

LA GÉNESIS DE UNA VERDADERA REVOLUCIÓN VERDE

Alejandra Covarrubias* y Federico Sánchez**

La publicación en 1953 del modelo de la doble hélice, por Watson y Crick, marca un hito en la ciencia al comprender cómo la hebra del ADN codifica la secuencia de aminoácidos de una proteína en el orden lineal de sus bases. Además, en la doble hélice se encuentra la explicación de cómo el código genético se replica y se preserva para las siguientes generaciones. Con este hecho surge la revolución del ADN. Poco tiempo después se pudo descifrar el enigma del código genético y de cómo se transcribe y se edita en el núcleo de la célula en la molécula mensajera de ARN y de cómo ésta viaja del núcleo al citoplasma para ser traducida en una proteína. Tan sólo dos décadas después del modelo de Watson y Crick surge la ingeniería genética con el descubrimiento de las enzimas de restricción que permiten cortar y manipular el ADN. Encadenado con este hallazgo se genera la primera unión de dos fragmentos de ADN de origen distinto, uno bacteriano y el otro eucariótico, originando la clonación de la información genética.

En 1995 se gesta un nuevo paradigma en las ciencias biológicas –la era genómica– como consecuencia de la publicación de la secuencia completa del genoma del primer organismo vivo, la bacteria y agente causal de la gripe, *Haemophilus influenzae*. Este logro titánico –el descifrar el libro de la vida– sin duda se debe a la sinergia entre varias disciplinas tales como las matemáticas, la computación y la biología. En estos días es difícil llevar la cuenta de los organismos cuya secuencia de genomas se ha concluido o

está en vías de ser descifrada (en el orden de 450). Entre éstos se encuentran algunas bacterias patógenas y benéficas para plantas y animales; dos levaduras; protozoarios; un nemátodo (*Caenorhabditis elegans*); la mosca de la fruta (*Drosophila melanogaster*); una planta silvestre, el organismo modelo por excelencia de la genética de las plantas, *Arabidopsis thaliana*, y la secuencia de dos variedades de un cereal del cual depende la alimentación de más de la mitad de la población del planeta, el arroz (*Oryza sativa japonica* y *O. s. indica*). Asimismo, la secuencia de dos organismos modelo de vertebrado están muy cercanas a concluirse, la del ratón (*Mus musculus*) y la del pez cebra (*Danio rerio*). Hace unas cuantas semanas fue

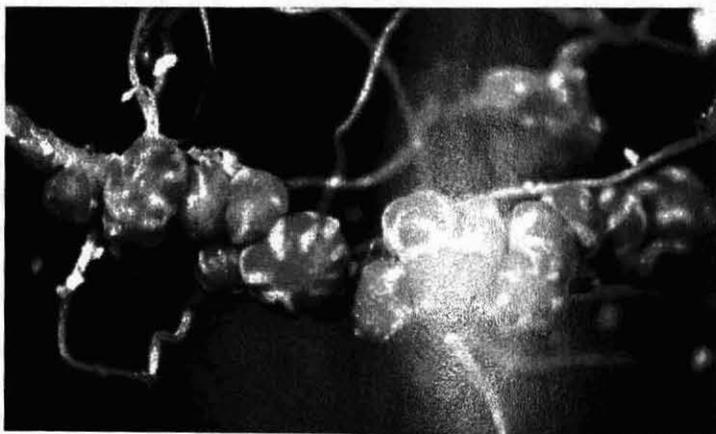
publicada la secuencia del pez globo (*Fugu rubripes*), el vertebrado con el genoma más pequeño que se conoce hasta la fecha, apenas un octavo del tamaño del genoma humano. Es interesante mencionar que el número de genes identificados en sus 365 millones de bases de ADN es aproximadamente el mismo, entre 35 mil y 40 mil, que los de los geno-

mas de humano y de ratón (Aparicio *et al.*, *Science* 1301-1310, 2002).

