



# SECUENCIADORES DE NUEVA GENERACIÓN

El tricorder a nuestro alcance

## en ocasiones los biólogos

*no pueden identificar los organismos con total precisión en el campo, bien por falta de tiempo, de medios o por alguna dificultad propia del grupo de estudio en cuestión. No han sido pocas las ocasiones en las que he oído a especialistas bromear en esta tesitura sobre la idea de tener algún tipo de dispositivo que, por ejemplo, pudiese secuenciar una parte del genoma del organismo en cuestión de minutos para dar una respuesta rápida a su identificación, un aparatito portentoso al más puro estilo del “tricorder” que Spock llevaba en Star Trek. Hasta hace unos años, la consideración sería de una tecnología capaz de hacer algo así me parecía tecnológicamente impensable teniendo en cuenta la dirección hacia la que estaba evolucionando la tecnología y las propias limitaciones de un objetivo tan ambicioso; sin embargo los espectaculares cambios que ha experimentado y continúa haciendo la industria de la secuenciación genética nos obliga a reconsiderar esa opinión.*

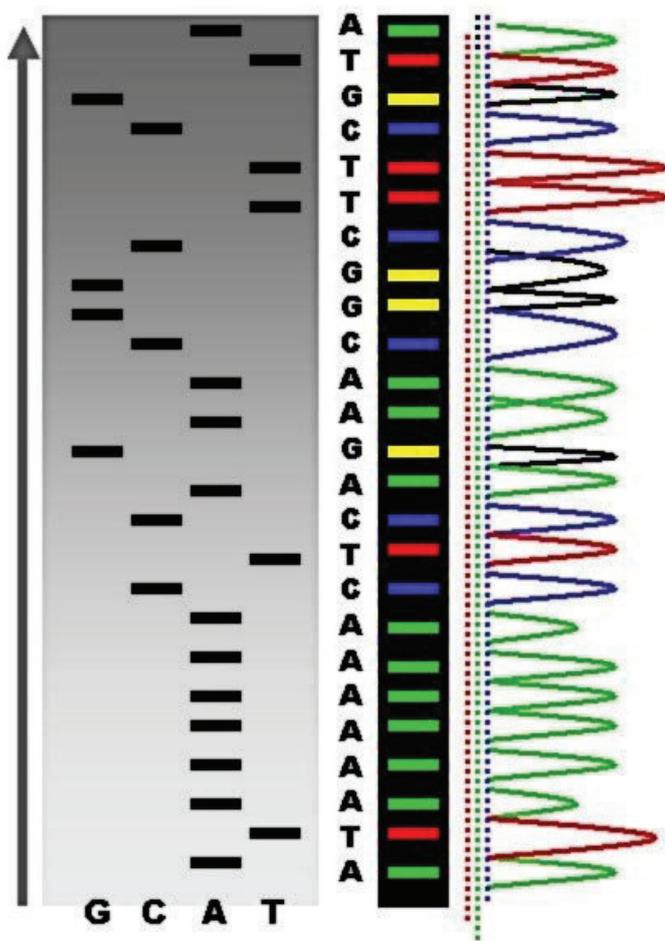
**Se cumplen sesenta años** desde el descubrimiento de la estructura en doble hélice del ADN. Sesenta años no es tanto tiempo, no ya a escala geológica, sino incluso en términos meramente históricos. Sin embargo todos estamos de acuerdo en que los avances de la ciencia durante las últimas décadas se han sucedido a un ritmo especialmente vertiginoso. Por eso sorprende mirar con algo de perspectiva cómo, prácticamente desde el desconocimiento absoluto (no ya de la estructura del material genético, sino de su más básico funcionamiento) hemos sido capaces de desarrollar tecnologías que nos permiten secuenciar genomas enteros en sólo unos días y con un coste más que reducido. ¿Dónde está el límite? La idea

de un aparato portátil con el que poder secuenciar fragmentos del genoma en el momento ya no parece sólo un elemento de la ciencia-ficción.

