

EL USO DE LA GENÓMICA EN AGROBIOTECONOLOGÍA

Carla Schommer, Ramiro Rodriguez y Javier Palatnik *

Un acercamiento a la genómica y su impacto en la producción agropecuaria por investigadores del Instituto de Biología Molecular y Celular de Rosario (IBR) y la Universidad Nacional de Rosario (UNR); una herramienta en auge cuya potencialidad recién estamos descubriendo.

¿QUÉ ES LA GENÓMICA?

El constante avance tecnológico de nuestra sociedad no deja de sorprendernos día a día, y las ciencias no son una excepción. Una nueva rama de la ciencia que ha revolucionado la manera en la que vemos la biología y genera un moderno abanico de aplicaciones tecnológicas impensado hace unos pocos años, es la genómica.

Los genes son las unidades de información que definen a los seres vivos. Los genes tienen la información codificada para el color de ojos y la altura de los seres humanos, pero también para el tamaño que va a tener un grano de soja o el sabor de un fruto de tomate. Los seres humanos tenemos unos 25.000 genes, una mosca tiene 12.000 y una bacteria unos 5.000 o menos. Las plantas suelen tener más genes que los animales. Por ejemplo, el maíz tiene unos 40.000, mientras que

* Carla Schommer es investigadora en el IBR (Instituto de Biología Molecular y Celular de Rosario, CONICET/UNR) y docente del Centro de Estudios Interdisciplinarios de la UNR.

Ramiro Rodriguez es investigador en el IBR (Instituto de Biología Molecular y Celular de Rosario, CONICET/UNR) y docente de la UNR. Javier Palatnik es licenciado en Biotecnología, Doctor en Ciencias Biológicas por la UNR. Posdoctorados en el Instituto Salk en La Jolla, California, Estados Unidos, y en el Max Planck de Tuebingen, Alemania. Investigador principal del CONICET en el Instituto de Biología Molecular y Celular de Rosario (IBR, CONICET-UNR), donde preside la Fundación del IBR. Profesor Asociado – Jefe del Laboratorio Biología del ARN IBR (CONICET-UNR).



la soja tiene 45.000 genes. Hasta hace poco tiempo, los científicos sólo podían estudiar la función de un gen a la vez. Pero debido a una revolución en métodos y equipamientos tecnológicos, ahora se pueden estudiar todos los genes en forma simultánea. Esto es lo que hace la genómica.

La genómica puede aplicarse a todos los sistemas biológicos, desde bacterias y virus, hasta las plantas y los seres humanos. Los estudios genómicos muchas veces encuentran aplicaciones directas en la medicina, tanto en el diagnóstico como en el tratamiento de enfermedades. Mientras que en agrobiotecnología se usa para conseguir la mejora de variedades cultivadas para aumentar la productividad o la respuesta al estrés. La genómica se encarga tanto de cartografiar todos los genes de un individuo, como de identificar sus funciones.

SECUENCIACIÓN DE ALTO RENDIMIENTO O "NEXT GENERATION SEQUENCING"

En 1977, el bioquímico Frederick Sanger desarrolló un método para secuenciar el ADN y su aplicación resultó en la publicación del primer genoma completo. En este caso, fue el genoma del virus Φ X174, que resultó estar compuesto por poco más de 5.000 bases organizadas en 11 genes. Aunque fue un avance enorme para la

época, la secuencia del virus es minúscula comparado con la de organismos más complejos. El genoma de un ser humano contiene aproximadamente 25.000 genes, pero para conocerlos todos es necesario secuenciar las más de 3.000 millones de bases que forman su ADN, un número muy diferente a las 5.000 bases de aquel virus.

