

FRANCISCO MESTRES NAVAL

Profesor Titular de Genética y Profesor de Genética Forense,
Departament de Genètica, Facultat de Biologia, Universitat de Barcelona

JOSEP VIVES-REGO

Catedrático Emérito de Microbiología. Forense medioambiental
Departament de Microbiologia, Facultat de Biologia, Universitat de Barcelona

Bancos y bases de datos genéticos para usos forenses

RESUMEN

Los bancos y bases de datos con los perfiles de ADN son herramientas poderosas en forensia. En los bancos de datos se almacenan los perfiles genéticos individuales de personas relacionadas con actos delictivos, mientras que en las bases se guarda información genética poblacional útil para interpretar estadísticamente los resultados. Los bancos de datos son útiles en la resolución de casos policiales y en la identificación de víctimas. Se describe como evitar las falacias de la acusación y de la defensa.

ABSTRACT

Databanks and databases containing the DNA profiles are powerful tools in forensics. Individual genetic profiles of people related to criminal activity are stored in databanks, whereas population genetic information useful to statistically understand the results is kept in databases. Mainly, they are useful to investigate criminal cases and to victim identification. Avoiding the prosecutor and defendant's fallacies are described.

Palabras clave

ADN, perfiles genéticos, bancos de datos, bases de datos, identificación de personas, falacia de la acusación, falacia de la defensa, identificación de víctimas.

Keywords

DNA, genetic profiles, databanks, databases, individual identification, prosecutor's fallacy, defendant's fallacy, victim identification.

SUMARIO: I. Introducción. II. Antecedentes históricos de la Genética Forense. III. Diferencias entre bancos y bases de datos. 1. Banco de datos. 2. Base de datos. IV. Solidez científico-técnica de los perfiles de ADN. V. Identificación de un perfil genético en un banco de datos y cálculo de la probabilidad de encontrar este mismo perfil por azar en la población: protocolo de referencia. VI. Argumentaciones científico técnicas sobre la utilización e interpretación de la información forense basada en el ADN: falacias de la acusación y falacias de la defensa. 1. Falacia de la acusación: interpretación errónea de la información obtenida a través del ADN. 2. Falacia de la defensa: uso inadecuado de la información proporcionada por el ADN. VII. Condiciones para la inclusión de un determinado perfil genético en un banco de datos. VIII. Situaciones en las que los bancos de datos de perfiles genéticos son útiles y en las que no lo son. IX. Conclusiones. X. Bibliografía.

I. INTRODUCCIÓN

La utilidad y las aplicaciones finales a nivel policial y jurídico de los bancos y bancos de datos policiales son básicamente tres: 1) La investigación y averiguación de delitos en el sentido más amplio de la palabra. 2) El control de las personas que han sido condenadas y presentan un alto riesgo de reiteración delictiva y que puedan acceder al tercer grado penitenciario o a la libertad condicional o definitiva. 3) Su utilización en los procesos de identificación de restos de cadáveres o de personas desaparecidas. Estas finalidades constituyen en sí objetivos éticos y jurídicos. Lo que puede ser discutible es en que medida la inclusión de cierto tipo de datos en los bancos o bases policiales es ética si no se trata de datos que puedan conducir a las tres finalidades policiales previamente definidas.

En líneas generales el llevar a cabo la denominada prueba del ADN en los procedimientos policiales y penales consta básicamente de cuatro fases: i) la toma de muestras en el lugar de los hechos previa inspección ocular u obtención por extracción directa del detenido, sospechoso o imputado; ii) el análisis del ADN contenido en las muestras que se hacen llegar al laboratorio; iii) los datos obtenidos se contrastan con los perfiles genéticos disponibles en los bancos de datos; iv) interpretación de los resultados (gracias a las bases de datos) que siempre deben serlo en términos de probabilidad y v) incorporación de los datos e interpretaciones obtenidos en las actuaciones del proceso penal para su utilización como prueba. Recientemente, se ha publicado un artículo (MESTRES Y VIVES-REGO, 2009a) en el que se revisan y discuten los aspectos científicos, periciales, procesales, sociales y éticos de la utilización forense de la huella genética (secuencia del ADN).

En este artículo analizamos, desarrollamos y hacemos propuestas forenses que pretenden resolver los problemas recientemente aparecidos a nivel internacional a la hora de interpretar correctamente los datos de identificación a través del ADN. De manera especial hemos intentado facilitar la resolución de dos problemas

potenciales en la aplicación de la genética forense en el ámbito jurídico español y que son: 1) Proponer el uso diferenciado de los conceptos de “bancos” y “bases” de datos a nivel policial y jurídico, en la línea de los usos jurídicos internacionales y 2) Evitar las denominadas falacias de la acusación y la defensa en la utilización de la prueba del ADN.

II. ANTECEDENTES HISTÓRICOS DE LA GENÉTICA FORENSE

En la actualidad nadie duda de la importancia de la información proporcionada por los perfiles genéticos para resolver diferentes cuestiones del ámbito legal, jurídico y forense. Muchas veces dicha información puede ser incluso la piedra angular para la correcta resolución de algunos casos. A pesar de que la Genética forense es una especialidad que pasa por un momento álgido debido a su aparición casi constante en los medios de comunicación y en productos cinematográficos y televisivos, no sería del todo correcto afirmar que se trata de una especialidad nueva. Sus orígenes se pueden remontar a principios del siglo XX, cuando Landsteiner identificó en los humanos el grupo sanguíneo ABO. Mediante él, los individuos de una población humana pueden clasificarse en las clases discretas que son los grupos: O, A, B o AB. Una persona sólo puede pertenecer a uno de estos grupos. Es importante darse cuenta que existe variabilidad para dicho carácter, no todas las personas son idénticas. Es más, la frecuencia (o abundancia relativa, que normalmente se presenta en forma de porcentaje) de cada uno de estos grupos puede ser característica de determinadas poblaciones humanas. Sin embargo un único grupo sanguíneo, por si sólo, no nos permite identificar a personas individuales, puesto que en nuestro país o región o ciudad existen muchos individuos que pertenecen a un mismo grupo sanguíneo. Sin embargo, el grupo sanguíneo ABO no es el único presente en los humanos. Por ejemplo también se utiliza el MN, que permite clasificar a todos los seres humanos en tres clases discretas: M, N o MN. Si añadimos un grupo más en nuestro análisis vamos perfilando más a los individuos de una población. Fijémonos que si consideramos el grupo ABO y el MN conjuntamente podremos clasificar ahora a los individuos en las siguientes categorías: OM, ON, OMN, AM, AN, AMN, BM, BN, BMN, ABM, ABN y ABMN. Es decir, cuantos más grupos sanguíneos podamos analizar en un mismo individuo, mejor podremos acercarnos más a un perfil que será individual para cada persona de la población humana. Lo único que necesitamos es una cantidad elevada de grupos sanguíneos u otros marcadores moleculares para poder identificar con precisión a un individuo en concreto. El trabajar con grupos sanguíneos y otras proteínas similares tenían un problema grave, y es que las proteínas se degradaban con bastante rapidez, ya sea por culpa de los efectos ambientales o de los microorganismos.

