primeras recuperaciones de DNA de restos óseos (HAGELBERG y SYKES 1989; HORAI *et al.* 1989; YOKOI *et al.* 1989; THUESEN y ENGBERG 1990). El primero, publicado en la prestigiosa revista *Nature*, alcanzó gran repercusión, pues se logró recuperar fragmentos de DNA mitocondrial de tamaño superior a 150pb a partir de huesos de entre 300 y 5000 años de antigüedad. La recuperación de fragmentos de elevado peso molecular evidenciaba la óptima preservación del DNA en este tipo de tejido.

El 1990 se consiguió obtener con éxito DNA a partir de piezas dentales (HÄNNI *et al.* 1990). Posteriormente se comprobó que este tipo de material es uno de los que ofrece mayores garantías de preservación del DNA (KUROSAKI *et al.* 1993; DEGUSTA *et al.* 1994; WOODWARD *et al.* 1994a; OOTA *et al.* 1995; ZIERDT *et al.* 1996).

Aunque la mayoría de estudios de aDNA se centran en el mtDNA, la capacidad de recuperar información genética nuclear ha sido documentada desde el inicio de la disciplina (PÄÄBO 1985b). Entre las aplicaciones del análisis de nDNA al estudio de muestras antiguas destacan la asignación de sexo (STONE *et al.* 1996; HUMMEL y HERRMANN 1997; BURGER *et al.* 1999) y el estudio de potenciales relaciones familiares entre individuos (HUMMEL y HERRMANN 1991; JEFFREYS *et al.* 1992; GILL *et al.* 1994; KUROSAKI *et al.* 1993; BÉRAUD-COLOMB *et al.* 1995). Con todo, el empleo de microsatélites para el estudio de restos antiguos ha sido criticado cuestionándose la posibilidad de replicación por la extensiva degradación del material genético (PÄÄBO *et al.* 1989; RAMOS *et al.* 1995). Este tema se tratará más extensamente dentro del apartado “aplicaciones de los estudios de aDNA” incluido en este capítulo.

3.2.2.1. DNA antediluviano

El desarrollo de la técnica de la *PCR* produjo gran entusiasmo en el incipiente campo del aDNA. Debido en parte al desconocimiento de la problemática asociada a esta técnica y a las características físico-químicas del DNA antiguo, se publicaron secuencias de DNA de millones de años de antigüedad en prestigiosas revistas. Estos casos han sido sin duda los más espectaculares de toda la historia del aDNA no sólo por su antigüedad sino porque fueron los más controvertidos y rebatidos en posteriores publicaciones.

Uno de los primeros trabajos fue la publicación de las secuencias de DNA procedentes de una hoja de magnolia, de 17-20 millones de años de antigüedad, depositada en arcilla en el fondo de un lago al norte de Idaho, Estados Unidos (GOLENBERG *et al.* 1990). Dos años más tarde se informó de la obtención de un

fragmento de mtDNA de un ciprés del género *Taxodium* procedente del mismo depósito (SOLTIS *et al.* 1992).

Un conjunto de trabajos que merece la pena destacar dado su impacto mediático son aquellos que informan de la recuperación de fragmentos de DNA de organismos preservados en ámbar. Numerosos estudios han demostrado que el ámbar ofrece un ambiente único para la conservación del DNA (LINDAHL 1993b), al favorecer la eliminación rápida del agua, y actuar como protector ante el ataque microbiano y el daño oxidativo (POINAR *et al.* 1996; AUSTIN *et al.* 1997a). La preservación tisular, celular e incluso bioquímica así lo atestiguan (GRIMALDI *et al.* 1994; POINAR *et al.* 1994, 1996).

Las secuencias de mayor antigüedad publicadas proceden de plantas y animales embebidos en este material. Entre los trabajos más destacables cabe citar el de Poinar y colaboradores, que, en 1993, obtuvieron secuencias del gen *rbcL* de cloroplasto de una hoja de un árbol extinto de la especie *Hymenaea protera* (POINAR *et al.* 1993); el de Cano y colaboradores de ese mismo año, en el que se documentaba la recuperación de dos secuencias del gen que codifica para el rRNA 18S de un coleóptero de la familia *Nemonychidae* de entre 120 y 135 millones de años de antigüedad (CANO *et al.* 1993); o el de DeSALLE *et al.* 1992, donde también se caracterizaba a un insecto, en este caso una termita fósil -*Mastotermes electrodominicus*-, de hace 25-30 millones de años. El equipo de Raúl Cano, tras aislar DNA de tejido abdominal de varios especímenes de la especie de abeja *Proplebeia dominicana*, preservada en ámbar de 25-40 millones de años, recuperó una región del rDNA 16S de un endosimbionte del género *Bacillus* (CANO *et al.* 1994). En un trabajo posterior consiguieron “revivir” esporas de este mismo microorganismo incubándolas en su laboratorio (CANO y BORUCKI 1995). Aislaron y secuenciaron el mismo fragmento de rDNA 16S de su anterior trabajo y comprobaron que las secuencias recuperadas en ambos eran prácticamente idénticas.

En 1994 se publicaron varias secuencias de DNA mitocondrial de dos huesos, posiblemente de dinosaurio, del cretáceo (WOODWARD *et al.* 1994b). Las secuencias presentaban una gran heterogeneidad, que los autores atribuyeron al daño molecular. A pesar de ello, reconstruyeron una secuencia consenso que compararon con una amplia base de datos de secuencias de numerosas especies animales. El resultado reveló que la

secuencia difería tanto de las aves como de los mamíferos, probablemente debido al reducido tamaño del fragmento obtenido.

El entusiasmo inicial creado con la publicación de estos y otros muchos trabajos similares se vio rápidamente sustituido por una ola de escepticismo que aún hoy continúa. El conocimiento de las características físico-químicas del DNA puso en duda su pervivencia durante períodos de tiempo tan grandes (LINDAHL *et al.* 1993a; PÄÄBO y WILSON 1991). La falta de preservación de otras moléculas como la lignina en las plantas del Mioceno, o la quitina en los insectos preservados en ámbar (LOGAN *et al.* 1993, 1995; STANKIEWICZ *et al.* 1998), junto con el estudio de racemización de aminoácidos presentado por Poinar y colaboradores en 1996, contribuyeron a reforzar esta tesis.

Otro argumento en contra de la posibilidad de recuperar DNA endógeno tan antiguo lo constituyó el fracaso en la reproducción de estos trabajos por parte de otros investigadores en laboratorios diferentes.

El primer intento de replicación fue llevado a cabo por SIDOW y colaboradores en 1991. Este equipo investigador no logró reproducir la información genética de otras hojas del depósito miocénico de Idaho.

Los resultados de las primeras evidencias de recuperación de DNA a partir de organismos preservados en ámbar se examinaron más tarde. En 1997 WALDEN y ROBERTSON intentaron replicar el trabajo de CANO *et al.* 1992a y amplificaron el mismo segmento de 18S rRNA en nueve himenópteros preservados en ámbar de 15-20 millones de años de antigüedad de la misma especie que el trabajo original (*Proplebeia dominicana*). Ninguna de las secuencias obtenidas resultó ser endógena. Paralelamente Austin y colaboradores (AUSTIN *et al.* 1997b) intentaron obtener, sin éxito, DNA de 10 individuos de esta misma especie y de otras tres especies de insectos conservadas en ámbar y copal. Las secuencias obtenidas a partir de diferentes extractos de un mismo espécimen, o de especímenes distintos, o procedentes de otras amplificaciones de un mismo extracto, resultaron diferentes en todos los casos y el análisis filogenético reveló que se agrupaban con secuencias de vertebrados y de hongos. Estos mismos autores señalaron, además, que la longitud de los fragmentos de DNA recuperado en los

trabajos de CANO *et al.* 1992a y 1993 excedía los 500pb, lo que contradecía anteriores observaciones acerca de que la fragmentación del aDNA hace improbable la amplificación de segmentos superiores a 200pb (PÄÄBO *et al.* 1989; COOPER 1994; HANDT *et al.* 1994a).

Los resultados obtenidos en los especímenes más antiguos fueron reexaminados por otros investigadores, que revelaron notables errores en su consecución e interpretación.

GUTIÉRREZ y MARÍN en 1998 reexaminaron los datos empleados por CANO *et al.* 1993 para la comparación de la secuencia supuestamente procedente de un insecto de la especie *Lebanorhinus succinus* de entre 120-135 millones de años de antigüedad. Una estima de la divergencia que debería existir entre la secuencia antigua y las especies actuales hizo pensar a Gutiérrez y Marín que la secuencia recuperada por Cano no era antigua, apuntando la posibilidad de que podría tratarse de una contaminación con DNA exógeno. Tras realizar una búsqueda en bases de datos de secuencias de diversas especies, estos autores concluyeron que la secuencia antigua mostraba gran homología con secuencias de hongos. Además, las secuencias de *Aedes albopictus* y *Drosophila melanogaster*, supuestamente homólogas a la secuencia recuperada, no se correspondían con las publicadas por Cano y colaboradores. Aún más, la secuencia correcta de *Aedes albopictus* correspondiente al segmento analizado por Cano, fue introducida en la base de datos de comparación en 1994, un año después de que se publicara su trabajo. La secuencia de *Drosophila melanogaster* publicada en CANO *et al.* 1993 era un compendio de diferentes fragmentos de la secuencia real. Las secuencias actuales empleadas como marco de comparación de los resultados obtenidos por este mismo autor en otro trabajo (CANO y BORUCKI 1995) se cuestionaron también (BECKENBACH 1995).

El trabajo de WOODWARD *et al.* 1994b aparece rebatido por diversos autores en el mismo número de la revista *Nature* (HEDGES y SCHWEITZER 1995; HENIKOFF 1995; ALLARD *et al.* 1995; ZISCHLER *et al.* 1995b). En los tres primeros estudios se critican los datos utilizados para la comparación de la secuencia consenso de mtDNA (HEDGES y SCHWEITZER 1995), y la de aminoácidos deducida a partir de ésta (HENIKOFF 1995). También se critica el método de comparación

empleado (ALLARD *et al.* 1995). La ampliación de las bases de datos y la aplicación de una reconstrucción filogenética basada en el método de la parsimonia reveló, en todos los casos, la existencia de mayores homologías entre las secuencias de dinosaurio con los mamíferos, y más concretamente con la especie humana. ZISCHLER *et al.* 1995b descubrieron que dichas secuencias correspondían a integraciones nucleares del mtDNA humanas, procentes probablemente de la contaminación de los extractos de dinosaurio con pequeñas cantidades de DNA humano. Woodward, por su parte, se defiende en el mismo número de la revista (WOODWARD 1995) argumentando que las secuencias individuales de su trabajo no pueden ser empleadas para un análisis filogenético puesto que probablemente proceden de cadenas de DNA dañadas y que es la secuencia consenso la que representa la secuencia original. Pone en duda, además, que las secuencias obtenidas pudieran proceder de un inserto genómico de mtDNA dada su heterogeneidad y a la ausencia de contaminación directa con mtDNA, un material mucho más abundante que sus insertos en el nDNA. La similitud entre las secuencias del Cretáceo y el grupo mamífero podría ser debida a la convergencia evolutiva del gen del citocromo b o de segmentos del mismo entre ambos grupos.

3.3. Aplicaciones de los estudios de aDNA

3.3.1. Caracterización genética de especies extintas

Siguiendo los pasos de Higuchi y colaboradores (HIGUCHI *et al.* 1984) la tecnología del aDNA fue rápidamente aplicada a otros animales extintos. El interés de estos estudios radica en su capacidad para establecer relaciones filogenéticas entre especies del pasado y las actuales, lo que, a su vez puede esclarecer las relaciones filogenéticas dudosas entre especies actuales. Así, el estudio de un segmento del gen del citocromo C mitocondrial del lobo marsupial (THOMAS *et al.* 1989) reveló su relación con otros marsupiales australianos pero no con los marsupiales sudamericanos como se había sugerido.

Otro trabajo destacable fue el desarrollado por el equipo de Alan Cooper (COOPER *et al.* 1992). Se estudió la relación de las *moas* –aves gigantes no voladoras oriundas de Nueva Zelanda–, con otras especies voladoras y no voladoras de Australia y Sudamérica. Mediante la obtención de secuencias de DNA mitocondrial de especímenes

de cuatro especies de moas –una de ellas de 3500 años de antigüedad– se demostró que las moas estaban más relacionadas con los emús Australianos y con los casuarios que con los kiwis. Ello indicaba que Nueva Zelanda había sido colonizada inicialmente por aves no voladoras y de forma independiente por los ancestros de las moas y por los ancestros de los kiwis. La secuencia completa del mtDNA de las moas ha sido publicada recientemente lo que ha permitido datar de una forma más precisa las divergencias dentro de este grupo y ha sugerido un origen en el Cretáceo tardío para estas aves no voladoras (COOPER *et al.* 2001).

Entre los especímenes extintos más antiguos analizados se encuentran los mamuts (HAGELBERG *et al.* 1994a; HÖSS *et al.* 1994; YANG *et al.* 1996; OZAWA *et al.* 1997; NORO *et al.* 1998; GREENWOOD *et al.* 2001; THOMAS *et al.* 2000; DEBRUYNE *et al.* 2003); los perezosos (HÖSS *et al.* 1996a; GREENWOOD *et al.* 1999; HOFREITER *et al.* 2003); el león de las cavernas (BURGER *et al.* 2004); el oso de las cavernas (HÄNNI *et al.* 1994; LOREILLE *et al.* 2001; ORLANDO *et al.* 2002; HOFREITER *et al.* 2002, 2004); y el oso pardo del Pleistoceno (LEONARD *et al.* 2000; BARNES *et al.* 2002).

Por otra parte, entre los estudios que han tenido mayor relevancia dentro del campo de la Antropología destaca la caracterización genética del hombre de Neandertal. Los Neandertales son un grupo de homínidos extintos que habitaron Europa y el oeste de Asia, entre hace 300.000 y 30.000 años aproximadamente. Durante parte de este tiempo coexistieron con los humanos modernos –*Homo sapiens sapiens*– originarios del continente africano. La relación filogenética de este clado con el humano actual ha sido ampliamente discutida barajándose tres hipótesis:

1. Los Neandertales fueron los ancestros directos de los europeos actuales.
2. Los Neandertales fueron totalmente reemplazados por los humanos modernos sin mezclarse.
3. A pesar de ser finalmente reemplazados por el hombre moderno, existió cierta mezcla entre ambas subespecies durante el período de convivencia, por lo que potencialmente habría cierta contribución genética de los Neandertales al *pool* genético humano actual.

El análisis de la variabilidad genética actual en los genomas nuclear y mitocondrial ha contribuido a apoyar la segunda de estas hipótesis (CANN *et al.* 1987; VIGILANT *et al.* 1991; HAMMER 1995; ARMOUR *et al.* 1996; TISHKOFF *et al.* 1996).

En un trabajo que causó gran impacto dentro de la comunidad científica, el equipo investigador de Svante Pääbo en Munich, en colaboración con el de Mark Stoneking en Pennsylvania, consiguió recuperar una secuencia de 378pb del espécimen tipo de Neandertal (KRINGS *et al.* 1997). La secuencia se reconstruyó a partir de múltiples clones de diferentes amplificaciones solapantes de aproximadamente 100pb cada una. La secuencia consenso obtenida exhibía 27 diferencias (24 transiciones, dos transversiones y una inserción) respecto a la secuencia humana de referencia ó *CRS* (ANDERSON *et al.* 1981). El patrón mutacional relativo al tipo de cambio (ratio transiciones/transversiones) y a las posiciones mutadas (25 son posiciones polimórficas entre los humanos actuales) era el típico de los humanos modernos y de los chimpancés (**Tabla I5**).

Cuando se comparó la secuencia con una amplia base de datos de secuencias de mtDNA actuales de individuos de los cuatro continentes, se vio que el número medio de sustituciones entre ambos (27.2 ± 2 , intervalo 22-36) ocupaba un valor intermedio entre el número de diferencias internas entre la humanidad actual (8.0 ± 3.1 , intervalo 1-24) y entre ésta y los chimpancés (55 ± 3 , intervalo 46-67). Esto suponía que existían tres veces más diferencias entre la secuencia Neandertal y las secuencias humanas que entre la humanidad actual. Si los Neandertales descendieran de los europeos –tal como se ha postulado–, se esperaría que la secuencia obtenida estuviera relacionada de una forma más próxima con este grupo humano. Sin embargo, las comparaciones con la humanidad actual revelaron que la secuencia Neandertal no mostraba afinidad con ninguno de los grupos humanos actuales, siendo prácticamente equidistante de todos ellos. Estas conclusiones y el análisis filogenético mediante el método de Neighbor-Joining –que situaba la putativa secuencia Neandertal fuera de la variabilidad genética humana–, se utilizaron como argumento en contra de la continuidad genética entre Neandertales y humanos actuales.

Basándose en esta única secuencia, los mismos autores dataron entre 550000-690000 el momento de la divergencia entre el linaje mitocondrial humano y el Neandertal. Teniendo en cuenta que la divergencia de los genes precede a la de las poblaciones, esta fecha está en concordancia con los datos arqueológicos y paleontológicos que sitúan la divergencia entre los Neandertales y los humanos anatómicamente modernos hace 250000-300000 años aproximadamente.

Dos años después, el mismo equipo investigador obtuvo un fragmento de 340 pb de la Región Hipervariable II del mtDNA del mismo espécimen, (KRINGS *et al.* 1999b). Los resultados mostraron de nuevo, gran divergencia entre la secuencia encontrada y las secuencias humanas actuales de la misma región mitocondrial. En la reconstrucción filogenética con las dos regiones hipervariables del mtDNA, la secuencia neandertal constituía un *outgroup* respecto al resto de secuencias humanas, sugiriendo, una vez más, que la línea Neandertal divergió mucho antes de la existencia del ancestro común del pool genético mitocondrial humano actual (**Tabla I4**). El número de diferencias entre la secuencia neandertal y el conjunto humano es mucho mayor a la observada entre diferentes subespecies de chimpancé y bonobo para la misma región mitocondrial, lo que sugiere que los Neandertales y los humanos modernos evolucionaron como entidades separadas durante un período de tiempo considerable.

Desde la publicación de estos primeros trabajos se han publicado otras ocho secuencias, cuatro completas y cuatro parciales, de la HVSI del mtDNA de otros tantos especímenes de Neandertal de diferentes procedencias (OVCHINNIKOV *et al.* 2000; KRINGS *et al.* 2000; SCHMITZ *et al.* 2002; SERRE *et al.* 2004; LALUEZA-FOX *et al.* 2005) (**Figura FI6**).

Las cinco secuencias de Neandertal completas de la HVSI del mtDNA son muy similares entre sí y muy diferentes de las secuencias humanas actuales. De todas ellas la más divergente es la secuencia del espécimen tipo de Feldhofer (KRINGS *et al.* 1997) (**Tabla I5**). La divergencia podría deberse a la presencia errores en la asignación de las mutaciones en posición 16107, 16108, 16111 y 16112 (SCHMITZ *et al.* 2002). El TMRCA estimado entre estas cinco secuencias es de 195000 ± 43000 A.P. considerando las mutaciones potencialmente artefactuales del espécimen tipo y de 162000 ± 41700 si no se tienen en cuenta (LALUEZA-FOX *et al.* 2005).