A pesar de esto, motivados por la importancia de conocer el genoma completo de un organismo y contando con avances continuos en las técnicas de secuenciación, los científicos armaron consorcios de decenas de grupos de trabajo para secuenciar los genomas de organismos más complejos. Llevó 20 años hasta que en 1996 el genoma del primer organismo eucariota, el de la levadura *Saccharomyces cerevisiae* (un organismo muy utilizado industrialmente para la fabricación de pan, cerveza y vino) fue completado después un esfuerzo de 10 años por parte de 100 laboratorios en tres continentes. Este microorganismo tiene 5.700 genes distribuidos en 12 millones de bases.

El gran desafío, la secuenciación del primer genoma humano, fue terminado en el año 2003, demandando un esfuerzo coordinado de múltiples países durante 15 años. El costo de este primer genoma fue de más de 3 billones de dólares subsidiados fundamentalmente



EL GENOMA Y EL LIBRO DE LA VIDA

Hoy en día, es común escuchar el término ADN. ADN es la abreviatura de Ácido Desoxirribonucleico, que es el nombre de la molécula encargada de almacenar la información de un ser vivo. El ADN es una molécula muy larga, y contiene regiones que almacenan información específica y concreta. Cada una de estas regiones que almacenan información recibe el nombre de gen. Y el genoma es el conjunto de todos los genes de un ser vivo.

Para comprender qué es un genoma, podemos hacer una analogía literaria. Un libro que comprende un cúmulo de información particular está compuesto por una serie de miles de palabras, cada una de ellas generadas a su vez por una secuencia particular de algunas de las 23 letras del alfabeto.

De forma análoga, un genoma (el "libro" en nuestra analogía) contiene toda la información necesaria para construir un ser vivo, y está compuesto por miles de genes (las "palabras"), cada uno de ellos formado por una secuencia particular de letras. En lugar de las 23 letras del alfabeto, los genes son secuencias de 4 letras llamadas "bases", denominadas G, A, T y C, por sus propiedades químicas.

La genómica trata de entender cómo funciona un organismo utilizando la información que contiene guardada en todo su genoma. Su objeto de estudio no son sólo unas pocas palabras o frases del libro, sino que analiza en simultáneo la totalidad de las palabras que forman el libro. Esto ha sido muy útil para iden-

tificar la causa de enfermedades hereditarias en los seres humanos. Si se compara la secuencia del genoma de seres humanos que tienen cierta enfermedad con aquellos que son sanos, se puede llegar a identificar las bases (letras) cambiadas en los genomas de los pacientes enfermos y que son responsables de la enfermedad. El cambio de una base por otra es lo que se conoce como mutación. En biomedicina, conocer las mutaciones que causan enfermedades es muy importante porque permite identificar la causa exacta de una enfermedad y cuál es el gen que está afectado. Esto permite desarrollar nuevos tratamientos terapéuticos, y para el paciente en particular también da la posibilidad de tener una medicina personalizada. Las mutaciones no siempre son perjudiciales. Hay veces que cambiar una base (o letra) de un gen por otra hace que el gen funcione mejor, o que dé nuevas propiedades a un ser vivo. Hoy en día se conocen mutaciones que hacen que las plantas sean más tolerantes a la sequía o produzcan semillas más grandes.



por agencias gubernamentales de los Estados Unidos, como NSF (National Science Foundation) y DOE (Department of Energy). El proyecto permitió exitosamente cartografiar todos los genes de un ser humano por primera vez, pero tuvo un éxito no pensado adicional. Durante el proyecto de secuenciación del genoma humano se diseñaron nuevos métodos y equipamiento para la secuenciación de ADN, que revolucionaron la forma y potencia de los estudios genómicos.

Estas nuevas tecnologías son conocidas como "Next Generation Sequencing" (NGS) y están basadas en la posibilidad de secuenciar simultáneamente millones de fragmentos de ADN contenidos en volúmenes muy pequeños. Estas nuevas herramientas bajaron significativamente los costos y esfuerzos necesarios para secuenciar el ADN. Hoy en día es razonable secuenciar un genoma humano por 5.000 dólares en vez de los 3 billones que costaban 15 años atrás. Estos costos accesibles permitieron difundir el uso de la genómica a un gran número de laboratorios, lo que a su vez disparó una explosión del uso y aplicaciones de la genómica.