Un salto cualitativamente importante se produjo en 1984, cuando Jeffreys y el *Forensic Science Service* del Reino Unido introdujeron los análisis del ADN¹ en los estudios forenses (Jeffreys et al. 1985). Dicha molécula es mucho más estable que las proteínas (por ejemplo, se puede recuperar ADN a partir de momias o de colillas de cigarrillos o incluso de los intestinos de ciertos insectos necrófagos que se han alimentado de un cadáver humano). El ADN presenta una ventaja adicional y es que el número de variantes (que científicamente se denominan polimorfismos) es mucho mayor que el que se obtiene, por ejemplo, con los grupos sanguíneos. Así por ejemplo, para el grupo sanguíneo ABO, como ya hemos visto, tan sólo existen cuatro tipos de variantes (grupos A, B, AB y O). Sin embargo, con un único marcador de ADN, de los muchos que existen (del orden de millares), se pueden obtener más de 70 tipos diferentes de variantes. El ADN es una molécula que se encuentra en todos los seres vivos y cuyo cometido es guardar y transmitir la información genética de los individuos. El ser humano tiene 23 pares de cromosomas en cada una de las células de su cuerpo y cada uno de ellos contiene una molécula lineal de ADN extremadamente larga. Además, el orgánulo celular denominado mitocondria también contiene ADN. El ADN humano es una macromolécula formada por una secuencia de millones de subunidades químicas denominadas nucleótidos. Existen cuatro tipos diferentes de nucleótidos simbolizados por letras: A, T, C y G. Sus combinaciones a lo largo de las cadenas de ADN contienen de forma codificada la información genética de los organismos. Los genes pueden definirse, de forma simplificada, como secuencias de nucleótidos dentro de las cadenas de ADN que especifican una función del ser vivo. Para identificar a un individuo de forma similar a como se hacía mediante los grupos sanguíneos no es necesario conocer toda la secuencia completa de nucleótidos de un organismo (que contiene millones de nucleótidos). Para usos forenses es suficiente secuenciar determinados trozos de ADN humano (marcadores genéticos) que son lo suficientemente informativos para caracterizar una persona de forma individual y por lo tanto poder identificar a cualquier persona, animal o planta desde el punto de vista forense. Realizar dicho proceso no es más que obtener el perfil genético de los individuos. Existen diversos marcadores genéticos, tales como los minisatélites, los microsatélites (o STR del inglés, *short tandem repeats*, es decir, secuencias cortas repetidas en tandem), regiones concretas del mtDNA (ADN mitocondrial), SNP (*single nucleotide polymorphism* que puede traducirse como polimorfismo de nucleótido único) y otros. Sin embargo en la actualidad se trabaja fundamentalmente con los STR, los SNP y el mtDNA, este último para algunas situaciones concretas (para más detalles puede consultarse BUTLER 2005 o GOODWIN et al. 2007). Seleccionando una batería adecuada de dichos marcadores se obtiene un poder de discriminación tal que pueden identificarse personas de

¹ Aunque la nomenclatura científica correcta es DNA, en este trabajo utilizaremos ADN por ser la denominación habitualmente utilizada.

forma individual con una probabilidad de error despreciable. Se puede obtener el perfil genético de cada persona, que es lo que popularmente se ha denominado la “huella genética” al realizar un paralelismo terminológico poco afortunado con las huellas dactilares, pues el perfil genético personal no guarda relación con los dermatoglifos o huellas dactilares.

III. DIFERENCIAS ENTRE BANCOS Y BASES DE DATOS

La información referente a los perfiles genéticos y la variabilidad molecular que es útil en forensia puede almacenarse de forma digital o como documentos en papel. Estos ficheros que contienen los perfiles genéticos habitualmente se denominan bancos o bases de datos. Aunque tradicionalmente ambos términos (bancos o bases) se utilizan como sinónimos, se está extendiendo cada vez la diferenciación de significado entre ellos (RUDIN AND INMAN 2002; MICHAELIS ET AL. 2008). Creemos que es muy importante ir adoptando esta terminología también en nuestro país. Vamos a definir a continuación que se entiende por cada uno de ellos en el ámbito de la forensia científica.

1. Banco de datos

Se define un banco de datos como un conjunto amplio de perfiles genéticos pertenecientes a personas individuales que han llevado a cabo alguna actividad delictiva. Es decir, después de la detención a que haya lugar y tras la obtención de la muestra y análisis adecuado del ADN para los marcadores genéticos comentados con anterioridad, se obtiene el perfil genético. Este perfil se ha incorporado a un fichero que es el banco de datos. De forma general los bancos de datos policiales se utilizan principalmente para investigar delitos e identificar sospechosos. Por lo tanto el banco de datos se nutrirá de delincuentes y/o de la población reclusa. También existen bancos de datos en los que se almacenan perfiles genéticos individuales de muestras biológicas recogidas en el lugar de los hechos (sangre, semen, saliva, etc.) y que no pertenecen a ninguna persona identificada o asociada a la actividad policial o jurídica. Muchas veces esta información puede ser importante para la resolución de un caso y se guarda en un fichero con la esperanza de que quizá algún día se podrá saber a quien pertenecen. Por ejemplo, si el sospechoso es un delincuente habitual y en un futuro se conoce su perfil genético al realizar un nuevo acto delictivo, se le podrá relacionar con el caso antiguo. Por último existe una tercera categoría de bancos de datos, los de interés humanitario, es decir, que contienen el perfil genético de personas desaparecidas o de familiares suyos, con la esperanza de poder ser identificadas algún día. También existen bancos de datos particulares para situaciones determinadas, como el elaborado por las Fuerzas Armadas de los Estados Unidos donde se almacenan los perfiles genéticos de todos

sus miembros, para poder identificar adecuadamente los cadáveres caso de fallecer y quedar el cuerpo maltrecho dificultando su identificación por los métodos habituales (REILLY 2001). Dicho banco se instituyó antes de la Primera Guerra del Golfo (1991) y se demostró especialmente útil para la identificación de los cadáveres en el ataque suicida del 11-S al Pentágono. En la actualidad es ya habitual generar un banco de datos con los perfiles genéticos de las víctimas y de sus familias después de un accidente aéreo o una gran catástrofe (ya sea un desastre natural o un ataque terrorista).

El primer banco de datos de interés policial fue desarrollado por el *Forensic Science Service* del Reino Unido en 1995 y recibe el nombre de *UK National DNA Database* (por cierto, es de destacar que cuando se le puso el nombre no existía la diferenciación entre banco y base de datos; en realidad en este caso se trata de un banco de datos y no de una base de datos). El perfil genético de los individuos se obtiene mediante el análisis de 10 STR, en concreto los denominados: FGA, TH01, VWA, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433 y D21S11. Además se analizan el gen Amilogenina, que tiene una variante para el sexo femenino y otra diferente para el masculino, permitiendo averiguar dicho carácter en una muestra biológica. Estos marcadores son también el núcleo de los utilizados a nivel de bancos de datos de diferentes países europeos y de la Interpol. Pero sin duda, el banco de datos más famoso es el CODIS (*Combined DNA Index System*) que fue desarrollado a partir de 1996 por el laboratorio del FBI con la ayuda de científicos externos. Se preseleccionaron 17 STR para finalmente escogerse 13 en 1997. Algunos de estos marcadores genéticos son comunes a los descritos anteriormente, mientras que otros son nuevos. En concreto el sistema CODIS está formado por los STR siguientes: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 y D21S11. También está incluido el gen Amilogenina para poder identificar el sexo de la muestra. Estos marcadores STR presentan unas características de transmisión genética, fiabilidad de análisis, estudio simultáneo además de automático y unos elevados niveles de discriminación que los hacen idóneos para los bancos de datos. Muchos países utilizan uno de estos dos sistemas de STR o combinaciones parecidas. Así, tanto la Policía Nacional como la Guardia Civil utilizan el sistema CODIS. Si se quiere información adicional respecto a estos marcadores genéticos puede consultarse a BUTLER (2005) o visitar la página web: <http://www.cstl.nist.gov/biotech/strbase/index.htm>.

2. Base de datos

Una base de datos no es más que la recopilación de las frecuencias (abundancias relativas expresadas como porcentajes) estimadas en una población para cada variante de un marcador genético concreto y para todos los sistemas genéticos

considerados. Es importante resaltar que no contienen informaciones individuales, sino las abundancias en términos de porcentajes de las diferentes variantes genéticas a nivel poblacional. Las bases de datos se utilizan para calcular las estimas sobre cuan probable es encontrar por azar un perfil concreto en una determinada población. Se entiende por población cualquier grupo humano que definamos, tal como un país, una región o una localidad. Se puede hacer una distinción por etnias concretas dentro de una de estas unidades geográficas, que en algunos casos puede ser muy importante. Una buena base de datos precisa la información genética de una muestra de un número elevado de individuos de la población. Se recomienda que el tamaño de la muestra (número de individuos analizados) sea como mínimo de 100. Además, dichos individuos no deben tener relaciones de parentesco entre ellos, sino algunas variantes genéticas estarían más representadas de lo que están en la población y otras poco representadas o incluso ausentes. Como se verá más adelante es fundamental en Genética forense contar con buenas bases de datos.