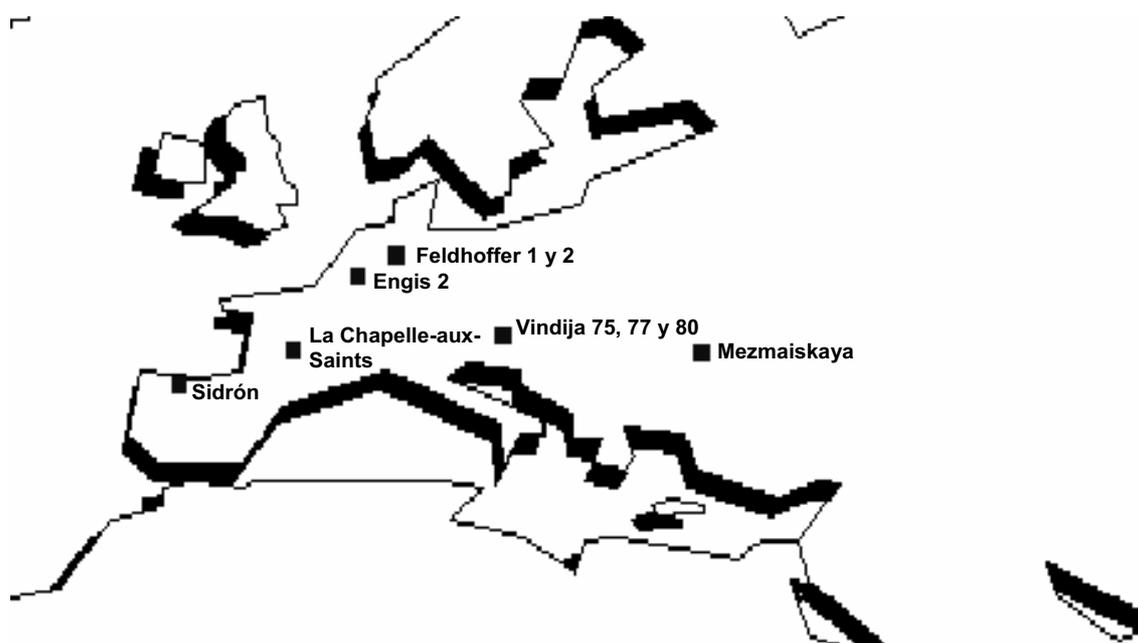


Figura FI6. Procedencia geográfica de los especímenes de Neandertal de los que se han recuperado secuencias de mtDNA.

Tabla I4. Características de las secuencias de Neandertal recuperadas hasta la fecha. (*) N° cambios mínimo y máximo respecto al resto de secuencias neandertales.

Referencia	Procedencia	Región estudiada	Posiciones estudiadas	N° cambios	
				CRS	Resto*
Krings <i>et al.</i> 1997	Feldhoffer (Alemania)	HVSI	16023-16400	28	6-13
Krings <i>et al.</i> 1999b	Feldhoffer (Alemania)	HVSI	00057-00396	12	-
Ovchinnikov <i>et al.</i> 2000	Mezmaiskaya (Norte Cáucaso)	HVSI	16078-16393	23	0-13
Krings <i>et al.</i> 2000	Vindija (Croacia)	HVSI y HVSI	16023-16378	24	0-7
Schmitz <i>et al.</i> 2002	Feldhoffer (Alemania)	HVSI	16023-16378	23	0-7
Serre <i>et al.</i> 2004	Vindija (Croacia)	HVSI	16023-16378	24	0-7
Serre <i>et al.</i> 2004	Vindija (Croacia)	HVSI	16231-16261	4	0-1
Serre <i>et al.</i> 2004	Engis (Bélgica)	HVSI	16231-16261	3	0-1
Serre <i>et al.</i> 2004	La Chapelle-aux-Saints (Francia)	HVSI	16231-16261	3	0-1
Lalueza-Fox <i>et al.</i> 2005	El Sidrón (España)	HVSI	16231-16277	6	0-1

El elevado número de diferencias observadas entre las secuencias de Neandertal recuperadas hasta la fecha y las secuencias humanas actuales ha sido interpretado como evidencia de ausencia de contribución genética significativa de los Neandertales al acervo genético de la humanidad actual (KRINGS *et al.* 2000, 1997; OVCHINNIKOV *et al.* 2000; SCHMITZ *et al.* 2002; SERRE *et al.* 2004).

No obstante, el resultado ha sido puesto en duda por GUTIÉRREZ *et al.* 2002. Estos últimos autores señalan que ni el método de reconstrucción filogenética ni las medidas de divergencia empleadas en KRINGS *et al.* 1997 y OVCHINNIKOV *et al.* 2000 han sido las adecuadas. Los análisis filogenéticos aplicados a la comparación de las secuencias de Neandertal y secuencias humanas actuales no tienen en cuenta la variabilidad de la tasa de mutación en las diferentes posiciones de la HVRI. Cuando se emplea un modelo de mutación que tiene en cuenta la elevada homoplasia de esta región la separación filogenética entre los humanos y los Neandertales no es estadísticamente significativa. GUTIÉRREZ *et al.* 2002 apuntan, por otra parte, que las diferencias detectadas por KRINGS *et al.* 1997 entre las secuencias de Neandertal y las secuencias humanas, en cuanto al número medio de diferencias por parejas (*pairwise differences*), se deben a un sesgo en la base de datos de comparación empleada, que incluye un alto número de individuos de las mismas poblaciones. A la vista de los resultados obtenidos concluyen que las diferencias entre las secuencias de Neandertal y las secuencias humanas actuales no son tan grandes como se había señalado inicialmente.

KNIGHT 2003 identificó las posiciones sinapomórficas entre los Neandertales y humanos actuales. Tras eliminar las que coincidían entre miembros de ambos grupos y las que presentaban una elevada tasa de mutación, quedaban aún 8 de las posiciones sinapomórficas. Según Knight, este número es suficiente para considerar que existió una historia separada entre los humanos y los Neandertales y permite excluir la posibilidad de una contribución significativa de los últimos al acervo genético de los primeros.

Tabla 15. Secuencias de la HVSI del mtDNA de nueve especímenes de Neandertal. Sólo se muestran las posiciones variables en el conjunto respecto a la secuencia consenso (CRS) (ANDERSON *et al.* 1981). (·) Posiciones coincidentes con la secuencia consenso. (-) Posiciones no analizadas. ¹KRINGS *et al.* 1997; ²OVCHINNIKOV *et al.* 2000; ³KRINGS *et al.* 2000; ⁴SCHMITZ *et al.* 2002; ⁵SERRE *et al.* 2004; ⁶LALUEZA-FOX *et al.* 2005.

	1	6	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3
	3	7	8	9	0	0	1	1	2	3	4	5	5	6	8	8	8	8	0	2	3	3	4	5	5	6	6	7	9	1	2	4	6						
	7	8	6	3	7	8	1	2	9	9	8	4	6	9	2	3	9	9	3	0	4	4	6	8	2	3	8	9	1	0	4	2							
CRS	A	A	T	T	C	C	C	C	G	A	C	T	G	C	A	A	T	T	C	A	C	G	C	A	C	-	C	A	T	C	C	T							
Feldhoffer ¹	G	G	·	C	T	T	T	T	A	T	T	C	·	T	·	C	C	C	T	G	T	A	A	G	T	A	T	G	C	T	·	C							
Mezmaiskaya ²	·	·	C	·	·	·	·	·	A	T	T	·	A	T	C	C	C	C	T	G	T	A	A	·	T	A	T	G	C	T	T	C							
Vindija-75 ³	G	G	·	·	·	·	·	·	A	T	T	C	·	T	C	C	C	C	T	G	T	A	A	G	T	A	T	G	C	T	·	C							
Feldhoffer 2 ⁴	G	G	·	·	·	·	·	·	A	T	T	C	·	T	C	C	C	C	T	G	T	A	A	·	T	A	T	G	C	T	·	C							
Vindija-80 ⁵	G	G	·	·	·	·	·	·	A	T	T	C	·	T	C	C	C	C	T	G	T	A	A	G	T	A	T	G	C	T	·	C							
Vindija-77 ⁵	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Engis 2 ⁵	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
LCAS ⁵	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Sidrón ⁶	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		

Los resultados obtenidos de las secuencias de mtDNA de los individuos de Neandertal no permiten excluir totalmente la posibilidad de que ciertos miembros de las dos especies se mezclaran en el pasado. De hecho, de las extremas diferencias nucleotídicas encontradas entre ambas, no se deduce que fueran especies distintas como explicita el propio Svante Pääbo (KRINGS *et al.* 2000).

El profesor Milford H. Wolpoff –principal defensor de la teoría multiregional– mantiene que la ausencia de genes Neandertales en la humanidad actual puede deberse a un fenómeno de deriva genética, de selección natural o a una combinación de ambos (WOLPOFF 1998). De hecho –como muchos genetistas de poblaciones reconocen– los humanos actuales son sorprendentemente homogéneos genéticamente hablando, lo que apunta a un origen común reciente a partir de un grupo pequeño de individuos. Según esto, la variabilidad genética de las poblaciones del pasado debió de ser mucho mayor a la observada actualmente, pudiendo incluir genes de origen neandertal. Otra posibilidad es que estos “genes neandertales” hubieran evolucionado dentro de la línea humana a través de mutación (KRINGS *et al.* 1997).

Por otra parte, puesto que el mtDNA proviene exclusivamente de la madre, sería posible que los Neandertales hubieran contribuido con otros genes nucleares de origen paterno al *pool* genético de la humanidad actual, como sugiere la variabilidad genética de ciertos marcadores nucleares (HAMMER 1995; ARMOUR *et al.* 1996; TISHKOFF *et al.* 1996). Muchos investigadores piensan que este intercambio unilateral es improbable, aunque resulta prácticamente imposible probarlo, puesto que las posibilidades de recuperación de material genético nuclear de especímenes tan antiguos son prácticamente nulas, como reconoce el propio Svante Pääbo (KAHN y GIBBONS 1997).

3.3.2. Estudios poblacionales

Tradicionalmente se ha estudiado la historia de las poblaciones mundiales proyectando su variabilidad genética actual sobre el pasado. La posibilidad de recuperar información genética directamente de poblaciones del pasado permite, sin embargo, analizar directamente esta variabilidad pretérita. Esta nueva aproximación al estudio de las poblaciones hace posible el estudio temporal de los cambios genéticos en las

mismas. Permite verificar, además, las hipótesis planteadas desde otras disciplinas relativas a migraciones antiguas y mezcla entre poblaciones.

El DNA antiguo se ha empleado para tratar de profundizar en los orígenes de la población americana nativa (STONE y STONEKING 1993, 1998, 1999; ROGAN y SALVO 1990b, 1994; MONSALVE *et al.* 1996; RIBEIRO DOS SANTOS *et al.* 1996; PARR *et al.* 1996; LALUEZA *et al.* 1997; O' ROURKE *et al.* 2000a; KAESTLE 2001; GARCÍA-BOUR *et al.* 2004); de las poblaciones polinesias (HAGELBERG y CLEGG 1993, HAGELBERG *et al.* 1994b; HAGELBERG 1997); de los primeros pobladores de Japón (HORAI *et al.* 1989, 1991); o de China (OTA *et al.* 1995, 1999).

3.3.3. Aplicaciones en el campo de la arqueología y la medicina forense

La posibilidad de recuperar información genética a partir de restos óseos y piezas dentales ha abierto nuevas perspectivas dentro del campo de la arqueología y en determinados ámbitos de la medicina forense. Mediante el estudio de este material, abundante en el registro fósil, se pueden solventar incógnitas que difícilmente podrían ser resueltas de otra manera. Entre las posibles aplicaciones del DNA antiguo en arqueología conviene destacar tres, ampliamente documentadas en la bibliografía: el cálculo del número mínimo de individuos de una necrópolis, el diagnóstico de sexo, y el establecimiento de relaciones familiares. En el ámbito de la genética forense, además de las dos últimas aplicaciones mencionadas se ha recurrido a la tecnología del DNA antiguo para la identificación personal.

3.3.3.1. Diagnóstico de sexo

En arqueología el diagnóstico genético de sexo tiene especial interés cuando la preservación de los restos no permite una atribución anatómica, como sucede con individuos infantiles o piezas esqueléticas sueltas.

Para determinar el sexo en restos óseos o dentales se recurre comúnmente al gen de la amelogenina. Situado en los cromosomas sexuales X e Y, este gen tiene ciertas regiones que no son 100% homólogas en ambos y cuya amplificación constituye un método de diagnóstico de sexo sencillo y rápido (AKANE *et al.* 1992; SULLIVAN *et*

al. 1993; HAAS-ROCHHOLZ y WEILER 1997). Así se ha podido determinar el sexo de diversos restos humanos antiguos (STONE *et al.* 1996; HUMMEL y HERRMANN 1997; BURGER *et al.* 1999).

3.3.3.2. Relaciones familiares

La determinación de las relaciones familiares existentes entre individuos exhumados puede ayudar a entender los patrones sociales de enterramiento de la población en cuestión. Es de especial interés en los enterramientos múltiples – individuos procedentes de la misma tumba– (CLISSON *et al.* 2002) y en necrópolis (FILY *et al.* 1998; STONE y STONEKING 1999; SCHULTES *et al.* 2000; KEYSER-TRACQUI *et al.* 2003).

Para establecer las relaciones de parentesco se emplean habitualmente marcadores nucleares de apreciable poder de discriminación, por su elevado número de alelos y alta heterocigosidad. La menor proporción de copias del DNA nuclear, comparada con el DNA mitocondrial, limita su estudio sólo a restos excelentemente preservados. En muchos casos se suele combinar la información de los marcadores citados con la de la HVSI del mtDNA, a fin de conseguir una mayor resolución (GILL *et al.* 1994; KEYSER-TRACQUI *et al.* 2003)

3.3.3.3. Identificación personal.

En medicina forense la tecnología del DNA antiguo ha sido aplicada con éxito para la identificación personal a partir de restos óseos y dentales (JEFFREYS *et al.* 1992; HAGELBERG *et al.* 1991). El material objeto de análisis suele ser mucho más reciente y, en la mayor parte de los casos, preservado en mejores condiciones que los restos propiamente arqueológicos.

Uno de los primeros trabajos en los que se documenta una identificación personal de este tipo es el de HAGELBERG *et al.* 1991. Se consiguió identificar un cadáver de mujer de 8 años de antigüedad que había sido víctima de un asesinato. Sin embargo, el caso de identificación a partir de restos óseos que más difusión mediática

ha tenido es sin duda el de la familia imperial rusa (GILL *et al.* 1994). Para la identificación de los 9 cadáveres encontrados en 1991 en una fosa se estudiaron diferentes marcadores genéticos: DNA mitocondrial, secuencias cromosómicas de microsatélites y DNA de cromosomas sexuales (gen de la amelogenina). El estudio reveló que dos de los esqueletos eran padres de tres de los otros. Los tres individuos restantes no estaban genéticamente relacionados con ninguno de los otros restos de la fosa, por lo que se atribuyeron a los sirvientes asesinados junto a la familia real. La comparación del DNA mitocondrial de los restos mencionados con el de descendientes actuales vivos de los zares por línea materna, apoya la hipótesis de que los restos se correspondían con los del crimen de Ekaterimburgo. Este mismo estudio permitió aclarar el enigma de la supuesta princesa Anastasia –Ana Anderson– que desde 1921 decía ser hija del Zar y que finalmente presentó una secuencia mitocondrial distinta a la de los restos encontrados. A pesar de todo, esta investigación ha recibido numerosas críticas desde su publicación (KNIGHT *et al.* 2004 a y b).

Otra vertiente interesante de este tipo de estudios es la caracterización genética de personajes históricos, en algunas ocasiones con fines identificativos. Así, en 2001 Vernesi y colaboradores publicaron una secuencia de la HVRI del mtDNA de los restos atribuidos al evangelista Lucas procedentes de un sepulcro de Padua. A pesar de la escasa resolución de la HVRI, los autores mantienen que la secuencia obtenida es compatible con un origen sirio del evangelista, aunque no descartan la posibilidad de que los restos pudieran haber sido sustituidos durante su traslado a Padua. De la misma forma se están analizando actualmente los restos procedentes de las dos putativas tumbas de Cristóbal Colón, una en la Catedral de Sevilla y la otra en Santo Domingo, y de algunos de sus parientes contemporáneos más cercanos con la finalidad de determinar el lugar en que se encuentran sus auténticos restos.

A veces ha sido necesaria la caracterización genética de algunos descendientes actuales de personajes ilustres para su identificación, tal y como se hizo con la familia real rusa. Este recurso está siendo empleado hoy para la identificación de los restos de un individuo infantil encontrado en un sarcófago que pudieran pertenecer a uno de los hijos de Pedro I el Cruel o de Juan II de Castilla. Para ello se están estudiando actualmente diversos marcadores de Cromosoma Y entre sus posibles descendientes actuales, portadores del apellido “Castilla” (LÓPEZ-PARRA *et al.* 2004).

3.3.4. Epidemiología

Otra interesante aplicación del DNA antiguo es la caracterización genética de virus y bacterias. Este tipo de estudios permiten, por una parte, establecer la evolución de los patógenos a lo largo del tiempo y, por otra, determinar la incidencia de determinadas enfermedades infecciosas en las poblaciones del pasado. En varios trabajos se documenta la recuperación de información genética de bacterias como *Mycobacterium tuberculosis* (SPIGELMAN y LEMMA 1993; ARRIAZA *et al.* 1995; DIXON *et al.* 1995; BARON *et al.* 1996; DONOGHUE *et al.* 1998; FLETCHER *et al.* 2003a y b; HAAS *et al.* 2000; SALO *et al.* 1994; ZINK *et al.* 2001, 2003); *Mycobacterium leprae* (RAFI *et al.* 1994; MONTIEL *et al.* 2003); *Treponema palidum* (MAROTA 1995) y *Yersinia pestis*; parásitos como *Plasmodium falciparum* (TAYLOR *et al.* 1998) y *Tripanosoma cruzi* (GUHL *et al.* 1997); y virus como el de la gripe (REID *et al.* 1999).

3.3.5. Dieta

Los coprolitos o heces halladas en los yacimientos arqueológicos constituyen otra fuente potencial, alternativa al material óseo, a partir de la cual obtener DNA. Permiten conocer los hábitos alimenticios de determinadas especies extintas y proporcionan, además, información acerca del entorno que las rodeaba. El estudio del DNA de los coprolitos recibe el nombre de “coproscopia molecular” y ha sido aplicado con éxito a diversas especies animales (POINAR *et al.* 2003, 1998; POINAR y EGLINTON 2000), incluido el hombre (POINAR *et al.* 2001).

3.3.6. Identificación de flora y fauna

Otra interesante aplicación de la tecnología del DNA antiguo es la identificación de la flora y la fauna de diferentes estratos de un yacimiento determinado. La extracción de material genético de sedimentos y su análisis mediante la amplificación de fragmentos de genes o regiones no codificantes del DNA muy conservadas entre especies, supone una alternativa al análisis directo del material excavado.