LOS GENOMAS DE LAS PLANTAS

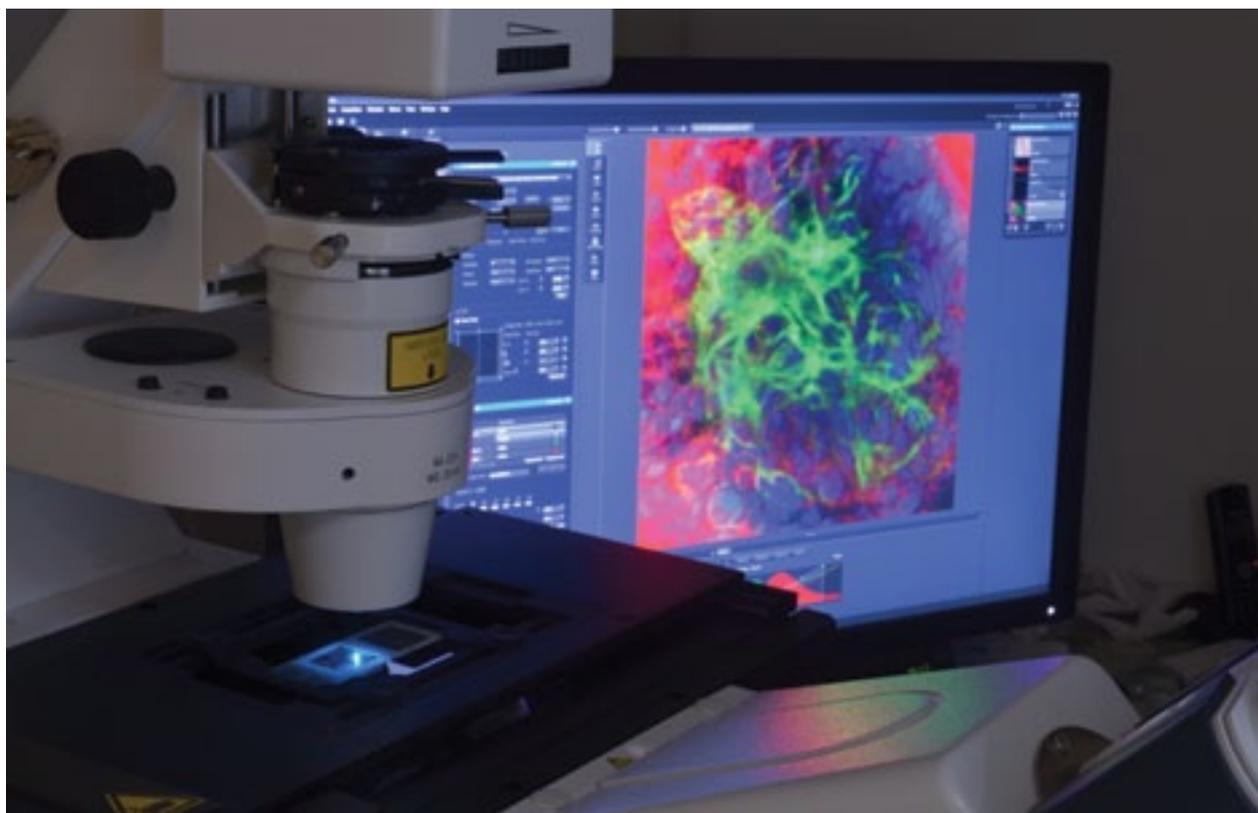
Cada especie vegetal utilizada en producción agrícola existe en múltiples variedades. Un ejemplo extremo lo constituyen el brócoli, el repollo en sus diferentes colores, la coliflor y los repollitos de Bruselas. Todas estas plantas, si bien de formas, colores y sabores muy diferentes, pertenecen a una misma especie de plantas conocida con el nombre común de col o el científico *Brassica oleracea*. Siendo todas de la misma especie, tienen genomas casi idénticos, y sólo algunas pequeñas variaciones en algunos pocos genes son responsables de las enormes variaciones que presentan.

