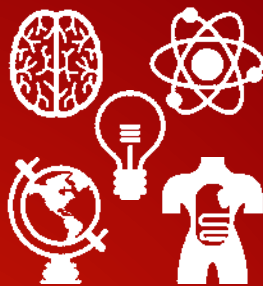


PINT OF SCIENCE



Edición Especial - Sevilla 2019



Secuencias de ADN: el asesino vivía en Sacramento

Joseph James DeAngelo tenía 72 años cuando la policía lo arrestó el año pasado, en su domicilio de Sacramento, como sospechoso de la autoría de 50 violaciones y 12 asesinatos cometidos hace cuatro décadas.

FUE UN CASO de gran impacto social entonces, que mantuvo en vilo a California y al país entero. Pese a esto, nunca se había podido identificar al asesino. Lo único claro era el modus operandi, y que actuó en varios condados. Con el tiempo, el caso se olvidó, excepto para Paul Holes, el agente que le perseguiría durante décadas hasta dar con él. Todo este desenlace comenzó en el año 2017, cuando habló con la genetista Bárbara Rae-Venter, que le ayudó a descubrir una posible pista... preguntando a una molécula, al ADN.

SI NOS REMONTAMOS a los años 70, en aquella época no se conocía la secuencia del genoma humano, que se completaría en el 2003, ni tampoco existían las herramientas para obtener en el laboratorio miles de copias de genes determinados (que se inventarían en la década de los 80). Afortunadamente, las muestras

obtenidas en escenas forenses, se recogían, y se almacenaban. **¿Y QUÉ ES LO QUE HA** ocurrido para que se pudiese identificar a un asesino tantas décadas después? Pues principalmente tres cosas: 1) por un lado la mejora en las técnicas experimentales de extracción y secuenciación del ADN (actualmente es posible secuenciar ADN muy degradado, etc.), 2) la reducción tremenda del coste de secuenciación del mismo, y 3) el desarrollo tremendo de la Bioinformática y el análisis de datos. Gracias a estos avances se ha popularizado y comercializado la tecnología para uso personal y/o recreativo creando la denominada "genómica personal", que se encarga de la secuenciación y análisis de la información genética de las personas.

ESA CURIOSIDAD y necesidad que tenemos los humanos de saber de dónde venimos, de trazar nuestros ancestros, ha

sido hábilmente explotada por la industria genómica, donde cientos de empresas ofrecen estos servicios, que producen un informe final indicando qué parte de nuestro ADN viene digamos del Cáucaso, qué proporción de mi genoma es Neandertal, o cuántas mutaciones tenemos en alguna parte de nuestro ADN que nos predispone a cierto tipo de enfermedades. El tremendo potencial de esta información, tampoco ha pasado inadvertida en el entorno policial, donde el nivel de resolución existente hasta entonces era muy limitado, ya que llegar hasta un nombre y unos apellidos a partir de un perfil genético era impensable.



Ana Maria Rojas Mendoza
 @amrojasmendoza
 Científica Titular CSIC,
 Centro Andaluz de
 Biología del Desarrollo
 (CABD)



¿Y DÓNDE SE ALMACENA esta información? En principio son las propias empresas que comercializan las pruebas, las que desde luego almacenan los perfiles genéticos de sus usuarios. Pero existen algunos repositorios de acceso públicos con información genética de miles de personas. Uno de esos sitios es GEDmatch, donde cualquier usuario con conexión a internet y datos proporcionados por las empresas que ofertan genómica puede subir su información genética y compararla con otros perfiles. Normalmente, estos usuarios quieren buscar parientes lejanos, hermanos perdidos, etc... Según un estudio reciente [1], más de 15 millones de personas han usado estos servicios con distintos propósitos. Con estos números en las bases de datos, una estimación a la baja predice que básicamente

cualquier persona con una herencia genética europea (porque la mayoría de las personas que se hacen esos pruebas son blancos de descendencia europea), podría tener un primo en tercer grado presente en esa base de datos. Y, realmente, un primo en tercer grado es todo lo que se necesita, en principio, para identificar a alguien con nombre y apellidos. Veamos cómo.