Si bien es cierto que este pez pequeño, considerado como un manjar en la cultura culinaria japonesa, no es aún un organismo modelo, ha contribuido a la identificación de cerca de mil genes putativos del humano. Además, ha sido un modelo de la eficiencia y coordinación de esfuerzos que permiten concluir la secuencia de su genoma con un tiempo récord de tan sólo 18 meses y un costo de 12 millones de dólares.

Desde luego, el reto más ambicioso ha sido hasta el momento la secuencia del genoma humano (*Homo sapiens*) por la trascendencia, las implicaciones y el esfuerzo compartido que ha significado descifrar un genoma de tales dimensiones (Baltimore, 2001).

Las técnicas de secuenciación del ADN han evolucionado a una velocidad vertiginosa. Tan sólo en dos décadas



* Investigadora titular C. Investigadora nacional nivel III. Cuenta con más de 50 publicaciones y es experta en biología molecular y genética de plantas

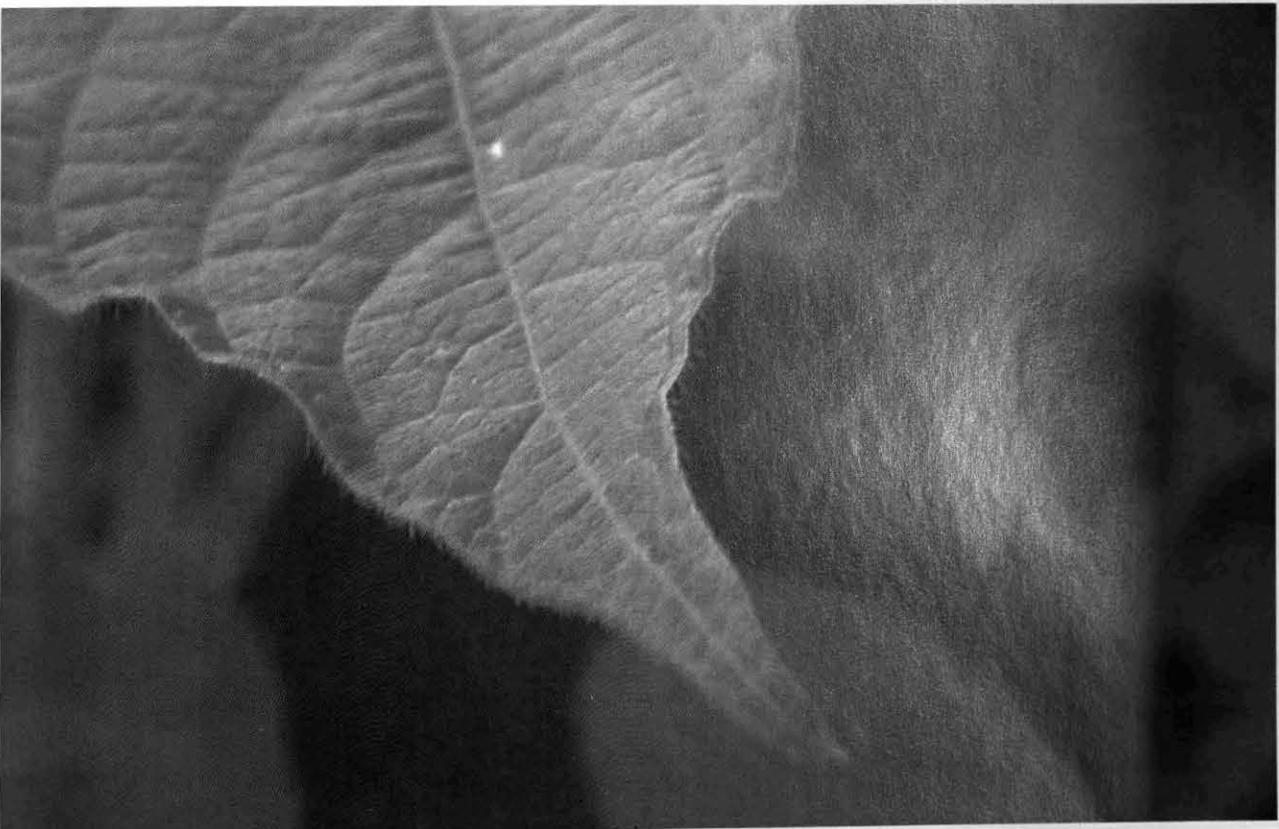
** Investigador titular C. Investigador nacional nivel III. Cuenta con más de 50 publicaciones y es experto en biología molecular y señalización entre microorganismos y plantas