Diferencias más sutiles, pero no menos importantes, se encuentran en todas las especies de interés agrónomo, y son responsables de la diferencia en productividad de distintas variedades. Existen muchas variedades de los cultivos económicamente importantes como el maíz y la soja. Variedades que van desde las silvestres -que

en algunos casos aún se encuentran en la naturaleza-, a las variedades de elite de alto rendimiento obtenidas luego de cientos de años de mejoramiento. De la misma manera que las coles, estas variedades tienen genomas muy similares entre sí, y nuevamente algunas pequeñas variaciones son las responsables de sus diferentes características.

La explosión de la genómica de los últimos años se aprecia si analizamos unos simples números. En el año 2000 se completó la secuenciación del genoma de la primera planta, *Arabidopsis thaliana*. Esta especie se utiliza como modelo de laboratorio, resultando ser algo así como el "ratón" de los biólogos vegetales. Se eligió la secuenciación del genoma de esta planta como proyecto inicial por su pequeño genoma y genética simple: apenas 120 millones de letras que se combinan en secuencias específicas que dan lugar a 33.602 genes. La segunda especie vegetal cuyo genoma fue secuenciado fue el arroz. De hecho, en 2002 fueron publicados en simultáneo los genomas de dos sub-especies del arroz, Japónica e Indica. En este caso, el genoma es más grande y complejo, con unos 370 millones de letras organizadas en al menos unos 50.000 genes. Al día de hoy, se conocen los genomas de decenas de especies vegetales, incluyendo maíz, soja y uno de las más recientes, el de la quinoa, un cultivo regional con alta proyección económica mundial.

Pero la secuenciación de genomas no se limita sólo a nuevas especies, sino que se extiende para analizar los genomas de variedades dentro de una misma especie.





Por ejemplo, se ha completado como parte de un proyecto multinacional la secuenciación de 3.024 genomas de arroz de distintas variedades provenientes de 84 países.