Por último recordar que gracias a la disponibilidad de ordenadores de gran potencia y velocidad de proceso (*hardware*) y al desarrollo de programas (*software*) especializados en la utilización de grandes ficheros de datos es posible en la actualidad manejar de forma muy eficiente tanto los bancos como las bases de datos. Sin las herramientas informáticas sería imposible procesar adecuadamente el volumen de información proporcionado por el ADN, como también es imposible hoy día procesar los otros tipos de criterios de identificación e información forense sin la informática. Por tanto, y a modo de resumen, debe quedar claro que un banco de datos es un fichero en el que se almacenan perfiles genéticos (entre otros datos) de personas individuales y concretas cuya utilidad es conocer si una muestra biológica pertenece a una determinada persona o no. En cambio, una base de datos resume la información genética global de toda una población, no conteniendo por tanto información individualizada. Su utilidad es cuantificar de forma estadística cual es la probabilidad de encontrar por azar un determinado perfil genético en dicha población, sea o no de interés policial o jurídico.

IV. SOLIDEZ CIENTÍFICO-TÉCNICA DE LOS PERFILES DE ADN

Una pregunta que nos puede venir fácilmente al pensamiento es ¿los análisis de los perfiles de ADN son fiables?, es decir, siempre que analicemos el ADN de una misma muestra, ¿obtendremos el mismo perfil genético? Es obvio que la respuesta a esta pregunta es crucial para la utilidad probatoria y por tanto jurídica de los perfiles de ADN. Durante los primeros años en que se empezó a utilizar el ADN con fines forenses no existía ninguna regulación y se aceptaba la validez de este tipo de pruebas. Muchos de los análisis de aquella época fueron realizados con evidentes limitaciones técnicas, sin procedimientos estandarizados a nivel de laboratorio y llevados a cabo por personal mal adiestrado. El primer caso donde se puso en

evidencia las carencias de los análisis de ADN fue el de *El Pueblo contra Castro* de 1987 (PATTON 1990; MICHAELIS ET AL. 2008). Tanto los testigos forenses de la acusación como de la defensa, admitieron que la calidad de las pruebas presentadas (es decir de los resultados de los análisis de ADN) era tan deficientes, que se consideraron inadmisibles. Este caso actuó de catalizador para la obtención de una buena estandarización y el FBI instituyó en 1988 el *Technical Working Group on DNA Analysis Methods* (abreviado TWGDAM). Desde entonces se han sucedido en los Estados Unidos diferentes agencias reguladoras del control de calidad de los laboratorios encargados de procesar las muestras de ADN. La introducción de los marcadores STR en genética forense condujo a que en varios juicios que tuvieron lugar en Estados Unidos (*El Pueblo contra Bokin* (California, 1999), *El Pueblo contra Shreck* (Colorado, 2000) y *El Estado contra Pfenning* (Vermont, 2000)) la defensa intentase invalidar las pruebas basadas en el ADN con el argumento de que los kits de laboratorio para dichos análisis (conjunto de reactivos que permiten analizar de forma precisa y simultánea varios STR a la vez) producidos por varias compañías bioquímicas no eran públicos ni estaban validados de forma suficiente (BUTLER 2005). Estos kit comerciales estaban protegidos por patentes y en principio las compañías eran reacias a exponer sus detalles técnicos en público. En vista de los problemas judiciales una de las empresas hizo públicas las características de sus protocolos técnicos en una revista científica de gran repercusión y otra compañía suministró los detalles bajo orden de protección judicial en cuantos casos hiciese falta. Desde el año 2000 existen multitud de trabajos científicos que avalan la fiabilidad de los kits comerciales de análisis y nadie pone ya en tela de juicio los resultados que se obtienen. Sin embargo surge otra cuestión, los análisis de ADN a partir de muestras biológicas tiene un coste y actualmente se debate en Estados Unidos, que sigue siendo el país de referencia también en este caso, quien debe cargar con los gastos (HOUCK AND SIEGEL 2006). A su vez se valora que tipo de análisis forense debe realizarse en función de la información que produce respecto a los costes que conlleva. En cualquier caso, lo que si está claro en la actualidad es que los análisis del ADN deben realizarse, tanto si son usados por la acusación como por la defensa, en laboratorios debidamente acreditados y que superen una serie de controles de calidad. Estas acreditaciones y controles de calidad ya hace años que se exigen también en España.

V. IDENTIFICACIÓN DE UN PERFIL GENÉTICO EN UN BANCO DE DATOS Y CÁLCULO DE LA PROBABILIDAD DE ENCONTRAR ESTE MISMO PERFIL POR AZAR EN LA POBLACIÓN: PROTOCOLO DE REFERENCIA

El protocolo a seguir es que el oficial especialista recoja las muestras biológicas (sangre, semen, saliva, etc.) que se consideren relevantes para el caso. Estas deben

ser cuidadosamente empaquetadas y rotuladas iniciándose así una cadena de custodia que culmina con la recepción de las muestras por parte del laboratorio especializado (que como ya hemos comentado debe estar convenientemente acreditado y homologado). En él se obtiene el perfil genético del individuo al cual pertenece la muestra biológica. Este perfil se compara con los perfiles que se guardan almacenados en los correspondientes bancos de datos policiales. Puede ser que dicho perfil no coincida con ninguno almacenado en los bancos de datos, con lo cual es almacenado en otro banco de datos (perfiles biológicos recogidos en lugares de los hechos que no coinciden con ningún perfil existente en un banco de datos policial). Como ya se ha comentado, es importante guardar estos perfiles pues algún día puede ser que si sean identificados. En el caso de que el perfil genético obtenido a partir de la muestra biológica coincida con uno idéntico en el banco de datos policial, se puede identificar al individuo en cuestión. Claro, podría ser que en la misma población del sospechoso hubiese otra persona con el mismo perfil genético. En esta situación debe calcularse cual es la probabilidad de obtener este mismo perfil (el del sospechoso) por azar en la población. Para realizar la estima de dicha probabilidad es necesario tener conocimientos de dos disciplinas científicas: la Genética de Poblaciones y la Estadística. Definidas de manera elemental, la primera estudia la variabilidad y distribución de los genes y sus variantes (los denominados alelos) en las poblaciones naturales y los mecanismos responsables de sus cambios en el tiempo. Por otra parte, la Estadística se dedica a estudiar los datos numéricos obtenidos de fenómenos naturales y a estimar sus probabilidades. En el caso que nos ocupa, una descripción detallada de cómo calcular la probabilidad de obtener por azar el mismo perfil que el de un sospechoso en una población puede obtenerse en BUTLER (2005), GOODWIN ET AL. (2007), FUNG AND HU (2008) Y LI (2008). Sin embargo nos gustaría mostrar de manera sencilla la forma en como se llevan a cabo las estimas de estas probabilidades. Vamos a suponer un caso simple, el del grupo sanguíneo MN del cual hemos hablado al inicio de este trabajo. Este gen presenta únicamente dos variantes (o alelos): M o N. Nosotros podemos conocer las abundancias relativas (frecuencias) de estos alelos M y N en cualquier población humana. Tan solo necesitamos muestras de sangre de los individuos y analizarlas bioquímicamente. Vamos a suponer que en una determinada población encontramos lo siguiente:

Grupo sanguíneo	Número de individuos
M	119
MN	76
N	13
Total	208

¿Cuántos alelos M y cuántos alelos N tendremos en esta población? Muy fácil, sabemos que todos nuestros genes están repetidos dos veces, es decir, somos diploides. Además sabemos que el modelo de herencia es sencillo, el grupo M esta formado únicamente por individuos que son genéticamente MM (homocigotos para M), análogamente los del grupo MN solamente pueden ser MN (heterocigotos) y por último los del grupo N sólo pueden ser NN (homocigotos para N). Por lo tanto, lo único que hemos de hacer es contar:

Grupo M: presenta dos alelos M, por tanto $119 \times 2 = 238$

Grupo MN: presenta 76 alelos M y otros 76 N

Grupo N: presenta dos alelos N, por tanto $13 \times 2 = 26$

Total de alelos M en la población estudiada: $238 + 76 = 314$

Total de alelos N en la población estudiada: $26 + 76 = 102$

Fijémonos que el total de alelos que hay en la población para el grupo sanguíneo MN es $314 + 102 = 416$. Este mismo valor se obtiene si multiplicamos el número total de individuos analizados por dos ($208 \times 2 = 416$), puesto que cada individuo es portador de dos alelos de este grupo sanguíneo (es diploide).

