



**Universitat de les
Illes Balears**

Facultat de Ciències

Memòria del Treball de Fi de Grau

Aplicaciones forenses de la secuenciación masiva: situación actual, ventajas y limitaciones.

Micaela Ariadna Arango Colonna

Grau de Bioquímica

Any acadèmic 2018-19

Treball tutelat per: Antònia Picornell Rigo.

*Departament de Biologia, Biologia Fonamental i Ciències de la Salut, Física, Química

S'autoritza la Universitat a incloure aquest treball en el Repositori Institucional per a la seva consulta en accés obert i difusió en línia, amb finalitats exclusivament acadèmiques i d'investigació	Autor		Tutor	
	Sí	No	Sí	No
	X		X	

Paraules clau del treball:

Genética Forense, Secuenciación Masiva (*Next Generation Sequencing, Massive Parallel Sequencing*), perfil genético, marcadores genéticos.

	PÁGINA
1. Resumen.	4
2. Abreviaturas.	4
3. Introducción.	4
3.1. <i>Contexto histórico.</i>	4
3.2. <i>Polimorfismos de DNA como marcadores forenses.</i>	5
3.3. <i>Etapas del análisis forense.</i>	6
3.4. <i>Técnicas de análisis utilizadas en Genética Forense</i>	6
3.5. <i>Técnicas de secuenciación.</i>	7
3.6. <i>Objetivos.</i>	8
4. Materiales y métodos.	8
4.1. <i>Búsqueda bibliográfica.</i>	8
4.2. <i>Diseño de la encuesta.</i>	8
5. Resultados.	10
5.1. <i>Búsqueda bibliográfica: Tablas y evolución.</i>	10
5.2. <i>Resultados de la encuesta.</i>	12
6. Estado actual de la secuenciación masiva en Genética Forense.	16
6.1. <i>Ventajas e inconvenientes de la secuenciación masiva.</i>	16
6.1.1. <i>Elaboración de perfiles genéticos</i>	17
6.1.2. <i>Muestras en Genética Forense.</i>	17
6.1.3. <i>Estudio de SNPs.</i>	17
6.1.4. <i>Epigenética, transcritpoma, mtDNA y Medicina Forense.</i>	18
6.1.5. <i>Inconvenientes.</i>	18
6.1.6. <i>Resumen.</i>	19
6.2. <i>Comercialización y validación.</i>	20
6.2.1. <i>Kits para NGS en Genética Forense.</i>	20
6.2.2. <i>Validación de la técnica.</i>	21
6.3. <i>Software de análisis.</i>	21
6.3.1. <i>Análisis de datos.</i>	21
6.3.2. <i>Software de alineamiento, clasificación y almacenamiento.</i>	22
6.4. <i>Bases de datos y nomenclatura.</i>	22
6.4.1. <i>Bases de datos.</i>	22
6.4.2. <i>Nomenclatura.</i>	22
6.5. <i>Consideraciones éticas.</i>	23
7. Conclusiones.	24
8. Bibliografía.	24

1. RESUMEN

La Genética Forense ha permitido la identificación de individuos y el establecimiento de parentescos entre individuos gracias al estudio de marcadores biológicos. Fue a finales del siglo XX cuando esta rama de la ciencia llegó a su máximo esplendor debido a la introducción del uso de polimorfismos de DNA como marcadores. Hasta la fecha, la técnica estrella para el análisis de estos marcadores ha sido la de Reacción en cadena de la Polimerasa (PCR) con una posterior Electroforesis Capilar (CE).

Sin embargo, la creación de técnicas de secuenciación, en especial en la última década con la Secuenciación Masiva (NGS), ha creado debate sobre si la técnica PCR-CE debería seguir siendo considerada como el *golden standard* para el análisis de material genético en el contexto forense. Las NGS son técnicas que ofrecen resultados más refinados que los obtenidos actualmente con el método estándar, a pesar de ello, debido a que son relativamente nuevas todavía se carece de la existencia de métodos validados y estandarizados que puedan ser implantados de forma rutinaria en el laboratorio.

Forensic Genetics have allowed the identification and establishment of kinship between individuals due to the study of biological markers. At the end of the 20th Century this science branch reached its highest peak because DNA polymorphisms were started to be used as markers. To date, the main technique for DNA analyses has been the Polymerase Reaction Chain (PCR) followed by a Capillary Electrophoresis (CE).

Even so, the emerging of sequencing techniques, specially during the last decade where Massive Parallel Sequencing was developed, has created a debate on whether or not the PCR-CE technique has to be maintained as the golden standard for analyzing genetic material. Next Generation Sequencing techniques can offer more precise results than the ones obtained with the standard method, despite so, they are quite novel techniques that lack of validated and standard methods that can be used at a laboratory on a daily basis.

2. ABREVIATURAS

NGS <i>Next Generation sequencing</i>	MPS <i>Massive Parallel Sequencing</i>	PCR Reacción en cadena de la polimerasa
CE Electroforesis Capilar	RT-PCR Reacción en cadena de la polimerasa con retrotranscriptasa inversa	cDNA DNA complementario
nDNA DNA nuclear		
RLFP <i>Restriction Fragment Length Polymorphism</i>	VNTR <i>Variable Number of Tandem Repeats</i> o DNA minisatélite	STR <i>Short Tandem Repeats</i> o DNA microsaelite
SNP <i>Single Nucleotide Polymorphism</i> o Polimorfismo de secuencia	InDels Polimorfismo de longitud	HLA Sistema mayor de histocompatibilidad
mtDNA DNA mitocondrial	AIM <i>Ancestry-informative marker</i>	
CoDIS Combined DNA Index System	Pb Pares de bases	ddNTP dideoxinucleótidos
ESS <i>European Standard Set</i>		
SAM Sequence Alignment Map	BAM Binary Alignment/Map	VCF Variant Calling Format

3. INTRODUCCIÓN

3.1. Contexto histórico.

La Genética Forense es la ciencia encargada de la resolución de procesos jurídicos mediante el análisis del material genético de los involucrados¹. Las áreas más tratadas en este ámbito son las de parentesco (*tests* de paternidad, linajes, inmigración...), criminalística (asesinatos, asaltos, delitos sexuales...) e identificación de personas desaparecidas y restos humanos¹.

La clave reside en el estudio de elementos que puedan ir variando entre individuos y con ello distinguir a una persona de otra. Es aquí cuando entró en juego el término polimorfismo, acuñado por Ford en 1940²: “Se trata de ocurrencia en un mismo ambiente de dos o más formas discontinuas de una especie en proporciones tales que la más rara de ellas no puede ser mantenida en la población simplemente por mutación recurrente”.

El primer marcador forense utilizado fue el sistema ABO de las células sanguíneas, descrito por primera vez a principios del siglo XX por Karl Landsteiner³. Posteriormente, el resto de sistemas de antígenos eritrocitarios como podían ser el Rh, MNS o Duffy, se sumaron a la cola de los marcadores genéticos utilizados en forense¹. Fue en la década de los 70 cuando diversos investigadores, destacando entre ellos a Wolfgang Mayr, introdujeron a estas técnicas el análisis de antígenos del sistema mayor de histocompatibilidad (HLA), cuyo uso fue muy destacado para casos de paternidad⁴.

El principal problema con aquellos marcadores era su análisis en muestras degradadas como pueden ser huesos y restos biológicos como podrían ser pelos o semen¹. Pero todo ello cambió cuando en 1980 se introdujo el estudio del DNA como nuevo marcador genético¹.

3.2. Polimorfismos de DNA como marcadores forenses.

Un polimorfismo de DNA se define como aquellas variables de un mismo gen (alelos) que se encuentran en una determinada posición fija del cromosoma (*locus*)⁵. A mayor cantidad de alelos que haya en un *locus*, habrá un mayor polimorfismo y una menor probabilidad de encontrar dos individuos genéticamente iguales⁵. Los polimorfismos de DNA son divididos en cuatro clases principales¹:

- a) Polimorfismos de secuencia (SNP). Son variaciones en una sola base de la secuencia del DNA, producidas por mutaciones puntuales. También se pueden incluir los resultados de diversas mutaciones puntuales en una misma secuencia.
- b) Polimorfismos de longitud (InDels). Se producen por inserciones o deleciones de material genético en la secuencia de DNA.
- c) DNA minisatélite (*Variable Number of Tandem Repeats- VNTRs*). Fragmentos de unos 300-20.000 pares de bases (pb) de longitud cuya secuencia se basa en repeticiones de 15-50 pb.
- d) DNA microsatélite (*Short Tandem Repeats- STR*). Fragmentos de 50-500pb con secuencias repetitivas de 2-6pb de longitud.

Los polimorfismos de DNA se pueden encontrar tanto en cromosomas autosómicos como sexuales (X, Y)⁶. El cromosoma Y es pequeño (60Megabases) con un modo de herencia paterna, del que se pueden estudiar tanto STRs como SNPs, localizados en la región no-pseudoautosómica y, por tanto, al contrario de lo que ocurre en cromosomas autoómicos, son polimorfismos libres del proceso de recombinación⁶. Además, el DNA localizado en esta región se hereda de forma haploide, es decir que sus *loci* se heredan siempre de forma conjunta, formando un haplotipo⁷. Las características de los polimorfismos del cromosoma Y han sido de mucha utilidad en casos de delitos sexuales, relaciones de parentesco y el trazo de la evolución y linajes paternos⁷.

Para la identificación de individuos también cabe destacar el papel del gen de la amelogenina, ya que, debido a su posición en los cromosomas sexuales, permite la distinción de individuos XX e XY⁸. Esto se debe a que los productos de PCR de amelogenina se diferencian en que en el caso de AMELX existe una deleción de 6pb en el tercer intrón, cosa que no ocurre con AMELY⁸.

Por su parte, el genoma mitocondrial es un buen sujeto de análisis cuando uno se encuentra ante muestras que carecen de DNA nuclear (nDNA), como podrían ser pelos sin bulbo o muestras degradadas. Dependiendo del tipo celular, existen entre 100 y 10000 copias de mDNA por cada copia de nDNA⁹, además presenta una elevada tasa de mutación, es haploide y se transmite

mayoritariamente por vía materna¹⁰. Aunque sea menos usado que el nDNA para la distinción de individuos debido a su menor variabilidad, ha demostrado ser clave a la hora de establecer linajes maternos y para la identificación de restos óseos¹¹.

3.3. Etapas del análisis forense.

Las muestras utilizadas en Genética Forense son de carácter biológico, desde fluidos como la sangre, saliva o semen, hasta restos de tejidos o huesos. En primer lugar, se aplican las pruebas preliminares con el objetivo de determinar la naturaleza y el organismo de procedencia de una muestra¹². Estas pruebas se engloban en: pruebas orientativas, que permiten descartar ciertas procedencias de la muestra; las de certeza, usadas para definir el tipo de resto biológico y las específicas para determinar a qué organismo pertenece¹².

