

Ana Mondon
Mónica Poverene

La licenciada en Ciencias Biológicas Ana Mondon es becaria doctoral de CONICET. La ingeniera agrónoma Dra. Mónica Poverene es docente del Departamento de Agronomía.
Contacto: almondon@criba.edu.ar

La secuenciación de última generación en genomas de plantas y sus aplicaciones

El genoma es la dotación básica de ADN de un organismo, ya sea almacenado en el núcleo, plásmidos, mitocondrias o cloroplastos y contiene toda su información hereditaria. La molécula de ADN tiene una estructura lineal, dispuesta en doble hélice y sus unidades estructurales son los nucleótidos, nombrados usualmente A, C, T y G de acuerdo con la base nitrogenada que contienen, adenina, citosina, timina y guanina, respectivamente. Secuenciar implica obtener el tipo y el orden de los nucleótidos que componen una molécula de ADN.

Trece años han pasado desde que se secuenciara el primer genoma vegetal completo, el de la crucifera *Arabidopsis thaliana*. A partir de entonces, el uso de la secuenciación y sus aplicaciones ha crecido a una velocidad increíble. Entre 2005 y 2011 la capacidad de secuenciación se incrementó en cinco órdenes de magnitud gracias a los sistemas automáticos, superando el crecimiento estimado

para el desarrollo de procesadores de computadoras.

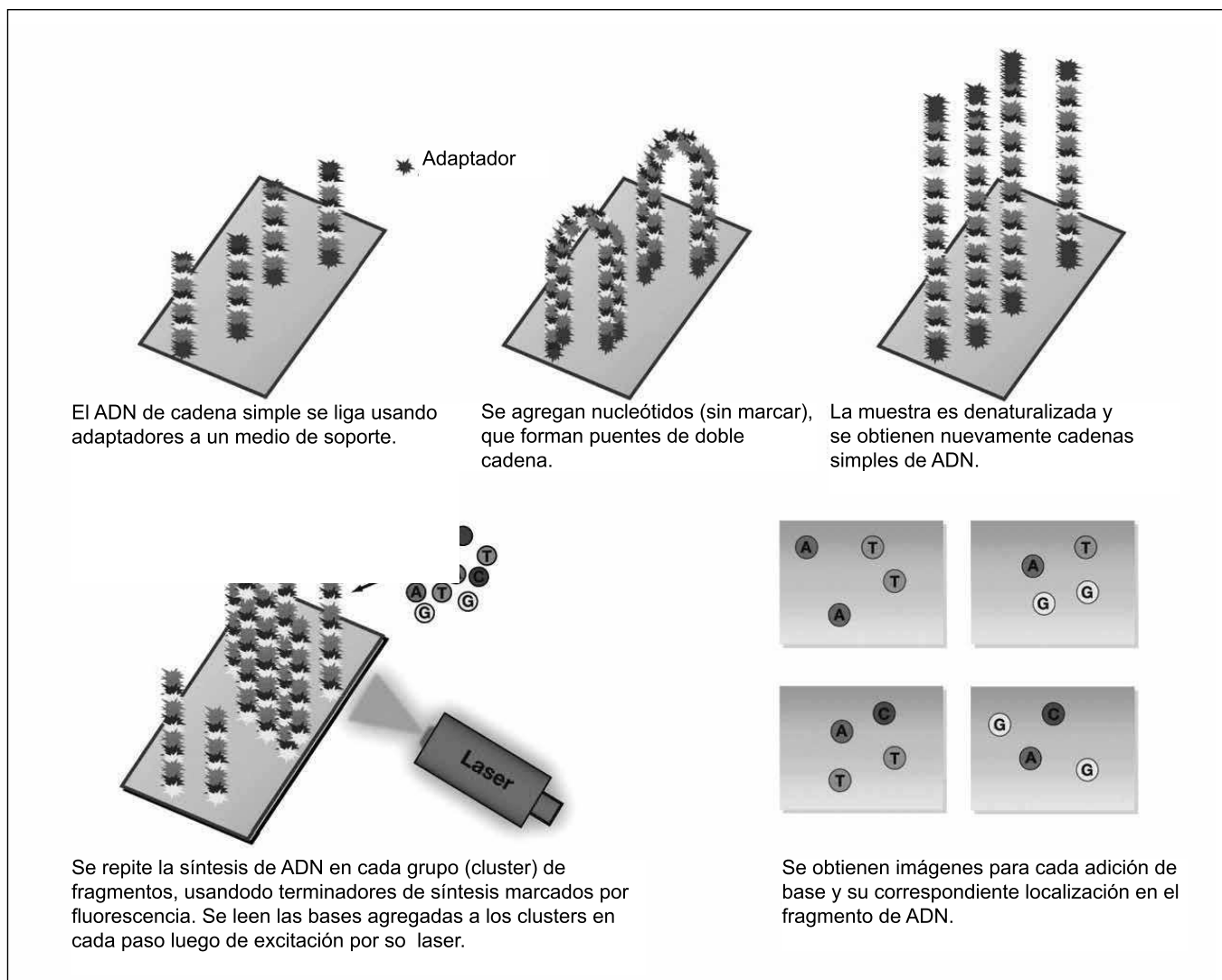
El primer método de secuenciación de ADN fue desarrollado en 1976 por Allan Maxam y Walter Gilbert. Sin embargo, para secuenciar el genoma humano, *Arabidopsis* y arroz, la técnica de Fred Sanger (dos veces Premio Nobel) fue la más usada. A lo largo de 30 años la secuenciación de Sanger se hizo más efectiva, en gran parte gracias a la automatización de los procesos, reducción de costos y la posibilidad de obtener lecturas de fragmentos de mayor longitud. El genoma completo de *Arabidopsis thaliana* es de aproximadamente 120 Mbp (millones de pares de bases) y el promedio de un genoma de angiospermas es mayor a 6000 Mbp, el doble que el genoma humano.