La abundancia relativa de cada alelo (lo que técnicamente se denomina frecuencia relativa) se obtiene dividiendo el número alelos de un tipo por el total de alelos de la población. Es decir:

$$\text{fr (M)} = 314 / 416 = 0,754$$

$$\text{fr (N)} = 102 / 416 = 0,246$$

Evidentemente, la suma de fr (M) y de fr (N) da 1, puesto que solamente hay dos frecuencias ($0,754 + 0,246 = 1$). En Genética de Poblaciones, es habitual representar la frecuencia de un alelo por la letra p y la del otro alelo por la letra q , de forma que en nuestro caso a la fr (M) la llamaríamos p y a la fr (N) la denominaríamos q .

Muchas poblaciones humanas presentan valores diferentes para estas frecuencias, por ejemplo:

Población	fr (M)	fr (N)
Barcelona	0,49	0,51
Blancos U.S.A.	0,54	0,46
Aborígenes australianos	0,25	0,75
Indios Puebla	0,76	0,24

Estas frecuencias pueden servir para tener una base de datos respecto al grupo sanguíneo MN en diferentes poblaciones humanas. Queremos mostrar un modelo sencillo de cómo procederíamos, recordando que estamos ante una situación muy simple, pero que puede ser muy ilustrativa del procedimiento lógico a seguir. Vamos a suponer que intentamos esclarecer un caso acaecido en Australia entre los aborígenes. Se ha producido una reyerta y el agredido ha sido capaz de clavar un objeto punzante en el agresor y por tanto tenemos una muestra de sangre de dicho agresor. La analizamos y vemos que pertenece al grupo sanguíneo M. ¿Qué probabilidad hay de que otro individuo de la misma población presente también dicho grupo sanguíneo?

Evidentemente usaríamos la base de datos de las frecuencias para el grupo sanguíneo MN provenientes de la población de aborígenes australianos. ¿Cuál es la probabilidad en esta población de encontrar individuos MM o MN o NN? Si suponemos que los genes se combinan al azar la abundancia de cada alelo viene reflejada por su frecuencia. De manera que la probabilidad de obtener un individuo MM (probabilidad que llamaremos P (MM)) dependerá de coger un alelo M de la población (p) y otro alelo M de la población (p). Como tomar dichos alelos de la población son sucesos independientes, entonces la probabilidad de tener un individuo MM en esta población será el producto de p por p , es decir:

$$P(\text{MM}) = p \times p = p^2$$

De forma análoga, la probabilidad de obtener un individuo del grupo N en la población de aborígenes australianos será:

$$P(\text{NN}) = q \times q = q^2$$

En cambio, para tener un individuo del grupo MN o bien tomamos primero un alelo M (con probabilidad p) y luego uno N (con probabilidad q) o al revés, cogemos primero uno N (con probabilidad q) y después uno M (con probabilidad p). En este caso se trata de una combinación doble, así el cálculo resultante es:

$$P(\text{MN}) = 2 \times p \times q = 2pq$$

El caso de nuestro ejemplo, el encontrar un individuo con el grupo sanguíneo M en la población de aborígenes sería:

$$P(\text{MM}) = p \times p = p^2, \text{ es decir, } 0,25^2 = 0,0625.$$

Si la población está formada por 10000 aborígenes, entonces posiblemente esperaríamos encontrar entre ellos unos 625 ($0,0625 \times 10000$) individuos con el grupo sanguíneo MM. Por lo tanto habría un número muy elevado de sospechosos.

Este ejemplo era tan sólo para ilustrar el método, pero ya hemos comentado que los grupos sanguíneos ya no se usan y además un único marcador es muy poco

informativo. Como ya se ha expuesto con anterioridad, la mayoría de bancos y bases de datos que se utilizan en la actualidad están basados en los marcadores microsatélites (también llamados STR). Cada una de estas secuencias de ADN no tiene dos variantes, sino muchas. Además no se utiliza un único microsatélite, sino que se utilizan muchos (el sistema CODIS del FBI de basa en 13 microsatélites). Vamos a presentar otro ejemplo numérico para analizar las diferencias. Supongamos un caso de violación en una ciudad del sur de California en que la víctima identifica al agresor como “hispano”. Se recogen muestras de semen y por los procedimientos que ya se han descrito obtenemos el siguiente perfil para todos los microsatélites del sistema CODIS. Hay que recordar que los cromosomas humanos están por duplicado, por tanto habrá dos alelos posibles para cada microsatélite. El perfil genético obtenido a partir de la muestra del semen recogido es:

Microsatélite	Primer alelo	Segundo alelo
CSF1PO	11	11
FGA	20	21
TH01	7	9
TPOX	12	12
VWA	15	16
D3S1358	17	17
D5S818	11	12
D7S820	10	13
D8S1179	10	16
D13S317	8	12
D16S539	10	13
D18S51	12	20
D21S11	30	30

Para calcular cuál sería la probabilidad de encontrar al azar en la población este mismo perfil genético, en primer lugar se escogería la base de datos adecuada. En este caso, por lo que sabemos del sospechoso, usaríamos la de “Hispanos de Estados Unidos”, que se puede consultar en el Apéndice II del libro BUTLER

(2005). ¿Cómo se realizaría el cálculo? Siempre que tengamos un único alelo (variante) para un microsatélite concreto (el individuo es homocigoto), tomaríamos su frecuencia (p) de la base de datos y la elevaríamos al cuadrado (p^2). Por ejemplo, para el primer marcador (CSF1PO) el sospechoso es homocigoto 11, 11 y por tanto cogeríamos la frecuencia de dicho alelo de la base de datos (0,29286) y la elevaríamos al cuadrado (0,08577). En cambio, cuando el sospechoso tenga dos alelos diferentes para un mismo marcador (heterocigoto) debe multiplicarse: dos por la frecuencia del primer alelo (p) y por la frecuencia del segundo alelo (q), es decir $2pq$. Por ejemplo, para el segundo marcador (FGA) el sospechoso es heterocigoto 20, 21 luego tomaríamos la frecuencia del alelo 20 (0,08929) y del alelo 21 (0,16786) de la base de datos y realizaríamos la operación $2 \times 0,08929 \times 0,16786$ obteniendo el valor 0,02998. Se iría procediendo de la misma forma con todos los marcadores utilizados para encontrar el perfil genético. Una vez se tiene el valor para cada marcador, como todos ellos son independientes entre si, el valor final de la probabilidad de encontrar el mismo perfil genético por azar en la población no es más que el producto de todos ellos. En el ejemplo el valor final sería $2,81988 \times 10^{-19}$, es decir una probabilidad extremadamente baja. Es un valor que cuesta de entender, pero es casi 0, en concreto 0,000000000000000000281988, luego la probabilidad de encontrar dicho perfil por azar en la población es casi nula. Existen diferentes formas de presentar dicho valor en ambientes legales y judiciales para hacer más comprensible su significado y no incurrir en errores (GOODWIN ET AL. 2007).

Hay que hacer una serie de consideraciones respecto a los cálculos que se han mostrado. En primer lugar no siempre puede conocerse el grupo racial del sospechoso. En estas situaciones se utilizan todas las bases de datos disponibles y se toma como valor de referencia el que produzca un resultado más elevado. Generalmente, este resultado será también un valor extremadamente cercano a cero, así si se utiliza la base de datos para caucásicos (blancos) de los Estados Unidos (BUTLER 2005) el valor de la probabilidad es de $7,6274 \times 10^{-21}$ y si se usa la base de datos para los afro-americanos de dicho país el valor que se obtiene es de $7,4545 \times 10^{-24}$. Ambos valores son menores (más cercanos a cero) que en caso de utilizar la base de datos de los hispanos. Si en la población conocemos aproximadamente el porcentaje de cada grupo racial, se puede realizar el cálculo mediante otra aproximación (MICHAELIS ET AL. 2008). Supongamos que en nuestra población se ha estimado que existe un 60% de caucásicos, un 30% de hispanos y tan sólo un 10% de afro-americanos. Podríamos tener en cuenta estas proporciones y ajustar nuestro cálculo de la probabilidad de encontrar por azar el mismo perfil genético en la población del siguiente modo:

$$P = 0,6 \times 7,6274 \times 10^{-21} + 0,3 \times 2,81988 \times 10^{-19} + 0,1 \times 7,4545 \times 10^{-24} = 0,8917359 \times 10^{-19}$$

valor que sigue siendo más bajo que cuando se utiliza únicamente la base de datos de los hispanos. Es muy importante escoger bien la base de datos para realizar los cálculos que se han presentado, pero a veces es complicado (CHO AND SANKAR 2004).