En cuanto a las técnicas de análisis de DNA para la identificación individual, el *golden standard* es el estudio simultáneo de 15 STRs distribuidos por todo el genoma¹³. Estos STRs deben tener alto grado de heterocigosidad y que la combinación de éstos sea tal que facilite la discriminación entre individuos¹³. Además, suelen tener longitudes entre 90-200pb y están formados por bloques de repetición de 4-5pb ya que eso les otorga una mayor estabilidad y con ello disminuyen la aparición de artefactos, facilitan la reproducibilidad y garantizan la robustez de los resultados¹⁴.

De este estudio se obtiene un patrón ordenado de STRs denominado como perfil genético, el cual es característico de cada individuo, aunque existen ciertas excepciones como los gemelos monocigóticos¹⁵. El perfil genético puede ser transformado en un código numérico y almacenado en una base de datos para su posterior manejo y comparación¹³.

El material genético necesario para crear una huella de DNA puede venir de una muestra biológica recogida en una escena del crimen y por tanto su origen es desconocido (DNA dubitado); o provenir de alguno de los implicados en el crimen (DNA de referencia o indubitado)¹². Los resultados obtenidos con la muestra dubitada son comparados con los de referencia, además las bases de datos permiten comparar perfiles generados por delitos anteriores con los que se crearon a partir de muestras de la escena del crimen¹².

Para rechazar la compatibilidad o coincidencia de un perfil se precisa de al menos dos marcadores, siendo tres en el caso de STRs debido a la mayor tasa de mutación de éstos, la incompatibilidad se considera concluyente¹². Por el contrario, si el resultado es coincidente o compatible, será necesario una valoración estadística para comprobar la probabilidad de que se haya dado por fruto del azar¹². Una vez establecido el resultado, se elabora un informe médico-legal que será utilizado en el juicio¹.

3.4. Técnicas de análisis usadas en Genética Forense.

Los primeros marcadores de DNA forense fueron los VNTRs o minisatélites, analizados mediante la técnica de análisis de polimorfismos de longitud mediante fragmentos de restricción (*Restriction Fragment Length Polymorphism-RFLP*)¹⁶. La técnica consiste en generar fragmentos de DNA mediante el uso de enzimas de restricción, separación de éstos mediante una electroforesis en gel y transferencia a una membrana donde se realizará un *Southern Blot*, obteniendo un patrón de bandas¹⁶. Los grandes problemas de esta metodología eran la gran cantidad de DNA de partida requerido (100-200ng) y la dificultad para realizar comparaciones entre laboratorios¹⁶.

A día de hoy, para el análisis de DNA consiste en el uso de la técnica PCR-CE¹². Como las muestras en forense suelen ser escasas o estar degradadas, es necesario un paso que permita la amplificación del material de partida y es por ello que se aplica la PCR, ya que se obtienen muchas copias de DNA a partir de una sola molécula inicial¹². Además, la técnica de PCR sólo requiere de 0,2-2ng de material de partida, y la introducción de PCR *multiplex* ha permitido la amplificación de diferentes regiones de DNA de manera simultánea mediante la adición de distintos cebadores⁹. En el caso de disponer sólo de RNA, se procederá a realizar una RT-PCR (Reacción en cadena de la polimerasa con transcriptasa inversa), obteniendo así cDNA¹⁷. Una vez generados los fragmentos, se separan por electroforesis capilar, distinguiéndolos por tamaño y se comparan con patrones alélicos de cada uno de los marcadores utilizados¹².

3.5. Técnicas de secuenciación.

La secuenciación es un conjunto de técnicas cuya finalidad es determinar el orden de nucleótidos de una cadena de DNA¹⁸. La primera técnica de secuenciación fue descrita por Maxam y Gilbert (1974) y se trata de la secuenciación por hidrólisis química específica, la cual se basaba en el marcaje con P³², tratamiento químico y revelado por autorradiografía¹⁹.

En 1977, Sanger estableció el método de secuenciación por dideoxinucleótidos (ddNTP)¹⁸. Los ddNTPs provocan la terminación de la elongación de la cadena generando fragmentos de DNA que pueden ser separados por electroforesis y revelados por fluorescencia¹⁷. El principal problema es la incapacidad de este método para analizar múltiples *loci* de manera simultánea y junto con la aparición del Proyecto Genoma Humano, se llegó a la conclusión de que se debía crear una nueva tecnología de mayor rendimiento²⁰.

Es por ello que a principios del siglo XXI empezaron a salir nuevas técnicas de secuenciación basadas en el procesamiento paralelo masivo, es decir, el análisis simultáneo de secuencias que fueron denominadas con técnicas de Segunda Generación (*Next Generation Sequencing* (NGS) o *Massive Parallel Sequencing* (MPS))²¹.

Las NGS comparten un paso previo de preparación que es la realización de librerías²²: el DNA se fragmenta y sus extremos son reparados, se añade una cola de poli A y los fragmentos se ligan a una plataforma donde serán amplificados por PCR y con ello conseguiremos *clusters* o conjuntos de fragmentos de las zonas de DNA seleccionadas²². A continuación, se procede a la secuenciación y en función a cómo se realice, existen toda una serie de plataformas de NGS:

- a) *454 Genome Sequencing System* (2005-2007): Pirosecuenciación o secuenciación por síntesis. A los fragmentos del *cluster* se les incorpora la DNA polimerasa y cuando ésta añade el nucleótido se libera un pirofosfato que es captado por la ATP sulfúrilasa y transformado en ATP²³. Posteriormente, se acopla la a la luciferasa dando lugar a oxiluciferina la cual otorga una señal luminosa cuantificable que distinguirá qué nucleótido se ha incorporado²³.
- b) SOLiD (2006): Secuenciación por ligación. Una vez producida la amplificación un *primer* universal se une y a continuación se procede al ligamiento de octámeros fluorescentemente marcados²². La unión provoca la escisión de los 3 últimos nucleótidos del octámero, generando fluorescencia y el pentámero restante se une al *primer* por la DNA ligasa. Después de cada ciclo, el *primer* se elimina y se añade otro nuevo en la posición n-1²².
- c) *Illumina Solexa* (2007): Secuenciación por hibridación. Algunas bases añadidas por la polimerasa se encuentran marcadas por fluorescencia en el extremo 3'-OH por lo que cuando es incorporado, el proceso se detiene y se escinde el marcaje de la base provocando fluorescencia²³. Posteriormente, el extremo 3'-OH de la base incorporada queda libre para continuar con la enlongación²³.
- d) Ion Torrent (2010): Secuenciación por semiconducción. La plataforma en la que se produjo la amplificación se deposita en pocillos de un microchip equipados con sensores de voltaje, los cuales son capaces de detectar los protones liberados durante el proceso de elongación y transformados en señales eléctricas que distinguirán qué base ha sido incorporada²⁴.

Durante estos últimos años la base de las técnicas de secuenciación ha ido cambiando, saliendo a la luz las técnicas de Tercera Generación¹⁷. Éstas no secuenciarían el DNA amplificado, sino que serían capaces de secuenciar la molécula de forma directa¹⁷.

Entre estas técnicas destaca *MinION platform*, introducida por *Oxford Nanopore Technologies* en 2014²⁵. La doble cadena de DNA es separada mediante adaptadores y a continuación se procede a su paso por un nanoporo embebido en una bicapa lipídica o polímero sintético gracias a la aplicación de un campo eléctrico²⁵. El paso a través del nanoporo provoca un cambio en la corriente de iones que se encuentra en su interior, quedando ésta parcialmente bloqueada a medida que el DNA viaja. Este cambio en la amplitud de la corriente es detectado y usado para determinar la secuencia

nucleotídica a medida que el fragmento pasa a través del poro²⁵. Las dos hebras se vuelven a unir una vez se finaliza el recorrido²⁵.

Las *Third Generation Sequencing* permitirán eliminar los *bias* de amplificación y reducirán el tiempo de secuenciación a horas²⁶. Además, serán capaces de realizar lecturas de hasta kilobases de longitud y diferenciar de manera efectiva variantes estructurales en el material genético²⁶. Sin embargo, el principal problema que se enfrentan es la presencia de *ratios* de error mayores que las NGS²⁶.

3.6. Objetivos.

Este trabajo busca analizar la situación actual de las técnicas de secuenciación de segunda generación en el ámbito de la Genética Forense. Para ello, se realizará una búsqueda bibliográfica sobre el estado de arte tanto de la técnica como del contexto en el que se quiere explicar. También se pretende definir las ventajas que ofrecen las NGS si sustituyesen al *golden standard* actual y a qué obstáculos se tendrían que enfrentar para poder conseguirlo, destacando la aplicabilidad, estandarización y validación y comercialización.

4. MATERIALES Y MÉTODOS

4.1. Búsqueda bibliográfica.

La primera etapa de la búsqueda fue de carácter más general, se usaron palabras clave como "*Forensic Genetics*", "*Next Generation Sequencing*" y "*Massive Parallel Sequencing*", con el objetivo de ir conociendo diferentes aspectos de la Genética Forense, cómo se trabaja y qué aplicaciones tiene. El motor de búsqueda utilizado fue "PubMed" y para obtener una visión más amplia del tema, no se aplicaron restricciones cronológicas ni de especie.

El siguiente paso fue proceder a la combinación de las palabras claves y aplicar filtros con el objetivo de acotar la búsqueda y relacionarla con el tema de trabajo. En el caso de PubMed, se aplicaron los filtros de restricciones cronológicas de 10 y 5 años y en especie humana. Además, en esta parte de la búsqueda se añadió la consulta de las principales revistas especializadas en Genética Forense como *Forensic Science International*, *Forensic Science International: Genetics*, e *International Journal of Legal Medicine*.

A continuación, se elaboraron tablas (**Tablas 1 y 2**) cuantificando los artículos encontrados en cada caso y determinando cuántos de estos fueron escogidos siguiendo los criterios de selección. A la hora de escoger artículos, se recopilaron todos ellos cuyos títulos y *abstracts* se relacionaban y adecuaban a la temática del trabajo.

Los documentos elegidos en un principio no reflejan el número total de artículos que van a ser incluidos en este trabajo. Hay que tener en cuenta que de estos primeros artículos se seleccionaron mediante la lectura del *abstract* y aunque inicialmente hayan podido parecer de utilidad, una lectura profunda de ellos puede haberlos llevado a descartar. También hay que tener en cuenta los artículos que se hayan podido encontrar a partir de los seleccionados inicialmente y cuya información sí haya sido de utilidad.

4.2. Diseño de la encuesta.

Como parte del proyecto se incluye una encuesta cualitativa, se han elaborado una serie de preguntas relacionadas con la temática del trabajo y se han enviado a un conjunto de expertos en el tema. El objetivo de la encuesta es poder conocer el punto de vista de personas que trabajan de manera rutinaria en laboratorios de Genética Forense, ya que sus opiniones ayudarán a comprender mejor el estado actual de las técnicas de secuenciación masiva dentro de este ámbito y cómo éste podría evolucionar con el paso de los años.