**Mencionar el hito del descubrimiento** de la estructura del ADN no es sólo una referencia tópica. La peculiar doble hélice y su casi autoexplicativo sistema de replicación nos hicieron ver que la información genética está contenida en una cadena lineal, en una secuencia unidimensional, como si fuese un texto escrito. Como los distintos nucleótidos (las “letras” de ese texto que es la secuencia) son siempre químicamente iguales, el único conocimiento que importa es el de esa estructura primaria, el orden y disposición de los nucleótidos. Ser capaces de leer y de reproducir esa secuencia abrió las puertas de la biotecnología y propició una revolución sin precedentes en campos como la sistemática y la medicina.

**Sin intención de realizar** un repaso exhaustivo de los distintos métodos de secuenciación que han existido y existen, creo que es un ejercicio interesante comparar algunos de los más relevantes para entender las diferencias, no sólo tecnológicas, sino también de planteamiento que tienen unas y otras.

**El método Sanger**, desarrollado en los años 70, se convirtió en poco tiempo en un ingenioso sistema para poder leer la secuencia de un fragmento dado de ADN, y se sigue utilizando con frecuencia con multitud de aplicaciones diferentes. Este sistema suele necesitar como punto de partida una solución muy pura del fragmento de ADN que se quiera secuenciar, por lo que es necesario una amplificación previa, que puede



Principio de secuenciación por el método Sanger identificando el nucleótido terminal capado de los fragmentos ordenados por tamaños.

la replicación Sanger resultará en una sopa de copias parciales de esa secuencia finalizadas con un nucleótido capado. Basta con ordenar esos fragmentos por su longitud (normalmente mediante una electroforesis en un gel capilar) para recuperar la información de la naturaleza y orden de cada uno de los nucleótidos de la secuencia. La visualización habitual del método Sanger son esos inconfundibles cromatogramas que permiten trabajar directamente con la secuencia individual.

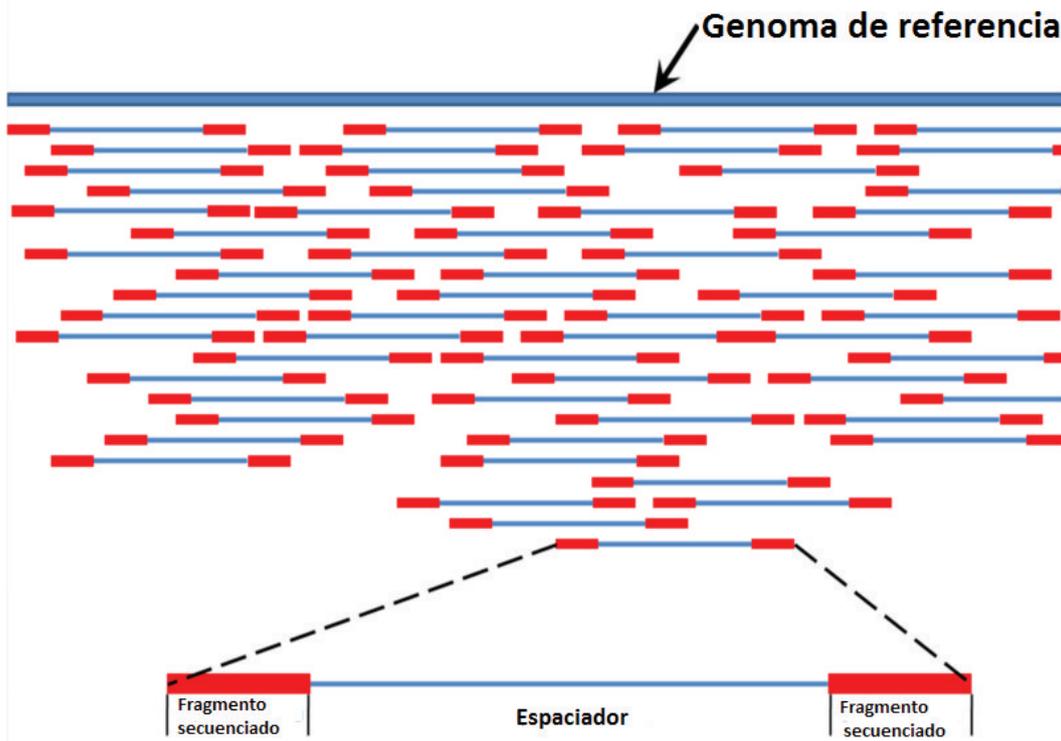
*Sin embargo, la inmensidad de los genomas* puso de manifiesto en un momento dado que este tipo de aproximaciones resultan muy ineficientes si lo que se quiere es conseguir toda la información posible, y en un momento dado, secuenciar genomas completos, y no sólo de los organismos modelo. El volumen de una reacción Sanger puede parecerse diminuto, pero 25 microlitros es una inmensidad de espacio dedicado sólo a una secuencia de unos pocos centenares de pares de bases, y aunque un secuenciador de este tipo puede trabajar con varias soluciones a la vez, seguía quedándose pequeño.