Otro detalle importante es que en el ejemplo se ha supuesto que los genes se combinan totalmente al azar y a veces existen ciertas poblaciones que hacen que dicha premisa no sea exactamente cierta. En estos casos se utilizan una serie de correcciones numéricas basadas en modelos de la Genética de Poblaciones que no se comentarán aquí, pero que están descritas de forma sencilla en GOODWIN ET AL. (2007).

También debe recordarse que los perfiles genéticos no son útiles para el caso de los hermanos gemelos idénticos (monocigóticos). Dichos hermanos gemelos son genéticamente iguales puesto que provienen de una única fecundación. Aunque sean personas distintas tendrán el mismo perfil para los marcadores genéticos utilizados. Existe un ejemplo reciente en el que la policía berlinesa no puede saber cual de los dos hermanos gemelos monocigóticos es el responsable de un robo, puesto que los marcadores genéticos para ambos son idénticos (POCH 2009). En este tipo de hermanos sí puede reconocerse de manera individual a cada uno de ellos mediante las huellas dactilares, que son diferentes para los dos gemelos monocigóticos. Otra situación especial es el caso de las personas que son quimeras, es decir, que tienen en su cuerpo células con un perfil genético y células con otro distinto. Un caso frecuente es el de los trasplantes, donde las células del órgano recibido tendrán el perfil genético del donante, no el del receptor. También debe tenerse en cuenta las transfusiones de sangre, donde los glóbulos blancos recibidos de un donador, tendrán un perfil genético diferente al del receptor. También pueden aparecer quimeras como resultado de procesos naturales poco frecuentes durante el desarrollo embrionario (YU ET AL. 2002).

VI. ARGUMENTACIONES CIENTÍFICO TÉCNICAS SOBRE LA UTILIZACIÓN DE LA INFORMACIÓN FORENSE BASADA EN EL ADN: FALACIAS DE LA ACUSACIÓN Y FALACIAS DE LA DEFENSA

Es muy importante realizar una valoración correcta de la información proporcionada por las pruebas de ADN. En general, la opinión pública acepta las evidencias generadas por el análisis de ADN como algo definitivo, seguramente debido a la influencia de los medios de comunicación, las series televisivas y las producciones cinematográficas. Debe recordarse siempre que tipo de información y por tanto que tipo de preguntas pueden tener respuesta y cuales no, tras la obtención de un perfil genético basado en el ADN. El resultado ideal tendría que permitirnos determinar a quien pertenece y con que probabilidad una determinada muestra

biológica (sangre, saliva, semen, etc.) encontrada en el lugar de los hechos. En cambio, los perfiles de ADN, no aportarán ninguna información en lo referente a por qué estaba dicha muestra biológica en aquel lugar, ni tampoco si el sospechoso llevó a cabo la actividad delictiva. Estas dos cuestiones deben responderlas otros peritos expertos en otras ciencias forenses.

Desde el punto de vista científico técnico ya se ha comentado que la recolección de las muestras es un trabajo delicado y deben realizarlo especialistas. Estas deben etiquetarse de manera correcta para evitar errores y deben seguirse los protocolos de custodia hasta la recepción de la muestra por el laboratorio que deberá analizarlas. En la actualidad los análisis para obtener los perfiles genéticos a partir de las muestras de ADN son muy precisos y reproducibles. Ya se ha comentado que existen toda una serie de protocolos técnicos que minimizan los posibles (aunque improbables) errores. Los laboratorios acreditados y homologados siguen toda una serie de pautas que hacen extremadamente fiable el resultado. Si el perfil genético obtenido ha de contrastarse con los perfiles almacenados en los bancos de datos policiales, ya se ha visto que con el número y tipo de marcadores genéticos utilizados (los STR) la probabilidad de encontrar ese mismo perfil por azar en la población es extremadamente baja. De todas formas se debe ser prudente y una vez obtenida la identificación del individuo en cuestión hay que analizar de nuevo el caso a la luz de la información facilitada por el ADN. A este respecto debe recordarse el error que se produjo en el Reino Unido en 1999, cuando se utilizaban solamente 6 marcadores STR (GOODWIN ET AL. 2007). Se pensaba que era una cantidad suficiente de marcadores, puesto que las estimas de la probabilidad de encontrar por azar el mismo perfil en la población rondaban alrededor de 10^{-8} . En dicho año se produjo un robo, se obtuvo el perfil genético de una muestra biológica, se realizó la búsqueda en el banco de datos policial y se identificó al individuo como el Sr. Raymond Easton. El robo se produjo en Bolton, a 200 Kms. de la residencia del señor Easton, que padecía Parkinson, era incapaz de vestirse o ir al baño por sí solo y no podía desplazarse más de 10 metros sin ayuda. A pesar de ello fue arrestado por la Policía de Swindon y estuvo desde las nueve de la mañana hasta las cuatro de la tarde en las dependencias policiales. Su abogado solicitó pruebas de ADN más detalladas (utilizando más marcadores STR), se demostró su inocencia y fue indemnizado con 2000 libras esterlinas por el error.

Sin duda alguna una fuente importante de error proviene de la interpretación de la probabilidad de encontrar por azar un perfil genético en la población (BALDING AND DONNELLY, 1994; GOODWIN ET AL. 2007; MICHAELIS et al. 2008). Posiblemente muchos de los especialistas en temas legales no lo son ni en Genética de Poblaciones ni en Estadística. Dichos errores pueden agruparse en dos categorías, las denominadas falacias de la acusación y falacias de la defensa.

1. Falacia de la acusación: interpretación errónea de la información obtenida a través del ADN

Cuando se detecta una coincidencia entre el perfil genético del lugar de los hechos y el perfil genético de un sospechoso (por ejemplo detectado en la búsqueda de un banco de datos policial), el cálculo de la probabilidad de encontrar por azar dicho perfil en la población responde a la pregunta siguiente: ¿cuál es la probabilidad que el perfil genético de cualquier otra persona que se tomase al azar de una población de referencia coincidiese con el perfil genético obtenido en la muestra biológica recogida en el lugar de los hechos? Dicha pregunta asume que el defendido es inocente y que la coincidencia entre su perfil genético y el obtenido en el lugar de los hechos es pura casualidad. La falacia de la acusación consiste en interpretar dicha probabilidad como si representase la probabilidad de que el defendido fuese realmente inocente a pesar de la coincidencia de los perfiles genéticos. Entonces se asume que la coincidencia de perfiles de ADN es una prueba de culpabilidad y que el valor de la probabilidad que se calcula (habitualmente extremadamente pequeña) representa la probabilidad de que el defendido pueda no ser culpable. Por tanto dicha interpretación es una falacia y además es claramente perjudicial para el defendido. La falacia consiste en una mala interpretación de la probabilidad por parte de la fiscalía. Explicado mediante un ejemplo sencillo, si un perro se describe como un animal de cuatro patas no podemos hacer la transposición de que todo animal con cuatro patas es un perro. Fijémonos en estas dos frases: “la probabilidad de obtener este perfil genético si proviniese de cualquier otra persona de la población que no fuese el sospechoso sería de 1 en 10 millones” (correcto) no es equivalente a decir que “la probabilidad de que las pruebas provenga de cualquier otra persona de que no sea el sospechoso es de 1 en 10 millones” (incorrecto). Finalmente se debe recordar que la coincidencia entre los perfiles genéticos es una evidencia más que permite establecer la culpabilidad del sospechoso. Así por ejemplo, si el defendido puede probar que en el momento en que sucedieron los hechos estaba en otro lugar, él es inocente del delito que se le imputa. Entonces existe otra explicación para la coincidencia entre su perfil genético y el obtenido en el lugar de los hechos, a pesar de que la probabilidad de encontrar dicho perfil por azar fuese extremadamente baja.