WILLERSLEV *et al.* 2003 estudiaron el gen del cloroplasto que codifica para la ribulosa-bisfosfato carboxilasa (*rbcL*), y varias porciones de genes mitocondriales (citocromo *b* (*cyt b*), *rDNA* 16S y *rDNA* 12S), además de la región control. Tras clonar los productos de *PCR*, se compararon las secuencias y se concluyó que las idénticas en un 96% pertenecían al mismo taxón. Para la identificación de los taxa se compararon las secuencias recuperadas con bases de datos de la misma región mediante BLAST y análisis filogenético.

Así pueden paliarse ciertas limitaciones que acompañan a los estudios convencionales de aDNA, como la destrucción del material, o su contaminación con DNA exógeno durante la excavación. Con todo, la contaminación reciente resulta difícil de evaluar.

3.4. Preservación del DNA

3.4.1. Procesos de degradación del material genético

Cuando un organismo muere comienza el proceso de degradación de sus biomoléculas, entre ellas el DNA. La muerte celular comporta la liberación de las enzimas autolíticas contenidas en los lisosomas celulares. La pérdida de la regulación enzimática celular aumenta la actividad de estas enzimas, entre las que se incluyen tanto las *endo* como las *exonucleasas*, que degradan el material genético en un proceso conocido como autólisis (PÄÄBO *et al.* 1989). Las endonucleasas cortan la cadena de DNA por diversos puntos, mientras que las exonucleasas van eliminando nucleótidos por sus extremos. Cuando la actividad nucleásica cesa, se produce la colonización del organismo por bacterias, hongos y, en algunos casos, insectos que continúan el proceso de descomposición. Finalmente, son procesos químicos, principalmente de hidrólisis y oxidación, los que completan la degradación del material genético (LINDAHL 1993a).

Por tanto, la hidrólisis y la oxidación son los procesos principales responsables de la degradación del DNA en la célula, tras la muerte del organismo. Describiremos, a continuación, estos procesos con más detalle.

3.4.1.1. Hidrólisis

La fragmentación del DNA puede producirse mediante la rotura directa de los enlaces fosfodiéster –que unen las diferentes desoxirribosas entre sí– o de los enlaces N-glicosídicos –que conectan éstas a las bases nitrogenadas–. Los primeros son altamente estables (LINDAHL 1993a), mientras que los segundos resultan muy lábiles incluso en condiciones fisiológicas (LINDAHL y NYBERG 1972; LINDAHL y KARLSTROM 1973; LOEB y PRESTON 1986).

La rotura del enlace glicosídico del DNA –normalmente por protonación de la base nitrogenada–, origina una posición sin base (*baseless*) en un proceso conocido como *depurinización* en el caso de rotura de una base púrica o *depirimidización*, si se trata de una pirimidina (LINDAHL 1993a, HÖSS *et al.* 1996b; BADA *et al.* 1999). La depurinización tiene lugar a una tasa muy superior a la depirimidización, estimada en

4×10^{-7} /seg a 95°C y pH 7.4 (LINDAHL y NYBERG 1972; LINDAHL y KARLSTROM 1973). Estimaciones realizadas “*in vivo*” sugieren que este proceso tiene lugar cada 10 horas, aunque factores como la temperatura, la fuerza iónica, el pH y la presencia de metales pesados contribuyen a su aceleración. Una vez se ha generado un sitio apurínico (*AP site*), el DNA queda debilitado y se produce su rotura, conocida como β -eliminación, en ese punto.

En las células metabólicamente activas, los mecanismos de reparación intracelulares reparan los sitios apurínicos impidiendo la rotura del DNA. En el caso del material antiguo, la ausencia de estos mecanismos de reparación provoca que el DNA se hidrolice en fragmentos pequeños, en unos miles de años a temperatura moderada, principalmente por depurinización (LINDAHL 1993a). Sin embargo, –como veremos a seguidamente– ciertas condiciones medioambientales pueden “proteger” el DNA de la depurinización y consiguiente fragmentación, ralentizando la tasa de esta última reacción.

Otro mecanismo de rotura común en el DNA es el generado por desaminación hidrolítica, principalmente de la Citosina y su homólogo 5-Metil-Citosina a Uracilo y Timina respectivamente (LINDAHL 1993a). La vida media de un residuo de Citosina se ha estimado en 200 años a 37°C y pH 7.4 en DNA de cadena sencilla, y en 30000 años si se encuentra en un DNA de doble cadena (FREDERICO *et al.* 1990). Aunque los residuos de Citosina son más susceptibles de sufrir desaminación que los de 5-Metil-Citosina, son esencialmente estos últimos los responsables de la mutagénesis espontánea del DNA antiguo. Los sustratos resultantes de la desaminación de la Citosina son rápidamente reparados por la abundante y ubicua enzima Uracil-DNA glicosilasa (UDG). Esta enzima, sin embargo, no es capaz de actuar sobre el derivado de la 5-Metil-Citosina, la Timina. Aunque la pareja G-T constituye un sustrato para los sistemas de reparación que detectan apareamientos incorrectos de bases (*mismatch correction systems*), estos sistemas son mucho más lentos que el mecanismo de reparación de la UDG, por lo que el riesgo de fijación de la mutación C→T resulta elevado (LINDAHL 1993a).

La desaminación hidrolítica de las bases púricas, *i.e.* Adenina y Guanina, ocurre a una tasa mucho menor que la desaminación hidrolítica de las pirimidinas (KARRAN y LINDAHL 1980). La desaminación de la Adenina a Hipoxantina se repara de forma análoga a la del Uracilo en el DNA mediante la Hipoxantina-DNA glicosilasa, aunque a un ritmo mucho menor debido a la baja concentración de esta enzima en la célula (HILL-PERKINS *et al.* 1986). El par Hipoxantina-Citosina resulta más estable que el Hipoxantina-Timina, por lo que esta desaminación tiene un efecto final mutagénico. No existen sistemas de reparación celular para la desaminación de la Guanina a Xantina. No obstante, puesto que la Xantina se aparea preferentemente con Citosina, este proceso raramente tiene un efecto final mutagénico. Además, el puente que une la desoxiribosa a la Xantina es particularmente susceptible a la hidrólisis espontánea, generándose un *AP site* que puede ser reparado por las *AP endonucleasas* celulares.

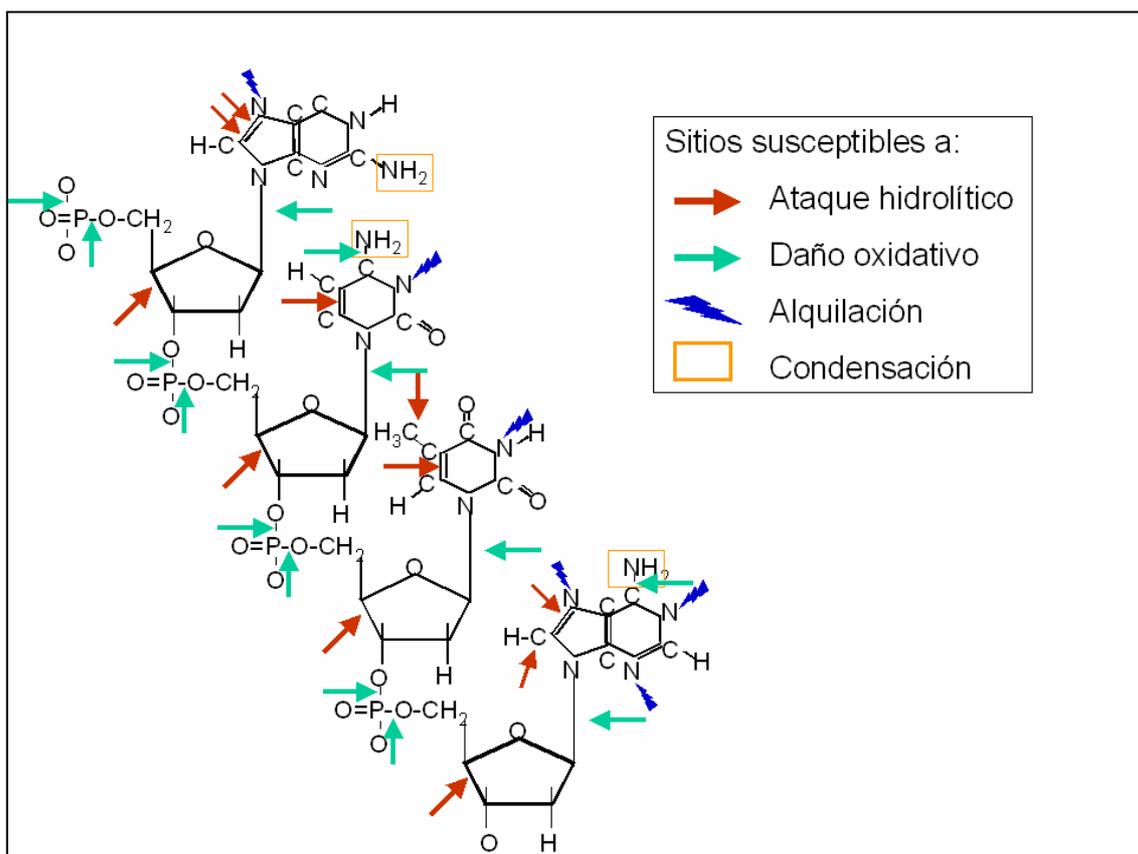


Figura F17. Lugares de la cadena del DNA susceptibles a diferentes mecanismos de degradación: ataque hidrolítico, daño oxidativo, alquilación y condensación. Modificado de POINAR 2002.

3.4.1.2. Daño oxidativo

El daño oxidativo en el DNA está producido principalmente por radicales libres superóxido (O_2^+), peróxido de hidrógeno (H_2O_2) e hidroxilo (OH^\cdot). Todos ellos son esencialmente subproductos de reacciones celulares endógenas, aunque también pueden ser generados tras la exposición de la célula a radiaciones ionizantes o luz ultravioleta o por degradación bacteriana y/o fúngica.

Las principales manifestaciones del daño oxidativo sobre el DNA son modificaciones de los residuos de azúcar, transformaciones de las pirimidinas Citosina (C) y Timina (T) en Hidantoínas, así como aparición de posiciones sin base (*baseless*) y puentes intermoleculares (PÄÄBO 1989). La mayoría de estas lesiones bloquean la DNA polimerasa inhibiendo, por tanto, la *PCR*.

Entre los productos de degradación de las *pirimidinas* presentes en DNA antiguo se encuentran: la 5-Hidroxi-5-Metil Hidantoína (5-OH-5-MeHid), la 5- Hidroxi-Hidantoína (5-OH-Hid), el 5-Hidroxi-Uracilo (5-OH-Ura) y el 5,6-Dihidroxi-Uracilo (5,6-diOH-Ura); y entre los de las *purinas*: la 4,6 Diamino-5-Formamidopirimidina (Fapy-Ade), la 2,6 Diamino-4-Hidroxi-5-Formamidopirimidina (Fapy-Gua), la 8-Hidroxi-Adenina (8-OH-Ade) y la 8-Hidroxi-Guanina (8-OH-Gua) (**Figura FI8**).

De todas ellas, las más comunes son la 5-OH-5-MeHid, 5-OH-Hid (pirimidinas) y la 8-OH-Gua (purina). La 5-OH-5-MeHid es uno de los productos mayoritarios resultantes de la degradación de Timina en DNA expuesto a radiaciones γ (BREIMER y LINDAHL 1985), mientras que la 8-OH-Gua se genera tras el ataque del DNA por radicales hidroxilo (DIZDAROGLU 1992). La cantidad de las pirimidinas modificadas (5-OH-5-MeHid y 5-OH-Hid) en los extractos de DNA parece guardar una relación inversa con la eficiencia de su amplificación, probablemente causada por el bloqueo que provocan estos productos en la *PCR*.

Sin embargo la 8-Hidroxi-Guanina tiene un efecto premutagénico directo sobre el DNA dado que se aparea preferentemente con la Adenina provocando la incorporación de un nucleótido incorrecto si se amplifica por *PCR* (HÖSS *et al.* 1996b).

Además de las especies reactivas derivadas del oxígeno antes mencionadas, existen en el interior de la célula otras moléculas capaces de modificar el DNA. Una de las más estudiadas “*in vivo*” es la S-Adenosil Metionina o SAM, un agente alquilante que actúa principalmente sobre los anillos de nitrógeno de las purinas. Entre los productos derivados de esta reacción se encuentra la 3-Metil-Adenina, que puede bloquear la replicación del DNA. En las células vivas esta lesión cuenta con su correspondiente enzima reparador, que rápidamente escinde la base alterada generando un *AP site* que será rellenado posteriormente (LINDAHL 1993a).

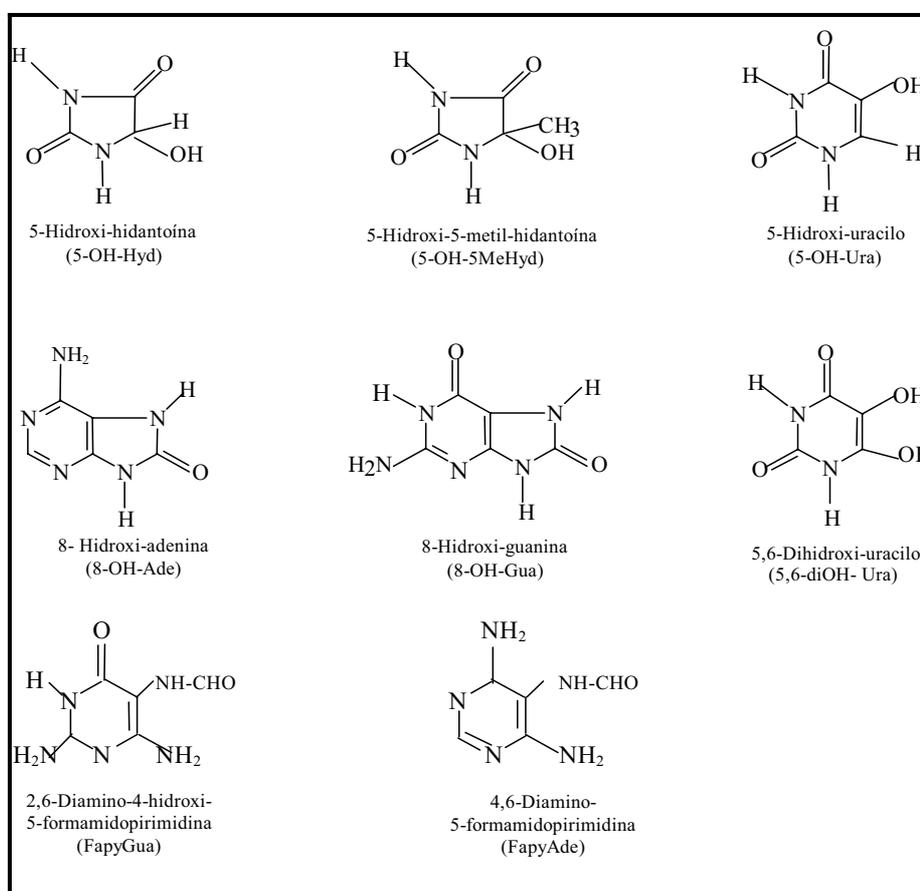


Figura F18. Estructura molecular de las bases nitrogenadas modificadas más frecuentes en DNA antiguo. Modificado de HÖSS *et al.* 1996b.

Se ha observado tanto daño hidrolítico como oxidativo en diversas macromoléculas extraídas de tejidos antiguos (HÖSS *et al.* 1996b; POINAR *et al.* 1996; POINAR 1999). Según Svante Pääbo, alrededor de 1 de cada 10 pirimidinas están oxidadas en el DNA antiguo (PÄÄBO 1989). Ello afecta más frecuentemente a la Timina,

que en este tipo de material representa tan sólo un 5% respecto a la cantidad presente en un DNA de una célula viva. Por el contrario, las desaminaciones en las Citosinas son menos importantes.

3.4.2. Factores que influyen en la conservación del DNA

La velocidad y el grado de descomposición del material genético de un resto dependen de diversos factores tanto endógenos como exógenos. Puede decirse que el éxito en la obtención de información genética a partir de material antiguo está condicionado por tres factores que desglosamos a continuación:

- El tipo de material de estudio.
- Las condiciones del enterramiento.
- Las condiciones de almacenamiento.

3.4.2.1. Material de estudio

3.4.2.1.1. Tipo de tejido

Desde el desarrollo del DNA antiguo como disciplina, se ha comprobado que el material genético se conserva mejor en tejidos duros –como el hueso o el diente– que en tejidos blandos (HAGELBERG y SYKES 1989; HÄNNI *et al.* 1990). La causa puede ser el bajo contenido en agua y enzimas en tales tejidos, aunque también se ha propuesto que el DNA podría resultar “secuestrado” por la hidroxiapatita de los mismos, lo que lo protegería de la acción de las enzimas degradadoras (SAMBROOK *et al.* 1989; LINDAHL 1993b).

De los dos tipos de tejidos duros mencionados, el diente es el que ofrece mayores garantías a la hora de recuperar información genética endógena. Ciertos estudios han demostrado que el DNA puede permanecer en el interior de los dientes en condiciones extremas de humedad, temperatura y acidez (SCHWARTZ *et al.* 1991; ÁLVAREZ-GARCÍA *et al.* 1996). Más aún, el DNA recuperado de piezas dentales parece ser de mejor calidad que aquel procedente de huesos o tejidos blandos (KUROSAKI *et al.* 1993; DEGUSTA *et al.* 1994; WOODWARD *et al.* 1994a; OOTA *et al.* 1995; ZIERDT *et al.* 1996). La mejor preservación del DNA en dientes podría

deberse a la extrema dureza de los tejidos que envuelven la pulpa (esmalte y dentina) (PFEIFFER *et al.* 1999), a su elevado contenido en hidroxiapatita (DEGUSTA *et al.* 1994) o, más probablemente, a la combinación de ambos factores.

3.4.2.1.2. Tipo de material genético

El material genético más estudiado en restos antiguos es sin duda el DNA mitocondrial. Esencialmente por su mayor abundancia en la célula, que hace más factible la recuperación de fragmentos de DNA intactos.

Sin embargo, es de esperar que este tipo de material genético esté mucho más degradado que el DNA nuclear, al haber mayor concentración de radicales libres en la mitocondria (ROGAN y SALVO 1990a). En contraposición, el núcleo es un compartimento celular pobremente oxigenado (LINDAHL 1993a), por lo que, durante los primeros estadios de la muerte celular, cuando la membrana nuclear todavía está intacta, presenta una baja concentración de radicales libres (ROGAN y SALVO 1990a). Por otra parte, el DNA cromosómico está asociado a histonas, las cuales podrían ejercer un efecto protector contra el ataque de los radicales libres de oxígeno tras la degradación de la membrana nuclear (LINDAHL 1993a).