Esta encuesta comprende un total de 8 preguntas, ha sido creada con *Google Documents* y se accede por: https://docs.google.com/forms/d/1vg7z3fMLqqF30pTlvMnw5R7AulBRbuxZuQUZ3H259xk/viewform?edit_requested=true#responses

1. ¿Su laboratorio utiliza de manera rutinaria técnicas de secuenciación masiva?

- Sí
- No
- En casos especiales.

2. En el caso de haber respondido Sí o en casos especiales:

a) *¿En qué situaciones lo utiliza?*

- Paternidades u otras relaciones de parentesco.
- Identificación individual en criminología.
- Descripción de las características fenotípicas.
- Otros: (especifique)

b) *¿Qué plataforma y kits utiliza? (Puede nombrarse más de un kit según el marcador a analizar)*

c) *¿Qué ventajas técnicas le ofrecen las NGS ante la PCR-EC? Puede seleccionar más de una opción.*

- Mayor rendimiento de la técnica.
- Obtención de resultados más precisos.
- La menor tasa de error de la técnica.
- Capacidad para trabajar con varias muestras a la vez.
- Otros: (especifique)

d) *¿Cuáles han sido los principales inconvenientes a la hora de implantar la técnica en el laboratorio? Puede seleccionar más de una opción.*

- El coste económico.
- La formación del personal.
- La duración de la obtención de resultados.
- Otros: (especifique)

3. En el caso de haber respondido "No":

a) *¿Qué desventajas técnicas encuentra? Puede seleccionar más de una opción.*

- El coste económico.
- El entrenamiento del personal.
- La duración de la obtención de resultados.
- La cantidad de datos innecesarios/irrelevantes producidos.
- La falta de estandarización/validación.
- Otros: (especifique)

4. La introducción de NGS al ámbito rutinario de un laboratorio forense permitiría la creación de perfiles genéticos más refinados, ¿cree usted que vale la pena la sustitución de la PCR-CE teniendo en cuenta los cambios que se deberían realizar en procesos de detección y almacenamiento de datos?

5. Aunque las técnicas de secuenciación de segunda generación permitan la secuenciación de genomas completos, ¿cree que la secuenciación dirigida es una herramienta más útil en el ámbito de la Genética Forense?

6. ¿Se debería invertir más en el desarrollo de los denominados "*personal sequencers*" o en las plataformas de secuenciación de tercera generación? ¿Por qué?

7. ¿Se debería crear una base de datos en la nube con acceso a los laboratorios forenses para que cada uno de los laboratorios pudiese introducir los nuevos alelos descubiertos en los marcadores?

8. Con las NGS se pueden detectar la susceptibilidad de un individuo a padecer cierta enfermedad ¿Cree que en algún caso sería conveniente la comunicación del resultado?

- Se debería comunicar el resultado siempre.
- Se debería comunicar el resultado en caso del que el individuo lo desee.
- No se debería comunicar nunca el resultado.

5. RESULTADOS

5.1. Búsqueda bibliográfica: Tablas y evolución.

Las tablas presentadas a continuación son reflejo de la búsqueda bibliográfica realizada el 2 de noviembre de 2018, tanto en el motor de búsqueda PubMed como en las diferentes revistas de Medicina Forense. Las tablas elaboradas (**Tablas 1 y 2**) muestran las palabras clave utilizadas, los filtros de búsqueda aplicados, el número total de artículos que ha dado la búsqueda en sí, cuáles de éstos son *reviews* y cuántos han sido finalmente seleccionados.

Tabla 1. Representación de la búsqueda bibliográfica en PubMed.

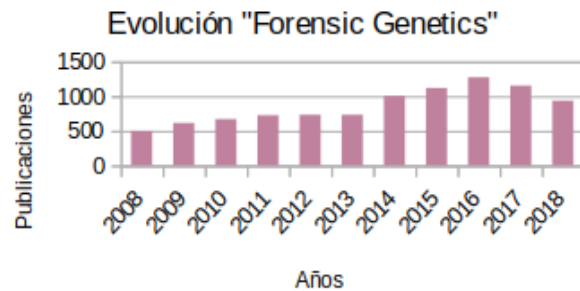
PALABRAS CLAVE	AÑOS	FILTROS	ARTÍCULOS	SEL.	REVIEW	SEL.
Forensic Genetics + Next Generation Sequencing	Últimos 10 años	<i>Best Match</i>	403	21	49	10
		<i>Best Match Species: Human</i>	326	3	37	1
	Últimos 5 años	<i>Most recent</i>	339	3	33	2
		<i>Most Recent Species: Human</i>	286	5	27	3
Forensic Genetics + Massive parallel sequencing	Últimos 10 años	<i>Best Match</i>	28	4	1	0
		<i>Best Match Species: Human</i>	24	0	1	0
	Últimos 5 años	<i>Most recent</i>	28	0	2	1
		<i>Most recent Species: Human</i>	24	0	1	0

Tabla 2. Representación de la búsqueda bibliográfica en revistas.

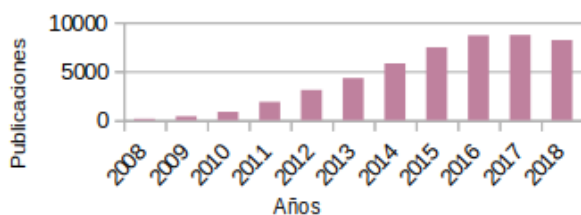
REVISTA	PALABRAS CLAVE	AÑOS	ARTÍCULOS	SEL.	REVIEW	SEL.
Forensic Sciences International: Genetics	<i>Massive parallel sequencing</i>	Últimos 10 años	224	4	10	3
		Últimos 5 años	226	2	10	0
	<i>Next Generation sequencing</i>	Últimos 10 años	317	4	19	2
		Últimos 5 años	265	0	16	0
International Journal of Legal Medicine	<i>Massive parallel sequencing</i>	Últimos 10 años	2	0	0	0
		Últimos 5 años	2	0	0	0
	<i>Next Generation Sequencing</i>	Últimos 10 años	40	2	4	1
		Últimos 5 años	38	1	4	0
Forensic Sciences International	<i>Massive parallel sequencing</i>	Últimos 10 años	17	0	4	1
		Últimos 5 años	12	0	3	0
	<i>Next Generation Sequencing</i>	Últimos 10 años	42	1	5	2
		Últimos 5 años	37	0	4	0

Para determinar cómo ha ido evolucionando la aplicación de las técnicas de secuenciación masiva en Genética Forense, se ha querido encontrar la relación entre el número de publicaciones que incluyen las palabras clave *Forensic Genetics* y *Next Generation sequencing/Massive parallel sequencing*. Para ello, se ha ido anotando la cantidad de artículos publicados en cada uno de los campos respectivamente y luego se ha hecho una búsqueda conjunta de ambas palabras clave. Los resultados se muestran a continuación:

Gráfica 1. Representación en gráfico de barras de la cantidad total de artículos sobre "Forensic Genetics" publicados en los últimos 10 años y encontrados en el motor de búsqueda PubMed.



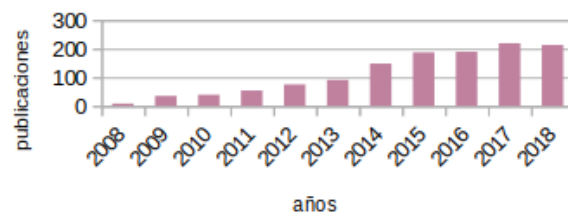
Gráfica 2. Evolución de "Next Generation Sequencing" y "Massive Parallel sequencing"



Gráfica 2. Representación en gráfico de barras de la cantidad total de artículos sobre "Next Generation Sequencing" o "Massive parallel sequencing" publicados en los últimos 10 años y encontrados en el motor de búsqueda PubMed.

Gráfica 3. Representación en barras de la cantidad de artículos sobre "Forensic genetics + ("Next Generation Sequencing + Massive Parallel Sequencing") publicados en los últimos 10 años y encontrados en el motor de búsqueda PubMed.

Gráfica 3. Evolución de "Forensic genetics + ("Next Generation Sequencing + Massive Parallel Sequencing")



Por lo que se puede ver, durante la última década ambos términos de manera independiente han sufrido un crecimiento bastante elevado (**Gráficas 1 y 2**). En el caso de la evolución de *Forensic Genetics*, se puede apreciar un aumento del doble en el número de publicaciones entre los años 2008 y 2014, siendo su máximo en el año 2016 y descendiendo en los siguientes años (**Gráfica 1**). Por otro lado, el número de publicaciones que incluían los términos *Next Generation Sequencing* o *Massive Parallel Sequencing*, ha crecido de manera exponencial durante los últimos 10 años, llegando a su máximo entre 2016 y 2017 y bajando en 2018 (**Gráfica 2**).

También, cabe destacar que la Genética Forense como tal apareció durante el s.XX y es por tanto normal que en 2008 ya hubiese una gran cantidad de artículos publicados relacionados con dicha temática. En cuanto a las técnicas de secuenciación de segunda generación, vieron la luz a principios del s.XXI por lo que el número de publicaciones no era tan elevado en un principio.

Para finalizar, se hizo una representación relacionando ambos términos con el fin de determinar la evolución de la Genética Forense gracias a la introducción de técnicas de secuenciación de segunda generación (**Gráfica 3**). Como se puede ver en un principio, el aumento del número de artículos también ha sido de manera exponencial, aunque se puede apreciar un ligero retraso en comparación con las publicaciones que sólo se relacionaban con la secuenciación masiva. Esto es debido a que las investigaciones relacionadas con la aplicación de estas técnicas en Genética Forense tardaron varios años en poder iniciarse.

5.2. Resultados de la encuesta.

La encuesta realizada se envió a un total de 7 expertos de diferentes instituciones relacionadas con la Genética Forense, éstos serán identificados por sus iniciales. A continuación se incluyen todas las respuestas de cada uno de los entrevistados.

1. ¿Su laboratorio utiliza de manera rutinaria técnicas de secuenciación masiva?

Tabla 3. Respuestas a la pregunta 1.

NOMBRE	INSTITUCIÓN	RESPUESTA
J.F.F.	Universitat de les Illes Balears	No
O.G.F.	Laboratorio Genética Forense, Policía Vasca-Ertzaintza, Departamento de Seguridad, Gobierno Vasco	Sí
L.P.	Laboratorio de ADN de la Comisaría General de Policía Científica	No
M.A.	Instituto de Medicina Legal de Valencia	No
M.P.	Universidad del País Vasco (UPV)	No
C.T.M.	Departamento de Medicina Forense. Universidad de Copenhague	Sí
A.B.	Departamento de Genética. SIMEF (Studio Indagini Mediche E Forensi, Calabria, Italia)	No

2. En el caso de haber respondido “Sí o en casos especiales”:

2.A. ¿En qué situaciones lo utiliza? Puede seleccionar más de una opción.