2. Falacia de la defensa: uso inadecuado de la información proporcionada por el ADN

Consiste en presentar que en la población existen varios individuos con el mismo perfil que el del defendido. De hecho se basa también en una mala interpretación del concepto de la probabilidad de encontrar un perfil genético por azar en la población. Por ejemplo, si en un caso concreto se ha estimado dicha probabilidad como 1 en 8 millones, la defensa puede realizar (erróneamente) el

siguiente cálculo: “si la población española se estima en 40 millones de habitantes, entonces habrá en España 5 personas con el mismo perfil genético que el obtenido en la muestra biológica recogida en el lugar de los hechos, luego la probabilidad de que pertenezca al defendido es tan sólo de 1/5, la cual genera una duda razonable respecto a su culpabilidad”. Esta línea de razonamiento es errónea puesto que nada se conoce de estas supuestas cinco personas, incluso no se sabe si realmente existen. Caso de existir, no se conoce ni donde viven, ni su edad, ni su género ni mucho menos, que relación tienen con los hechos que se investigan. Como en el caso anterior, la falacia de la defensa proviene de una mala interpretación del concepto de la probabilidad de encontrar un perfil genético por azar en la población y ello lleva al error de interpretación. Que exista una cierta probabilidad de encontrar un cierto perfil genético en la población no implica que dicho perfil exista.

A pesar de todos los problemas y dificultades que se han presentado, lo que es innegable es que la búsqueda de perfiles genéticos en bancos de datos ha sido y es de suma importancia para la resolución de muchos casos. Incluso se puede afirmar que alguno de ellos no se hubiese podido solucionar si no hubiese sido por estos procedimientos. Existen muchos e interesantes ejemplos que pueden encontrarse en BUTLER (2005), HOUCK AND SIEGEL (2006) y GOODWIN ET AL. (2007). Alguno de estos casos pudieron resolverse muchos años después gracias a los bancos de datos. Así por ejemplo, en 2001 pudo demostrarse la culpabilidad de Brian Field del asesinato del niño Roy Tutill acaecido en 1969. Estos casos antiguos, que han permanecido durante años sin resolver, se denominan ‘cold cases’ en inglés, que pueden traducirse como ‘casos congelados’ o también “casos abiertos”. Finalmente debe recordarse que los perfiles genéticos pueden servir para inculpar a sospechosos, pero no es menos cierto que también sirven para exculpar a presuntos causantes de un delito. En el histórico caso Pitchcock, el primero en que se usaron procedimientos de la Genética forense moderna, se pudo demostrar la inocencia del primer sospechoso del caso al no coincidir su perfil genético con el hallado en el lugar de los hechos.

VII. CONDICIONES PARA LA INCLUSIÓN DE UN DETERMINADO PERFIL GENÉTICO EN UN BANCO DE DATOS

Existe un gran debate a todos los niveles (nacional e internacional) sobre que perfiles genéticos deben incluirse o no en un banco de datos (para ver ejemplos concretos pueden consultarse las referencias: RELLY 2001; CHO AND SANKAR 2004; JOBLING AND GILL 2004; GAMERO ET AL. 2007; MICHAELIS ET AL. 2008). De entrada se ha de diferenciar lo que son bancos de datos de material biológicos de lo que son bancos de datos de fichas (físicas o informáticas) con los perfiles genéticos. Esta diferenciación es muy importante por las implicaciones que puede tener. El mantener el material biológico implica que es posible el analizar de

nuevo del genoma (contenido genético total de un individuo). Es importante pues puede permitir verificar si un perfil es correcto (confirmación), o analizar más marcadores o estudiar marcadores nuevos. Por ejemplo, los marcadores utilizados a lo largo de la historia de la Genética forense han ido cambiando a lo largo de los años buscando la simplificación, seguridad de los resultados y eficiencia (incluyendo el coste) del estudio (JOBILING AND GILL 2004). Mantener las muestras biológicas permite obtener el perfil genético del mismo individuo respecto a nuevos marcadores genéticos. Pero ello presenta un dilema ético importante: al poder estudiar todo el genoma, las muestras biológicas podrían utilizarse para otras finalidades (por ejemplo, conocer la posibilidad de ser susceptible a ciertas enfermedades, con las implicaciones laborales y de seguros que presentan). En diversos países se almacena este tipo de muestras biológicas mediante las fichas FTA (BUTLER 2005; GOODWIN ET AL. 2007). Estas son unas tarjetas con unos productos químicos (estabilizantes del ADN y que evitan su degradación) que permiten almacenar el ADN a partir de células sanguíneas o de las epiteliales de la boca. El ADN se une a la matriz de papel de la tarjeta y puede mantenerse así durante años (hasta ahora se sabe que más de 17 años). Si se quiere analizar genéticamente la muestra tan solo debe obtenerse una pequeña muestra con una perforadora de papel (de las de uso en oficinas) y llevar a cabo los protocolos clásicos de extracción y análisis de ADN. Por lo tanto pueden obtenerse mucha información de tipo genético/biológico/médico de un individuo.

El otro tipo de banco de datos no es más que una ficha con la información escrita de las variantes (alelos) de los marcadores genéticos analizados. Ya se ha mostrado la utilidad de esta información para resolver diversas cuestiones de interés legal. También existen toda una serie de cuestiones de tipo ético/legal respecto a las condiciones de almacenamiento de dicha información y a su uso. Por ejemplo, ¿qué personas deben incluirse en dichos bancos de datos?, ¿es lícito solicitar a una persona que done su ADN para un banco de datos?, ¿si de joven alguien cometió algún acto delictivo menor, debe permanecer su perfil genético durante toda su vida en el banco de datos?, ¿estos archivos deben borrarse transcurridos unos años?, ¿deben borrarse los registros de las personas una vez fallecidas?, ¿quién tiene acceso a dichos bancos, en qué condiciones y cómo deben utilizarse adecuadamente? Estas son algunas de las muchas preguntas que se pueden plantear al respecto. En cuanto a posibles respuestas, también existe un amplio espectro de opiniones. Hay quien propone que los bancos de datos sean universales, es decir, que se incluya a toda la población (GOODWIN ET AL. 2005; Gamero et al. 2007). En general los miembros de las fuerzas policiales están a favor, como también lo están algunos políticos (es interesante ver la fotografía del Sr. Tony Blair donando sus células epiteliales en 1999, fotografía que puede verse en REILLY 2001) y científicos especialistas en Genética forense. Ante las reticencias expresadas por muchos sectores sociales su respuesta muchas veces es del tipo: “si alguien no

realiza ningún acto delictivo no tiene por que temer nada por tener su perfil genético almacenado en un banco de datos”. Dichas reticencias tienen su razón de ser pues muchas veces en el curso de una investigación se ha pedido “voluntariamente” que diversas personas den su consentimiento para que se analice su ADN. No queda clara la voluntariedad y además no ser voluntario se ha interpretado como una cierta culpabilidad. Esto sucedió por ejemplo en el famoso caso Pitchfork. Además, la legislación que regule dichos bancos de datos ha de dejar muy claro quien tiene autorización para acceder a ellos y con que finalidad. La otra posibilidad es incluir únicamente a los autores de ciertos delitos, cuya gravedad es también motivo de debate. En algunos países la detención por cualquier motivo implica la inclusión del perfil en un banco de datos de ADN. Si posteriormente se demuestra la inocencia del detenido, en ciertos países conlleva el borrar su registro del banco de datos, pero en otros no. En unas naciones las fichas personales se borran al cabo de unos años después de la muerte del individuo, mientras que en otras nunca se eliminan. De todo ello se deduce que es un tema complejo y cada país (o administraciones locales como los Estados de los Estados Unidos) tiene una normativa particular. Es fundamental tener una legislación precisa sobre el tema.