*La respuesta a esta necesidad* la ha dado una explosión de las denominadas técnicas de nueva generación (NGS, en inglés *Next Generation Sequencing*). Cuando conocemos los detalles de estas nuevas técnicas descubrimos que algunos principios fundamentales (la replicación por PCR, el uso de nucleótidos capados fluorescentes, e incluso la longitud de los fragmentos a secuenciar) siguen siendo los mismos, y que lo que en realidad permite obtener un volumen asombroso de información genética es el desarrollo de la nanotecnología (la manipulación casi a nivel individual de fragmentos de ADN) y del tratamiento informático de dicha información.

*Tomemos por ejemplo* una de esas técnicas, la denominada **pirosecuenciación**. En ella no se parte de una solución muy pura de un único fragmento de ADN, sino que se pueden leer simultáneamente miles de secuencias a partir de una biblioteca de ADN (un conjunto de cadenas cortas que normalmente se obtiene tras la fragmentación controlada del genoma de un organismo). Los fragmentos de esta biblioteca pueden ser de unos pocos cientos de nucleótidos de largo, nada significativamente largo respecto a Sanger, pero el objetivo es hacer que cada una de estas

hacerse por clonación de fragmentos genéticos o, más frecuentemente, por PCR. En cualquiera de los dos casos, es necesaria una preparación previa de ese sustrato, de ese conjunto de millones y millones de copias de una misma región, contenidas generalmente en el volumen que ocupa un grano de arroz.

*El método Sanger* propiamente dicho consiste en utilizar esta secuencia como molde y completar la cadena complementaria usando una mezcla de nucleótidos normales y nucleótidos “capados” (con una ligera modificación química que impide a la enzima replicadora continuar su trabajo una vez ha incorporado ese nucleótido concreto). Como hay innumerables copias de la misma secuencia molde,



Los métodos de secuenciación de última generación suelen producir millones de pequeños fragmentos anónimos que pueden cartografiarse sobre un genoma de referencia



Considerando la impresionante producción de datos que permiten, muchos de los nuevos secuenciadores tienen un tamaño bastante discreto

cadenas se una físicamente a una cuenta microscópica en suspensión, para después verter la emulsión en una placa microperforada donde dichas cuentas acaban colocándose. Los procesos de amplificación y secuenciación son relativamente ordinarios, pero se consiguen miles de ellos simultáneos, contenidos en los volúmenes microscópicos de los pocillos de la placa, detectados en una sola imagen y procesados informáticamente para almacenar por separado la lectura de cada una de las secuencias.