Tabla 4. Respuestas a la pregunta 2.A.

NOMBRE	RESPUESTA
O.G.F.	- Descripción de las características fenotípicas del individuo a partir de muestra dubitada (SNPs) - SNPs para identificación individual y ancestralidad. - Secuenciación del genoma mitocondrial completo.
C.T.M.	- Relaciones de parentesco (SNPs y AIMs de forma ocasional).

2.B. ¿Qué plataforma y kits utiliza? (Puede nombrarse más de un kit según el marcador a analizar)

Tabla 5. Respuestas a la pregunta 2.B.

NOMBRE	RESPUESTA
O.G.F.	Ion Torrent
C.T.M.	Ion Torrent Identity panel/S5

2.C. ¿Qué ventajas técnicas le ofrecen las NGS ante la PCR-EC? (Puede seleccionar más de una opción).

Tabla 6. Respuestas a la pregunta 2.C.

NOMBRE	RESPUESTA
O.G.F.	Mayor número de lecturas generadas
C.T.M.	Mayor número de lecturas generadas

2.D. ¿Cuáles han sido los principales inconvenientes a la hora de implantar la técnica al laboratorio? (Puede seleccionar más de una opción).

Tabla 7. Respuestas a la pregunta 2.D.

NOMBRE	RESPUESTA
O.G.F.	- El coste económico que supone la implantación de la técnica al laboratorio. - La falta de estandarización/validación.
C.T.M.	- El coste económico que supone la implantación de la técnica al laboratorio. - La falta de estandarización/validación.

3. En el caso de haber respondido “No”:

3.A. ¿Qué desventajas técnicas encuentra? (Puede seleccionar más de una opción).

Tabla 8. Respuestas a la pregunta 3.A.

NOMBRE	RESPUESTA
J.F.F.	- El coste económico que supone la implantación de la técnica al laboratorio. - La formación del personal.
L.P.	- El coste económico que supone la implantación de la técnica al laboratorio.
M.A.	-El coste económico que supone la implantación de la técnica al laboratorio. - La formación del personal. - La cantidad de datos innecesarios/irrelevantes producidos (en el caso de secuenciación de genomas completos). - La falta de estandarización/validación. -La falta de nomenclatura alélica.
M.P.	- El coste económico que supone la implantación de la técnica al laboratorio. - La formación del personal. - La mayor cantidad de DNA de partida que requieren las técnicas de secuenciación frente a la usada en PCR-CE- - La falta de estandarización/validación. - La falta de nomenclatura alélica.
C.T.M.	El coste económico que supone la implantación de la técnica al laboratorio. - La cantidad de datos innecesarios/irrelevantes producidos (en el caso de secuenciación de genomas completos). -La falta de estandarización/validación.
A.B.	- El coste económico que supone la implantación de la técnica al laboratorio., - La mayor cantidad de DNA de partida que requieren las técnicas de secuenciación frente a la usada en PCR-CE. -La falta de estandarización/validación. - La falta de nomenclatura alélica.

4. La introducción de NGS al ámbito rutinario de un laboratorio forense permitiría la creación de perfiles genéticos más refinados, ¿cree usted la pena la sustitución de la PCR-CE teniendo en cuenta los cambios que se deberían realizar en procesos de detección y almacenamiento de datos?

Tabla 9. Respuestas a la pregunta 4.

NOMBRE	RESPUESTA
J.F.F.	Siempre que el laboratorio tenga los recursos, supondrá un avance disponer de nuevas bases de datos más refinadas. Aunque sin la validación necesaria, los protocolos actuales, son más que suficientes para casos forenses simples.
O.G.F.	No.
LP.	Si, merece la pena porque se obtiene más información.
M.A.	Por supuesto es mucho más informativo y con más aún más.
M.P.	Actualmente no, en un futuro es posible que la NGS sustituya a las técnicas actuales.
C.T.M.	Diría que no en general, pero depende de la complejidad del caso.
A.B.	Sí, pero el cambio de CE a NGS en el ámbito rutinario necesita tiempo.

5. Aunque las técnicas de secuenciación de segunda generación permitan la secuenciación de genomas completos, ¿cree que la secuenciación dirigida es una herramienta más útil en el ámbito de la Genética Forense?

Tabla 10. Respuestas a la pregunta 5.

NOMBRE	RESPUESTA
J.F.F.	Probablemente, ya que no tendremos tanta información "innecesaria".
O.G.F.	Sí.
LP.	La secuenciación de genomas completos es controvertida en el ámbito penal porque puede invadir la privacidad del individuo.
M.A.	Siempre. Pues necesitamos información identificadora no con valor clínico.
M.P.	Sí
C.T.M.	Sí, creo que es útil.
A.B.	Sí

6. ¿Se debería invertir más en el desarrollo de los denominados "*personal sequencers*" o en las plataformas de secuenciación de tercera generación? ¿Por qué?

Tabla 11. Respuestas a la pregunta 6.

NOMBRE	RESPUESTA
J.F.F.	Personalmente, primero tendríamos que tener validados e introducidos en rutina los de segunda generación, antes de ir más allá. Las plataformas de tercera generación hoy por hoy tienen como principal objetivo la obtención de secuencias muy largas, cosa que en las muestras forenses, no solemos poder obtener, y el factor limitante no es la técnica, sino la misma muestra.
O.G.F.	Sí. Es una tecnología con gran futuro.
L.P.	Creo que aún queda un largo camino en el ámbito de la genética forense para considerar el uso de <i>personal genomers</i> .
M.A.	En personal. Han de conocer técnica y saber interpretar. Las plataformas se pueden compartir y bajarán su coste previsiblemente.
M.P.	Sí, en el área de la secuenciación masiva paralela (MPS) dirigida específicamente al área de trabajo, en concreto en Genética Forense.
C.T.M.	No se si hay que invertir "más". Creo que el desarrollo suficientemente rápido. Hay que invertir en otros aspectos, como los procesos de estandarización y técnicas para un mejor almacenamiento de gran volumen de datos.
A.B.	Plataformas de secuenciación de bajo costo.

7. ¿Se debería crear una base de datos en la nube con acceso a los laboratorios forenses para que cada uno de los laboratorios pudiese introducir los nuevos alelos descubiertos en los marcadores?

Tabla 12. Respuestas a la pregunta 7.

NOMBRE	RESPUESTA
J.F.F.	Evidentemente, si se incluye en rutina la utilización de NGS, lo primero que tendremos que tener son bases de datos representativas.
O.G.F.	No
L.P.	Definitivamente sí.
M.A.	Es buena idea, pero siempre con controles restrictivos de calidad previos a la introducción de datos.
M.P.	Sí
C.T.M.	Sí
A.B.	Sí

8. Con las NGS se pueden detectar la susceptibilidad de un individuo a padecer cierta enfermedad ¿Cree que en algún caso sería conveniente la comunicación del resultado?

Tabla 13. Respuestas a la pregunta 8.

NOMBRE	RESPUESTA
J.F.F.	No se debería comunicar nunca el resultado
O.G.F.	No se debería comunicar nunca el resultado
L.P.	Se debería comunicar el resultado en caso del que el individuo lo desee.
M.A.	Se debería comunicar el resultado en caso del que el individuo lo desee.
M.P.	Se debería comunicar el resultado en caso del que el individuo lo desee.
C.T.M.	No se debería comunicar nunca el resultado.
A.B.	Se debería comunicar el resultado siempre.

9. Anote aquí cualquier aclaración que desee realizar.

Tabla 14. Respuesta pregunta 9.

NOMBRE	RESPUESTA
J.F.F.	En cuanto a la comunicación de resultados, en forense, ninguna información que no haya sido solicitada por el juez, debe en principio ser revelada.
M.A.	El uso de NGS planteará dilemas técnicos éticos profesionales y jurídicos en menos de 3 años.
M.P.	La MPS en el ámbito de la Genética Forense es ya una realidad. Sin embargo, de momento, su aplicación es limitada, entre otras cosas porque sólo hay un desarrollo comercial que es fácil de implementar en los laboratorios forenses pero no cubre todos los marcadores que son de interés.
C.T.M.	La encuesta parece ir muy enfocada a secuenciación de STRs, cuando se trabaja con pruebas de parentesco. No siempre es así. Nuestro laboratorio trabaja, de momento, con SNPs (y STRs-CE). Sólo quería hacer esa observación.

De los 7 expertos entrevistados, sólo 2 usan de manera rutinaria NGS en su laboratorio. Principalmente aplican esta metodología en el estudio de SNPs para identificación fenotípica, individual y ancestral y la secuenciación del genoma mitocondrial. Ambos utilizan la plataforma *Ion Torrent PGM*, y en un caso concreto, el kit *HID-Ion AmpliSeq™ Identity Panel for Human Identification*, enfocado al estudio de SNPs para la identificación de individuos¹⁷.

Además, estos dos entrevistados coinciden en que la gran ventaja de las técnicas de secuenciación masiva son el mayor número de lecturas generadas por estas plataformas pero que el coste y la falta de estandarización/validación son el mayor impedimento que han encontrado al haberlas implantado en su laboratorio. Estos dos argumentos a su vez son los más utilizados por el resto de expertos que no utilizan las NGS dentro de su laboratorio. Ellos, por su parte, también han afirmado que otras de las razones por las cuales no se han llegado a introducir estos métodos en sus laboratorios son la necesidad de formar al personal, la mayor cantidad de DNA de partida que requiere la secuenciación además del exceso de datos producidos por dichas técnicas y la falta de nomenclatura alélica.

A la hora de responder sobre si en el futuro las técnicas de NGS podrían sustituir a la PCR-CE como método estándar para la creación de perfiles genéticos, sólo uno respondió "No" de forma rotunda. El resto coincidía en que las tecnologías más novedosas y todo lo que sirviera para la obtención de resultados más refinados sería bienvenido pero que no es algo que crean que puede llegar a ocurrir a corto plazo.

En cuanto al tipo de secuenciación a aplicar, todos afirman que una secuenciación de tipo dirigida favorece mejor el análisis de muestras forenses, ya que te permite elegir zonas específicas del genoma para secuenciar y así obviar datos en exceso o innecesarios¹⁷. También se les preguntó sobre el uso de los *personal sequencers*, secuenciadores de segunda generación con una capacidad operativa menor²³ o el paso directo a técnicas de secuenciación de tercera generación, pregunta en la cual los encuestados tienen claro que antes de intentar aplicar cualquier técnica en el ámbito de la

Genética Forense, primero se debería conseguir establecer una metodología estandarizada y validada.