ha pasado de ser una técnica de electroforesis en gel, complicada y lenta, a una altamente sofisticada, totalmente automatizada y muy eficiente. De tal suerte que la cantidad de información generada y disponible en las bases de datos es tan grande que se ha requerido avanzar hacia nuevos horizontes en la estadística y la informática, para dar origen a la bioinformática. Esto ha permitido comparar, extraer y dar sentido a la información, de tal forma que ésta se traduzca en nuevo conocimiento. No sólo es importante conocer la secuencia contenida en el libro de la vida –el genoma– sino también estudiar cómo esta información se expresa en un momento particular de la vida de un organismo. El estudiar el universo de moléculas de ARN –el transcriptoma– que se producen en la célula, significa un incremento en varios órdenes de magnitud en la complejidad de la información, pues las moléculas de ARN se editan y procesan, es decir, un mismo gen puede dar origen a muchas versiones de diferentes copias de RNA.

Asimismo, el universo de proteínas de una célula –el proteoma– que se genera cuando el ARN es traducido a una secuencia de aminoácidos, significa otro salto cuántico en la complejidad de la información que se obtiene, pues es común que a las proteínas se les adicionen moléculas complejas de lípidos o azúcares y otras modificaciones que alteran sus propiedades fisicoquímicas y, por lo tanto, su función en la célula.

Finalmente, un reto de proporciones descomunales es el estudio del perfil metabólico que produce un organismo en un momento dado –el metaboloma–. Esta tecnología poderosa y emergente permite incrementar con rapidez nuestro entendimiento de los procesos biológicos fundamentales y tendrá un impacto industrial, terapéutico y social de proporciones muy grandes.

Esta cadena de paradigmas que se originan de conocer la secuencia del genoma es reconocida como una nueva revolución en las ciencias biológicas:



LA ERA POSGENÓMICA

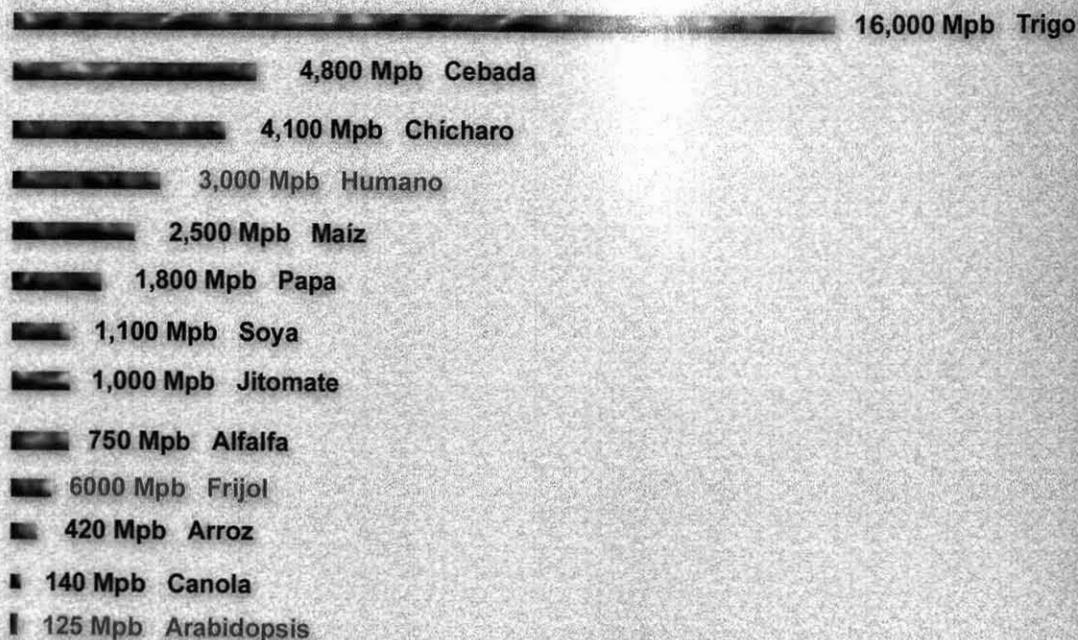
En este ensayo revisaremos el impacto que está generando el advenimiento de la era posgenómica en las plantas, en lo que se ha denominado la nueva revolución verde, porque permite entender y aplicar el conocimiento desde una perspectiva holística y no exclusivamente enfocada al incremento en la producción de los cultivos sin tomar en cuenta el impacto en el medio ambiente y la salud.