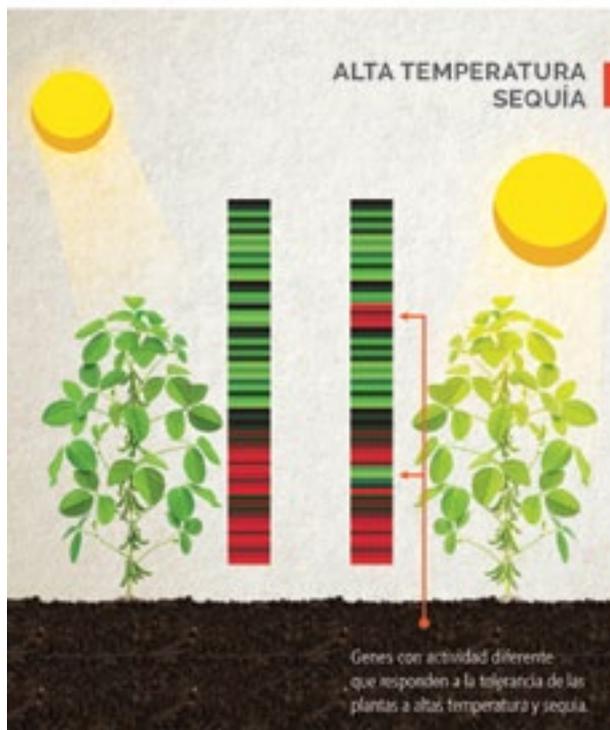
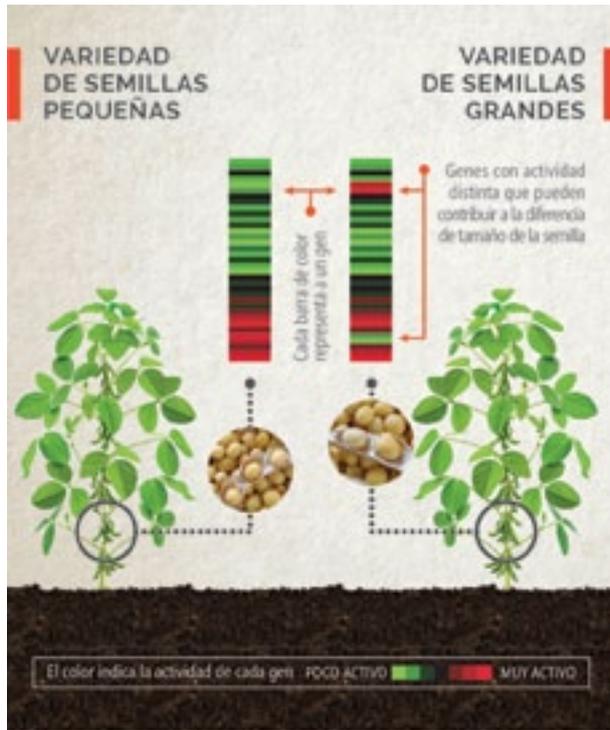
Ahora bien, ¿para qué sirve este vasto cúmulo de información? O puesto en otros términos, ¿para qué nos sirve conocer cuántos y qué genes tiene cada especie y sus variantes? Una de las aplicaciones viene justamente de las variedades que se encuentran para cada especie. Por ejemplo, de los cientos de variedades de arroz, cada una tiene características particulares que le dan mejor rendimiento, calidad nutricional, tolerancia a sequía o patógenos y otras características determinantes del rendimiento. Comparando los genomas de estas variedades se puede determinar cuáles son los genes responsables de las características deseadas. Por ejemplo, si se analizan los genomas de arroz de algunas variedades con semillas grandes, una característica determinante del rendimiento de este cultivo. Estudios realizados en el Instituto de Biología Molecular y Celular de Rosario (IBR) determinaron qué variantes de un gen llamado GRF promueven el crecimiento de la planta modelo en *Arabidopsis thaliana*, haciendo que la misma sea más grande. Cuando se analizaron las plantas de arroz con semillas grandes se vio que tenían el mismo tipo de cambio en su ADN al que se había sido descrito por investigadores del IBR. Este cambio en apenas 2 "letras" en el genoma de arroz es suficiente para cambiar significativamente el tamaño del grano. Con este conocimiento disponible, es posible diseñar nuevas variedades rápidamente, o bien

buscar este mismo tipo de cambio en otras especies, como soja y maíz.

GENOMAS EN ACCIÓN: LA GENÓMICA FUNCIONAL

Como hemos mencionado más arriba, el genoma es el conjunto de genes de un organismo dado, siendo los genes las unidades fundamentales de almacenamiento y transmisión de la información de dicho organismo. Esta visión algo estática de los genomas dista de ser completa. Los genomas son altamente dinámicos. De los 25.000 genes de un ser humano, no todos están prendidos al mismo tiempo; si eso pasara seguramente las células morirían. Entonces, una célula del músculo se caracteriza por tener prendidos un grupo de genes, mientras que una célula del hígado tiene prendidos otros. Este patrón de genes define justamente la identidad de cada célula de nuestro cuerpo.

En plantas pasa lo mismo: las células de la raíz y la hoja tienen el mismo genoma (es decir, los mismos genes), pero un 13% está prendidos solo en raíces y un 18% solo en hojas. Algunos genes están prendidos en raíces y hojas, pero otros son muy específicos de un solo órgano de la planta. Los genes también cambian su expresión frente a las condiciones ambientales. Estos cambios le permiten a la planta desarrollar mecanismos de defensa frente a condiciones ambientales adversas como sequía, salinidad, frío, calor o el ataque de patógenos. Experimentos de genómica han permitido determinar que en respuesta a suelos