Con respecto a la validación y estandarización, a la posibilidad de crear una base de datos *on-line* de carácter internacional, todos los expertos menos uno coincidieron en que sería una gran herramienta para la obtención de datos de referencia de carácter global, pero que se tenía que tener especial cuidado con los resultados introducidos y quién los introduce.

Para acabar, se les preguntó por aquellos datos “innecesarios” que podrían suponer la obtención de información relacionada con la salud de los individuos de los cuales proceden las muestras y cuál sería la acción más correcta a la hora de comunicar la predisposición de estos a padecer cierto tipo de enfermedad. En este caso las respuestas han sido ambiguas, con 2 respuestas en contra de la comunicación, otra en contra, pero con la posibilidad de comunicarlo si el juez así lo pide, 3 a favor siempre que el paciente lo desee y otra a favor, en cualquier caso.

Como conclusiones, los expertos han remarcado que la aplicación actual de las NGS es limitada, principalmente por la falta de una metodología común para todos los laboratorios, pero que en el futuro si se desarrolla de forma completa, podría llegar a ser una técnica rutinaria. Además, cabe destacar que en septiembre de 2017 el Grupo de Habla Española y Portuguesa de la *International Society for Forensic Genetics* (GHEP-ISFG) realizó una encuesta relacionada con las aplicaciones de la secuenciación masiva en la Genética Forense a laboratorios de América Latina y Europa en la cual se llegó a conclusiones similares que en la encuesta realizada en este trabajo²⁷.

6. ESTADO ACTUAL DE LA SECUENCIACIÓN MASIVA EN GENÉTICA FORENSE

6.1 Ventajas e inconvenientes de la secuenciación masiva.

La creación de tecnologías de segunda generación pretende mejorar 3 aspectos importantes con respecto a técnicas anteriores: la preparación de librerías de DNA mediante procesos que no requieran el uso de células (clonación bacteriana), producción de reacciones en paralelo y detección del producto final sin necesidad de electroforesis²⁸. Además, han superado con creces el rendimiento de sus predecesoras siendo capaces de llegar a los 6 billones de lecturas por vez, en contraste con los 30 millones de la secuenciación Sanger²⁹.

El rendimiento de las técnicas de secuenciación se determina con la cantidad de datos que pueden producir: la profundidad de la secuenciación o cobertura (*coverage*) es el promedio de veces que cada base en un genoma es secuenciada. La extensión de la secuenciación es el porcentaje del genoma que puede ser cubierto durante el proceso de lectura²³. Ambos términos son inversamente proporcionales, una menor cantidad de genoma a leer (mayor extensión) tendría como resultado una menor cobertura²¹.

Además, la estrategia de secuenciación puede dividirse en dos si se tiene en cuenta la selección previa de secuencias diana: si no se realiza ninguna selección previa y se amplifican fragmentos del genoma aleatorios, tendrá lugar una secuenciación al azar o *shotgun sequencing*. Ésta ha sido muy utilizada para secuenciar genomas completos por lo que ha supuesto una revolución en el campo de la investigación biomédica²¹.

Por otro lado, si se seleccionan secuencias diana previas, se está ante una secuenciación dirigida o *target sequencing*²¹. Será pues la que se utilice más en el ámbito de Genética Forense ya que sus estudios se basan principalmente en el análisis de conjuntos de marcadores y no de genomas completos. Además, el estudio sólo de ciertas zonas de interés da lugar a una extensión mucho menor que si se quisiera secuenciar el genoma completo, y por tanto daría resultados con mayor cobertura²¹.

6.1.1. Elaboración de perfiles genéticos.

El *golden standard* actual a la hora de analizar el material genético y generar perfiles genéticos es la técnica PCR-CE³⁰. Al igual que en las NGS, existen pasos de amplificación del material genético, por lo que las dos técnicas presentan errores técnicos característicos como el patinaje de la polimerasa³⁰.

A pesar de esta similitud, la competencia entre ambas técnicas reside en el tipo de información que pueden dar cuando se crean los perfiles genéticos³¹. En el caso de la PCR-CE, los fragmentos se distinguen por su longitud, la cual cosa tiene poder de discriminación suficiente para muchas aplicaciones forenses³¹. Sin embargo, las NGS permitirían diferenciar alelos tanto por longitud como por secuencia de bases, produciendo un refinamiento de los perfiles ya establecidos y propiciando el descubrimiento de nuevos polimorfismos³¹.

Por otro lado, es necesario considerar el tiempo que se tarda en obtener los resultados del análisis, ya que es de vital importancia en el ámbito de la Genética Forense¹⁷. Se requieren buenos resultados en el menor tiempo posible y los 2-3 días laborables que dura el proceso de las NGS no parecen ofrecer ninguna mejora frente al único día necesario para obtener resultados mediante PCR-CE¹⁷. No obstante, las NGS presentan una ventaja frente al actual *golden standard* y es su mayor capacidad de multiplexación, es decir, el análisis de más marcadores de forma simultánea¹⁷.

6.1.2. Muestras en Genética Forense.

Las evidencias obtenidas en escenas del crimen suelen presentar un alto grado de contaminación o degradación debido a la exposición a factores ambientales como humedad, calor, radiación UV y contaminaciones de diversos tipos³².

El material necesario para la obtención de resultados reproducibles y fiables por PCR-CE es de 0,2-2ng, sin embargo, para las NGS se necesita de media más cantidad, entre 10ng-1ug²⁸. A pesar de parecer que se encuentran en desventaja, un punto a favor de las NGS es que no requieren de electroforesis, y por tanto quedan eliminados pasos de purificación de la columna que son una gran fuente de pérdida de material²⁸.

Otro aspecto a tener en cuenta es que muchas de las muestras contienen mezclas de DNA³³. Los errores de amplificación pueden llegar a enmascarar material de contribuidores menores, por lo que es un problema en ambas técnicas. Sin embargo, a la hora de interpretar los resultados, las NGS permiten una mejor resolución de mezclas ya que la técnica PCR-CE únicamente puede distinguir alelos a nivel de longitud mientras que la secuenciación de segunda generación permite además inferir a nivel nucleotídico.

6.1.3. Estudio de SNPs.

Los SNPs son marcadores bi-alélicos, por lo que presentan una menor variabilidad que los STRs. Esto supone que, si se quieren estudiar para establecer perfiles genéticos, se deben utilizar un mayor número de marcadores³⁴. Por el contrario, esta menor variabilidad facilita el establecimiento de relaciones de parentesco cercanas³⁴.

El estudio de conjuntos de SNPs determinados, puede relacionarse con la determinación de la etnia y el linaje ancestral de los individuos³⁵. Además, muchos SNPs están relacionados con genes de pigmentación u otras características relacionadas con la apariencia y por tanto son una buena herramienta para inferir en las características físicas de un individuo³⁶.

Por otra parte, el pequeño tamaño de sus amplicones (150pb aprox.) los convierten en marcadores ideales a la hora de analizar muestras muy degradadas o contaminadas³². El uso de NGS mejoraría el estudio de estos marcadores ya que facilitaría la genotipación simultánea de un gran número de SNPs, la distinción de alelos e impulsaría el descubrimiento de nuevas mutaciones³⁵.

6.1.4. Epigenética, transcriptoma, mtDNA y Medicina Forense.

Las marcas epigenéticas permiten la regulación de la expresión génica, la cual va variando según el tipo celular y el tiempo y en el ámbito forense permite la diferenciación de gemelos monocigóticos, la dilucidación de la edad del individuo o tipo de muestra que se está analizando. Las NGS facilitarían el análisis de este tipo de marcaje³⁷.

La diferenciación de gemelos monocigóticos mediante huella de DNA es un gran reto debido a que éstos tienen el mismo perfil de microsatélites³⁸. No obstante, a partir del nacimiento, los gemelos comienzan a sufrir deriva epigenética, por la cual las marcas químicas establecidas en su genoma variarán según el tiempo de forma diferente en cada uno de ellos³⁹.

Por otra parte, la determinación del tejido de origen de una muestra se realiza mediante análisis de la expresión del mRNA utilizando PCR cuantitativas o la medida del grado de metilación del DNA mediante el método del bisulfito o tratamiento con endonucleasas⁴⁰. Sin embargo, son altamente susceptibles a la contaminación ambiental, por lo que la introducción del NGS para el estudio de marcas epigenéticas serviría para reducir las tasas de errores producidos⁴⁰. Además, la secuenciación del transcriptoma permite de forma más eficaz la distinción de isoformas y poderlas relacionar con el tipo celular del que provienen⁴⁰.

Cómo ya se ha establecido, el genoma mitocondrial es una buena fuente de análisis para muestras en las cuales el material genético nuclear no se encuentra disponible⁴¹. El principal obstáculo del uso de mtDNA es la presencia de heteroplasmia, es decir, la coexistencia dos o más tipos mitocondriales por célula o en el individuo⁴². Esto podría llevar a errores de identificación por diferencias de coincidencia entre dos indicios biológicos de distinto tejido, pero de la misma persona⁴¹. Por suerte, las NGS han contribuido de forma muy positiva en la detección y distinción de la heteroplasmia⁴².

Otra aplicación de la secuenciación masiva en el campo de la Medicina Forense es el estudio de marcadores genéticos que facilitan el *screening* de genes implicados en diferentes enfermedades, permitiendo un diagnóstico prematuro, y los relacionados con casos de muerte súbita⁴³.

6.1.5. Inconvenientes.

Las NGS son técnicas relativamente nuevas y presentan una serie de problemas que deben solucionarse antes de poder ser aplicadas al ámbito rutinario de un laboratorio. Lo primero es el coste extra que implica la maquinaria y el entrenamiento requerido a los futuros usuarios de la plataforma⁴⁴.

En el caso del ámbito forense, la gran cantidad de datos que se obtienen generarían costes añadidos para su interpretación. Es por ello, que a partir de 2010 se empezaron a distribuir tecnologías con menor capacidad, los *personal sequencers: Genome Sequencer Junior System (Roche 454 Technologies)* y *MiSeq (Illumina)*²³.

Las lecturas generadas por NGS rondan los 100-400pb⁴⁵, lo que provoca problemas a la hora de analizar STRs de forma completa y dificulta el ensamblaje de genomas *de novo* ya que no se pueden abarcar grandes estructuras repetitivas del genoma y por tanto muchas quedarían obviadas⁴⁶. Además, las secuencias obtenidas son más costosas de analizar y almacenar que los datos producidos por PCR-CE, por lo que se requiere de una actualización en equipos bioinformáticos⁴⁶.

El *software* diseñado para las técnicas de secuenciación debe de tener una serie de objetivos principales: ser capaz de detectar de forma correcta la secuencia y reducir el ruido de fondo generado por errores durante los diferentes pasos del proceso de secuenciación⁴⁷, ser capaz de realizar un correcto alineamiento de las secuencias y por tanto una buena identificación de éstas, poder distinguir de forma correcta los alelos detectados y ser compatible con las bases de datos ya establecidas⁴⁴. Por otro lado, también se requiere crear un algoritmo capaz de adaptarse a la nueva nomenclatura utilizada para distinguir los diferentes alelos generados que se vayan generando mediante secuenciación⁴⁸.