LA NUEVA REVOLUCIÓN VERDE

Las plantas son la base de la vida sobre la tierra. En muchos sentidos las plantas son organismos más exitosos que los animales: fueron los primeros en colonizar la superficie de nuestro planeta. Aún ahora son capaces de crecer en lugares donde ningún animal podría existir. Las plantas alcanzan tamaños mucho mayores que cualquier animal, pueden vivir por periodos más largos y son capaces de producir su propio alimento, en tanto que los

animales dependen totalmente de ellas. Por muchos criterios, podríamos decir que la vida vegetal domina la Tierra. De estos organismos, las plantas con flores (angiospermas) constituyen un éxito en la evolución reciente; a pesar de haber surgido hace 200 millones de años, hoy día existen 250 mil especies que predominan en casi cualquier ecosistema terrestre y acuático, y entre las que podemos encontrar a la mayoría de las que nos sirven de alimento. Un hecho asombroso es la gran diversidad de formas y de compuestos químicos que se pueden encontrar en todas ellas. Sin embargo, debido a su reciente diversificación, uno podría esperar que la mayoría de los genes identificados en alguna de estas especies tengan alguna contraparte en las otras. Por ello, conocer la información contenida en el genoma de algunas de estas especies sería de gran valor para entender su evolución y los mecanismos involucrados en su crecimiento y reproducción, así como en sus estrate-

COMPARACIÓN ENTRE LOS TAMAÑOS DE LOS DIFERENTES GENOMAS



gias adaptativas (defensa contra plagas, tolerancia a ambientes adversos, etcétera).

En diciembre de 2000 se concluyó la secuencia del genoma de la primera planta, conocida como *Arabidopsis thaliana*, pariente cercana de la mostaza y de la coliflor (*The Arabidopsis Genome Initiative*, 2000). Tomó cuatro años descifrar la secuencia del genoma nuclear, que consta de aproximadamente 125 millones de pares de bases. La *Arabidopsis* ofrece muchas ventajas (es un organismo modelo); el tamaño de su genoma es de los más pequeños que se conocen entre las plantas y en general de otros eucariontes, alrededor de 32 veces más pequeño que el genoma humano (tabla I). El tamaño de la planta y ciclo vital de entre dos y tres meses ofrece muchas ventajas. Así, la mayoría de los procesos del desarrollo, como la formación de hojas, tallos, flores y semillas, así como los procesos fisiológicos como la fotosíntesis, tienen su contraparte en los genes que controlan estos procesos en otras plantas como el maíz, el arroz o el trigo.

Una de las herramientas más poderosas con las que se cuenta en la biología vegetal desde hace 17 años ha sido precisamente la *A. thaliana*, una planta que, además de lo pequeño de su genoma y de su talla, ha resultado ser un buen modelo para realizar experimentos de genética molecular. Las ventajas de esta planta llevaron a obtener la secuencia completa de los nucleótidos contenidos en su genoma. El conocimiento que este logro trajo consigo ha representado el avance más importante de este milenio en la biología vegetal.