El DNA mitocondrial no presenta una estructura de empaquetamiento tan compleja como la del DNA nuclear. Sin embargo, *in vivo* se encuentra unido a ciertas proteínas (*binding proteins*) formando un complejo denominado nucleoide. Algunos autores han sugerido que estas proteínas podrían tener cierta función protectora similar a la de las histonas en el núcleo (NEWMAN *et al.* 1996; FISHER *et al.* 1992).

3.4.2.2. Condiciones del enterramiento

Las características del ambiente en el que se encuentra depositado un resto pueden ralentizar e incluso detener el proceso de degradación. La “calidad” y “cantidad” del DNA recuperado del mismo dependerá, por lo tanto, de cuáles sean estas características. Entre las condiciones del enterramiento que más afectan la degradación del DNA conviene destacar la temperatura, la humedad, el pH o la presencia de ciertos compuestos en el suelo.

3.4.2.2.1. Temperatura

Es el factor que más condiciona la preservación del material genético. Las temperaturas bajas durante el período de deposición de un resto favorecen su conservación óptima (LINDAHL 1993a; BURGER *et al.* 1999). Además, si la temperatura es elevada no hay garantía de una correcta preservación del material genético aunque se den otras condiciones medioambientales favorables, como la sequedad y ausencia de microorganismos (BURGER *et al.* 1999).

Ello se debe a que, a bajas temperaturas, se produce una ralentización de las reacciones químicas responsables de la degradación orgánica. Se ha calculado que una reducción de 20°C en la temperatura provoca una disminución de entre 10 y 25 veces en la tasa de dichas reacciones (HÖSS *et al.* 1996b).

La mayoría de las muestras más antiguas de las que se ha podido recuperar DNA de forma reproducible proceden de lugares con baja temperatura. Este es el caso de los ejemplares de mamut lanudo (HÖSS *et al.* 1994; HAGELBERG *et al.* 1994a; TAYLOR 1996; YANG *et al.* 1996), algunos de los especímenes de Neandertal estudiados hasta la fecha (KRINGS *et al.* 1997; SCHMITZ *et al.* 2002) o el “hombre del Tirol” (HANDT *et al.* 1994a), por citar algunos ejemplos.

Se ha postulado, también, que la presencia de elevadas temperaturas tras la deposición de un resto puede favorecer la deshidratación parcial del DNA, deteniendo por tanto los procesos de hidrólisis. Este efecto, en ausencia de oxígeno, puede actuar sobre el material genético protegiéndolo de la degradación (LINDAHL 1993a). Ambas condiciones podrían ser las responsables del excelente estado de preservación de los organismos conservados en ámbar.

3.4.2.2.2. Humedad

La humedad ejerce un efecto adverso sobre la preservación del material genético. Permite la penetración de las sustancias orgánicas del sedimento en el interior del resto, lo que incrementa las posibilidades de que el extracto de DNA presente moléculas inhibitoras. Adicionalmente, la humedad favorece las reacciones de

degradación hidrolítica y oxidativa, provocando una reducción en la cantidad y especialmente en la calidad del *DNA template* (BURGER *et al.* 1999).

3.4.2.2.3. pH

La presencia de un pH neutro o ligeramente alcalino en el enterramiento favorece la preservación del DNA (LINDAHL, 1993a). Sin embargo, HERRMANN y NEWSELY 1982 en un estudio *in vitro*, demostraron que una disminución paulatina del pH provocaba la degradación de la hidroxiapatita de los huesos o dientes.

3.4.2.2.4. Compuestos del suelo

Se ha apuntado la posibilidad de que ciertos compuestos minerales del suelo como la montmorilonita, la kaolinita, el feldespato o el cuarzo podrían asociarse al DNA protegiéndolo frente a la acción endonucleásica (LORENZ y WACKERNAGEL 1987; ROMANOWSKI *et al.* 1991; ÁLVAREZ *et al.* 1998). Se ha documentado también la posibilidad de unión del DNA con los ácidos húmicos del suelo, unión que causaría el mismo efecto protector (CRECCHIO y STOTZKY 1998).

3.4.2.3. Almacenamiento

Podría considerarse el almacenamiento como una segunda fase de deposición, en la que el proceso de degradación del material genético continúa bajo nuevas condiciones.

En general, el almacenamiento frío de un resto (-20°C) inmediato tras su exhumación favorece una mejor preservación de su material genético y da lugar a resultados reproducibles (BURGER *et al.* 1999). El grado de alteración de los materiales almacenados a temperatura ambiente es mayor y depende sobremanera del tiempo de permanencia en estas condiciones. Los restos depositados durante un tiempo breve sufren principalmente modificaciones moleculares que afectan a su calidad y a la capacidad de obtener resultados reproducibles a partir de los mismos. La degradación es aún mayor si el almacenamiento se produce durante un período largo a temperatura ambiente.

El material genético ya extraído de un resto antiguo puede seguir degradándose. En un estudio llevado a cabo en 1997, HERRMANN y HUMMEL demostraron que los extractos frescos de DNA antiguo experimentaban un proceso de degradación rápida (entre seis y cuarenta semanas) que era independiente de la temperatura de almacenamiento. En su estudio se sugería que la estabilidad del DNA en los extractos podría estar relacionada con la composición del medio en el que se encuentra diluido. Los medios más habituales –el agua o el etanol–, no favorecían la preservación a largo plazo del DNA. Sin embargo, aquel DNA almacenado en una matriz de sílica y etanol presentaba una mayor estabilidad.

3.4.3. Límite de preservación del DNA

Las secuencias recuperadas de organismos de millones de años de antigüedad (CANO *et al.* 1993; WOODWARD *et al.* 1994b; DeSALLE *et al.* 1992; GOLENBERG *et al.* 1990; SOLTIS *et al.* 1992) evidencian que el DNA puede preservarse durante largos períodos de tiempo si el ambiente de deposición es el adecuado.

Sin embargo, los estudios de preservación del DNA y otras biomoléculas celulares *in vitro* proponen la fecha de 100000 años como su límite máximo de pervivencia (LINDAHL *et al.* 1993a; POINAR *et al.* 1996).

Para tratar de explicar la contradicción aparente entre ambos resultados se han propuesto dos explicaciones:

1. El DNA recuperado de millones de años de antigüedad no es endógeno.
2. La degradación del DNA en el registro fósil es imposible de reproducir en el laboratorio.

La primera de las explicaciones planteadas fue sugerida ante la imposibilidad de reproducir los resultados obtenidos en algunos de los estudios anteriormente mencionados (SIDOW *et al.* 1991; WALDEN y ROBERTSON 1997; AUSTIN *et al.* 1997b), y al detectarse errores en la reconstrucción filogenética empleada que sugerían la posibilidad de una contaminación con DNA humano o de otras especies actuales

(GUTIÉRREZ y MARÍN 1998; HEDGES y SCHWEITZER 1995; HENIKOFF 1995; ALLARD *et al.* 1995; ZISCHLER *et al.* 1995b).

A pesar de la dudosa fiabilidad de algunos de estos estudios la posibilidad de que el DNA pueda sobrevivir a períodos superiores a 100000 años no puede descartarse totalmente. Varios autores han sugerido que, tras la deposición, el DNA podría interaccionar con moléculas circundantes que le confirieran estabilidad posibilitando su preservación durante largos períodos de tiempo.

Entre las evidencias de recuperación de DNA de mayor antigüedad se encuentran los procedentes de organismos preservados en diversos tipos de resinas vegetales. Existen diferentes tipos de resinas, dependiendo de su antigüedad y grado de polimerización. El *copal* es la variante más joven y parcialmente polimerizada (250-5000 años), seguida por la *resina sub-fósil* (5000-40000 años) y finalmente por el *ámbar*, todavía más antigua. Se ha comprobado de manera fiable que el ámbar ofrece un ambiente único para la conservación del DNA (LINDAHL 1993b), lo que se debe a la rápida deshidratación de los tejidos y la protección que proporciona al organismo ante los agentes meteorológicos y los ataques microbiano y fúngico (POINAR *et al.* 1996; AUSTIN *et al.* 1997a).

Algunos autores mantienen que la excelente preservación tisular observada en diversos depósitos fósiles en ámbar (GRIMALDI *et al.* 1994) es extensible al nivel celular e incluso bioquímico (POINAR y HESS 1982; GRIMALDI *et al.* 1994; POINAR *et al.* 1994, 1996). Sin embargo, STANKIEVICZ *et al.* (1998) estudiaron la pervivencia de moléculas resistentes –ligno-celulosa en hojas y quitina en insectos–, en organismos preservados en ámbar mediante *flash-pirólisis* y *GC/MS*. Los resultados mostraron que estas moléculas presentaban un gran nivel de alteración, lo que llevó a estos autores a concluir que la degradación de estos materiales tan resistentes en los organismos preservados en ámbar no resultaba en principio compatible con la supervivencia de una molécula mucho más lábil como es el DNA.

Otro ambiente que ofrece grandes posibilidades para la preservación del DNA durante períodos de tiempo largos son las sales de cristal. Los depósitos de evaporita –normalmente de sal de halita– son los restos desecados de antiguos mares o lagos

salinos. Hay evidencias de que los microorganismos pueden sobrevivir en el interior de cristales de halita (McGENITY *et al.* 2000; NORTON y GRANT 1988). Recientemente se han publicado secuencias de rDNA 16S de varios microorganismos conservados en este material de entre 11 y 425 millones de años de antigüedad (FISH *et al.* 2002). De ser sus resultados ciertos, se trataría de la evidencia más antigua de preservación del material genético. Sin embargo –y como los propios autores reconocen–, durante su historia las evaporitas pueden sufrir varios procesos de fractura y recristalización, durante los cuales pueden resultar contaminadas con microorganismos actuales o de otras épocas.

3.5. Problemas metodológicos en la obtención de DNA antiguo

La obtención de material genético genuino a partir de restos antiguos tropieza con una serie de dificultades de tipo técnico, relacionadas sobre todo con el estado de preservación del DNA en estos restos.

El DNA antiguo presenta un conjunto de características fisico-químicas que lo diferencian, a grandes rasgos, del DNA procedente de organismos vivos también denominado “DNA fresco” y que podríamos resumir en cuatro:

- Escasez
- Fragmentación de las cadenas.
- Modificaciones moleculares.
- Presencia de inhibidores de la *PCR* en los extractos.

Es la suma de estas características la que impone ciertas limitaciones técnicas a la hora de recuperar DNA de restos antiguos, y que obliga, en muchos casos, a la implementación de protocolos específicos para su exitosa consecución.

3.5.1. Escasez y fragmentación de las cadenas

La escasez y la fragmentación de las cadenas características del DNA antiguo son manifestaciones de un mismo fenómeno: la degradación *post-mortem* del material genético. De hecho, una degradación extrema del material genético de un tejido antiguo podría ser interpretada como ausencia total o escasez de su DNA, puesto que muchos de los métodos de cuantificación convencionales no detectan fragmentos pequeños. Inversamente, aunque en algunos casos la cantidad total de DNA antiguo de una muestra sea, en principio, suficiente para garantizar su recuperación mediante una *PCR* exitosa, la fragmentación del material solo permite la recuperación de fragmentos pequeños.

La longitud de los fragmentos amplificables depende de las condiciones de preservación del resto antiguo (BURGER *et al.* 1999). Sin embargo son muchos los autores que señalan que el DNA antiguo suele encontrarse degradado en fragmentos, de

uno a unos pocos cientos de bases (PÄÄBO 1985a y b, 1986; HAGELBERG y SYKES 1989; HÄNNI *et al.* 1990; ROGAN y SALVO 1990a; GOLOUBINOFF *et al.* 1993; HANDT *et al.* 1994a).

Algunos científicos mantienen que no es posible la amplificación por *PCR* de fragmentos de DNA superiores a 150 pares de bases (pb) en tejidos blandos (PÄÄBO 1989). Este límite es sensiblemente superior en tejidos sólidos –como hueso o diente–, siendo de 300pb en DNA mitocondrial y de poco más de 100pb en material nuclear (HAGELBERG *et al.* 1994a; HAUSWIRTH *et al.* 1994). La amplificación de fragmentos muy grandes de DNA en restos antiguos corresponde generalmente a DNA contaminante (HANDT *et al.* 1994b). Se ha observado, también, que la longitud del fragmento amplificado guarda una relación inversamente proporcional con la cantidad de producto obtenido tras la amplificación (PÄÄBO 1987), lo que ha sido propuesto como criterio de autenticidad.

Para solventar el problema de la fragmentación del DNA antiguo se han propuesto diversas estrategias. En primer lugar, existe una tendencia generalizada a analizar aquel material genético que *a priori* ofrece mayores probabilidades de éxito debido, por ejemplo, a su mayor abundancia en la célula, como es el caso del mtDNA. Se analizan, además, los segmentos pequeños que pueden suministrar más información, dado que presentan un elevado nivel de polimorfismo y han sido muy estudiados en la población actual. Este es el caso de la región Hipervariable I (HVSI) del mtDNA, con mucho la más estudiada en DNA antiguo.

Cuando se desea reconstruir la información genética de un fragmento largo a partir de un material muy degradado, se suele amplificar por *PCR* pequeños fragmentos solapantes, a veces de menos de 50pb. La estrategia ha sido empleada con éxito para la completa reconstrucción de la HVSI del mtDNA de cinco especímenes distintos de Neandertal (KRINGS *et al.* 1997, 2000; OVCHINNIKOV *et al.* 2000; SCHMITZ *et al.* 2002; SERRE *et al.* 2004). Sin embargo, en algunos casos la preservación irregular de los diferentes fragmentos de DNA impide realizar una reconstrucción completa de la región deseada mediante este procedimiento (SERRE *et al.* 2004; LALUEZA-FOX *et al.* 2005).

3.5.2. Modificaciones moleculares

Algunas de las modificaciones moleculares del DNA antiguo –consecuencia de su degradación *post-mortem*–, pueden limitar e incluso impedir totalmente su recuperación. No obstante, existen otro tipo de lesiones que hacen factible la recuperación aunque introducen errores en la secuencia final.

3.5.2.1. Lesiones que bloquean la polimerasa

3.5.2.1.1. Modificaciones oxidativas de las bases nitrogenadas y los residuos de azúcar

Algunas lesiones del DNA, habitualmente inducidas por la acción de radicales libres como el peróxido de hidrógeno (H_2O_2), el hidroxilo (OH \cdot) o el ión superóxido (O_2^-), pueden bloquear la *Taq* polimerasa impidiendo la elongación de las cadenas e inhibiendo, por tanto, la *PCR*. Estos compuestos atacan habitualmente los dobles enlaces de las bases nitrogenadas o de las pentosas del DNA, provocando, en ambos casos, la rotura de los anillos correspondientes. Entre los residuos más frecuentes en DNA antiguo que se ha demostrado bloquean la *Taq* polimerasa, se encuentran la 5-Hidroxi-5-Metil Hidantoína (5-OH-5-MeHyd) y la 5-Hidroxi-Hidantoína (5-OH-Hyd) (HÖSS *et al.* 1996b).

La solución para hacer frente a este tipo de “daño molecular” es la misma que se emplea cuando el material genético está muy fragmentado: amplificación de pequeños fragmentos solapantes altamente informativos.

3.5.2.1.2. Puentes cruzados (*Cross-links*)

Otro tipo de manifestación del daño molecular en el DNA antiguo son los puentes cruzados o *cross-links*, visibles en muchas ocasiones por microscopía electrónica (PÄÄBO 1989). Los puentes cruzados pueden originarse entre cadenas del DNA (*puentes intramoleculares*) o entre el DNA y las proteínas del medio (*puentes intermoleculares*). Habitualmente este tipo de manifestaciones se producen por condensación entre los azúcares de la cadena del DNA y los grupos amino de las bases nitrogenadas o de las proteínas (Reacción de Maillard). Como en el caso anterior, este

tipo de lesiones actúan bloqueando la *Taq* polimerasa e impidiendo el copiado de los fragmentos de aDNA por *PCR*.

El tratamiento del DNA –previamente a su amplificación– con un agente químico, el bromuro de N-Fenilacil Tiazolio –conocido comúnmente como PTB–, se ha mostrado efectivo para la amplificación de aquellas moléculas afectadas por puentes cruzados (VASAN *et al.* 1996). Este compuesto actúa rompiendo los productos resultantes de las reacciones de Maillard y posibilitando la amplificación de las cadenas dañadas. El PTB fue empleado con éxito en la recuperación de una de las secuencias de Neandertal (KRINGS *et al.* 2000).

3.5.2.2. Lesiones que provocan la incorporación incorrecta de bases: “*miscoding lesions*”

Algunas reacciones de hidrólisis, como la desaminación, pueden provocar la modificación de algunas bases del DNA sin llegar a inhibir la amplificación, generando la incorporación de bases incorrectas en estos lugares por parte de la *Taq* polimerasa.

Estas lesiones en la cadena original de DNA han sido bautizadas como *miscoding lesions* (HOFREITER *et al.* 2001; HANSEN *et al.* 2001; GILBERT *et al.* 2003a y b). Su estudio ha despertado un gran interés en la comunidad científica puesto que una incorrecta evaluación de este “daño molecular” puede comportar una interpretación errónea de los resultados obtenidos.

Las principales evidencias de la existencia de *miscoding lesions* en el DNA proceden del análisis de la distribución, tipo y frecuencia de las diferencias observadas entre las secuencias de diferentes clones de diferentes amplificaciones de diversas muestras (HOFREITER *et al.* 2001; GILBERT *et al.* 2003a y b).

En 2001 Hofreiter y colaboradores, tras observar un patrón mutacional diferente entre los clones procedentes de diversas amplificaciones, clasificaron estas diferencias en tres tipos (ver **Figura F19**):

- *Sustituciones consistentes*: las presentes en todos los clones de una misma amplificación pero no en aquellos de otras amplificaciones del mismo extracto.
- *Sustituciones no consistentes*: las presentes en algunos de los clones de una misma amplificación, pero no en otros.
- *Sustituciones puntuales o singletons*: presentes en algunos clones con sustituciones consistentes.