El año 2005 marcó un punto de inflexión dada la aparición de los que se conoce como Secuenciación de Próxima, Segunda o Última Generación (NGS por su sigla en inglés - Next Generation

Sequencing). Mediante esta tecnología se pueden secuenciar millones de fragmentos de ADN.

Varias empresas han lanzado al mercado sistemas automáticos de NGS que difieren en la longitud de las lecturas y el tiempo necesario para obtenerlas, lo que determina su rendimiento. El surgimiento de estos sistemas permite analizar varias muestras en paralelo y disminuye significativamente los costos. Anteriormente, cada proyecto de obtención de genoma solía llevarse a cabo por consorcios de instituciones públicas y privadas a costos de millones de dólares; hoy en día los costos de la secuenciación se calculan en miles. La secuenciación de novo es el análisis inicial de un genoma, para obtener por primera vez su secuencia. La re-secuenciación implica la secuenciación de parte o todo el genoma de un organismo que se compara con un genoma de referencia, como por ejemplo, el de *Arabidopsis thaliana*. NGS ha sido usado básicamente en re-secuenciación.





¿Por qué secuenciar genomas vegetales?

La secuenciación de genomas de plantas ha estado impulsada por especies que funcionan como modelos en biología general, o por interés agronómico o evolutivo en la genómica de cultivos. El contar con esos datos permite descubrir genes de interés, evaluar diversidad genética en las poblaciones, realizar estudios evolutivos, detectar variaciones epigenéticas, comprender vías metabólicas y una infinidad de otras aplicaciones. En el plano agronómico, el mayor interés está puesto en el mejora-

miento de los cultivos. Esta área es una de las que presenta mayores desafíos frente a un rápido crecimiento de la población humana, la disminución de la superficie arable y distintos tipos de estreses bióticos y abióticos.

Durante años los marcadores moleculares han sido una gran herramienta en genética vegetal. Hoy en día el foco está centrado en marcadores SNP (sigla inglesa para indicar diferencias de un solo nucleótido entre individuos). Su principal limitante es el tiempo que requiere identificarlos, pero el advenimiento de NGS y la posibilidad de realizar

re-secuenciación han facilitado la tarea. Uno de los primeros resultados que se obtienen a partir del descubrimiento de esos marcadores es una estimación de la variabilidad genética cuando se analizan varios individuos o poblaciones; en el caso de cultivos y sus parientes silvestres se espera encontrar mayor variabilidad en estos últimos.

La posibilidad de localizar genes de importancia agronómica o porciones del genoma asociadas a ciertos caracteres es de gran utilidad, ya que deriva en lo que se conoce como selección asistida por marcadores. La mejora de los culti-

vos fue llevada a cabo hasta la década de 1980 principalmente basada en el fenotipo. La disponibilidad de marcadores hace que muchos pasos en la selección se aceleren. Por ejemplo, teniendo identificados marcadores asociados con la resistencia a un determinado estrés, se pueden evaluar los individuos sin necesidad de recurrir (en principio) a ensayos de campo.

A lo largo de la historia, la domesticación de especies para llegar a variedades cultivadas ha significado la pérdida de variabilidad genética. NGS permite cuantificar rápidamente esta variabilidad e identificar caracteres en especies silvestres emparentadas con los cultivos, que permiten la mejora de estos últimos. La mayoría de los proyectos actuales de secuenciación en plantas se aplican a cultivos y el contar con un genoma de referencia facilita la re-secuenciación de genotipos o especies silvestres. Durante una pasantía de cuatro meses en la Universidad de British Columbia, Vancouver, Canadá, la primera autora de este artículo procesó 220 muestras de *Helianthus annuus* y *H. petiolaris* colectadas en Argentina, en el laboratorio del Dr. Loren H. Rieseberg. El objetivo de este proyecto es analizar las consecuencias de la hibridación y evolución conjunta de estas especies en el país. Actualmente las secuencias están siendo analizadas.