La información codificada en el genoma de esta hierba nos deja ver que las plantas son mucho más complejas que lo que muchos biólogos se habían imaginado. Se ha vaticinado que la *Arabidopsis* posee un número sorprendentemente alto de genes —aproximadamente 26 mil—. Sin embargo, también como producto del análisis de la secuencia obtenida resulta asombroso el hecho de que cerca de 70 por ciento del genoma de la *Arabidopsis* se ha duplicado, lo cual permite predecir que sólo hay alrededor de 15 mil genes diferentes. Asimismo, gracias a un análisis más exhaustivo de la secuencia de su genoma, se ha podido comprobar que este pequeño genoma está organizado en cinco cromosomas, como resultado de la pérdida de genes y de múltiples arreglos de sus cromosomas. Así, este estudio permite afirmar que el número de genes diferentes en la *Arabidopsis* sólo es ligeramente mayor al

de la mosca de la fruta —13 mil 600— y menor que el calculado para el gusano nemátodo *Caenorhabditis elegans* —18 mil 424—. Por otro lado, la diversidad funcional entre los genes identificados parece ser muy similar a la encontrada en la mosca y en el *C. elegans*, ya que los genes de estos tres organismos se pueden clasificar en 11 mil tipos diferentes de proteínas. Esto indica que ésta es la complejidad mínima que se requiere para que diferentes organismos multicelulares se desarrollen de manera normal y respondan adecuadamente a los ambientes que los rodean. Resulta interesante el hecho de que a pesar de que la *Arabidopsis* comparte la gran mayoría de las familias de proteínas con estos organismos multicelulares, también presenta muchas familias de proteínas nuevas y carece de varias familias de proteínas comunes. Se ha especulado que el tamaño pequeño de los genes de la *Arabidopsis* representa una ventaja selectiva, ya que esto le ha permitido tolerar una mayor reorganización genómica. Entre más pequeños sean los genes, es menos probable que éstos se lleguen a interrumpir en este proceso. La presencia de genes compactos también parece ser común en otras plantas que poseen genomas de mayores dimensiones. Sin embargo, la distancia entre estos genes suele ser de diez a cien veces mayor que en el caso de la *Arabidopsis*.

¿Por qué la multiplicación de ciertos genes se ha seleccionado en las plantas? Aunque aún nos falta mucho por descubrir, la información genómica y genética nos habla de que muchos de los genes duplicados en algunas especies vegetales tienen funciones únicas. Esto se ha explicado a través de mutaciones que se han generado en las regiones regulatorias, de tal manera que se pueden tener genes duplicados que se expresen diferencialmente durante el desarrollo, o bien, en respuesta a cambios ambientales. También es posible que mutaciones en estos genes hayan dado como resultado proteínas con funciones ligeramente diferentes. Así se puede imaginar que la evolución de las plantas que florecen pudo haber comenzado con material genético duplicado, el cual se modificó y seleccionó paulatinamente para producir estructuras y compuestos nuevos, algunos de ellos específicos para cada especie.

Aproximadamente 40 por ciento de los genes de la *Arabidopsis* (y de otros genomas) todavía no tienen asignada una función. El descifrar esto puede tomar varios años,

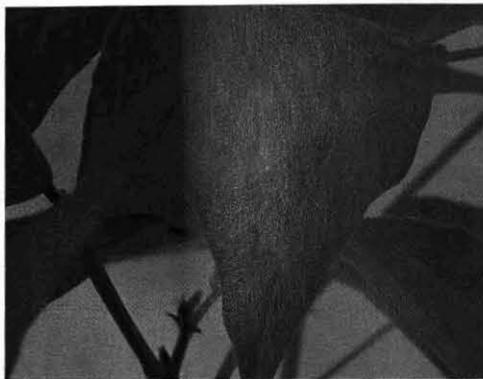
por lo que la genética, la fisiología y la bioquímica, en conjunción sinérgica, en lo que se conoce como la genómica funcional, es vital para alcanzar este objetivo