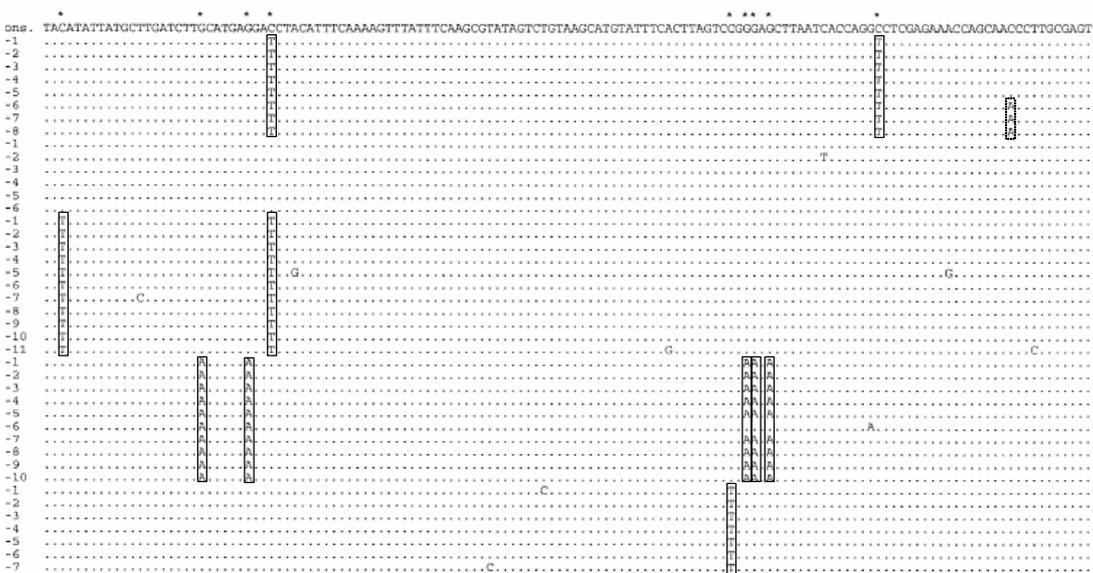


Figura F19. Secuencias de diferentes clones de cinco amplificaciones diferentes (1→5) de una misma muestra de oso de las cavernas. El primer número identifica la amplificación y el segundo el clon. Se han enmarcado con línea continua las *sustituciones consistentes* de una misma amplificación, y con línea de puntos aquellas *sustituciones inconsistentes*. Modificado de HOFREITER *et al.* 2001.

En principio se pensó que la heterogeneidad de los clones podía ser debida a errores introducidos por la *Taq* polimerasa durante el proceso de copiado de las cadenas. Sin embargo, ni la tasa de error de la *Taq* polimerasa ni el patrón de error proporcionado por ésta resultaban compatibles con las diferencias observadas entre los clones. El número de diferencias era demasiado elevado y parecía afectar a determinadas posiciones preferentemente, y no al azar, como se esperaría en el caso de errores de la *Taq* polimerasa (FATTORINI *et al.* 1999; HANSEN *et al.* 2001; GILBERT *et al.* 2003a y b). Se sugirió, por tanto, que podría tratarse de lesiones en determinados puntos de la cadena original de DNA reproducidas en las secuencias finales (HANSEN *et al.* 2001; HOFREITER *et al.* 2001; GILBERT *et al.* 2003 b).

En las secuencias de amplificaciones de DNA antiguo se dan cuatro tipos de cambios predominantes respecto a la secuencia consenso: las transiciones $C \rightarrow T$, $G \rightarrow A$, $A \rightarrow G$ y $T \rightarrow C$. El hecho de que nunca se observaban simultáneamente sustituciones del tipo $C \rightarrow T$ y $G \rightarrow A$ en una misma secuencia, hizo suponer a Hofreiter y colaboradores (2001) que ambos eventos se generaban por un único tipo de modificación molecular originaria, y que el cambio final observado dependía de la cadena (L o H) dañada inicialmente. Como puede verse en la **Figura FI10**, ambos cambios pueden darse mediante dos procesos: transformación de C en T (*esquema a*) o bien transformación de G en A (*esquema b*). Si la amplificación se originara a partir de una única cadena L donde la C ha sido sustituida por T, las moléculas finales contendrían el cambio $C \rightarrow T$, mientras que si ésta tuviera lugar a partir de una cadena H modificada de la misma manera, las moléculas finales contendrían el cambio $G \rightarrow A$ (*esquema a*, **Figura FI10**). Por el contrario, si la modificación original fuera $G \rightarrow A$, de ocurrir ésta en la cadena L se produciría el cambio final $G \rightarrow A$, mientras que si tuviera lugar en su complementaria se generaría el cambio $C \rightarrow T$ (*esquema b*, **Figura FI10**). Ambos cambios – manifestaciones de un mismo evento molecular –, fueron bautizados por GILBERT y colaboradores (2003b) como “*Transiciones de tipo 2*” y representados como $C \rightarrow T/G \rightarrow A$.

Para comprobar cuál de los dos cambios originales era el responsable, Hofreiter y colaboradores (HOFREITER *et al.* 2001) amplificaron con un sólo cebador durante los primeros ciclos de PCR diferentes muestras que exhibían el patrón $C \rightarrow T$ ó $G \rightarrow A$ (*Single-Primer PCR* ó *SP-PCR*) con el fin de reproducir los errores de la cadena copiada. Los clones analizados presentaban mayoritariamente la sustitución $G \rightarrow A$, sugiriendo que en las cadenas de DNA *template* los residuos de Citosina habían sido sustituidos por Timina.

Una modificación molecular que puede explicar este proceso es la desaminación hidrolítica de la Citosina (C) y de su homólogo 5-Metil-Citosina a Timina (T) y Uracilo (U). Este proceso es especialmente rápido y abundante en células vivas (HOFREITER *et al.* 2001; HANSEN *et al.* 2001).

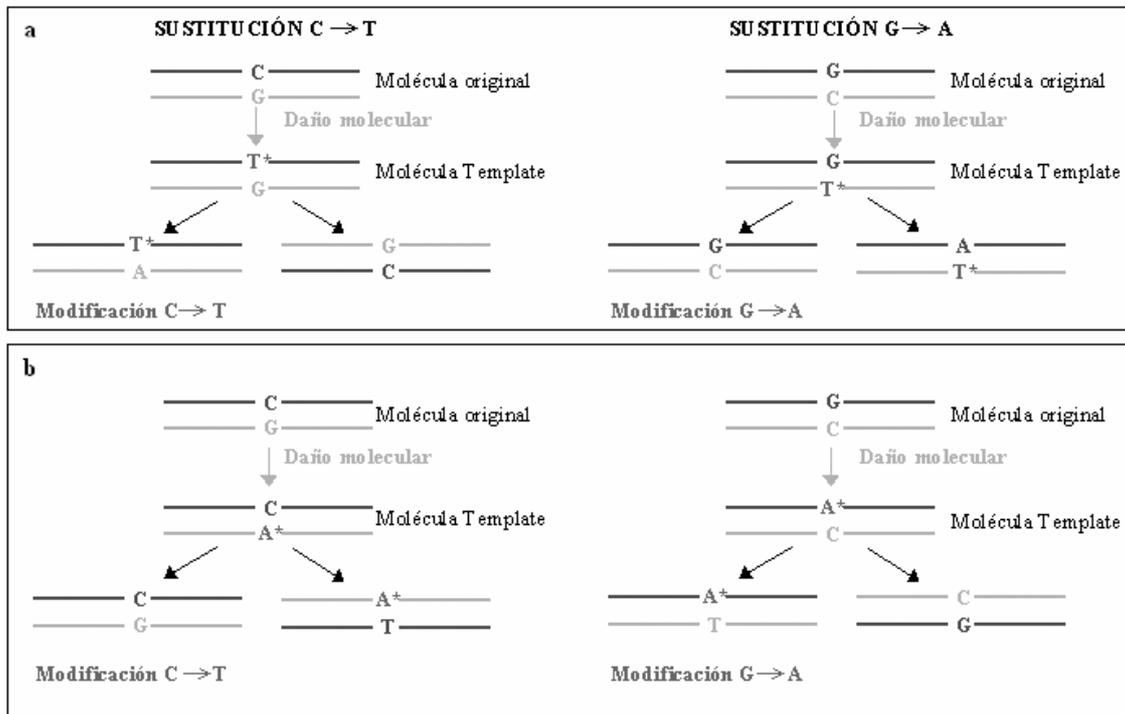


Figura FI10. Mecanismos posibles para explicar la aparición de sustituciones C→T y G→A (Tipo 2) en una misma secuencia. Modificado de GILBERT *et al.* 2003b.

Una evidencia a favor de esta explicación es que el tratamiento de los extractos con la enzima Uracil-DNA-Glicosilasa (UDG), previamente a su amplificación, conseguía eliminar todas las sustituciones del tipo C→T y G→A presente en los productos de amplificación. La enzima actúa eliminando las moléculas de Uracilo del DNA y creando en su lugar una posición sin base (*baseless* o *AP site*), que es posteriormente suprimida por β eliminación.

Hay otro mecanismo que permite explicar las sustituciones del tipo C→T/G→A. Si tras la incorporación de un residuo de Deoxi-Adenosina, al final de la cadena de DNA por la *Taq* polimerasa, se produjera un *jumping PCR* que apareara esta A final con una C presente en otra molécula de DNA y prosiguiera el proceso de copiado, se acabaría formando una secuencia híbrida con una sustitución errónea (KWOK *et al.* 1990) (Figura FI11). La eliminación de las sustituciones de Tipo 2 tras el tratamiento con UDG sugiere, sin embargo, que el principal mecanismo responsable es la desaminación de los residuos de Deoxi-Citosina.

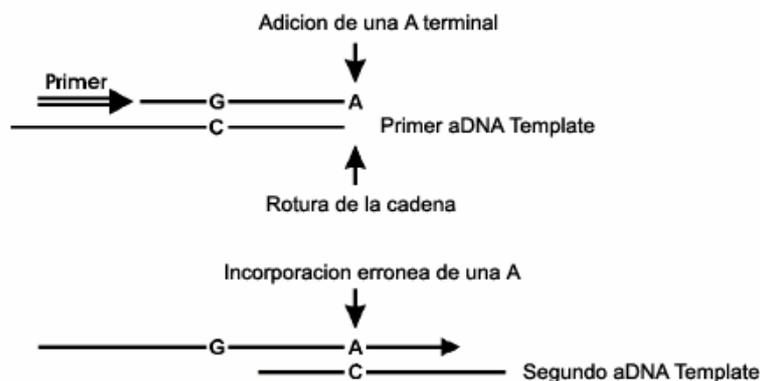


Figura F111. Mecanismo alternativo mediante el cual pueden producirse lesiones de Tipo 2 en las cadenas de DNA antiguo. Modificado de HOFREITER *et al.* 2001.

El segundo tipo de cambio más común en muestras antiguas son las transiciones A→G y T→C, que GILBERT y colaboradores (2003b) pasaron a denominar “*Transiciones de tipo 1*”. Como en las Transiciones de tipo 2, se comprobó que los cambios estaban causados por un único mecanismo molecular; en este caso, la desaminación de la Adenina (A) a un análogo de la Guanina, la Hipoxantina (HX) (GILBERT *et al.* 2003a).

Las transversiones C→A/G→T aparecen, también, en forma de *sustituciones consistentes* en los productos de amplificación de algunas muestras de DNA antiguo, como el espécimen tipo de Neandertal (KRINGS *et al.* 1997). El cambio puede explicarse por la transformación de los residuos de deoxi-Citidina en 8-Hidroxi-deoxi-Guanosina en la cadena de DNA *template*, que provoca la introducción de residuos de deoxi-Adenosina.

3.5.2.2.1. Hot-spots de daño molecular

Aunque el trabajo pionero de HOFREITER *et al.* 2001 sugirió que el daño molecular se distribuía al azar a lo largo de la región control del mtDNA, estudios más recientes mostraron que hay puntos de especial incidencia en ese tracto. Muchos de estos lugares constituyen *hot-spots* mutacionales o posiciones hipermutables en poblaciones actuales (GILBERT *et al.* 2003a). En la **Tabla I6** se expone la ocurrencia

de daño molecular y las tasas de mutación de determinadas posiciones de la HVSI del mtDNA.

Como algunas de tales posiciones son identificadoras de haplogrupos, una incorrecta evaluación del daño molecular de una muestra podría conducir a una asignación errónea de haplogrupo. La frecuencia de este tipo de asignaciones erróneas depende del número de mutaciones que definen el haplogrupo así como de la tasa de daño en cada una de ellas (GILBERT *et al.* 2003a).

La estructura secundaria y terciaria del DNA en estas posiciones –relacionada por algunos autores con la presencia de *hot-spots*–, podría explicar también su susceptibilidad a las modificaciones *post-mortem* (HEYER *et al.* 2001).

Adicionalmente, parece que existen ciertas restricciones funcionales comunes a la mutación y al daño molecular. Lo demuestra el hecho de que las tasas de daño molecular y de mutación *in vivo* sean superiores en la HVRI que en regiones codificantes (GILBERT *et al.* 2003a).

En la región 7sDNA HVRI hay tres subregiones de función conocida: 7sDNA (16106→16195), MT5 (16194→16208) y TAS (16157→16172) (DODA *et al.* 1981; OHNO *et al.* 1991). La región 7sDNA está formada por un fragmento pequeño de la cadena H que proporciona la característica triple cadena del DNA mitocondrial, la cual se extiende desde el codón STOP, en la posición 16104, hasta, al menos, la posición 00110 (DODA *et al.* 1981; MEYER *et al.* 1999). La *Secuencia Asociada a la Terminación* (TAS) interacciona con factores de unión específicos de secuencia (DODA *et al.* 1981; WALLACE *et al.* 1995) mientras que se ha postulado que la región MT5 es un lugar de unión de proteínas (MEYER *et al.* 1999).

Hay estudios que no han detectado ni mutaciones ni daño molecular en el codón STOP de la región 7S (EXCOFFIER y YANG 1999; MEYER *et al.* 1999; GILBERT *et al.* 2003a). Por su parte, las regiones funcionales TAS y MT5 presentan unas tasas de daño molecular y de mutación *in vivo* muy bajas (MEYER *et al.* 1999). Estos resultados refuerzan la hipótesis de la existencia de restricciones funcionales comunes a la mutación y al daño molecular.

Tabla 16. Tasas de mutación estimadas *in vivo* y tasa de mutación *post-mortem* de diferentes posiciones de la Región Hipervariable I del DNA mitocondrial. Las tasas han sido estandarizadas para su comparación. Se señalan en negrita las posiciones con diferencias entre las tasas de mutación *in vivo* y la tasa de daño molecular *post-mortem*. Modificada de GILBERT *et al.* 2003a.

Posiciones HVRI mtDNA	Tasas de mutación “in vivo”		Tasa daño molecular
	Excoffier y Yang 1999	Meyer <i>et al.</i> 1999	Gilbert <i>et al.</i> 2003a
16093	4	3	4
16126	0	4	3
16129	4	4	1
16148	0	3	1
16163	2	3	3
16172	4	3	1
16182	4	0	1
16183	4	3	2
16187	0	3	1
16189	4	4	1
16192	4	3	1
16209	4	0	1
16219	0	3	1
16223	4	4	4
16230	0	4	1
16234	3	0	2
16265	4	2	1
16270	4	3	4
16274	0	3	2
16278	4	4	2
16290	2	2	3
16291	4	2	1
16293	4	4	1
16294	4	4	2
16298	0	2	4
16304	4	0	1
16309	4	4	1
16311	4	4	2
16319	0	3	2
16320	3	2	1
16327	0	2	3
16343	3	2	1
16355	3	2	2
16362	4	4	1

Otro tracto de la HVRI, la región LDR1 (*Low Damage Region 1*) (16365→16395) –al que no se le ha atribuido ninguna función específica– presenta niveles muy reducidos de daño molecular. Una posible explicación sería que esta subregión resultara protegida mediante la asociación del mtDNA a determinadas proteínas (GILBERT *et al.* 2003b).

3.5.2.2.2. Posibles soluciones frente a las “*miscoding lesions*”

Una de las soluciones más efectivas para superar el efecto de las lesiones moleculares en las cadenas de DNA *template* es su incubación con la enzima Uracil-DNA-glicosilasa antes de la *PCR* (HOFREITER *et al.* 2001). El tratamiento resulta

especialmente útil cuando no está dañado todo el DNA endógeno; entonces, pueden eliminarse selectivamente las cadenas dañadas, impidiendo su actuación como sustratos de amplificación y respetando las cadenas intactas, que actuarán como *template*.

Para estimar la probabilidad de que las lesiones presentes en las cadenas de DNA originales aparezcan en la secuencia final, es recomendable determinar el número de copias de DNA *template* en los extractos. Se ha demostrado que cuando la amplificación se inicia a partir de una o pocas moléculas de DNA original es mucho más probable que las modificaciones químicas, si las hubiera, afecten al resultado final. No obstante, cuando la amplificación se inicia a partir de más de 1000 moléculas es difícil que se observen sustituciones consistentes en los clones derivados de la misma (HANDT *et al.* 1996).

Alternativamente, cuando la cantidad de moléculas de DNA original es baja se recomienda realizar múltiples amplificaciones –preferiblemente a partir de diferentes extractos–, y secuenciar varios clones de cada una de ellas. Las sustituciones que no varíen entre las diferentes amplificaciones probablemente sean de DNA endógeno.

3.5.3. Inhibidores de la PCR

En los primeros estudios de DNA antiguo ya se observó que, en algunos casos, los extractos de DNA de tejido, diente o hueso, presentaban otras moléculas de origen desconocido (PÄÄBO 1985b, 1987; BROWN y BROWN 1992; DICKEL y HAUSWIRTH 1992). Posteriormente, y con la generalización del uso de la *PCR*, esta observación inicial fue relacionada con la presencia, en algunos de estos extractos, de sustancias inhibitoras de la reacción de amplificación (PÄÄBO *et al.* 1988; PÄÄBO 1989; HAGELBERG y CLEGG 1991).

La naturaleza y mecanismos de acción de dichos inhibidores resultan todavía hoy desconocidos. Sin embargo, se han recopilado un conjunto de observaciones empíricas que orientan bien al respecto:

- El poder inhibitor está presente en extractos de diferente naturaleza, procedencia y antigüedad.
- Muchos extractos con capacidad inhibitor presentan una coloración marrón oscura (PÄÄBO 1989; HAGELBERG y CLEGG 1991; KOLMAN y TUROSS 2000) y, en algunos casos, al visualizar una alícuota de los mismos en un gel de electroforesis se observa la emisión de fluorescencia (PÄÄBO 1990; HUMMEL y HERRMANN 1991; PERSSON 1992; KOLMAN y TUROSS 2000, SCHOLZ *et al.* 1998), incluso en ausencia de Bromuro de Etidio (HÄNNI *et al.* 1995).
- El patrón de absorción de luz UV de algunos extractos con poder inhibitor difiere de los que no inhiben (ARROYO-PARDO *et al.* 2002).
- El protocolo de extracción convencional de DNA (método Fenol-Cloroformo) no consigue eliminar las moléculas inhibitoras puesto que aparecen en los extractos finales.
- Los inhibidores son retenidos por el filtro de 30000 daltons (Centricon 30 o Centriplus 30000), utilizado para concentrar el DNA en el último paso de la extracción (SCHOLZ *et al.* 1998).

3.5.3.1. Naturaleza de los inhibidores

Las moléculas que se han propuesto como responsables de la inhibición de la *PCR* son básicamente compuestos del suelo, o subproductos de degradación orgánica. A continuación se describen brevemente algunos de ellos.

3.5.3.1.1. Ácidos húmicos y fúlvicos

La posible acción inhibitora de los ácidos húmicos y fúlvicos sobre las enzimas de biología molecular fue señalada, por vez primera, por Svante Pääbo (PÄÄBO 1989). Estos compuestos constituyen una superfamilia de moléculas muy abundantes en el suelo, que pueden acompañar al DNA en el proceso de extracción y que son inhibidores muy potentes de la reacción de *PCR*. Tan sólo 100ng de ácido fúlvico son suficientes para la inhibición total de una amplificación control de fago λ (TUROSS 1994). Además, las propiedades fluorescentes de esta familia de compuestos han sido

relacionadas con la fluorescencia observada en agarosa de algunos de los extractos de DNA antiguo.