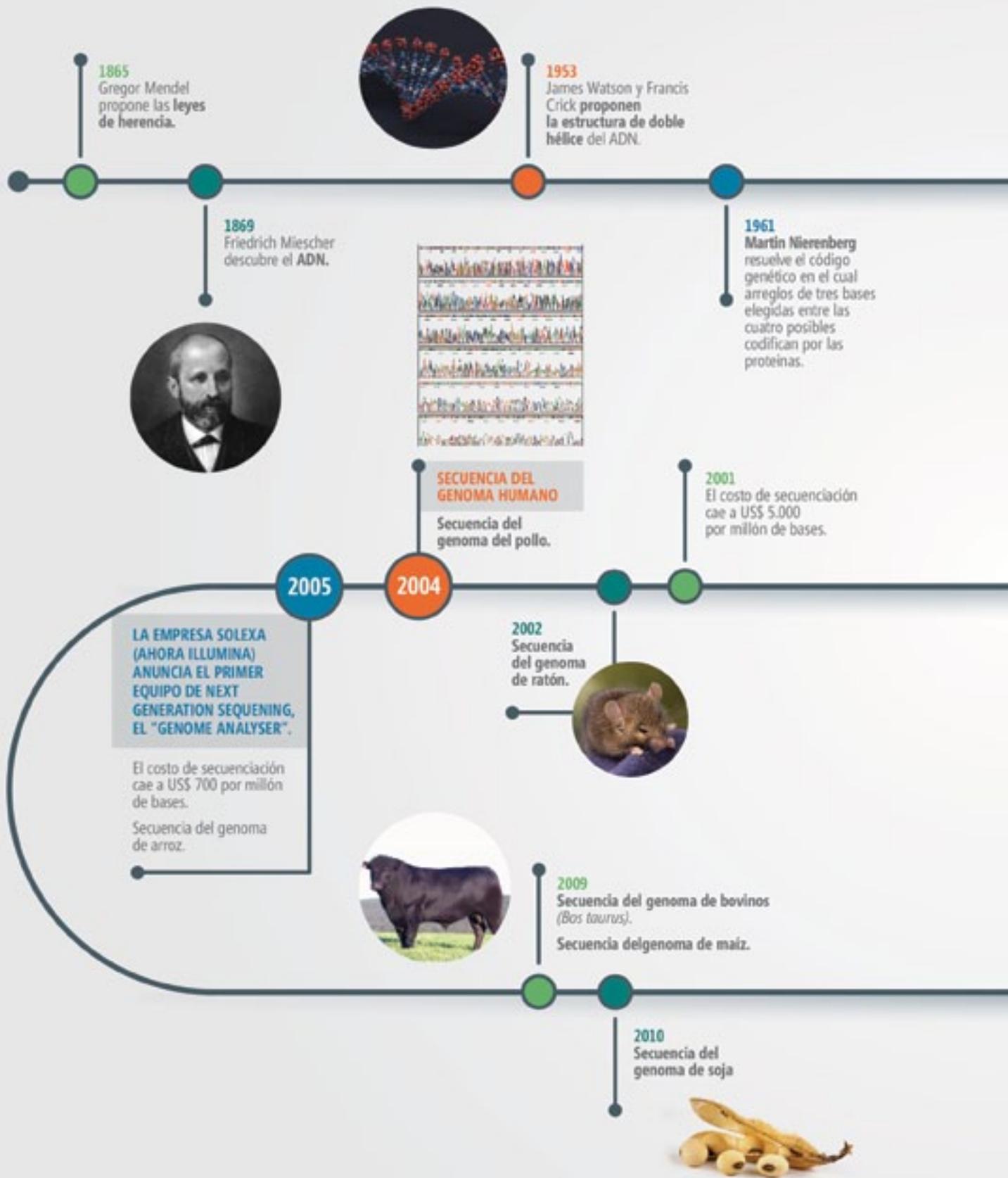


¿QUÉ SON Y QUÉ HACEN LOS GENES?