ASUMIENDO QUE UNA persona tiene a priori una probabilidad media de tener 850 parientes en tercer grado o menos (es decir, parientes que comparten un tatarabuelo), y que el número de usuarios que solicitan esos servicios aumenta, la probabilidad de que tengamos un primo en tercer grado que además haya depositado su información genética en alguna base de datos, es elevada. Si además esas empresas "ceden"

esta información a las fuerzas de seguridad (tipo FBI), el acceso a información genética aumenta sustancialmente. Bien, pues supongamos que un pariente en tercer grado nuestro es un criminal. Y supongamos además que se puede "adivinar" la edad del pariente (por información adicional que manejan los investigadores), lo que reduciría en un 90% el número de parientes sospechosos. Y, además, digamos que los investigadores pueden tener información realista de "donde viviría" nuestro pariente en un radio de aproximadamente 150 km. Esto ya excluiría a un porcentaje significativo de personas. Y desde luego, podríamos seguir bajando el número de sospechosos si además sabemos el sexo de la persona en cuestión. Por lo tanto, de una muestra inicial de 800 parientes lejanos, podemos reducir fácilmente la

cifra a unos diez, lo cual reduce significativamente el universo de búsqueda.

ESTE PROCESO, que parece tan sencillo y directo, es, de hecho, bastante complicado, ya que requiere de un gran conocimiento en genética y análisis de datos. El formato de la información genética procedente de una prueba dirigida al consumidor es diferente al formato de la información genética de un perfil forense, y hay que convertir un formato en otro. Básicamente es como una aplicación de un teléfono que sólo funciona en Android y no puede leerse en un iPhone, o viceversa. Esto nos lleva a preguntarnos ¿cuál es la razón de que esos formatos sean diferentes? Se trata de una cuestión legal. Precisamente porque el formato del perfil forense protege nuestra privacidad, ya que no se puede llegar de ahí a un nombre con apellidos, y, por eso mismo, estas pruebas se aceptan en los juicios. Ahora, ¿de qué depende ese formato? Del tipo de información que miremos en nuestro ADN. Nuestro ADN nos dice cosas distintas, en función de lo que le preguntemos, y como esa información genética no es directamente comparable, ahí es donde los analistas de datos con conocimiento del dominio genético, son esenciales.

NUESTRO ADN es el manual de instrucciones de nuestro cuerpo. Es como un libro hecho de una sola palabra con un abe-

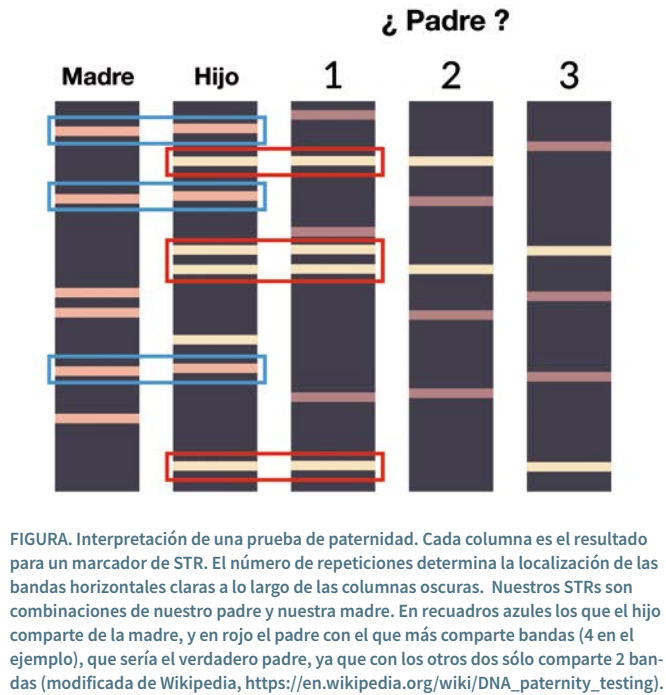


FIGURA. Interpretación de una prueba de paternidad. Cada columna es el resultado para un marcador de STR. El número de repeticiones determina la localización de las bandas horizontales claras a lo largo de las columnas oscuras. Nuestros STRs son combinaciones de nuestro padre y nuestra madre. En recuadros azules los que el hijo comparte de la madre, y en rojo el padre con el que más comparte bandas (4 en el ejemplo), que sería el verdadero padre, ya que con los otros dos sólo comparte 2 bandas (modificada de Wikipedia, https://en.wikipedia.org/wiki/DNA_paternity_testing).

cedario de 4 letras (ATGC) que se repiten de maneras diferentes. El ADN de todos los seres humanos que vivimos en el planeta se parece muchísimo, en un 99% de similitud en la secuencia. Lo que preguntamos a nuestro ADN en los casos de identificación, es ¿cómo de similar es ese 1% diferente que hace que yo sea más parecido a mis hermanos que a mis primos? Este proceso es el que hacen las empresas de genealogía (por ejemplo 23andMe, Ancestry, Family Tree, etc...), lo que se llama “genotipado”. Imaginemos que en ese libro imaginario hay índices que representan en conjunto ese 1%, y que esos índices son lugares del libro que nos diferencian a las personas. Es decir, imagi-