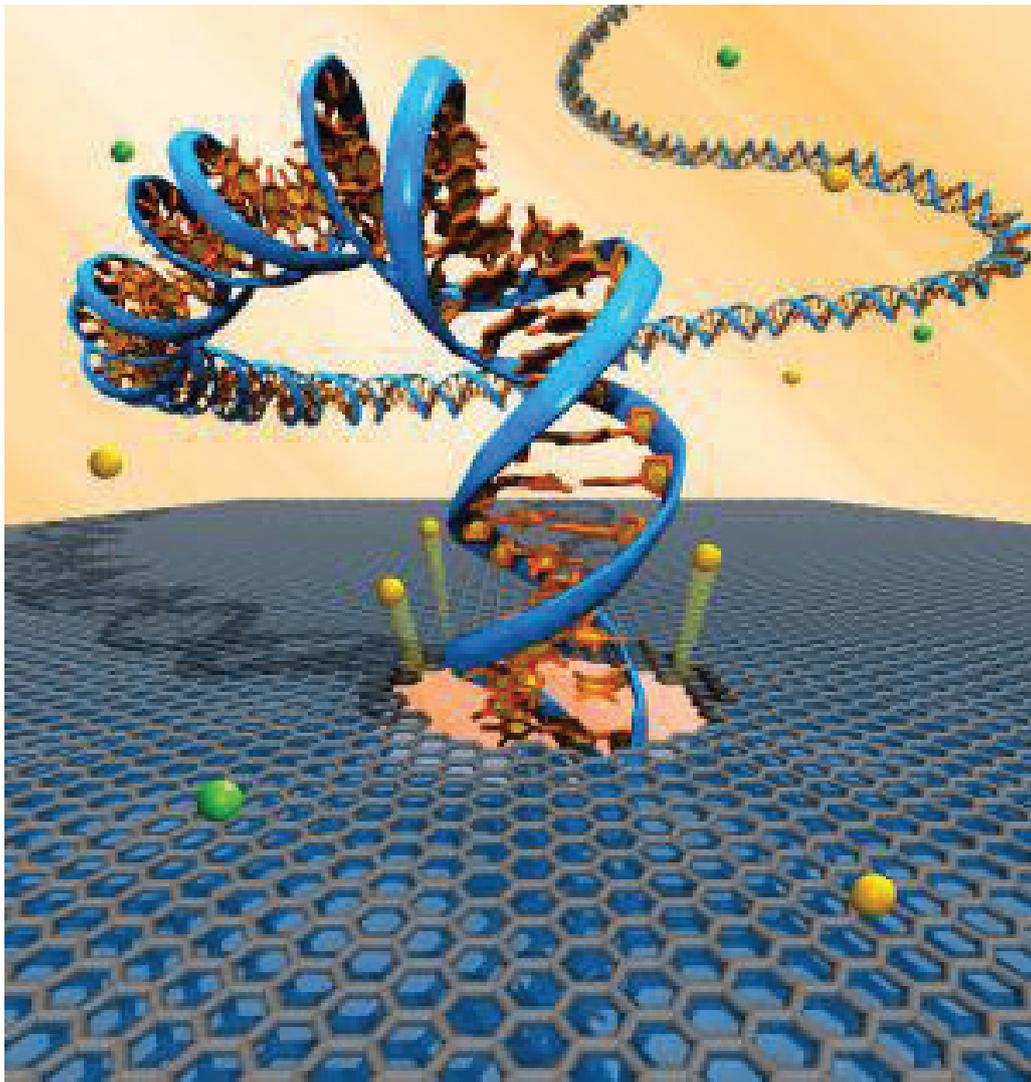
**En lugar de trabajar con las lecturas** individuales de secuencias de pocos cientos de nucleótidos, el producto de una sesión de pirosecuenciación puede ser del orden de centenares de millones de pares de bases en sólo unas horas. Esto hace totalmente imposible la revisión manual de las secuencias, aunque es innecesario, pues el margen de error es cada vez más bajo. Por otra parte, al contrario que ocurría antes, cada uno de los fragmentos es anónimo: no se sabe a qué lugar del genoma corresponde. Esta incertidumbre se ve paliada por el tremendo número de lecturas obtenidas: un tratamiento informático permite detectar áreas de solapamiento y construir una secuencia consenso ensamblando toda la información obtenida. Si el genoma no es muy extenso (por ejemplo, un genoma bacteriano o el de una mitocondria o un cloroplasto), una sola sesión puede secuenciarlo en su totalidad, algo que en el pasado hubiese llevado años de trabajo. Además, existiendo ya un número considerable de genomas de referencia, no resulta complicado avanzar en la investigación de los genomas de organismos convencionales (más allá de ratones, moscas de la fruta y *Arabidopsis*) a partir de fragmentos anónimos.