Si bien los genes están presentes en animales y plantas, los primeros indicios de su existencia fueron revelados por estudios en arvejas realizados por el botánico Gregor Mendel en 1865. Sus trabajos con cruces de diferentes variedades de arveja le permitieron enunciar las primeras leyes de herencia, sentando con ello las bases de la genética, una rama fundamental de las ciencias biológicas. Mendel postuló que cada planta lleva en los genes información que pasa de una generación a la próxima por herencia y que estos genes establecen en su combinación el aspecto de cada organismo. Poco después, en 1869, Friedrich Miescher descubrió el ADN como un componente nuevo en las células. Pero fue recién en 1953 que James Watson y Francis Crick determinaron la estructura de doble hélice del ADN, hecho que revolucionó el entendimiento sobre la información genética.

Esta información es interpretada usando un código universal descifrado por Marshall Nirenberg en 1961. Él descubrió que cada combinación de tres bases del ADN codifica para un aminoácido, los que combinados entre sí arman las proteínas. Así, un gen que tiene una secuencia de 3.000 bases tiene la información para construir una proteína de 1.000 aminoácidos. Las proteínas, a su vez, pueden tener distintas funciones biológicas: actúan como enzimas permitiendo las reacciones químicas, nos defienden de virus y bacterias, comunican unas células con otras y forman las fibras musculares, entre muchas cosas más. En definitiva, los genes tienen la información para armar proteínas y las proteínas cumplen con múltiples funciones en los seres vivos. Una mutación en un gen puede hacer que la proteína cambie también en su secuencia y funcione diferente o no funcione.

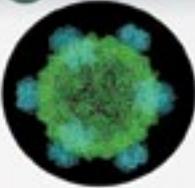




1977

Invento de la secuenciación por el método de Sanger.

Secuencia completa del DNA virus FX174.



1987

Primera máquina automatizada de secuenciación. Desarrollado por la empresa Applied Biosystems.

1984

Secuencia del virus Epstein-Barr, causante de la mononucleosis aguda infecciosa.

1990

COMIENZA EL PROYECTO DEL GENOMA HUMANO

El costo de secuenciación está en US\$ 750.000 por millón de bases.



1998

Secuencia del genoma del primer organismo multicelular, el gusano modelo *Caenorhabditis elegans*.



1995

Secuencia completa del primer genoma bacteriano (bacteria de la gripe, *Haemophilus influenzae*).



2000

Secuencia del genoma de la primera planta, *Arabidopsis thaliana*.

Secuencia del genoma de la mosca de fruta *Drosophila melanogaster*.



1996

Secuencia del primer eucariota, *Saccharomyces cerevisiae*, luego de 10 años de trabajo.

2011

Secuencia del genoma de papa.



2014

Secuencia del genoma de trigo. Secuencias de genomas de más de 3.000 variedades de arroz depositados en bancos de datos.



2012

Secuencia del genoma de cerdo. Secuencia del genoma de tomate.

Publicación del "Encode project": enciclopedia integrada de todos los elementos funcionales presentes en el genoma.

Los costos de secuenciación continúan bajando.



2016

Secuencia del genoma de la quinoa, café, zanahoria, trébol rojo, kalanchoe, ananá, etc.

salinos cambian su expresión más de 600 genes del genoma de *Arabidopsis* en muy corto tiempo. Dentro de ese grupo de genes hay al menos algunos que son esenciales para permitir el crecimiento en condiciones de salinidad. Una situación similar se da cuando las plantas son sometidas a otras situaciones adversas como la sequía, cambios de temperaturas o ataques de patógenos. Poder identificar los genes que responden frente a estas situaciones y comprender cómo funcionan da la posibilidad de generar nuevas variedades con mejorado rendimiento en ambientes adversos.