3.5.3.1.2. Residuos de porfirinas o productos derivados de su degradación

Los derivados de la porfirina o sus productos de degradación han sido señalados por algunos autores como responsables de la actividad inhibidora de la *PCR* de algunos extractos de DNA antiguo (HIGUCHI 1992; MONTIEL *et al.* 1997). La acción inhibidora de esas moléculas se basa en su capacidad para secuestrar iones metálicos, necesarios para el funcionamiento de la polimerasa.

Las porfirinas están presentes en algunas hojas vegetales –además de en la sangre y en los tejidos blandos–, y constituyen una familia molecular que puede considerarse como derivada de un compuesto tetrapirrólico original: la porfina. Las porfirinas más abundantes son las protoporfirinas y aunque pueden describirse quince protoporfirinas isómeras, solo una forma, la protoporfirina IX, está presente en la naturaleza. La protoporfirina IX forma complejos quelatos tetradentados con iones diversos, en los que el metal se une por enlaces coordinados a los 4 nitrógenos de sus anillos pirrólicos. El efecto inhibidor de la protoporfirina IX se ha relacionado con su capacidad de quelar los iones magnesio, necesarios para el funcionamiento de la *Taq* polimerasa (HIGUCHI 1984). Adicionalmente, los derivados porfirínicos tienden a formar agregados moleculares, lo que explicaría que quedarán retenidos junto al DNA en la membrana del Centriplus 30000.

3.5.3.1.3. Productos de la reacción de Maillard

La reacción de Maillard es sobradamente conocida en el campo de la tecnología de los alimentos desde mediados de los años 60. Consiste en la reacción de un grupo carbonilo con un grupo amino que da lugar a una serie de intermediarios como compuestos carbonílicos y deoxiosonas, estos últimos cuando hay azúcares implicados en la reacción.

Aunque los productos de Maillard son muchos y variados –desde grupos carbonilo unidos a cadenas cortas de carbono hasta heterociclos como furanos y

pirroles-, el producto final de la reacción de Maillard son las Melanoidinas, cuya estructura es en buena parte desconocida. Su coloración marrón y su elevado peso molecular (hasta 100.000 daltons) ha sido relacionada por muchos autores con la capacidad inhibidora de algunos extractos de aDNA (PÄÄBO 1989, 1990, 1993; PÄÄBO *et al.* 1989; LINDAHL 1993a; HÄNNI *et al.* 1995).

Se ha postulado que la propia fragmentación del DNA podría inducir la producción de reacciones de Maillard, dado que las pentosas liberadas tras la ruptura de sus enlaces fosfodiéster y N-glicosídico constituyen también un sustrato para esta reacción.

3.5.3.1.4. Productos de degradación del DNA

La degradación del material genético, sea por rotura de sus enlaces o por modificaciones químicas tales como la reducción de azúcares, genera ciertos productos capaces de inducir la inhibición directa de su propia amplificación (ROGAN y SALVO 1990; ABROL y CHAUDHARY 1993; PIKAART y VILLEPONTEAU 1993).

SCHOLZ *et al.* 1998 demostraron, *in vitro*, que el colágeno de tipo I degradado presentaba características compatibles con el comportamiento de las sustancias inhibidoras:

- exhibe un gran poder inhibidor sobre una *PCR* con DNA fresco en su forma nativa (colágeno S).
- proporciona una fluorescencia azul cuando se analiza en un gel de agarosa.
- presenta un máximo de absorbancia a 270-290nm.

En la mayoría de los casos, el tratamiento con colagenasa eliminaba la emisión de fluorescencia en gel de agarosa y la absorbancia a 270-290nm, consiguiéndose recuperar la amplificación de DNA fresco.

Sin embargo, no hay que confundir la capacidad inhibidora de un extracto de aDNA con la imposibilidad de su amplificación. El daño extensivo en el DNA, a causa de su fragmentación o modificación de bases, hace inviable en muchos casos la

amplificación. No obstante, a menos que éste genere productos capaces de interferir en la reacción de amplificación no podrá hablarse de inhibición *sensu strictu*.

3.5.3.2. Eliminación de los inhibidores

A pesar de que se desconocen la naturaleza y los mecanismos de acción de los inhibidores, se han descrito varios procedimientos para eliminar o atenuar el efecto inhibitor. Cabe distinguir los realizados **durante la extracción** –cuyo objetivo fundamental es separar el DNA del inhibidor–, y los llevados a cabo **durante la PCR**, – que tratan de contrarrestar el efecto inhibitor durante la propia reacción de amplificación–.

Muchos de los procedimientos llevados a cabo **durante la extracción** proceden de observaciones empíricas de que su empleo comportaba una disminución del efecto inhibitor, por lo que se incorporaron a los protocolos estándar de extracción de aDNA. Así sucedió con la observación de que los lavados cortos y sucesivos del polvo de hueso y/o diente con EDTA 0.5M –utilizados para su descalcificación– conseguían eliminar, al menos parcialmente, la coloración marrón que en algunos casos los acompañaba y el brillo azul de los extractos en la electroforesis en gel de agarosa (HAGELBERG y CLEGG 1991; KOLMAN y TUROSS 2000).

En otros casos la estrategia empleada consiste en el uso durante la extracción de agentes quelantes del DNA tipo Chelex[®], glass-milk o silica. Este método, sin embargo, no consigue eliminar totalmente los inhibidores (SCHNEE-GRIESE y LINDER 1994). En el trabajo de KALMÁR *et al.* 2000, la aplicación de Dextran Blue –una sustancia que se une selectivamente al DNA–, seguida de una precipitación con etanol, demostró ser efectiva para eliminar la fluorescencia azul de algunos extractos de aDNA en geles de agarosa, revirtiéndose su poder inhibitor sobre una *PCR* con DNA fresco.

En su trabajo de 1989, Svante Pääbo propone la centrifugación en un gradiente de sucrosa para la separación de los agentes inhibidores y de los ácidos nucleicos, seguida de una precipitación de los últimos con etanol. Sin embargo –y como él mismo reconoce en el mismo trabajo–, este método no es del todo recomendable cuando se desea llevar a cabo la amplificación de los extractos, pues puede comportar la pérdida

de grandes cantidades de DNA. En su lugar, Pääbo recomienda el empleo de microconcentradores Centricon 30 (Amicon). Por el contrario, HÄNNI *et al.* 1995 recomiendan sustituir estos microconcentradores –que se sabe no eliminan las sustancias inhibitoras–, por una precipitación con isopropanol.

Otras estrategias consisten en la purificación de los extractos –una vez concluida la extracción– mediante electroforesis en gel de agarosa, columnas Sephadex G-200, o kits comerciales de purificación. Según MILLER *et al.* 1999, los dos primeros son los tratamientos de purificación que más efectivamente eliminan los terminadores.

Para tratar de contrarrestar el efecto de la inhibición **durante la PCR** se han empleado diversas estrategias, de efectividad discutible, que pasaremos a enumerar a continuación.

El empleo de grandes cantidades de *Taq* polimerasa en la *PCR* (PÄÄBO *et al.* 1988) se basa en la suposición de que el mecanismo de inhibición se produce a través de la interacción del inhibidor con la polimerasa y en la existencia de una relación entre las cantidades absolutas de ambos. Dado el desconocimiento del mecanismo de acción de los inhibidores, la efectividad de este método es discutible. Además, es un método costoso que puede provocar la pérdida de especificidad (HIGUCHI 1992).

Otro método, comúnmente empleado a principios de los 90, consiste en la adición de Albúmina Sérica Bovina (BSA) a la mezcla de reacción (PÄÄBO *et al.* 1988; HAGELBERG *et al.* 1989).

Por otra parte, hay alternativas tendentes a diluir los inhibidores de los extractos, para posibilitar la amplificación (PÄÄBO 1990; WOODWARD *et al.* 1994a). Su inconveniente reside en que provocan una disminución paralela del DNA endógeno.

La mayoría de estas medidas resultan en un incremento del riesgo de contaminación con DNA exógeno de la muestra, pues aumentan la manipulación por parte del investigador. Sin embargo existen otras –como el almacenamiento de los extractos en frío durante varios días previamente a su amplificación–, que producen mejoras sustanciales en la amplificación y no presentan este inconveniente. Este

procedimiento produce la formación de un precipitado blanquecino que queda adherido a la pared del tubo y cuya eliminación por precipitación consigue revertir la capacidad de amplificación (HIGUCHI 1992; MONTIEL *et al.* 1997, 2001).

3.5.4. Contaminación

El primero en apuntar el riesgo de la contaminación con DNA exógeno cuando se emplea la *PCR* fue Svante Pääbo en su trabajo de 1989. La gran sensibilidad de esta técnica hace que, en ausencia o en bajas cantidades de DNA *template* endógeno, pueda producirse la amplificación preferente de DNA de otra fuente distinta.

3.5.4.1. Fuentes y etapas de riesgo de contaminación

Hay cinco momentos en los que puede producirse la contaminación de la muestra con DNA exógeno:

- Durante la deposición del resto, por traspaso de DNA entre organismos próximos.
- Durante la excavación por el personal arqueológico.
- Durante el depósito en un museo por el personal encargado de la conservación.
- Durante el análisis genético en el laboratorio por el personal investigador o por material genético extraído y/o amplificado con anterioridad en el mismo lugar.

Durante el tiempo que el resto permanece depositado puede producirse contaminación por DNA de microorganismos, hongos y fauna putrefactiva. En caso de enterramientos colectivos puede producirse también, al menos teóricamente, un traspaso de DNA de un espécimen a otro si existe mucha proximidad entre ellos. La contaminación con DNA de organismos de una especie diferente a la que se pretende estudiar no tiene por qué representar un gran problema si la región a caracterizar está lo suficientemente diferenciada a nivel de especie. Se consigue mediante el diseño de cebadores específicos y unas condiciones restrictivas de *PCR*. La contaminación con DNA de la misma especie durante este período puede resultar o no problemática dependiendo del objetivo final del estudio. Si se desea realizar un estudio poblacional o caracterizar a una especie extinta la existencia de traspaso de información entre

especímenes próximos no tiene por qué alterar las conclusiones, siempre y cuando exista contemporaneidad entre éstos. Ahora bien, si lo que se pretende es establecer relaciones de parentesco entre individuos de un mismo enterramiento esta posibilidad sí que debe de ser tenida en cuenta.

El DNA contaminante introducido durante la deposición del organismo estará, por tanto, sometido a los mismos factores diagenéticos que su DNA endógeno, por lo que es de esperar que transcurrido este período el nivel de degradación y alteración de ambos sea similar. Ello implica que ciertos criterios de autenticidad (ver siguiente apartado) –como la relación inversa observada entre la longitud del fragmento de amplificación y la intensidad de la banda de amplificación en el gel de agarosa–, no sirvan para la detección de este tipo de contaminación. El procedimiento más efectivo, en este caso, es la obtención de información genética a partir de porciones diferentes –y distantes– del organismo en cuestión.

La principal fuente de contaminación en los estadios siguientes a la deposición es el DNA actual de origen humano. Este DNA puede contaminar el resto antes, durante y después del aislamiento de su material genético en el laboratorio. El DNA contaminante puede introducirse directamente en la muestra mientras ésta es manipulada por los arqueólogos, el personal del museo o los propios investigadores, por ejemplo, a través de la descamación de la piel o de los aerosoles de la respiración. Otros vehículos de contaminación pueden ser el material y reactivos empleados para su análisis genético (tubos falcon, eppendorfs, puntas de pipetas, reactivos de extracción y *PCR*, aceite mineral...). De nuevo aquí, la fuente de contaminación más común son los propios investigadores. Sin embargo, se ha sugerido una posibilidad aún más inquietante como es la presencia en el material de laboratorio de cantidades importantes de DNA contaminante, introducido durante su fabricación y/o almacenamiento (SCHMIDT *et al.* 1995).

Hasta ahora nos hemos referido al DNA “fresco” –no tan modificado ni fragmentado como el de la propia muestra– como el factor contaminante. Precisamente por ello aumenta su potencial contaminante de manera que, incluso habiendo suficiente DNA endógeno, es de esperar que el DNA actual se amplifique preferentemente, dada su mayor calidad.

Pero, en el laboratorio, durante las fases de extracción y amplificación, puede darse también contaminación por otras fuentes mucho más difíciles de detectar y erradicar: el DNA extraído con anterioridad y los amplicones de anteriores experimentos. Estos contaminantes pueden flotar en forma de aerosoles en el ambiente, o quedar adheridos tras una amplificación y/o extracción a las superficies o al material de trabajo no desechable. Este tipo de contaminación se conoce comúnmente con el nombre de “contaminación de arrastre” o “*carry over*” (KWOK y HIGUCHI 1989).

3.5.4.2. Control de la contaminación

Para tratar de minimizar la contaminación introducida por traspaso directo de material genético desde el manipulador –arqueólogos, conservadores de museos, investigadores– a la muestra se emplean comúnmente un conjunto de medidas:

- Uso de ropa protectora: bata de laboratorio, guantes de látex o similar, mascarilla y reddecilla para el cabello.
- Eliminación de la capa superficial de la muestra antes de proceder a la extracción de su material genético.
- Realización de las extracciones y amplificaciones en el interior de una campana de flujo laminar.

Otras recomendaciones orientadas a reducir la contaminación en el laboratorio con DNA de experimentos anteriores son:

- Separación física de las áreas en las que se llevan a cabo los procedimientos de *pre* y *post-PCR*.
- Limpieza e irradiación con UV de las superficies de trabajo, antes y después de cada experimento.
- Empleo de material desechable de un solo uso.

Un procedimiento bastante efectivo, desarrollado recientemente para prevenir la contaminación por amplicones, es el uso combinado en las reacciones de amplificación de la enzima Uracil-DNA-glicosilasa (UDG) junto con el sustrato uracilo (dUTP) en

lugar de Timina (dTTP). La Uracil-DNA-glicosilasa rompe las cadenas de DNA en aquellos puntos en los que éstas presentan un uracilo (U). Si en las reacciones de amplificación se emplea rutinariamente en lugar de Timina (dTTP), Uracilo (dUTP), los amplicones generados podrán ser potencialmente eliminados por la Uracil-DNA-glicosilasa. La adición de esta enzima a la mezcla de reacción provoca la degradación selectiva del DNA contaminante (con U) respetando el DNA molde (sin U).

La ausencia de contaminación en todas las fases es prácticamente imposible de conseguir. Son numerosos los estudios que demuestran que, a pesar de emplear todas las precauciones y recomendaciones, siguen quedando cantidades traza de DNA contaminante (KRINGS *et al.* 1997; SCHMITZ *et al.* 2002; SERRE *et al.* 2004; LALUEZA-FOX *et al.* 2005). La detección de la contaminación resulta, por tanto, vital para descartar falsos positivos. Algunos de los procedimientos para la detección de la contaminación son:

- El uso de controles negativos que contienen todos los reactivos excepto la muestra, y que se procesan en paralelo a las muestras.
- La caracterización genética, para el marcador y fragmento que se desea estudiar, de todo el personal que haya tenido contacto con la muestra en algún momento.
- La realización de extracciones múltiples a partir de un mismo extracto, de diferentes extractos de una misma muestra, y/o de extractos de diferentes partes del organismo.
- La replicación del experimento en un laboratorio independiente por personal diferente.

3.5.5. *Jumping PCR*

Cuando la síntesis de una nueva cadena de DNA por *PCR* se ve interrumpida bien por la rotura o bien por la presencia de lesiones que bloquean la polimerasa en el DNA molde, la *Taq polimerasa* puede continuar la extensión de estas cadenas empleando otro fragmento de DNA diferente como molde. Este mecanismo, conocido como *jumping PCR*, tiene como consecuencia la formación de secuencias quiméricas que reúnen información genética de fragmentos de DNA no relacionados.

El *jumping PCR* es un fenómeno bastante común en las amplificaciones de DNA antiguo debido a su alto nivel de rotura y modificación molecular (PÄÄBO *et al.* 1989,

1990). La identificación de las secuencias generadas por *jumping PCR* puede resultar complicada, sobre todo si dentro de un mismo extracto conviven moléculas de DNA de diferente origen. En los extractos de DNA antiguo esta última situación es bastante frecuente y puede producirse por contaminación con DNA exógeno o por la existencia de cadenas modificadas molecularmente e intactas.

La probabilidad de que la secuencia directa reproduzca alguna de tales moléculas quiméricas dependerá de la relación entre el número de cadenas de *DNA template* modificadas e intactas presentes en el extracto. En caso de haber un número elevado de cadenas modificadas, el riesgo de que las secuencias finales obtenidas sean fruto del *jumping PCR* será también elevado. Entre las soluciones propuestas para detectar el *jumping PCR* están la estima del número de moléculas de DNA template, la realización de múltiples amplificaciones a partir de uno o varios extractos del mismo espécimen, y la secuenciación de diversos clones de cada amplificación.

3.6. Métodos de prospección molecular

El estado de preservación del material genético varía de un organismo a otro, por lo que es de esperar que no todos los especímenes antiguos contengan DNA amplificable. Dado el carácter destructivo de las técnicas de análisis, resulta conveniente el empleo de procedimientos que permitan evaluar *a priori* las posibilidades de éxito en la obtención de información genética fiable a partir de un determinado resto. Dichos procedimientos, conocidos como “de prospección molecular” (MONTIEL 2001) deben de ser de realización económica y sencilla, además de requerir poco material (HÖSS *et al.* 1996b). En el estudio del DNA antiguo se han propuesto diversos de estos métodos, algunos de los cuales se exponen a continuación.

3.6.1. Análisis macroscópico y/o microscópico

En el caso de los tejidos blandos la buena preservación de la estructura tisular y/o celular es un buen criterio para presuponer una buena preservación del material genético. Sin embargo, algunas estructuras pueden aparentarlo tan sólo –por el reemplazo de sus biomoléculas durante el proceso de mineralización–, de modo que, a un nivel prospectivo, este criterio resulta por sí solo insuficiente. Además, puede ocurrir que el DNA perviva fuera de la célula asociado a ciertas proteínas; por lo que no es condición necesaria la conservación de la estructura celular para su supervivencia. Este fue el caso del cerebro de Windover, en el que se pudo extraer DNA a pesar del escaso número de células completas observadas (DORAN *et al.* 1986).

En tejidos óseos, la relación entre la conservación de la estructura tisular y celular y la preservación del material genético no resulta evidente. La asociación del DNA con moléculas más resistentes de la célula –como la hidroxiapatita o ciertas proteínas nucleares– que lo protegen de la degradación, hace que no sea necesaria la preservación macroscópica para la pervivencia del DNA. Algunos autores, no obstante, mantienen que la buena preservación histológica del material esquelético es buen indicador de la presencia de DNA amplificable (COLSON *et al.* 1997). Por otra parte la coloración externa del hueso parece guardar cierta relación con su nivel de degradación orgánica, además de con el contenido de moléculas inhibitoras de la PCR (KOLMAN y TUROSS 2000). Por ejemplo, los huesos que han sido sometidos a una acción

constante de agua durante la deposición muestran una decoloración superficial oscura característica. Por su parte, la presencia de ácidos húmicos y derivados en el entorno del enterramiento puede generar una coloración marrón en la superficie ósea que, en la mayoría de los casos, permanece aún tras sucesivos lavados con EDTA.