La revolución posgenómica, en el estudio de la *Arabidopsis* y sus mutantes, ha permitido entender procesos fundamentales como la respuesta de la planta al medio ambiente (luz, temperatura, sequía) y a microorganismos patógenos como los virus, las bacterias o los hongos. El perfil metabólico durante el desarrollo de la planta y en respuesta a cambios fisiológicos, aun durante un día en la vida de la planta, es ya una realidad. Con la secuencia del genoma de la *Arabidopsis* es posible identificar muchos de los genes que participan en los procesos arriba mencionados de una manera más fácil para posteriormente buscarlos en una planta cultivada de interés agrícola.

Las plantas producen sus propias vitaminas y cofactores, y son el primer punto de concentración de compuestos esenciales como el fosfato, el hierro, el zinc, el magnesio, el potasio y todos los otros nutrientes minerales necesarios en la dieta animal. Por otro lado, el metabolismo complejo de las plantas les permite sintetizar compuestos secundarios, los cuales no son esenciales para la vida de la planta, y muchas veces son específicos de género o especie. Se ha estimado que las plantas que florecen son capaces de producir por lo menos cien mil compuestos secundarios diferentes. Muchos de estos compuestos no se encuentran en animales y son fuente de colores, sabores, fragancias y de la mayoría de los medicamentos que usamos. Aun cuando no todas las plantas son capaces de sintetizar todos estos compuestos o metabolitos, en la *Arabidopsis* existen los genes necesarios para producir las moléculas precursoras de estos productos secundarios.

LA EVOLUCIÓN Y ESTABILIDAD DEL GENOMA VEGETAL

Desde hace varios milenios, la acción dirigida del hombre ha tenido un papel fundamental en la domesticación y propagación de muchas plantas. La era posgenómica ha ayudado a entender por qué los genomas de las plantas son tan dinámicos. Algunos pastos y cereales, como el trigo



(*Triticum aestivum*) y la cebada (*Hordeum vulgare*) son enormes, cinco y 1.5 veces mayores que el del genoma humano (tres mil 200 millones de pares de bases), respectivamente. Hace unos 50 años, la doctora Barbara McClintock, después de observar la pigmentación inestable de los granos de maíz, propuso que un tipo de ADN era capaz de brincar, produciendo efectos visibles al integrarse cerca o dentro de algunos

genes. Esto fue el concepto que mereció el premio Nobel y que ella denominó como "elementos de transposición". Si bien primero fueron descritos en las plantas, están presentes en todos los eucariontes y han tenido un papel fundamental en el diseño y evolución del genoma. Las ciencias genómicas han cambiado la manera de estudiar estos elementos (Feschotte *et al.*, *Nature Review Genetics*, 329-341, 2002) permitiendo el estudio de un gran número de ellos, así como rastrear su origen y evolución. La actividad reciente de los elementos de transposición es responsable de la variación en el tamaño de los genomas de plantas relacionadas entre sí, por lo que no resulta sorprendente encontrar que los elementos de transposición que inducen nuevas inserciones, constituyen una fracción muy pequeña de todos los elementos en el genoma. Los elementos de transposición pueden constituir hasta entre 50 y 80 por ciento del genoma de algunos pastos, de ahí la importancia de despertar algunos de estos "genios dormidos" bajo ciertas condiciones de estrés y en ciertos fondos genéticos mutantes, precisamente por su gran potencial de reestructurar y acelerar la evolución de los genomas de las plantas.