La situación es aún más compleja, porque cuando se prenden, los genes pueden hacerlo con distinta intensidad. En los animales y los seres humanos pasa lo mismo, los genes no están simplemente prendidos o apagados, sino que pueden estar activos con distinta fuerza. Esto es importante, porque a veces las diferencias entre un cultivo sensible o tolerante a un estrés (o cualquier otra característica) se deben a la intensidad con que se activan los genes. Poder medir los genes prendidos y la intensidad con la que se encienden es muy importante para comprender las funciones de los genes. Justamente, es la Genómica Funcional la que se encarga de estudiar cómo se prenden y apagan los miles de genes de un individuo según el tipo de células y las condiciones ambientales particulares. Esto es posible nuevamente gracias a las nuevas técnicas de secuenciación de alto rendimiento NGS que permiten rápidamente y de forma económica analizar el estado funcional ("prendidos" o "apagados") de los miles de genes de un genoma.

CONSTRUYENDO REDES DE GENES

Entonces, en una planta hay miles de genes que se prenden y apagan con distinta intensidad dependiendo de las células que analicemos o las condiciones ambientales. Sin embargo, los resultados experimentales que se obtienen muestran que la mayoría de los genes no actúan solos, sino que interaccionan unos con otros formando redes. En estas redes, los genes aportan sus funciones específicas para crear complejos programas. De esta manera, el desarrollo de una semilla o la respuesta a la sequía de una planta no es simplemente el resultado de la acción de un único gen, sino de cientos o miles de genes que interaccionan unos con otros para desarrollar un programa concreto. En algunos casos, modificar un único gen, ya sea en su secuencia de ADN o por el nivel en que se activa, puede dar una mejora en el rendimiento de los cultivos. Sin embargo, la modificación de un único gen muchas veces no causa cambios apreciables en las plantas, y otras veces se obtienen resultados moderados. Una explicación a estos resultados limitados es que los genes actúan como parte de complejas redes, y un solo cambio en un gen no es suficiente para modificar el funcionamiento de todo el programa. La visión actual

es que para conseguir mejoras significativas es necesario comprender el funcionamiento de estas redes de genes, y a partir de ahí diseñar varios cambios en forma simultánea.

Grupos de investigación del IBR justamente trabajan para comprender las redes génicas que determinan el crecimiento de las plantas o su respuesta frente a condiciones de estrés o al ataque de patógenos. Esto permite primeramente la comprensión del funcionamiento de los programas de acción coordinada de los genes, y luego tiene aplicaciones biotecnológicas en el diseño de modificaciones de las redes de genes que permiten dar mejoras agro-biotecnológicas. La genómica, que propicia el estudio simultáneo de todos los genes de un organismo, brinda las herramientas para comprender las redes génicas.

Algo importante es que estos estudios generan un volumen de datos e información muy grande, por lo que son necesarias herramientas bioinformáticas y computadoras de alta capacidad para analizar este tipo de experimentos. A partir de aquí, estos estudios permiten obtener una cantidad masiva de información para diseñar estrategias orientadas a mejorar los cultivos, posiblemente a través de la modificación simultánea de distintos puntos de las redes de genes. La utilización de las herramientas genómicas está en auge en estos momentos y es posible que sólo estemos vislumbrando una pequeña parte de su potencialidad real ■

Glosario:

ADN: ácido desoxirribonucleico, molécula polimérica.

Base: unidad fundamental a partir de la que se construye el ADN. El ADN usa solo cuatro bases: A (adenosina), C (citosina), G (guanina) y T (timina).

Cromosoma: estructura construida por ADN que lleva la información de muchos genes y que puede ser observada microscópicamente durante la división una célula.

Gen: unidad de ADN que codifica para una proteína que cumple una función específica en el organismo. En algunos casos, los genes codifican para ARNs que cumplen funciones en las células.

Genética: rama de las Ciencias Biológicas que estudia la herencia.

Genómica: ciencia que trata de entender el conjunto de toda la información genética de un organismo.

Genómica Funcional: trata de entender cómo la información genética de un organismo está conectada con su fenotipo por el control de la actividad de los genes y a escala global.

Secuenciación: determinación la secuencia de bases "A", "C", "G", "T" están presentes en el ADN.