**Sorprendentemente**, incluso las técnicas de nueva generación se están sucediendo a un ritmo tan elevado que en poco tiempo parecen quedarse obsoletas. Sistemas alternativos como la **secuenciación por síntesis** consiguen resultados del orden de decenas de miles de millones de nucleótidos por sesión. Al igual que en el caso de la pirosecuenciación, requiera la preparación previa de una biblioteca de ADN. Cada fragmento de la biblioteca se adhiere de forma covalente a la superficie de unos canales transparentes

integrados en una célula de flujo y se amplifican *in situ*, formando un pequeño pincelillo de miles de copias de la misma cadena de ADN (fijadas a la superficie de la célula) y secuenciándose conforme se van agregando los nucleótidos complementarios. Al igual que en el caso anterior, la secuenciación simultánea de innumerables fragmentos anónimos no se gestiona de forma individual, sino que se integra o se ensambla mediante un tratamiento informático determinado. Es curioso que hayamos llegado a un punto en el que el mayor factor limitante en el estudio de genomas no es la propia inconmensurabilidad de los genomas en sí y su inabarcable longitud, sino la relativamente limitada capacidad de los investigadores de asimilar, ensamblar y cartografiar estas ingentes cantidades de datos.

**La secuenciación de última generación** es bastante eficiente. Las bibliotecas requieren cierta preparación que puede ser complicada, y los aparatos en sí siguen siendo carísimos, pero la gran cantidad de datos que producen hace que el precio por base secuenciada se reduzca de forma asombrosa. Sin embargo parecía que todo el progreso a la hora de secuenciar se estaba dirigiendo a grandes proyectos, a la búsqueda de fragmentos muy grandes de genoma. ¿Qué hay del *tricolor* que nos prometía la ciencia ficción?

**La respuesta puede estar** en una tecnología aún en desarrollo: **la secuenciación por nanoporos**. La idea básica es de una elegancia sobrecogedora: se trata de hacer pasar una única hebra de ADN a través de un angosto orificio de un nanómetro de diámetro en una superficie (por ejemplo de grafeno) que separa dos espacios llenos de fluido conductor. Cada nucleótido producirá un cambio característico en el flujo iónico a través del poro, permitiendo la lectura de la secuencia. Aunque obviamente esta técnica también es posible gracias a los avances asombrosos en el control de los materiales a pequeñísima escala, la revolución que puede producir algo así en la historia de la secuenciación de ácidos nucleicos puede hacer que todos los potentísimos sistemas que he comentado hasta ahora se queden en meras anécdotas.



Representación de la secuenciación por nanoporos en grafeno

**La secuenciación por nanoporos** sólo precisa una única hebra de ADN para leer, sin necesidad de enzimas amplificadoras ni complejas preparaciones. Teóricamente el coste energético podría ser ínfimo, así como el espacio necesario para hacerlo posible (obviamente), haciendo verosímil la idea de aparatos portátiles (de los que ya existen algunos modelos), así que aunque las posibilidades que se abren son innumerables, no puedo evitar pensar en su aplicación destinada a la identificación de organismos: un pequeño kit de campo en el que hacer una rápida extracción de ADN a la hoja de una planta con hidróxido sódico en el momento, se coloca una gota en el secuenciador por nanoporos (con conexión USB), se espera a tener una lectura de unos pocos cientos de pares de base y se consulta GenBank por internet: *voilà*, ahí tienes tu *tricorder*. Hace unos años esta idea me hubiese resultado un tanto ridícula, pero no puedo evitar creer que tiene todas las papeletas para hacerse realidad más pronto que tarde.

**Habría que añadir una reflexión** adicional al impacto que podría tener un juguete así en el desarrollo del estudio de la biodiversidad, que parece el sueño húmedo de los partidarios del *barcoding* (la iniciativa que pretende identificar cada especie de organismo por una secuencia indicadora de su genoma). De la misma manera que tener una calculadora no te convierte en matemático, este esperado *tricorder* no sustituirá el trabajo de los biólogos, sino que será el punto de partida para hacernos preguntas mucho más complejas y apasionantes.

**Rafael Medina**