nemos que en una persona el punto número dos del índice es una “A”, mientras que en otra es una “T”. Estos cambios en los índices, se llaman “polimorfismos de un solo nucleótido” o SNP (del inglés Single Nucleotide polymorphism, es decir variaciones en una única letra del ADN en una posición determinada). Las pruebas ofertadas comercialmente escrutan miles de SNPs. Por ejemplo, 23andMe analiza 650.000. Como los distintos grados de parentesco pueden expresarse en porcentajes de similitud en su ADN, se pueden buscar relaciones familiares. Así pues, dos hermanos, compartirán con sus padres cada uno un 50% del ADN de cada progenitor (pero ojo, que en cada hermano

ese 50% es diferente). Los tíos/sobrinos o abuelos/nietos, compartirían un 25%. Esta similitud va disminuyendo conforme aumenta la distancia de parentesco, así un primo en tercer grado compartiría un 0.78% (datos de 23andMe).

LA REGIÓN DEL ADN que se mira en entornos forenses o criminalistas, para, por ejemplo, realizar pruebas de paternidad (ver figura), no se basa en analizar variaciones en esos índices. En este caso preguntamos al ADN otra cosa, en particular: ¿cuál es mi código de barras? Fue Alec Jeffreys, ya en los 80 [2] quien llevara la Biología Molecular a la criminalística e inventó lo que se conoce como el “perfilado genético” del ADN. Los perfilados genéticos analizan otras regiones del ADN, que son muy variables y se llaman “ADN satélite” o STR (del inglés, Short Tandem Repeats). Los perfiles genéticos que podemos encontrar en las bases de datos de las fuerzas de seguridad incluyen información sobre unas 20 regiones STR de cada sospechoso. Estos son 20 lugares distribuidos por nuestro ADN (que podrían o no solapar con los índices del libro, o SNPs) donde una cadena de 2 a 6 bases de ADN tiende a repetirse. Por ejemplo: AAGGAAGGAAGG, sería esta cadena de cuatro letras (AAGG) repetida tres veces. Las personas tenemos un número diferente de repeticiones en estas zonas, lo que las convierte

como en un código de barras. Y todos los individuos tenemos un código único en estas regiones. Un ejemplo en la figura, que representa una prueba de paternidad.

¿CÓMO DE COMPARABLE es un perfil genético forense (basado en STRs) con uno comercial (basado en genotipado)? Un estudio reciente realizado en 872 personas encontró que para un 30% de ellas podrían encontrarse relaciones entre los dos tipos de perfiles empleando técnicas estadísticas muy complejas [3], por lo que para realizar estas comparaciones se requiere un conocimiento muy profundo en genética, bioinformática y estadística aplicada a la genealogía. ¿Y cómo es el proceso? Un buen ejemplo lo ilustra el llamado DNA Doe Project (<http://dnadoeproject.org/>), un proyecto sin ánimo de lucro que usa genealogía y genética para identificar personas, incluyendo víctimas de crímenes. Fundado por las investigadoras Colleen Fitzpatrick y Margaret Press, este proyecto selecciona casos no resueltos, de cuyas muestras se extrae de nuevo ADN. Este ADN es secuenciado, y, mediante bioinformática, se convierten los formatos y con técnicas estadísticas se buscan parecidos. Cuando los encuentran, contactan a las autoridades para realizar una prueba forense. En el caso de DeAngelo, probablemente tuvo que realizarse este mismo protocolo, aunque la policía no

se ha pronunciado oficialmente aún al respecto. El agente Paul Holes pidió a la genetista Bárbara Rae-Venter ayuda para resolver computacionalmente el caso. Lo único certero es que crearon un perfil genético compatible con GEDmatch para buscar primos en tercer grado del sospechoso y ... ¡voilà! Encontraron al primo con todos sus datos demográficos, lo que permitió realizar un árbol genealógico de la familia hasta el tatarabuelo, y a partir de ahí navegando por las ramas de vuelta (desde el tatarabuelo hasta los abuelos) se pudo identificar a potenciales sospechosos por alguna de las ramas de ese árbol. Una vez que DeAngelo fue considerado el sospechoso más probable, con una simple extracción de ADN de saliva, pelo, servilleta, o cualquier otra muestra, se obtuvieron los STRs, y se compararon con los STRs de la escena del crimen. Este último paso se tuvo que hacer porque existe cierto vacío legal en cuanto a si los test genealógicos pueden usarse como pruebas judiciales. Y así lo encontraron, preguntando al ADN. Por cierto, DeAngelo era policía. 🍷

Para saber algo más:
Erlich et al, 2018. Science. (DOI:10.1126/science.aau4832)
Jeffreys, et al. 1985. Nature. (DOI: 10.1038/31607)
Kim et al, 2018. Cell (DOI:10.1016/j.cell.2018.09.008)