Una visión general de los bancos de datos internacionales puede encontrarse en JOBLING AND GILL (2004), sin embargo nos gustaría comentar las características de algunos de ellos según su país. Uno de los bancos de datos pioneros fue el del Reino Unido, fundado en 1995 y denominado UK NDNAD (*The United Kingdom National DNA Database*). Como se ha explicado a lo largo de este trabajo, dicho banco de datos ha ido evolucionando incrementando el número de marcadores genéticos utilizados (GOODWIN ET AL. 2007). En el año 2008 contenía alrededor de 4 millones y medio de perfiles genéticos y el Primer Ministro Gordon Brown está a favor de incrementar el banco de datos. Sin embargo existe una gran preocupación ética por la posible discriminación de ciertos colectivos. Así por ejemplo, el 75% de los varones jóvenes negros entre 15 y 34 años están en el banco de datos, mientras que únicamente forman parte de él el 22% de los varones blancos en el mismo rango de edades. En los Estados Unidos, como ya se ha comentado, el FBI elaboró el sistema CODIS (*Combined DNA Index System*), que desde hace unos años está integrado informáticamente a nivel nacional, estatal y local, con lo que los diferentes cuerpos policiales pueden realizar búsquedas extensas. En el año 2006 incluía un número total de 3676971 perfiles genéticos, de los que 3528903 pertenecían a la población reclusa. La relación número de perfiles genéticos por habitante más elevada se encuentra en Nueva Zelanda, cuyo banco de datos nacional (NDD o *National DNA Database*) data de 1996. El banco de datos nacional de Australia (NCIDD o *National Criminal Investigation DNA Database*) se estableció en 1997 y ha permitido resolver varios casos difíciles, incluyendo algún ‘caso congelado’ o ‘caso abierto’. Una base de datos modificada del NCIDD sirvió para la identificación de 43 víctimas del atentado de la isla de Bali (2002).

También en 1997 se instauró el banco de datos de Holanda, que es operado por el NFI (*Nederlans Forensisch Intittut*, Instituto Forense Holandés) y utiliza el sistema CODIS. En el mismo año se inició el proyecto piloto del banco de datos nacional austriaco (uno de los primeros de Europa), en el se incluye a los sospechosos de delitos graves (incluyendo el robo) y se tiene especial cuidado en la confidencialidad (la identidad se salvaguarda con códigos de barras). Un año después se estableció el banco de datos de nacional de Alemania. En Bélgica el banco de datos está centralizado y sólo se incluyen en él a personas condenadas por delitos graves (asesinatos, violaciones, etc.). El banco de datos no contiene nombres, sino identificadores y únicamente el juez puede acceder a sus nombres. Los datos se borran a los diez años de la muerte del individuo. En 1998 se creó en Francia el *Fichier National Automatisé des Empreintes Génétiques* (Fichero Nacional Automatizado de las Huellas Genéticas) y está definido por ley su uso para la investigación de delitos graves. En España las primeras ideas sobre la creación de bancos de datos datan de 1994. Cada cuerpo policial elaboró sus propios bancos de datos, así la Guardia Civil tiene el ADNIC (ADN de Interés Criminal) y el FENIX (de interés humanitario) mientras que la Policía Nacional cuenta con el ADN-VERITAS (de interés criminal) y el ADN-HUMANITAS (de interés humanitario). Todos ellos utilizan los marcadores genéticos del sistema CODIS. Es de destacar que todos estos bancos de datos se interconectaron a finales de 2007, incluyendo el de los Mossos d'Esquadra. También se espera conectar todos los bancos de datos de las otras policías autonómicas para hacer más eficiente la labor policial.

A nivel internacional la Interpol ha realizado una combinación de los marcadores STR más comúnmente utilizados (un total de 10) para obtener el ISSOL (*Interpol Standard Set Of Loci*, que se puede traducir como Paquete de Marcadores Estandar de la Interpol) (Schuller et al. 2001). En un mundo donde existe el crimen organizado y el terrorismo internacional, la interconexión entre bancos de datos puede ser clave para resolver casos delictivos graves de alcance global. Para buscar una estandarización de los perfiles genéticos del ADN se creó en 1988 el EDNAP (*European DNA Profiling Group* o Grupo Europeo de Perfiles de ADN). Este colectivo está formado por científicos forenses de diferentes países europeos que buscan una armonización en las tecnologías del ADN para la investigación de los delitos. La idea es tener una interconexión de los bancos de ADN de los diferentes países, sin embargo el principal obstáculo para conseguirlo son las reticencias de tipo político.

VIII. SITUACIONES EN LAS QUE LOS BANCOS DE DATOS DE PERFILES GENÉTICOS SON ÚTILES Y EN LAS QUE NO LO SON

Indudablemente la utilización de los perfiles genéticos ha sido una herramienta nueva y útil en forensia. Sin embargo, las búsquedas en bancos de datos pueden ser

útiles en algunas situaciones y no serlo en otras. En el presente trabajo ya hemos presentado muchos ejemplos de su utilización para intentar resolver diferentes casos delictivos: asesinatos, violaciones, robos, etc. Identificar el perfil genético de una muestra biológica dejada en el lugar de los hechos (sangre, semen, saliva, etc.) en un banco de datos puede facilitar la identidad de la persona. También se ha comentado como proceder después de realizar la identificación y los errores potenciales que pueden comentarse si la información no se maneja correctamente. Otra situación donde son útiles los bancos de datos conteniendo perfiles genéticos es en la identificación de personas desaparecidas. Estos bancos de datos no son los mismos que los usados para fines de investigación criminal, sino que se componen de los perfiles genéticos de familiares de las personas desaparecidas o incluso del perfil de la propia persona. Este puede obtenerse, una vez ya no está la persona, por ejemplo recuperando su ADN a partir de células epiteliales de la boca que han quedado en su cepillo de dientes o de las células de la base del pelo con los cabellos que han quedado en su peine. Estos bancos de datos se denominan de interés humanitario y previamente ya se han citado algunos de ellos. Otra utilidad de los bancos de datos con perfiles genéticos es el de la identificación de las víctimas producidas como consecuencia de grandes catástrofes, ya sean naturales (terremotos o tsunamis, por ejemplo) o producidas por el hombre (accidentes aéreos, derrumbamientos o atentados terroristas, por citar algunos). Para ello se genera un banco de datos específico, que como en el caso anterior contendrá los perfiles de familiares o bien de las propias víctimas. Estos datos se cotejarán informáticamente con todos los perfiles procedentes de los restos. De hecho la tarea es doble: en primer lugar asignar todos los restos pertenecientes a una misma persona (si los cuerpos han quedado fragmentados) y después identificarla correctamente. Esta labor es de suma importancia para que las familias tengan identificados los correspondientes restos mortales para realizar un sepelio de sus seres queridos y además para temas legales como herencias o seguros. A lo largo de los años, la Genética forense ha realizado grandes avances en este campo al lograr obtener ADN a partir de restos muy degradados y además ser capaz de conocer su perfil genético. Algunas catástrofes en las que se han aplicado nuevas técnicas de Genética forense o se han desarrollado nuevos protocolos de actuación han sido, por ejemplo, el asalto del rancho Waco (Texas) (1993), los sucesivos accidentes aéreos de los vuelos TWA 800 (1996), Tupolev TU-154 (1996) y Swissair 111 (1998), los atentados del 11-S en Nueva York (2001) y del 11-M en Madrid (2004) y el tsunami del océano Índico (2004) (Butler et al. 2005; Alonso et al. 2005; Graham 2006; Goodwin et al. 2007). Desde el punto de vista de la elaboración de ficheros de ordenador con los bancos de datos de perfiles genéticos y de la generación de programas informáticos para la identificación rápida y precisa de los restos humanos, los atentados del 11-S (2001) son una referencia obligada (Butler 2005; Alonso et al. 2005).