En el campo de la Genética Forense la estandarización y validación son procesos clave ya que se requiere de una metodología que pueda facilitar el análisis e interpretación de resultados del propio laboratorio y que sean extrapolables y comparables con los de otros⁴⁹. El hecho de que las diferentes plataformas de NGS presenten distintos funcionamientos dan lugar a errores propios de la técnica que, sumados a que cada laboratorio escoge la plataforma que más le convenga, dificultan el establecimiento de una metodología estándar⁴⁴. En la **Tabla 15** quedan reflejadas las principales características de cada plataforma.

Tabla 15. Comparación entre las diferentes plataformas de secuenciación de segunda generación (5.1, 28, 45,84,88)

Empresa	Plataforma y año de lanzamiento	pb por lectura	Rendimiento (per run)	Tiempo	Cantidad de DNA	Ventajas	Desventajas
Roche 454	<i>GS FLX+</i> (2005)	1000	700Mb	23h	500ng	- Lecturas largas - Poco tiempo	- Error en zonas homopoliméricas - Bajo rendimiento
	<i>GS Junior</i> (2009)	1000	70Mb	18h	300ng	- Lecturas largas - Poco tiempo	- Error en zonas homopoliméricas
Illumina	<i>HiSeq2000</i> (2010)	2*150	1.5Tb	7h - 6 días	50ng-1ug	- Buena relación rendimiento/ coste - Mayor cantidad de protocolos estandarizados	- Lecturas cortas - Mucho tiempo
	<i>MiSeq</i> (2011)	2*300	15Gb	4-55h	50ng-1ug	- Fácil de usar - Poco tiempo	-Costoso
Thermo Fisher Scientific	<i>Life Technologies SOLiD 5500xl</i> (2007)	50-75	320Gb	10 días	1-3ug	- Baja tasa de error debido a que cada base se lee dos veces - Buen rendimiento	- Lecturas cortas - Fallos en secuenciación de genomas completos - Mucho tiempo - Poco usado (menos kits comerciales)
	<i>Ion Torrent PGM Ion 316 chip</i> (2010)	200-400	2Gb	4-7.3h	10ng	- Poco tiempo - Barato - Tecnología semi-conductora descarta el uso de detectores ópticos o fluorescentes	- Validación no consolidada - Gran tasa de error en zonas homopoliméricas

6.1.6 Resumen.

Las técnicas de secuenciación de segunda generación facilitarían el análisis de los diferentes marcadores de DNA y proporcionarían perfiles genéticos más refinados para la identificación de individuos. No obstante, aún se tendrán que enfrentar a una serie de retos como la falta de nomenclatura y compatibilidad con bases de datos ya existentes, la creación de *softwares* que puedan tratar almacenar el gran número de datos producidos de manera eficaz y el establecimiento de parámetros estándar para la técnica. Todas las ventajas e inconvenientes antes nombrados se resumen en la **Tabla 16**.

Tabla 16. Resumen de las ventajas e inconvenientes de PCR-CE y NGS.

TÉCNICA	VENTAJAS	INCONVENIENTES
PCR-CE	<ul style="list-style-type: none"> ✓ Validada y estandarizada. ✓ Resultados al mismo día o al siguiente. ✓ Bases de datos establecidas. ✓ Resultados aceptables en un tribunal de justicia. 	<ul style="list-style-type: none"> ✗ Diferencias entre fragmentos sólo por longitud. ✗ Menor resolución a la hora de trabajar con muchas muestras a la vez. ✗ Dificultad a la hora de analizar muestras degradadas o mixtas. ✗ Preparación compleja de la muestra.
NGS	<ul style="list-style-type: none"> ✓ Sensibilidad de nanogramos de muestra. ✓ Diferencias entre fragmentos por longitud y secuencia. ✓ Mejor capacidad para trabajar con varias muestras a la vez. ✓ Capacidad para analizar muestras degradadas o mixtas. ✓ Estudio simultáneo de marcadores de STR, SNPs, InDels, etc. ✓ Diferenciación de gemelos monocigóticos: Estudio de marcas epigenéticas. 	<ul style="list-style-type: none"> ✗ Dificultad a la hora de estandarizar. ✗ Ausencia de bases de datos. ✗ Errores técnicos específicos de cada plataforma. ✗ Mayor tiempo para obtener resultados: 2-3 días laborales. ✗ Dificultad a la hora de ensamblaje <i>de novo</i> del genoma. ✗ La preparación de librerías es costosa. ✗ Dificultad para secuenciar zonas homopoliméricas.

6.2. Comercialización y validación.

6.2.1. Kits para NGS.

Las técnicas de secuenciación contienen un paso de generación de librerías, o *clusters* del fragmento de DNA a secuenciar que serán anclados a superficies sólidas donde tendrá lugar la secuenciación¹⁸. Para la creación de librerías existen *kits* comerciales destinados al uso forense. Éstos suelen combinarse según la plataforma a utilizar, siendo común que las mismas casas comerciales que distribuyen los secuenciadores también comercialicen *kits*⁴⁴. En la **Tabla 17**, se recopilan las combinaciones de *kits* y plataformas más usadas en forense.

Tabla 17. Combinaciones de kits y plataformas para el análisis forense.

Plataforma	Kit para la preparación de librería	Marcadores	Kit de secuenciación
<p>Illumina MiSeq</p> <p>Illumina</p>	<p><i>Forenseq™ DNA Signature Prep Kit</i>⁵⁰</p> <p><i>Nextera XT DNA Library Prep Kit</i>⁵¹</p> <p><i>Nextera™ Rapid Capture Custom Enrichment Kit</i>⁵²</p>	<p>STRs: 29 STRs autosómicos, 9 X-STRs y 25 Y-STRs SNPs: 95 identificación, 56 linaje, 24 fenotipado. Amelogenina</p> <p>Genoma mitocondrial</p> <p>284 SNPs + InDels</p>	<p>MiSeq™ Reagent Kit⁵⁰</p>
<p>Ion Torrent PGM</p>	<p><i>HID-Ion AmpliSeq™ Identity Panel for Human Identification</i>⁵³</p> <p><i>HID-Ion AmpliSeq™ Ancestry Panel for Ancestry Estimation</i>¹⁷</p>	<p>SNPs: 124 autosómicos y 34 crY</p> <p>SNP: 165 autosómicos</p>	<p>Ion-PGM™ Hi-Q™ View⁵⁵</p>

Thermo Fisher Scientific	<i>Precision ID GlobalFiler™ NGS STR panel v2⁵⁴</i>	21 STRS de CODIS 3 STR del CR Y Amelogenina 9 STR adicionales no CODIS	
---	--	---	--

6.2.2. Validación de la técnica.

El establecimiento de una técnica de manera estándar y la comercialización de *kits* necesitan de un proceso previo de validación. Se trata de la evaluación de una práctica o procedimiento con el fin de determinar su eficacia y certeza para poder implantarse de forma rutinaria. En Europa, dicho proceso corre a cargo del *European Network of Forensic Science Institutes*⁵⁶.

Las metodologías se someten a dos tipos de validación: la externa o *developmental validation*, que se encarga de establecer las condiciones y limitaciones de una metodología o conocer si un marcador es adecuado en el contexto forense⁵⁷. Por otro lado, tenemos la validación interna que se encarga de demostrar que el método en pruebas da los resultados esperados ya que se compara con la técnica establecida y además incluye pruebas que comprueban el funcionamiento correcto del propio laboratorio⁵⁷.

Por tanto, en los estudios de validación tanto de posibles marcadores forenses o de combinaciones de plataformas y *kits* es muy importante tener en cuenta los pasos a realizar durante la técnica, el coste y los resultados obtenidos⁵⁸.

6.3. Software de análisis.

Las tecnologías de secuenciación de segunda generación requieren de programas informáticos cuyos algoritmos sean capaces de analizar y almacenar resultados de manera eficiente, permitiendo a su vez que estos datos puedan ser comparados entre laboratorios. Se debe garantizar una correcta determinación de cada nucleótido de la secuencia, eliminar o disminuir posibles *bias* y detectar patrones de error tanto en datos en bruto como de las secuencias consenso⁵⁹.

También hay que plantear que estas nuevas tecnologías afectarían no sólo a cómo se analizan los datos sino a cómo se deben nombrar los resultados para su correcto almacenamiento¹⁴. En el caso del *software* empleado en técnicas de PCR-CE, los archivos de datos contenían el nombre del locus, la longitud de sus alelos y las intensidades de fluorescencia detectadas por CE³⁰. En cambio, en el caso de la secuenciación lo que se generan son archivos que contienen el orden de bases de la secuencia³⁰.

El análisis de los distintos marcadores de DNA dan como resultado diferentes tipos de información y ya que las características de cada uno de éstos son diferentes, se requieren herramientas bioinformáticas específicas para su estudio. A pesar de ello, existen pasos comunes en el análisis de secuencias: lecturas en bruto (*raw reads*), alineamiento a secuencias y clasificación⁶⁰.

6.3.1. Análisis de datos.

La secuenciación genera lecturas no sólo de la secuencia del marcador sino de sus regiones flanqueantes, *primers* usados en la PCR, *tags* que diferencian las muestras, adaptadores de los fragmentos al soporte sólido y errores generados por fallos de la técnica⁶⁰. Todo ello se engloba en los datos en bruto de la secuenciación que requieren de un *software* que ayuda a interpretar las lecturas generadas y eliminar aquellas innecesarias o erróneas³⁰.

La mayoría de programas los otorga la misma compañía que distribuye las plataformas de secuenciación, destacando *MiSeq Reporter* de *Illumina* y *Torrent Suite Software* de *ThermoFisher*⁴⁴. En estos se incluyen las herramientas de lectura de secuencias (archivos FASTA/FASTQ), alineamiento y almacenamiento. Además, existen programas de corrección de errores por *base calling* como *Integrative Genomics Viewer* o *NextGENe*⁶⁰.

6.3.2. Software de alineamiento, clasificación y almacenamiento.

En el ámbito de la Genética Forense las lecturas se realizan mediante secuenciación por alineamiento en la cual los fragmentos analizados se comparan con los de un genoma de referencia³⁰. El alineamiento de secuencias se realiza mediante programas como *Bowtie 2*, *Novoalign* o *Burrows-Wheeler Aligner*, que originan datos SAM (*Sequence Alignment Map*) y BAM (*Binary Alignment/Map*)⁶¹. Como el proceso de alineamiento puede generar artefactos, también existen programas de re-alineamiento como *Genome Analysis Toolkit* o *FDSTools*⁴⁷.