3.6.2. Racemización de aminoácidos

Todos los aminoácidos, a excepción de la glicina, pueden presentarse en dos formas isoméricas *L* y *D*. Sin embargo, únicamente las formas *L* son empleadas en la biosíntesis proteica. Cuando un organismo muere sus aminoácidos sufren un proceso de *racemización*, en el que las formas *L* se transforman en formas *D* hasta que sus cantidades se igualan (POINAR *et al.* 1996).

La tasa de racemización difiere en cada aminoácido y, por otra parte, depende de factores tales como la presencia de agua, la temperatura, el pH y la quelación de ciertos iones metálicos a las proteínas. Muchos de estos factores afectan, también, a la principal reacción hidrolítica responsable de la degradación espontánea del DNA, la depurinización. Por esta razón se ha considerado el estudio de la racemización de los aminoácidos como un método útil a la hora de estimar la probabilidad de recuperar información genética fiable de un resto antiguo. De los diferentes aminoácidos, el ácido Aspártico presenta una alta energía de activación y una tasa de racemización muy rápida que es constante en un amplio rango de temperatura –a pH neutro– y similar a la tasa de depurinización del DNA.

En 1975, Helfman y Bada observaron que durante el envejecimiento en ciertos tejidos con bajo *turn over* (tasa de recambio) proteico –como la dentina, el esmalte dentario, los discos intervertebrales o el cristalino ocular–, se va produciendo una racemización gradual del ácido Aspártico de forma que el incremento de la concentración de *D*-Aspártico es dependiente de la edad con una determinada tasa de transformación. A partir de estos hallazgos, han sido muchos los equipos de investigación que han utilizado la tasa de racemización de este aminoácido para estimar la edad de un resto cadavérico (OGINO *et al.* 1985; OHTANI *et al.* 1990; OHTANI y YAMAMOTO 1991; RITZ-TIMME *et al.* 2003, 1990; FISHER *et al.* 1986; FUJII *et al.* 1989; GARNER *et al.* 1978; MAN *et al.* 1983; MASTERS *et al.* 1977, 1978).

Teniendo en cuenta que la racemización de los aminoácidos sigue una ley de tasa reversible de primer orden, la reacción de la racemización puede expresarse como:

$$\ln\left[\frac{(1 + D/L_{Asp})}{(1 - D/L_{Asp})}\right]_t - \ln\left[\frac{(1 + D/L_{Asp})}{(1 - D/L_{Asp})}\right]_{t=0} = 2kt$$

donde D/L es la relación (*ratio*) de formas D y L , y t es el tiempo. El término logarítmico $t=0$ describe la cantidad de ácido D -Aspártico formado mediante el procedimiento de hidrólisis con HCl 6 M a 100°C durante 6 horas, y k expresa la tasa de racemización.

La tasa de racemización no sólo depende del tipo de tejido analizado, sino de la fracción analizada del mismo. En general la fracción péptido-soluble (SP) presenta una tasa de racemización mucho mayor que la fracción colágeno-insoluble (IC) *in vivo* en determinados tejidos (OHTANI y YAMAMOTO 1991; MATSU'URA y UETA 1980). Ello se debe a que la molécula de colágeno está sujeta a menores remodelaciones, mientras que la fracción peptídica puede tener intercambio nutricional con la circulación.

Tras estudiar el grado de racemización de diversos aminoácidos en muestras de diferentes organismos y épocas, Poinar y colaboradores observaron que para relaciones D/L Aspártico mayores de 0.08 no era posible obtener una secuencia fiable de DNA (**Tabla I7**), estableciendo en D/L Asp=0.1 el límite para la recuperación de secuencias útiles. Parecía existir, además, una relación clara entre la *ratio* de las dos isoformas de este aminoácido y la longitud de la secuencia de DNA que podía recuperarse: en muestras con una D/L igual a 0.05 se obtenían secuencias de entre 140 y 340 pares de bases, mientras que en aquéllas con una tasa de racemización mayor tan sólo era posible la amplificación de fragmentos más cortos (**Figura FI12**).

En líneas generales el experimento no mostró la existencia de una correlación entre la edad del resto y su grado de racemización. Sin embargo, dado que la mayor parte de las muestras que produjeron resultados positivos de amplificación procedían de ambientes fríos, sí parecía existir una relación entre la procedencia del resto y su contenido en DNA amplificable. Teniendo en cuenta este último aspecto, los autores

estimaron que el máximo tiempo de pervivencia del DNA era de 1000 años en ambientes cálidos y de 100000 en ambientes fríos.

Tabla 17. Racemización de aminoácidos y amplificación de DNA. *D/L*: relación de formas D y L de los aminoácidos estimados. Asp: Ácido Aspártico, Ala: Alanina, Leu: Leucina. DNA: Tamaño en pares de bases (pb) de los productos de amplificación obtenidos, donde 0= ausencia de resultados de amplificación. Se señalan en negrita los casos en que las relaciones entre las ratios D/L de los tres aminoácidos sugieren contaminación con aminoácidos exógenos. Modificado de POINAR *et al.* 1996.

Muestra	Procedencia	Edad (años)	D/L Asp	D/L Ala	D/L Leu	DNA (p.b.)
<i>Equus sp.</i>	California	50	0.05	0.01	0.00	340
<i>Myiodon darwini</i>	Chile	13000	0.05	0.00	0.01	140
<i>Mammuthus primigenius</i>	Siberia	9700	0.05	0.00	0.00	200
<i>Equus ferus</i>	Siberia	42000	0.06	0.01	0.01	140
<i>M. Primigenius</i>	Siberia	50000	0.06	0.01	0.00	200
<i>M. Primigenius</i>	Siberia	35-40000	0.06	0.01	0.00	200
<i>E. Nemionus</i>	Alaska	27000	0.07	0.01	0.00	140
<i>Myiodon darwini</i>	Chile	13000	0.07	0.04	0.00	140
<i>Aptornis sp.</i>	Nueva Zelanda	3000	0.08	0.01	0.00	120
<i>Bos primigenius</i>	Europa	6500	0.11	0.12	0.11	0
<i>E. Ferus</i>	Alemania	5500	0.15	0.00	0.00	0
<i>Nothrotherium shastense</i>	Nuevo Méjico	13000	0.17	0.01	0.00	0
<i>Papio cf. Cynocephalus</i>	Egipto	2300	0.18	0.02	0.00	0
<i>E. Caballus</i>	Chile	20000	0.20	0.25	0.00	0
Fémur humano	Egipto	4500	0.21	0.02	0.00	0
<i>Megalonyx</i>	Florida	13000	0.24	0.85	0.00	0
Fémur humano	Egipto	4500	0.29	0.01	0.00	0
Fémur humano	Egipto	4500	0.29	0.12	0.00	0
Fémur humano	Egipto	4500	0.30	0.00	0.00	0
Fémur humano	Egipto	4500	0.31	0.02	0.00	0
<i>Megalonyx sp.</i>	Florida	13000	0.33	0.44	0.23	0
<i>Glossotherium sp.</i>	Cuba	15000	0.34	0.29	0.01	0
<i>Acratocnus odontrigonus</i>	Puerto Rico	15000	0.49	0.61	0.15	0
<i>Scelidon chiliense</i>	Perú	15000	0.51	0.81	0.15	0
<i>Eremotherium mirabile</i>	Perú	13000	0.60	0.27	0.14	0
<i>Megalocnus sp.</i>	California	15000	0.75	0.53	0.24	0
<i>Tyranosaurus rex</i>	Montana	65 Ma	0.23	0.59	SD	
Hueso dinosaurio	Utah	65 Ma	0.21	0.91	SD	
Hadrosaurio	Antártida	65 Ma	0.16	0.16	SD	
<i>Ornithomimidea</i>	Antártida	65 Ma	0.17	0.44	0.48	
Hoja actual, <i>Clarkia</i>	Idaho	0	0.05	0.02	0.01	
Hoja 1, <i>Clarkia</i>		17 Ma	SD	0.15	0.00	
Hoja 2, <i>Clarkia</i>		17 Ma	SD	0.29	0.00	
Hoja 3, <i>Clarkia</i>		17 Ma	SD	0.42	0.05	
Hoja 4, <i>Clarkia</i>		17 Ma	SD	0.32	0.55	
Sedimento 1 <i>Clarkia</i>		17 Ma	SD	0.15	0.00	
Sedimento 2 <i>Clarkia</i>		17 Ma	SD	0.14	0.00	
<i>Hymenoptera</i>	República Dominicana	100	0.08	0.01	0.00	
<i>Hymenoptera</i>	Colombia	3000	0.03	0.02	0.00	
<i>Hymenoptera</i>	Nueva Zelanda	30000	0.03	0.00	0.00	
<i>Hymenoptera</i>	República Dominicana	35 Ma	0.01	0.01	0.00	
<i>Diptera</i>	Mar Báltico	35 Ma	0.05	0.01	0.00	

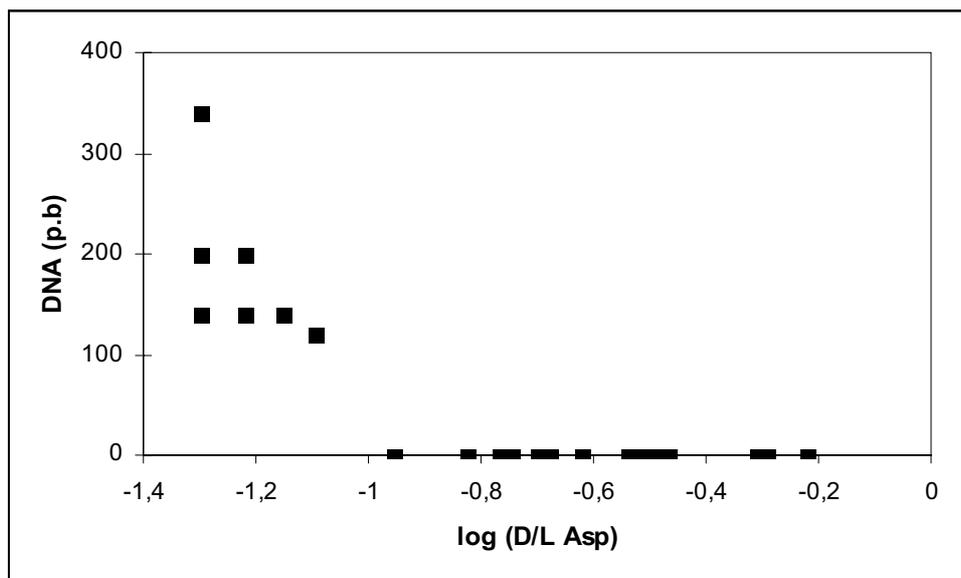


Figura F112. Relación entre la racemización del ácido Aspártico (expresada como $\log (D/L \text{ Asp})$) y el tamaño máximo de los fragmentos de DNA recuperados. Modificado de POINAR *et al.* 1996.

El grado de racemización de otros aminoácidos como la Alanina (Ala) o la Leucina (Leu) puede tomarse como indicador de la presencia o no de contaminación reciente de la muestra. Dado que la tasa de racemización de estos aminoácidos es mucho más lenta que la del ácido Aspártico, si los tres aminoácidos fueran contemporáneos la relación $D/L \text{ Asp} > D/L \text{ Ala} > D/L \text{ Leu}$ debería mantenerse.

Basándose en esta relación, POINAR *et al.* 1996 pusieron en duda algunos de los resultados de DNA antiguo obtenidos de muestras de millones de años de antigüedad. Este es el caso de las supuestas secuencias de dinosaurio (WOODWARD *et al.* 1994b), en las que la *ratio* de isoformas *D* y *L* de Alanina es claramente superior a la del ácido Aspártico, indicando por tanto contaminación de la muestra con aminoácidos exógenos. El contenido en Asp de las muestras procedentes del depósito de Clarkia (Idaho) era tan bajo que no fue posible la estima de su grado de racemización. La *ratio* de racemización de la Alanina de estas muestras proporcionó, por el contrario, un valor muy elevado, lo que según los autores resulta incompatible con la recuperación de información genética fiable. Las muestras preservadas en ámbar analizadas en ese estudio exhibían, sin embargo, una muy baja racemización de los aminoácidos estimados. Esta sorprendente preservación podría deberse al efecto protector de la

matriz de ámbar, cuyas condiciones casi anóxicas pueden ralentizar e incluso detener las reacciones de degradación del organismo.

Las principales ventajas del empleo de la racemización como método de prospección molecular son: la rapidez y la pequeña cantidad de muestra para el análisis. Su principal desventaja: la sofisticación del proceso –requiere de una infraestructura no habitual en un laboratorio convencional de biología molecular– y su coste. Por otra parte, al carecer el método de Poinar y colaboradores de una estandarización en cuanto al tipo de tejido estudiado, cabría plantearse si los valores de racemización propuestos son extrapolables a otros estudios de DNA antiguo realizados con un material diferente. Esta última cuestión se tratará más ampliamente en el apartado de discusión de la presente tesis doctoral.

3.6.3. *Flash-Pirólisis y GC/MS*

Otro método de prospección desarrollado, también por el equipo de Poinar (POINAR y STANKIEWICZ 1999), es la *flash-pirólisis* seguida de la separación de los diferentes subproductos en un Cromatógrafo de Gases (GC) conectado a un Espectrómetro de Masas (Py-GC/MS). La identificación de los productos separados por el GC se basa en las características de su espectro de masas y se realiza por comparación con los tiempos de retención y espectro de masas de pirolisados comerciales de aminoácidos: Histidina (His), ácido Aspártico (Asp), Valina (Val), Alanina (Ala), Prolina (Pro), Hidroxiprolina (HPro), Tirosina (Tyr), Triptófano (Trp), Cisteína (Cys), Serina (Ser), Glicina (Gly), Lisina (Lys), Arginina (Arg) y Leucina (Leu); dipéptidos: Pro-Ala, Pro-Val, Pro-Lys y Pro-Gly; polipéptidos (Thr-Pro-Arg-Lys y Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-His-Leu) y colágeno.

Tras someter a Py-GC/MS 11 muestras paleontológicas se observó que, en la mayor parte de los casos, los productos generados correspondían a derivados de aminoácidos y proteínas. Entre los productos más abundantes se encontraban los derivados de aminoácidos sencillos tales como la Prolina: pirrolina, pirrol y alquilpirroles C1 y C2; la Hidroxiprolina: diketodipirrol y pirrolina y la Fenilalanina: tolueno, estireno, etilcianobenceno y propilcianobenceno. Adicionalmente se detectaron algunas 2,5-diketopiperacinas, resultantes de la pirólisis de aquellos dipéptidos en los

que uno de sus componentes es la Prolina. Entre las 2,5-diketopiperacinas más abundantes de los pirolisados estudiados se encontraban las derivadas de Pro-Ala, Pro-Gly, Pro-Val y/o Pro-Arg, Pro-Hyp y Pro-Pro. La abundancia de productos derivados de Pro, Gly, Ala y HPro sugería la presencia de colágeno en la muestra original.

No existe una relación entre la abundancia de productos de pirólisis de aminoácidos simples y la preservación de DNA. Sin embargo, los resultados de este estudio apuntan a que esta relación se produce entre los productos de pirólisis derivados de dipéptidos (i.e. diketopiperacinas). La *ratio* entre la cantidad de productos derivados de aminoácidos simples y la de diketopiperacinas (AA/DKP) es una buena estima del grado de degradación de una muestra. El método predice que las muestras bien preservadas presentarán mayor cantidad de péptidos largos que darán lugar a diketopiperacinas por pirólisis. No obstante, las muestras peor preservadas que hayan sufrido una hidrólisis exhaustiva mostrarán mayor cantidad de aminoácidos libres que de diketopiperacinas. Los autores de este estudio advierten, en cualquier caso, que el método no garantiza por sí sólo la presencia de DNA amplificable en la muestra, por lo que recomiendan su empleo como complemento a la racemización de aminoácidos.

Entre las ventajas de este método se encuentran su sensibilidad –únicamente son necesarios unos miligramos de muestra para el análisis– y su rapidez. La sensibilidad es mucho menor que la de la racemización de aminoácidos, y el proceso puede generar productos resultantes de la reagrupación de las moléculas originales. La principal desventaja de la Py-GC/MS como técnica de prospección molecular es que, por sí sola, resulta insuficiente.

3.6.4. Detección de modificaciones en el DNA por GC/MS

Como se ha comentado en el apartado 3.4.2.1. de este mismo capítulo, cuando el daño oxidativo en el DNA es severo se produce un bloqueo de la *Taq* polimerasa que impide la amplificación. Otra de las manifestaciones del daño oxidativo –que, en este caso, no impide la amplificación–, es la modificación de las bases nitrogenadas (ver apartados 3.3.1.2. y 3.4.2.2.).

Este método de prospección molecular propone la detección de los productos resultantes del daño oxidativo del DNA –i.e. bases o residuos de azúcares modificados o agregados DNA-proteína–, previamente a su amplificación mediante cromatografía de gases asociada a espectrometría de masas, de forma similar a la Py-GC/MS.

A pesar de la elevada sensibilidad del método –que posibilita el análisis de cantidades mínimas de DNA–, no deja de ser una técnica destructiva que requiere una extracción del material genético para su estudio. Además –y como sucede con la racemización de aminoácidos–, se trata de un proceso caro y difícilmente reproducible en un laboratorio estándar de biología molecular, lo que cuestiona su utilidad como método de prospección molecular y la exigencia de su aplicación. En cualquier caso, la caracterización de los extractos de DNA que no hayan producido resultados de amplificación mediante GC/MS, puede proporcionar una valiosa información acerca de las causas del fracaso.

3.6.4.1. Detección de bases modificadas

Para la detección de bases modificadas la técnica más sensible y selectiva es la cromatografía de gases/espectrometría de masas (CG/MS), acompañada de una monitorización selectiva de iones (SIM). El espectrómetro de masas se programa para monitorizar un número característico de iones de un compuesto a medida que éste eluye de la columna del GC. Para una correcta identificación, por tanto, es necesario conocer el tiempo de elución y el espectro de masas característico de dicho compuesto. El cálculo de las concentraciones en las que se encuentran los productos se realiza, generalmente, añadiendo a la mezcla a analizar un isótopo marcado antes del análisis. Dicho isótopo es análogo a un analito de la mezcla, y actúa como estándar interno. En el caso de los productos del daño del DNA inducidos por radicales libres no existen analitos marcados disponibles, por lo que se emplean como estándares internos ciertos compuestos de estructura similar (WATSON 1985).

Para monitorizar un ión característico de cada compuesto, durante el análisis de CG/MS-SIM se construye, primero, la curva de calibración con la respuesta del espectrómetro de masas a cantidades conocidas del analito y el estándar. La *ratio* de corriente iónica de los iones se calibra, entonces, como una función del cociente de la

concentración molar del analito y del estándar interno. La pendiente de dicha curva representa el factor de respuesta molar relativa (RMRF).