En cambio los bancos de datos con perfiles genéticos no se utilizan en pruebas de paternidad, o en general, de parentesco. En estos casos tan sólo se precisa conocer las variantes de los marcadores genéticos de las personas implicadas y sobre dichos datos se realiza el estudio. Tampoco es especialmente interesante para el tema de la extranjería. Se podría elaborar un banco de datos con las personas entradas ilegalmente en un país. Si son expulsadas, y son detenidas de nuevo al intentar entrar otra vez en el país, con dicho banco de datos podría saberse. Pero dicha información no parece muy importante respecto al coste económico que implicaría el análisis de los marcadores genéticos y el tener un banco de datos específico. Se ha especulado mucho respecto a si a partir de la información genética se puede conocer el origen geográfico de una persona. Mediante el estudio de marcadores genéticos adecuados, en la actualidad tan sólo podemos tener una idea aproximada de a partir de que gran región geográfica procede una persona. Además, como todo estudio de este tipo, los resultados estarían sometidos al efecto de los errores estadísticos, que en este caso pueden ser bastante grandes. Por ejemplo, en el caso de los atentados del 11-S se pudo saber al analizar algunas regiones del mtDNA (ADN mitocondrial) que los restos de los presuntos terroristas que provenían del Oriente Medio, pues presentaban unas secuencias nucleotídicas abundantes en dicha área geográfica.

Finalmente, debe decirse con toda claridad que los perfiles genéticos no permiten determinar la edad de las personas ni se vislumbra que a través de los perfiles genéticos pueda obtenerse este tipo de dato en el futuro. Esta situación no excluye que los expertos en antropología y anatomía forense puedan valorar de modo presuntivo en la mayoría de los casos la edad de las personas.

IX. CONCLUSIONES

Es indudable que la genética molecular constituye una buena herramienta para la resolución de muchos temas forenses: identificación de muestras biológicas encontradas en el lugar de los hechos, de cadáveres después de una catástrofe, de personas desaparecidas, etc. El procedimiento analítico del ADN es fiable si se lleva a cabo correctamente, por tanto ha de realizarse en laboratorios debidamente acreditados. También es importante regular la forma en que debe almacenarse toda esta información basada en el ADN.

Existen dos tipos de ficheros con información genética de utilidad forense, los bancos y las bases de datos. Aunque originariamente se utilizaban de manera indistinta como si fuesen términos sinónimos, en la actualidad se usan con significados diferentes. Así se define un banco de datos policial, como un conjunto amplio de perfiles genéticos pertenecientes a personas individuales que han llevado a cabo alguna actividad delictiva. Se utiliza para almacenar perfiles genéticos de

delincuentes, de muestras biológicas recogidas en el lugar de los hechos o incluso para identificar a personas desaparecidas. En cambio, una base de datos resume la información genética (y otros tipos de datos) global de toda una población y su utilidad es cuantificar de forma estadística cual es la probabilidad de encontrar por azar un determinado perfil genético en dicha población. Es fundamental regular correctamente qué perfiles genéticos deben incluirse en los bancos de datos de interés forense y en que condiciones lo hacen. Deferentes países occidentales usan distintos criterios.

La información generada por el ADN debe interpretarse correctamente. Pueden producirse errores interpretativos a nivel de la fiscalía (falacia de la acusación) o de la defensa (falacia de la defensa). En este trabajo hemos presentado en qué consisten y como evitarlas.

Los bancos de datos son útiles en la resolución de ciertos casos delictivos y en la identificación de víctimas de grandes catástrofes o de personas desaparecidas. En cambio no parecen de utilidad para problemas relacionados con la extranjería. Los perfiles genéticos actuales no permiten conocer la procedencia concreta (país o región de origen) de una persona.

Sin lugar a dudas podemos decir hoy día que la genética forense ha estimulado la imaginación humana para resolver toda una serie de problemas policiales y jurídicos que sólo hace unas décadas era impensable su resolución. Es indudable que la genética forense ha puesto a prueba el ingenio y la imaginación humana para resolver situaciones concretas (Mestres y Vives-Rego, 2009b). Debemos por tanto congratularnos que la ciencia forense ha sido y es un ejemplo y un estímulo para el desarrollo tecnocientífico y que por tanto el aunar jurisprudencia y ciencia redunde en una sociedad más ética, próspera y estable.

X. BIBLIOGRAFÍA

- ALONSO, A., MARTÍN, P., ALBARRÁN, C., GARCÍA, P., FERNÁNDEZ de SIMÓN, L., ITURRALDE, M. J., FERNÁNDEZ-RODRÍGUEZ, A., ATIENZA, I., CAPILLA, J., GARCÍA-HIRSCHFELD, J., MARTÍNEZ, P., VALLEJO, G., GARCÍA, O., GARCÍA, E., REAL, P., ÁLVAREZ, D., LEÓN, A. and SANCHO, M. (2005). "Challenges of DNA Profiling in Mass Disaster Investigations". *Croat. Med. J.*, vol. 46, págs.540-548.
- BALDING, D. J. and DONNELLY, P. (1994). "The prosecutor's fallacy and DNA evidence". *Criminal Law Review*, PÁGS. 711-721.
- BUTLER, J. M. (2005). *Forensic DNA typing*. Elsevier Academic Press. Burlington (MA). USA.

- CHO, M. K. and SANKAR, P. (2004). "Forensic genetics and ethical, legal and social implications beyond the clinic". *Nature Genetics supp.* VOL. 36: PÁGS. s8-s12.
- FUNG, W. K. and HU, Y-Q. (2008). *Statistical DNA Forensic*. John Wiley and Sons Ltd. Chichester. U. K.
- GAMERO, J. J., ROMERO, J. L., PERALTA, J. L. CARVALHO, M. and CORTE-REAL, F. (2007). "Spanish public awareness regarding DNA profile databases in forensic genetics: what type of DNA profile should be included?". *J. Med. Ethics* vol. 33, págs. 598-604.
- GOODWIN, W., LINACRE, A. and HADI, S. (2007). *An introduction to Forensic Genetics*, John Wiley and Sons Ltd. Chichester, England.
- GRAHAM, E. A. M. (2006). "Disaster victim identification". *Forensic Sci. Med. Pathol.* Vol. 2, págs. 203-207.
- HOUCK, M. M. and SIEGEL, J. A. (2006). *Fundamentals of Forensic Science*, Elsevier Academic Press. Burlington (MA). USA.
- JEFFREYS, A. J., WILSON, V. and THEIN, S. L. (1985). "Individual-specific 'fingerprints' of human DNA". *Nature* vol. 316, págs. 76-79.
- JOBLING, M. A. and GILL, P. (2004), "Encoded evidence: DNA in Forensic analysis", *Nat. Rev. Genet.* Vol. 5, págs. 739-751.
- LI, R. (2008). "Forensic Biology". CRC Press. Boca Raton (FL). USA.
- MESTRES, F y J. VIVES-REGO. 2009a. "La utilización forense de la huella genética (secuencia del ADN o ácido desoxiribonucleico): aspectos científicos, periciales, procesales y sociales". *Revista La Ley Penal* nº 61, Junio.
- MESTRES, F. y J. VIVES-REGO. 2009b. "Genética forense: entre la tecnociencia y la imaginación". *Ludus vitalis* vol. 27, n. 32, págs. 447-450.
- MICHAELIS, R. C., FLANDERS jr., R. G. and WULFF, P. H. (2008). *A litigator's guide to DNA. From the laboratory to courtroom*. Elsevier Academic Press. Burlington (MA). USA.
- PATTON, S. M. (1990). "DNA fingerprint: the Castro case". *Harvard Journal of Law and Technology* vol. 3, págs, 223-240.
- POCH, R. (2009). "Dos gemelos, un robo y mucha confusión genética". *La Vanguardia* 20 Marzo 2009.
- REILLY, P. (2001). "Legal and public policy issues in DNA forensics". *Nature Reviews* vol. 2, págs.313-317.

- RUDIN, N. and INMAN, K. (2002). *An introduction to forensic DNA analysis*. CRC Press. Boca Raton (FL). USA.
- SCHULLER, W., FEREDAY, L. and SCHEITHAUER, R. (2001). *Interpol handbook on DNA data Exchange and practice. Recommendations from the Interpol DNA monitoring expert group*. First ed. Lyon, France.
- YU, N., KRUSKALL, M. S., YUNIS, J. J., KNOLL, J. H. M., UHL, L., ALOSCO, S., OHASHI, M., CLAVIJO, O., HUSAIN, Z., YUNIS, E. J., YUNIS, J. J. and YUNIS, E. J. (2002). “Disputed maternity leading to identification of tetragametic chimerism”. *N. Englan. J. Med.* Vol. 346, págs. 1545–1552.