Genética de poblaciones y evolución

NGS ha tenido también gran impacto en la genética de poblaciones. La variabilidad presente en distintas poblaciones puede evaluarse mediante el uso de marcadores moleculares en cantidades que antes eran impensadas. Más allá de lograrse un panorama ge-

neral sobre las especies o poblaciones en estudio, los marcadores, junto con el mapeo de regiones específicas de la secuencia de ADN, permiten establecer hipótesis filogenéticas y rastrearlas en el tiempo. Es decir, dados ciertos cambios en las secuencias, podemos conocer cuándo hubo eventos de diversificación, surgimiento o extinción de especies.

Diagnóstico de enfermedades

El uso de NGS para el diagnóstico de patologías vegetales está ligado al concepto de metagenómica (el estudio de varios genomas en conjunto). Con niveles de detalle crecientes, puede estudiarse el metagenoma de la hoja de una planta: el resultado es básicamente qué genomas se encuentran en dicha muestra, permitiendo la identificación de agentes patógenos y la posibilidad de tomar medidas de prevención o saneamiento. Otra de las aplicaciones de la metagenómica es el estudio de suelos e incluye además de la caracterización de patógenos, la localización de especies que pueden resultar beneficiosas para el desarrollo vegetal.

Uso de modelos biológicos

El modelo biológico es la herramienta que ha ayudado al desarrollo de muchas de las aplicaciones de NGS. Los primeros genomas en secuenciarse suelen ser los que se consideran de referencia. Se comenzó con el de *Arabidopsis thaliana*, típico modelo biológico por su simplicidad y a partir de allí otras especies, como es el caso del arroz como referente de las gramíneas. Los resultados obtenidos en estos modelos biológicos permiten, en principio y hasta que la accesibilidad a NGS sea mayor,

extrapolar resultados a especies con mayor complejidad.

Epigenética y transcriptómica

Los transcriptos (moléculas de RNA) a partir del ADN pueden variar por cambios denominados epigenéticos en la estructura del ADN. La secuencia en sí no cambia, pero se modifica su expresión por reacciones de metilación o modificaciones en las histonas y enrollamiento de la molécula. Aún cuando con el método de Sanger es posible detectarlas, NGS ha acelerado el proceso de detección de cambios epigenéticos. También puede secuenciarse el transcriptoma de un organismo utilizando lecturas de ARN mensajero o copias de moléculas de DNA, obtenidas a partir de este.

El rápido crecimiento de estas tecnologías trae consigo algunos problemas. La calidad de las secuencias que se obtiene con NGS es menor a la que puede obtenerse mediante métodos como el de Sanger. Sin embargo, la relación entre costos y cantidad de información generada favorece a NGS. La necesidad de contar con genomas de referencia para procesos de re-secuenciación también hace que NGS dependa en parte de información generada anteriormente, aunque cada día se utilice más NGS para secuenciar de novo. Pero el desafío más grande en NGS lo presenta el análisis de datos de las secuencias obtenidas; dado el volumen de información que se genera, hoy en día las principales limitantes están en la capacidad de procesamiento, ya sea mediante recursos informáticos o humanos, que si bien están en crecimiento, se han visto desbordados por la cantidad de resultados generados.

¿Existen formas de simplificar aún más la técnica, el costo y los tiempos de obtención de resultados? Hace menos de dos años se ha adoptado una forma de simplificar la secuenciación mediante un método denominado GBS (*genotyping-by-sequencing*). Este consiste en fragmentar los genomas y marcarlos con una secuencia específica, a modo de código de barras. Una vez hecho esto, las muestras pueden mezclarse, permitiendo obtener actualmente resultados de 384 muestras. Claramente no se obtiene el mismo detalle que al manipular muestras individuales, pero

permite optimizar los costos operativos de la secuenciación. La técnica es válida para identificación de marcadores como SNP, evaluación de variabilidad y filogenia entre otros, pero está basada en la información aportada por una fracción del genoma.

¿Qué depara el futuro? Claramente las empresas dedicadas al desarrollo de métodos de NGS continúan mejorando con el tiempo las prestaciones de sus instrumentos, logrando mayor cantidad de lecturas de secuencias, de mejor calidad y en menos tiempo.

Referencias bibliográficas

Delseny M., Han B., Hsing Y. I. 2010. High throughput sequencing: The new sequencing revolution. *Plant Science* 179: 407-422.

Edwards D., Batley J. 2010. Plant genome sequencing: applications for crop improvement. *Plant Biotechnology Journal* 8: 2-9.

Feulliet C., Leach J. E., Rogers J., Schnable P. S., Eversole K. 2011. Crop genome sequencing: lessons and rationales. *Trends in Plant Science* 16: 77-88.

Goldstein C., Caboche M. 2009. *High-throughput genotyping*. INRA Technological Worksheet.

Jackson S. A., Iwata A., Lee S-H, Schmutz J., Shoemaker R. 2011. Sequencing crop genomes: approaches and applications. *New Phytologist* 191: 915-925.

Metzker M. L. 2010. Sequencing technologies – the next generation. *Nature Reviews-Genetics* 11: 31-46.

Varshney R. K., Nayak S. N., May G. D., Jackson S. A. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27: 522-530