A continuación se procede a la clasificación de secuencias para determinar las variaciones de éstas y la formación que contienen⁶¹. Los archivos BAM generados anteriormente (o los archivos SAM convertidos a BAM) pasan por programas como *BCFtools*, *GATK* y *FreeBayes* y dan lugar a ficheros VCF (*Variant Calling Format*) donde se encontrarán las variantes genéticas de las secuencias del genoma de referencia⁶¹.

En cuanto al proceso de almacenamiento, actualmente se está estudiado un sistema de almacenamiento en la nube, *Amazon Web Services Cloud*⁶². Esto permitirá a los usuarios poder acceder a los resultados desde cualquier dispositivo a través de Internet y reducir drásticamente el volumen de memoria ocupado por éstos en los ordenadores, facilitando que equipos con una capacidad menor puedan trabajar con NGS⁶².

6.4. Bases de datos y nomenclatura.

El uso de técnicas NGS genera una mayor cantidad de datos y de más complejidad que los que resultan del uso de PCR-CE, es por ello que se necesitaría un cambio en el patrón establecido actualmente a la hora de nombrar cada una de las variaciones de una secuencia y establecer un nuevo algoritmo para almacenarlas⁶⁰.

6.4.1. Bases de datos.

Existen toda una serie de bases de datos de carácter general relacionadas con datos producidos en secuenciación como *Genome/exome Aggregation Databases* (genomAD, ExAC), y *1000 genomes project*⁶⁰. En cuanto al ámbito forense, existen bases específicas para cada uno de los marcadores que se utilizan, siendo en el caso de los STRs la base CODIS de Estados Unidos, en la cual los perfiles genéticos se basan en el análisis de 13 STRs y el gen de la amelogenina y el *European Standard Set* que analiza 15 STRs⁴⁵. Otros ejemplos son *Hirisplex*⁶³ y *Allele Frequency Database* (ALFRED)⁶⁴ para SNPs o *Phlyotree*¹¹ y *EDNAP mtDNA Population Database* (EMPOP)⁶⁵ para mtDNA.

6.4.2. Nomenclatura.

La distinción alélica proporcionada por las técnicas de secuenciación de segunda generación requiere de una nueva nomenclatura para la clasificación de datos obtenidos por la técnica que a su vez sea compatible con la ya establecida y permita listar de forma correcta cada una de las variantes genéticas existentes y por existir³¹.

En la metodología PCR-CE los alelos se nombran según el número de repeticiones totales seguido si es el caso por un punto y el número de nucleótidos sueltos⁶⁶. Así por ejemplo para un individuo con dos alelos 12 en el *locus* D13S317, la nomenclatura de PCR-CE sería D13S317 12-12 que incluiría como descripción del motivo [TATC]12.

La nueva nomenclatura para NGS debe ser sencilla, que se compagine con la anterior y facilite los estudios entre laboratorios. En 2015, K.J. van der Gaag y P. de Knijff⁶⁷ propusieron un tipo de nomenclatura basada en las consideraciones de la ISFG¹⁴, la cual engloba 4 elementos:

- a) El nombre del locus seguido por "CE" y la longitud del fragmento.
- b) Las coordenadas del marcador en el genoma de referencia, en el caso de los humanos el GRCh38.
- c) El motivo o tipo de repetición del que está compuesto el alelo.
- d) El tipo de variación que contiene el alelo. Primero se escriben las coordenadas cromosómicas del alelo, que pueden acortarse con una "x" seguido de los 3 últimos números de la coordenada. A continuación se añade ">" y el tipo de variación que sufre.

Ejemplo usando el *locus* anterior, sería:

D13S317-CE12 - chr13-GRCh38-g.82.148.025:82.148.068 – [TATC]12

En el caso de que por ejemplo hubiese un cambio G>A en posición 82.148.033 en el alelo, la nomenclatura sería:

D13S317-CE12 - chr13-GRCh38-g.82.148.025:82.148.068 – [TATC]12 - x.033G>A

6.5. Consideraciones éticas.

La identificación individual que proporciona el estudio de marcadores de DNA acarrea una serie de cuestiones éticas relacionadas con la privacidad individual, las cuales se tratan en la Declaración de Datos Genéticos Humanos establecida por la conferencia de la UNESCO en octubre de 2003⁶⁸.

La finalidad de esta declaración, es garantizar el respeto y dignidad humana y la protección y privacidad de los datos genéticos (Artículo 1). Ya que estudio del material genético puede dar datos relacionados con rasgos físicos, origen étnico y edad de los individuos, éstos nunca deberán ser usados como elemento de juicio, garantizando la igualdad (Artículo 3).

La creación de perfiles genéticos tiene como fines principales el diagnóstico, investigación o uso en un juicio (Artículo 5). La recolección de material biológico está sujeto al consentimiento expreso (Artículo 6) y así mismo cada uno puede decidir abstenerse a continuar con el proceso y en ese caso sus muestras y todo lo que han generado deben dejar de utilizarse (Artículo 9). Además, los individuos deben tener libre acceso a sus datos y tienen derecho a decidir si se les transmite o no la información relacionada con éstos, como la susceptibilidad a padecer enfermedades o resultados de una investigación (Artículo 10).

En el caso del proceso judicial, sólo se anunciará la compatibilidad o coincidencia del perfil a menos que haya una petición especial del juez. Para acabar, cuando los sospechosos sean señalados como inocentes, las bases de datos deben eliminar de manera automática de las bases los perfiles generados durante la investigación (Artículo 21). En el caso de que el individuo sea declarado culpable, se procede según la legislación de cada país, en España por ejemplo, es el juez el que decide si el perfil genético se incorpora o no a la base de datos según la naturaleza del delito (*Artículo 129 bis introducido por el número setenta del artículo único de la L.O. 1/2015, de 30 de marzo, por la*

que se modifica la L.O. 10/1995, de 23 de noviembre, del Código Penal («B.O.E.» 31 marzo). Vigencia: 1 julio 2015).

Así mismo, es de vital importancia que se controle el acceso a las bases de datos generadas y que se establezcan legislaciones estatales relacionadas con la transferencia de resultados entre naciones, manteniendo siempre la protección de dicha información (Artículos 14 y 18).

7. CONCLUSIONES

El uso de polimorfismos de DNA como marcadores genéticos en el ámbito de la Genética Forense tiene como objetivo la resolución de casos criminalísticos, investigación de personas desaparecidas, pruebas de paternidad e investigación de relaciones de parentesco.

A pesar de que el *golden standard* actual está basado en métodos de PCR-CE y proporcionaría la información necesaria para la correcta identificación individual, durante los últimos años las técnicas de secuenciación de segunda generación han sido consideradas como buenas candidatas para sustituir estos métodos. Las NGS proporcionarían muchísimas ventajas a la hora de trabajar con muestras degradadas o mezclas y permitirían el análisis de más de un tipo de marcador genético a la vez. Se trata de técnicas con mayor rendimiento que promoverían refinamiento de los perfiles genéticos y el estudio de forma directa de marcas epigenéticas.

Muchos de los laboratorios, sin embargo, no utilizan métodos de secuenciación de manera rutinaria debido a que éstas son costosas y no tienen financiación para cubrirlas. Además al ser técnicas bastante recientes, no se encuentran del todo desarrolladas y aún necesitan pasar requisitos de validación y estandarización imprescindibles en Genética Forense.

La falta de nomenclatura y de unos tests y metodologías comunes son requisitos clave para que los diferentes laboratorios forenses puedan trabajar con unos resultados fiables que permitan el análisis adecuado de las muestras sin gran tasa de error y se pueda optar por la comparación de resultados entre diferentes instituciones. Además se debe tener en cuenta que estos datos deben estar regulados bajo una legislación que permita manejarlos de forma correcta respetando los derechos humanos.

8. BIBLIOGRAFÍA

1. Lareu MV. Genética forense 1. Seminarios Universidad de Murcia 2008:1-16. https://www.um.es/biomybiotec/web/Seminarios/2008/papers/MV_Lareu_clase_genetica_forense.pdf.
2. Ford EB. Polymorphism and taxonomy. *Heredity (Edinb)*. 1955;9(2):255-264. doi:10.1038/hdy.1955.24
3. Landsteiner K. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralblatt für Bakteriologie Parasitenkunde und Infektionskrankheiten*. 1900;27:357-362. doi:10.1161/01.RES.25.4.500
4. Mayr W. Die genetik des HLA systems. *Hum Genet*. 1970:195-199.
5. Knight JC. *Human Genetic Diversity: Functional Consequences for Health and Disease*. Oxford University Press; 2009. doi:10.1093/acprof:oso/9780199227693.001.0001
6. Espuny MJF. *Introducción a La Tecnología Del ADN Aplicada En El Laboratorio Forense.*; 2005.
7. Kayser M. Forensic use of Y-chromosome DNA: a general overview. *Hum Genet*. 2017;136(5):621-635. doi:10.1007/s00439-017-1776-9

8. Francès F, Castelló A, Verdú F. El diagnóstico genético del sexo mediante el test de la amelogenina: Métodos y posibles fuentes de error. *Cuad Med Forense*. 2010;14(52):119-125. doi:10.4321/s1135-76062008000200002
9. Phillips C. Forensic genetic analysis of bio-geographical ancestry. *Forensic Sci Int Genet*. 2015;18:49-65. doi:10.1016/j.fsigen.2015.05.012
10. Holland MM, McQuillan MR, O'Hanlon KA. Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. *Croat Med J*. 2011;52(3):299-313. doi:10.3325/cmj.2011.52.299
11. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 2009;30(2):386-394. doi:10.1002/humu.20921
12. Solla LP. *Aplicaciones Forenses Del ADN*; 2004. http://www.cej.justicia.es/cej/html/publicaciones_01.htm.
13. Alonso AA. *Conceptos básicos de ADN forense*; 2004. http://www.cej.justicia.es/cej/html/publicaciones_01.htm.
14. Bodner M, Bastisch I, Butler JM, et al. Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). *Forensic Sci Int Genet*. 2016;24:97-102. doi:10.1016/j.fsigen.2016.06.008
15. Frumkin D, Wasserstrom A, Davidson A, Grafit A. Authentication of forensic DNA samples. *Forensic Sci Int Genet*. 2010;4(2):95-103. doi:10.1016/j.fsigen.2009.06.009
16. Balazs I. Forensic applications. *Curr Opin Biotechnol*. 1992;3(1):18-23. doi:10.1016/0958-1669(92)90120-8
17. Børsting C, Morling N. Next generation sequencing and its applications in forensic genetics. *Forensic Sci Int Genet*. 2015;18:78-89. doi:10.1016/j.fsigen.2015.02.002
18. Tucker T, Marra M, Friedman JM. Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine. *Am J Hum Genet*. 2009;85(2):142-154. doi:10.1016/j.ajhg.2009.06.022
19. Teijón Rivera JM, Garrido Pertierra A. *Fundamentos de Bioquímica Estructural*. Madrid: Editorial Tébar; 2006. https://books.google.es/books?id=avt8LFmp8q4C&printsec=copyright&redir_esc=y#v=onepage&q&f=false. Accessed May 22, 2019.
20. Garrigues F. NGS: Secuenciación de Segunda Generación - Genética Médica Blog. <https://revistageneticamedica.com/blog/ngs-secuenciacion/>. Published 2017.
21. Scheible M, Loreille O, Just R, Irwin J. Short tandem repeat typing on the 454 platform: Strategies and considerations for targeted sequencing of common forensic markers. *Forensic Sci Int Genet*. 2014;12:107-119. doi:10.1016/j.fsigen.2014.04.010
22. Alvarez-Cubero MJ, Saiz M, Martínez-García B, et al. Next generation sequencing: an application in forensic sciences? *Ann Hum Biol*. 2017;44(7):581-592. doi:10.1080/03014460.2017.1375155
23. Berglund EC, Kiiialainen A, Syvanen AC. Next-generation sequencing technologies and applications for human genetic history and forensics. *Investig Genet*. 2011;2:23. doi:2041-2223-2-23 [pii]v10.1186/2041-2223-2-23
24. Xu J. *Next-Generation Sequencing: Current Technologies and Applications*. Ontario: Caister Academic Press; 2014. https://books.google.es/books/about/Next_Generation_Sequencing.html?id=-QZQnQEACAAJ&redir_esc=y. Accessed April 10, 2019.
25. Mikheyev AS, Tin MMY. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour*. 2014;14(6):1097-1102. doi:10.1111/1755-0998.12324

26. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics Bioinforma.* 2016;14(5):265-279. doi:10.1016/j.gpb.2016.05.004
27. Grupo de Habla Española y Portuguesa de la International Society for Forensic Genetic. Resultados de la Encuesta GHEP sobre Aplicaciones Forenses de la Secuenciación Masiva.; 2017. <https://ghep-isfg.org/>.
28. Thermes C. Ten years of next-generation sequencing technology. *Trends Genet.* 2014;30(9):418-426. doi:10.1016/j.tig.2014.07.001
29. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nat Rev Genet.* 2012;13(1):36-46. doi:10.1038/nrg3117
30. de Knijff P. From next generation sequencing to now generation sequencing in forensics. *Forensic Sci Int Genet.* 2019;38:175-180. doi:10.1016/j.fsigen.2018.10.017
31. Parson W, Ballard D, Budowle B, et al. Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Sci Int Genet.* 2016;22:54-63. doi:10.1016/j.fsigen.2016.01.009
32. Gettings KB, Kiesler KM, Vallone PM. Performance of a next generation sequencing SNP assay on degraded DNA. *Forensic Sci Int Genet.* 2015;19:1-9. doi:10.1016/j.fsigen.2015.04.010
33. Budowle B, Onorato AJ, Callaghan TF, et al. Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed dna profiles in forensic casework. *J Forensic Sci.* 2009;54(4):810-821. doi:10.1111/j.1556-4029.2009.01046.x
34. Sobiah R, Syeda RH, Zunaira E, Nageen Z, Maria K, et al. Implications of Targeted Next Generation Sequencing in Forensic Science. *J Forensic Res.* 2018;09(02):1-8. doi:10.4172/2157-7145.1000416
35. Sampson JN, Kidd KK, Kidd JR, Zhao H. Selecting SNPs to Identify Ancestry. *Ann Hum Genet.* 2011;75(4):539-553. doi:10.1111/j.1469-1809.2011.00656.x
36. Kayser M. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Sci Int Genet.* 2015;18:33-48. doi:10.1016/j.fsigen.2015.02.003
37. Bartling CM, Hester ME, Bartz J, Heizer E, Faith SA. Next-generation sequencing approach to epigenetic-based tissue source attribution. *Electrophoresis.* 2014;35(21-22):3096-3101. doi:10.1002/elps.201400087
38. Weber-Lehmann J, Schilling E, Gradl G, Richter DC, Wiehler J, et al. Finding the needle in the haystack: Differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Sci Int Genet.* 2014;9(1):42-46. doi:10.1016/j.fsigen.2013.10.015
39. Martin GM. Epigenetic drift in aging identical twins. *Proc Natl Acad Sci.* 2005;102(30):10413-10414. doi:10.1073/pnas.0504743102
40. Lin MH, Jones DF, Fleming R. Transcriptomic analysis of degraded forensic body fluids. *Forensic Sci Int Genet.* 2015;17:35-42. doi:10.1016/j.fsigen.2015.03.005
41. Bandelt HJ, Salas A. Current Next Generation Sequencing technology may not meet forensic standards. *Forensic Sci Int Genet.* 2012;6(1):143-145. doi:10.1016/j.fsigen.2011.04.004
42. McElhoe JA, Holland MM, Makova KD, et al. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. *Forensic Sci Int Genet.* 2014;13:20-29. doi:10.1016/j.fsigen.2014.05.007
43. Brion M, Blanco-Verea A, Sobrino B, et al. Next generation sequencing challenges in the analysis of cardiac sudden death due to arrhythmogenic disorders. *Electrophoresis.* 2014;35(21-22):3111-3116. doi:10.1002/elps.201400148
44. Alonso A, Barrio PA, Müller P, et al. Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis.* 2018;39(21):2655-2668. doi:10.1002/elps.201800030

45. Phillips C, Gelabert-Besada M, Fernandez-Formoso L, et al. "New turns from old STaRs": Enhancing the capabilities of forensic short tandem repeat analysis. *Electrophoresis*. 2014;35(21-22):3173-3187. doi:10.1002/elps.201400095
46. Magi A, Benelli M, Gozzini A, Girolami F, Torricelli F, et al. Bioinformatics for next generation sequencing data. *Genes (Basel)*. 2010;1(2):294-307. doi:10.3390/genes1020294
47. Hoogenboom J, van der Gaag KJ, de Leeuw RH, Sijen T, de Knijff P, et al. FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. *Forensic Sci Int Genet*. 2017;27:27-40. doi:10.1016/j.fsigen.2016.11.007
48. Gelardi C, Rockenbauer E, Dalsgaard S, Børsting C, Morling N. Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles. *Forensic Sci Int Genet*. 2014;12:38-41. doi:10.1016/j.fsigen.2014.04.016
49. Sharma V, Chow HY, Siegel D, Wurmbach E. Qualitative and quantitative assessment of Illumina's forensic STR and SNP kits on MiSeq FGx™. *PLoS One*. 2017;12(11):1-21. doi:10.1371/journal.pone.0187932
50. Bruand J, Pham N, Pond SJK, et al. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. *Forensic Sci Int Genet*. 2017;28:52-70. doi:10.1016/j.fsigen.2017.01.011
51. Peck MA, Sturk-Andreaggi K, Thomas JT, Oliver RS, Barritt-Ross S, et al. Developmental validation of a Nextera XT mitogenome Illumina MiSeq sequencing method for high-quality samples. *Forensic Sci Int Genet*. 2018;34:25-36. doi:10.1016/j.fsigen.2018.01.004
52. Wendt FR, Warshauer DH, Zeng X, et al. Massively parallel sequencing of 68 insertion/deletion markers identifies novel microhaplotypes for utility in human identity testing. *Forensic Sci Int Genet*. 2016;25:198-209. doi:10.1016/j.fsigen.2016.09.005
53. Apaga DLT, Dennis SE, Salvador JM, Calacal GC, De Ungria MCA. Comparison of Two Massively Parallel Sequencing Platforms using 83 Single Nucleotide Polymorphisms for Human Identification. *Sci Rep*. 2017;7(1):1-6. doi:10.1038/s41598-017-00510-3
54. Li H, Zhao X, Ma K, et al. Applying massively parallel sequencing to paternity testing on the Ion Torrent Personal Genome Machine. *Forensic Sci Int Genet*. 2017;31(July):155-159. doi:10.1016/j.fsigen.2017.09.007
55. Zhang S, Niu Y, Bian Y, et al. Sequence investigation of 34 forensic autosomal STRs with massively parallel sequencing. *Sci Rep*. 2018;8(1):1-7. doi:10.1038/s41598-018-24495-9
56. European Network of forensic science institutes. Guidelines for the single laboratory: Validation of Instrumental and Human Based Methods in Forensic Science. 2014;(001):1-31. <http://enfsi.eu/wp-content/uploads/2017/06/Guidance-QCC-VAL-002.pdf>.
57. Scientific Working Group on DNA Analysis Methods. Validation Guidelines for DNA Analysis Methods. 2016;(December 2016):1-15. https://docs.wixstatic.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf.
58. Parson W, Strobl C, Huber G, et al. Reprint of: Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). *Forensic Sci Int Genet*. 2013;7(6):632-639. doi:10.1016/j.fsigen.2013.09.007
59. Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, et al. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res*. 2015;43(6):1-16. doi:10.1093/nar/gku1341
60. Liu YY, Harbison SA. A review of bioinformatic methods for forensic DNA analyses. *Forensic Sci Int Genet*. 2018;33(November 2017):117-128. doi:10.1016/j.fsigen.2017.12.005

61. Roy S, Coldren C, Karunamurthy A, et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagnostics*. 2018;20(1):4-27. doi:10.1016/j.jmoldx.2017.11.003
62. Bailey SF, Scheible MK, Williams C, et al. Secure and robust cloud computing for high-throughput forensic microsatellite sequence analysis and databasing. *Forensic Sci Int Genet*. 2017;31:40-47. doi:10.1016/j.fsigen.2017.08.008
63. Chaitanya L, Breslin K, Zuñiga S, et al. The HirisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Sci Int Genet*. 2018;35:123-135. doi:10.1016/j.fsigen.2018.04.004
64. Osier MV. ALFRED: an allele frequency database for diverse populations and DNA polymorphisms--an update. *Nucleic Acids Res*. 2002;29(1):317-319. doi:10.1093/nar/29.1.317
65. Institute of Legal Medicine Innsbruck Medical University. EDNAP MtDNA Population Database. https://gerichtsmedizin.at/empop_database.html. Accessed April 10, 2019.
66. Gettings KB, Aponte RA, Vallone PM, Butler JM. *STR Allele Sequence Variation: Current Knowledge and Future Issues*. Vol 18. Elsevier Ireland Ltd; 2015. doi:10.1016/j.fsigen.2015.06.005
67. van der Gaag KJ, de Knijff P. Forensic nomenclature for short tandem repeats updated for sequencing. *Forensic Sci Int Genet Suppl Ser*. 2015;5:e542-e544. doi:10.1016/j.fsigss.2015.09.214
68. UNESCO. *Declaración Internacional Sobre Los Datos Genéticos Humanos.*; 2008:1-11 http://portal.unesco.org/es/ev.php-URL_ID=17720&URL_DO=DO_TOPIC&URL_SECTION=201.html.