La CG/MS-SIM permite analizar la presencia de bases modificadas en el DNA a partir de cantidades mínimas –entre 0,1 y 0,4 μg – de éste (ARUOMA *et al.* 1989c). Con la técnica pueden detectarse un mínimo de 10^6 subunidades de nucleótido, pero se necesitan cantidades de 50 a 100 μg de DNA (DIZDAROGLU 1991) para el procesamiento de la muestra del DNA (hidrólisis y derivatización).

Debido a su elevada sensibilidad y selectividad, esta técnica permite evaluar el daño oxidativo sobre el DNA en cromatina, sin necesidad de su aislamiento previo. Ello permite reproducir *in vitro*, de una forma más real, el daño molecular inducido por los radicales libres *in vivo*. La organización del DNA dentro de la célula en unidades de empaquetamiento y su asociación con proteínas puede provocar un patrón de manifestación del daño molecular diferente al observado en DNA aislado. Por ejemplo, las histonas asociadas al DNA pueden reaccionar, también, bajo la acción de radicales libres afectando a las reacciones de oxidación-reducción de las bases del DNA (DIZDAROGLU 1991). Adicionalmente, las lesiones moleculares pueden favorecer la formación de puentes intermoleculares entre las proteínas y el DNA (*DNA-protein cross links* o *DPCs*).

La aplicación de esta técnica ha posibilitado el estudio del daño molecular sobre el DNA bajo condiciones muy diversas. Así, se ha probado el efecto de dicho daño en condiciones aeróbicas y anaeróbicas, observándose que algunos productos se forman en presencia de oxígeno, otros en ausencia, y otros en ambas condiciones (GAJEWSKI *et al.* 1990; DIRKSEN *et al.* 1988). Asimismo, esta técnica ha sido empleada para la caracterización y cuantificación del daño en el DNA inducido por Hipoxantina/Xantina oxidasa, por neutrófilos humanos activados, H_2O_2 en presencia de iones metálicos y por cobre acomplejado con 1,10-fenantrolina (ARUOMA *et al.* 1989a, 1989b, 1989c, 1991; DIZDAROGLU *et al.* 1990; BLAKELY *et al.* 1990).

3.6.4.2. Detección de modificaciones en la estructura de azúcar

Mediante la cromatografía de gases/espectrometría de masas (CG/MS) es posible detectar y analizar las modificaciones producidas en la estructura de azúcar de los nucleótidos del DNA. Estos azúcares pueden haber sido liberados al medio tras su modificación o encontrarse todavía unidos al DNA (BEESK 1979; DIZDAROGLU 1974).

El primero de los casos requiere del tratamiento de la muestra con NaBD₄ para convertir los azúcares libres en alcoholes. Este compuesto provoca su reducción a polialcoholes mediante la incorporación de uno o dos átomos de deuterio en los grupos carbonilo de la molécula de azúcar. Seguidamente, se somete la muestra a una trimetilsilización y se analiza por CG/MS. Los iones fragmentados de estos derivados de polialcoholes deuterados en el espectro de masas establecen con certeza la presencia o ausencia de grupos carboxilo, carbonilo y deoxi en el azúcar precursor y las posiciones originales de estos dos últimos grupos (DIZDAROGLU 1991).

Los azúcares aún ligados al DNA –tras la reducción con NaBD₄– son tratados con álcali –lo que provoca su liberación–, y posteriormente defosforilados con fosfatasa alcalina (DIZDAROGLU 1991). Una de las modificaciones del azúcar, el ácido 2-deoxipentoico, no requiere de la primera reducción puesto que, tras el tratamiento alcalino, resulta ya identificable (DIZDAROGLU *et al.* 1977).

3.6.5. Inhibición de una PCR con DNA fresco

La adición de diferentes cantidades de un extracto de DNA antiguo a una mezcla de reacción de PCR que contiene una cantidad estándar de “DNA fresco”, permite evaluar la presencia en dicho extracto de inhibidores de la PCR (GOODYEAR *et al.* 1994; WOODWARD *et al.* 1994a; FRANCALACCI 1995; HÄNNI *et al.* 1995; SCHOLZ *et al.* 1998; MILLER *et al.* 1999; KOLMAN y TUROSS 2000; ARROYO-PARDO *et al.* 2002; FERNÁNDEZ *et al.* 2003). El ensayo simultáneo de diversos volúmenes de extracto o de diluciones sucesivas del mismo, puede ofrecer información acerca de la concentración del inhibidor y de la cantidad de extracto a añadir para conseguir resultados positivos de amplificación (HÄNNI *et al.* 1995; SCHOLZ *et al.*

1998; MILLER *et al.* 1999; KOLMAN y TUROSS 2000; ARROYO-PARDO *et al.* 2002; FERNÁNDEZ *et al.* 2003).

Este método puede resultar un buen indicador de las probabilidades de éxito de la reacción de amplificación, o dicho de otra forma, de las causas de su fracaso. Se trata de un método barato, rápido, asequible y de fácil aplicación, aunque su empleo hace necesario la extracción previa del material genético y comporta, por tanto, la destrucción de la muestra.

3.6.6. Espectrofotometría-UV y HPLC

Este método de prospección combina el análisis de los extractos de DNA por espectrofotometría dentro del espectro de luz ultravioleta, y su análisis por Cromatografía Líquida de Alta Resolución (HPLC).

La medida de la absorbancia a 260nm ha sido empleada tradicionalmente como método de cuantificación del DNA de una muestra (SAMBROOK *et al.* 1989). El método no resulta adecuado, sin embargo, para la cuantificación de extractos de DNA antiguo dado su contenido en otras sustancias que pueden enmascarar la presencia del DNA endógeno (LASSEN *et al.* 1994; MONTIEL *et al.* 1997). En los extractos de DNA antiguo el pico de 260nm se encuentra ligeramente desplazado a la derecha, probablemente a causa de la absorción de otras moléculas y a la presencia de proteínas y productos de degradación del DNA.

Según ARROYO-PARDO *et al.* 2002 la ausencia de un pico de absorción a 260nm y la presencia en el cromatograma de HPLC de dos picos con áreas superiores a 15mV min^{-1} , a tiempos de retención de 0.8 y 1 min, caracteriza a aquellos extractos que no producen amplificación. La ausencia de absorción a 260nm puede corresponder a la ausencia de DNA en el extracto. El patrón cromatográfico de HPLC observado en los extractos es similar al obtenido para una mezcla de desoxiribonucleótidos (dNTPs), por lo que podría corresponder a productos de la degradación del DNA de la muestra.

3.7. Criterios de autenticidad

Desde el inicio de la disciplina del DNA antiguo se han sugerido un conjunto de criterios y medidas metodológicas para tratar de probar la autenticidad de la información genética recuperada. La necesidad de autenticar los resultados obtenidos se hizo patente a mediados de los años 90, cuando se concluyó que un conjunto de estudios eran irreproducibles (AUSTIN *et al.* 1997b). El caso más sonado fue quizás la publicación de una putativa secuencia de DNA de dinosaurio (WOODWARD *et al.* 1994b), que resultó ser una contaminación humana de una inserción nuclear del mtDNA (ZISCHLER *et al.* 1995b). En los años siguientes, expertos en el campo fueron introduciendo paulatinamente recomendaciones para garantizar la autenticidad de los resultados obtenidos. Los consejos fueron asimilados con rapidez por las revistas científicas de impacto, que, a partir de entonces, exigen la aplicación de los mencionados criterios como requisito de aceptación y publicación de los resultados obtenidos.

A continuación se exponen y comentan los diferentes criterios de autenticidad propuestos hasta la fecha en estudios de DNA antiguo. Los criterios han sido clasificados en cinco categorías:

- Infraestructura
- Metodología
- Interpretación de los resultados
- Reproducibilidad
- Pruebas adicionales

3.7.1. Infraestructura

- *Laboratorio exclusivo de DNA antiguo.* Los laboratorios en los que se realizan tareas de extracción y amplificación de DNA antiguo deben de estar separados de los laboratorios en los que se lleva a cabo cualquier otro tipo de estudios de DNA, especialmente si se trata de DNA humano. Mediante este procedimiento se pretende eliminar el DNA actual extraído, como fuente de contaminación. Algunos autores

recomiendan incluso minimizar o eliminar la introducción de controles positivos de DNA actual en los experimentos de amplificación (O'ROURKE *et al.* 2000b).

- *Separación física de los procesos de pre-PCR de la amplificación y los procesos post-PCR.* BÉRAUD-COLOMB *et al.* 1997 recomiendan el empleo de tres habitaciones separadas: la primera para la preparación de la muestra y extracción, la segunda para la preparación de las reacciones de PCR y la tercera para la amplificación, propiamente dicha, en un termociclador dedicado exclusivamente a DNA antiguo. Los extractos de DNA obtenidos en la primera habitación deben guardarse en una habitación diferente a todas las citadas.

- *Equipamiento de un laboratorio de DNA antiguo.* Resulta recomendable que las áreas dedicadas a la extracción y amplificación de DNA antiguo estén dotadas de un fluorescente de luz UV, accionable desde el exterior, para esterilizar la zona antes y después de cada experimento. Igualmente, se ha propuesto que la habitación destinada a la amplificación tenga presión positiva en su interior y filtros HEPA (PÄÄBO *et al.* 2004).

3.7.2. Metodología

- *Empleo de instrumental y equipamiento exclusivo para el análisis de DNA antiguo.*

Incluye:

- Uso de soluciones estériles, preferiblemente de fábrica. De no ser así, se pueden esterilizar mediante autoclavado y filtración.
- Empleo de material de un solo uso.
- Empleo de puntas de pipeta con filtro, a prueba de aerosoles.
- Limpieza del material reciclable con lejía o DNAsa, e irradiado posterior con luz UV.
- Manipulación de las muestras con guantes, en todo momento. Resulta recomendable también el empleo de mascarilla facial y de redecillas para el pelo por parte del personal investigador. Debe evitarse hablar durante la manipulación de las muestras para limitar la introducción de contaminación.

- Irradiación con luz UV de la habitación y del material antes de cada análisis, para destruir posibles moléculas de DNA contaminante remanentes del experimento anterior.

- *Eliminación de la capa superficial del hueso/diente* previamente a la extracción del material genético.

- *Alicuotado de los reactivos de PCR* y realización de reacciones de amplificación periódicas sobre estas alícuotas.

- *Análisis de cada espécimen por un único investigador*. En el caso de las determinaciones de sexo resulta recomendable la replicación de todo el proceso por un investigador de cada sexo.

- *Procesado de controles o blancos*, a los que se añaden todos los reactivos excepto el tejido antiguo, y que son procesados de forma paralela a las muestras objeto de estudio durante las fases de extracción del material genético (controles de extracción) y durante la reacción de amplificación (controles de amplificación o *PCR*). Cualquier producto de amplificación derivado de estos controles indicará la introducción de contaminación exógena al tejido antiguo durante la fase correspondiente de procesado de la muestra. PÄÄBO y colaboradores, en su revisión de 2004, recomiendan la introducción de un mínimo de tres blancos de amplificación en cada reacción. La contaminación durante esta fase suele ser sutil y no uniforme, por lo que el empleo de múltiples controles aumenta la probabilidad de detectar aquellos casos de contaminación esporádica. O'ROURKE y colaboradores (2000b) recomiendan la introducción de blancos de *PCR* "tapados" y "destapados". Los primeros permanecen tapados durante el experimento y permiten controlar la contaminación presente en los reactivos mientras que los segundos permanecen abiertos hasta el fin del ensayo, por lo que es posible detectar la contaminación introducida durante el proceso (*carry over*). Ninguno de los blancos –ni de extracción ni de *PCR*– deberá proporcionar resultado positivo de amplificación en ningún caso.

- Empleo de diferentes cebadores para amplificar el mismo segmento. Mediante este procedimiento se pretende evitar la amplificación preferente de las inserciones nucleares del mtDNA.

3.7.3. Interpretación de los resultados

- Correlación inversa entre el tamaño del amplicón y la intensidad del producto de amplificación visualizado en un gel de agarosa o la eficiencia de amplificación. Esta medida permite evaluar el grado de degradación del material genético y la presencia de lesiones que bloquean la *Taq* polimerasa. El DNA amplificado no suele ser endógeno en los casos en que la eficiencia de amplificación de los fragmentos pequeños es similar a la de los fragmentos de mayor tamaño.

- Las secuencias putativas de DNA antiguo deberán tener un *sentido filogenético*. Cuando se analiza una especie diferente a la humana, el DNA obtenido no debería ser de origen humano sino tener cierta relación con lo que se conoce sobre las relaciones filogenéticas y evolutivas de la especie en cuestión con otras especies próximas. Por ejemplo, si se analiza una especie extinta de insecto, el DNA recuperado deberá ser más similar al DNA de otros insectos actuales que al de los mamíferos o plantas. Cuando se trabaja con restos humanos antiguos, el sentido filogenético puede resultar más difícil de establecer. En este caso la secuencia o secuencias obtenidas deberían diferir de las del personal implicado en la manipulación de los restos. Resulta vital, en estos casos, la caracterización genética para los marcadores estudiados de todo el personal de laboratorio y, a ser posible, de todos quienes hayan manipulado la muestra en algún momento (personal de museo, arqueólogos...). Este criterio ha sido criticado argumentándose que un criterio de autenticidad no puede presuponer el resultado que se ha de obtener. Por ejemplo, en el caso de la primera secuencia de Neandertal recuperada, la ausencia de similitud entre ésta y las secuencias humanas actuales sirvió para demostrar su autenticidad.

3.7.4. Reproducibilidad

- Replicación del experimento dentro del mismo laboratorio y realización de, al menos, dos extracciones independientes del mismo espécimen –idealmente a partir de diferentes

elementos esqueléticos y/o distintas piezas dentales–, y varias amplificaciones a partir del mismo o de diferentes extractos. Este criterio permite detectar la contaminación de una extracción y/o amplificación particular así como la presencia de daño molecular en el DNA endógeno. Por otra parte, el análisis de diferentes extractos aumenta la probabilidad de amplificación de secuencias endógenas cuando el resto analizado contiene una baja cantidad de DNA.

- Replicación en laboratorios independientes.

Con ello se pretende eliminar la posibilidad de contaminación procedente del propio laboratorio no detectada en los blancos. El criterio plantea, sin embargo, una serie de problemas que en muchos casos impiden su aplicación. En primer lugar dado que la extracción de DNA es destructiva puede no existir muestra suficiente para más de un análisis. Por otra parte, muchos especímenes son frágiles y pueden dañarse durante el traslado de uno a otro laboratorio. La replicación en otro laboratorio puede dilatar el tiempo de publicación de los resultados. En otros casos pueden existir derechos de pertenencia del resto o disputa por la autoría de los resultados científicos derivados. Finalmente, si ya resulta difícil conseguir financiación para realizar un estudio de DNA antiguo en un laboratorio determinado, extenderla a dos o más resulta en muchos casos imposible. Se puede concluir que tan sólo unos pocos laboratorios de todo el mundo están en condiciones de cumplir este requisito.

3.7.5. Pruebas adicionales

- Ensayos bioquímicos de preservación de otras macromoléculas de la misma muestra.

Su objeto es establecer si estado de preservación general del fósil es lo suficientemente bueno como para permitir la preservación de su material genético. Estos métodos se emplean habitualmente con una finalidad prospectiva, para seleccionar los restos con mayor probabilidad de contener información genética genuina. El procedimiento más generalizado es el análisis de aminoácidos (contenido total y grado de racemización).

- Análisis de restos asociados de fauna. La obtención de información genética endógena de especímenes de fauna procedente de la misma excavación indicaría que la preservación de DNA en dicho yacimiento es plausible. La amplificación de tales restos con cebadores humanos puede servir para evaluar la incidencia de contaminación.

- *Clonación bacteriana de los productos de amplificación y secuenciación de múltiples clones.* Esta técnica permite detectar cualquier heterogeneidad presente en los productos de amplificación e identificar la fuente de la misma: contaminación, daño molecular y/o errores de la polimerasa.

- *Cuantificación del número de moléculas de DNA template amplificable presentes en un extracto.* Mediante la aplicación de este criterio puede establecerse la probabilidad de obtener sustituciones consistentes en la amplificación y la necesidad de realizar amplificaciones y/o extracciones múltiples. Si el número de moléculas de DNA originales es superior al millar (GILBERT *et al.* 2003 a y b) no es necesario realizar más amplificaciones puesto que la probabilidad de que se reproduzca alguna sustitución consistente en el producto final de *PCR* es muy baja. La amplificación múltiple resulta obligada para valores inferiores al citado. El método permite, también, estimar la *ratio* de secuencias endógenas y exógenas, y, en muchos casos, identificar la fuente de contaminación.

Idealmente, los criterios mencionados deberían ser considerados en el planteamiento de cualquier estudio de DNA antiguo. Sin embargo, puede no resultar posible por razones ajenas al potencial del laboratorio, tal como hemos comentado anteriormente. Por otra parte, el resultado de algunos de los citados criterios puede hacer innecesaria la aplicación de otros. Por ejemplo, si la cuantificación de DNA *template* proporciona un número de cadenas superior a 1000 no es preciso realizar extracciones y/o amplificaciones múltiples, puesto que raramente aparecerán sustituciones consistentes en el producto final de *PCR* (PÄÄBO *et al.* 2004). No obstante, la aplicación de otros criterios, como la realización de blancos de extracción y amplificación, es obligada en todos los casos.

PÄÄBO *et al.* 2004 recomiendan la aplicación exhaustiva de todos los criterios mencionados en los casos en que de una secuencia se deriven conclusiones de gran significación biológica. Un claro ejemplo lo constituye la primera secuencia de Neandertal publicada por KRINGS y colaboradores, en 1997. En este trabajo se aplicaron con éxito todos los criterios, incluyendo la replicación en un laboratorio independiente. En posteriores publicaciones de secuencias de otros especímenes de Neandertal, no se emplearon todos los criterios al constarse la gran similitud con la

aludida secuencia inicial (OVCHINNIKOV *et al.* 2000; KRINGS *et al.* 2000; SCHMITZ *et al.* 2002; SERRE *et al.* 2004).

Con todo, es importante hacer notar que la aplicación de los mencionados criterios no garantiza la autenticidad del resultado final. Por ejemplo, si una muestra –de hueso, diente o tejido– se hubiera contaminado previamente a su extracción, aunque cumpliera todos los criterios de autenticidad, el resultado final del análisis sería erróneo. En conclusión es la contextualización de la secuencia final (*i.e.* sentido filogenético) lo que “pesa” en cuanto a su autenticidad. Este criterio ha sido aplicado a las diferentes secuencias de Neandertal publicadas (KRINGS *et al.* 1997, 1999b, 2000; OVCHINNIKOV *et al.* 2000; SCHMITZ *et al.* 2002; SERRE *et al.* 2004; LALUEZA-FOX *et al.* 2005) y ha originado un intenso debate en torno a la autenticidad de la putativa secuencia paleolítica (CARAMELLI *et al.* 2003), idéntica a la secuencia más común en la actualidad (*CRS*) (ABBOTT 2003).