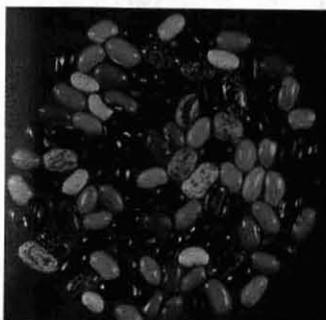
La disfunción completa de un gen es generalmente la manera más directa para entender cuál es su posible función. Ya que no existe un método de recombinación homóloga o de interrupción dirigida de un gen, como en la levadura y el ratón, la inserción al azar de elementos de transposición, como el *TDNA* y el sistema *cd/ds* del maíz, son las armas genéticas más poderosas con que se cuenta para generar tanto mutantes nulas como para activar genes, en lo que se conoce como las "trampas genéticas" (Page y Grossniklaus, 2002).

LA GENÓMICA COMPARATIVA

La genómica comparativa ha aportado grandes beneficios para entender el origen y evolución del genoma de la *Arabidopsis* y de otros genomas de plantas. Sin duda, las comparaciones previas que se hicieron entre el genoma de ésta y los de la levadura, la mosca de la fruta, un nemátodo y varios vertebrados han aportado datos funcionales de algo como 70 por ciento de los genes de la *Arabidopsis* a las distancias evolutivas que separan a estas especies. Las restringen la comparación a las regiones codificadoras exclusivamente.

Si se considera válido medir la importancia de la secuencia de un genoma por sus beneficios prácticos, sin lugar a dudas la secuencia del genoma del arroz (*Oryza sativa*), del cual depende la más de la mitad de la población mundial, ocupa un lugar destacado (Yu, *et al.*, *Science* 296, 80-92, 2002; Golf, *et al.*, *Science* 296, 92-100, 2002). Además, el arroz es un buen modelo para otros pastos, incluyendo los otros cereales importantes como el maíz y el trigo. Al hacer la comparación obligada con el genoma de la *Arabidopsis* se encontró que 85 por ciento de sus genes tienen sus homólogos correspondientes en el arroz. Sin embargo, sólo 50 por ciento de estos genes tienen homólogos en la *Arabidopsis*, sugiriendo un evento masivo de duplicación en la historia evolutiva del arroz. Alrededor de un tercio de los genes homólogos entre el arroz y la *Arabidopsis* son plantas específicas y aproximadamente 30 por ciento de los genes más altamente conservados entre ambas plantas no tienen asignada aún una función particular.

Otro ejemplo lo ofrece la reciente comparación del conjunto de genes únicos de una crucifera como la *Arabidopsis*, una solanácea como el tomate (*Lycopersicon esculentum*) y una leguminosa como la *Medicago truncatula*, lo que sugiere fuertemente que los genes muy conservados entre pares de especies tienen una alta probabilidad de estar conservados entre otras especies de plantas. Así, la genómica comparativa tiene aplicaciones valiosas para el mapeo de los genes en los cromosomas, para la generación de marcadores moleculares y para realizar estudios filogenéticos (Hall, *et al.*, 2002; Van der Hoeven, *et al.*, 2002).



Como se mencionó, el genoma de muchas plantas es tan grande que hace inviable en este momento obtener su secuencia completa. Tal es el caso del trigo, la cebada y el tomate. Si bien preguntas relevantes como la evolución de los genes y sus funciones no serán contestadas hasta que se cuente con la secuencia completa del genoma, por lo pronto la secuencia de los ARNs de genes expresados (ESTs) en todo el organismo (hoja, raíz, tallo, fruto, flor) per-

mite generar información muy valiosa. Por ejemplo, se puede tener un estimado del número de genes presentes en esa especie en particular y la comparación con el conjunto de genes únicos con otras plantas, lo que facilita encontrar sus equivalencias funcionales.

Del análisis de los diferentes grupos (11 mil familias) en los que se han clasificado los genes en el genoma de la *Arabidopsis*, se concluye que los genes biosintéticos y metabólicos son mucho más numerosos en esta planta que en otros organismos eucarióticos conocidos, los cuales comprenden aproximadamente 10 por ciento del genoma. Esto no resulta sorprendente si pensamos en que las plantas tienen que llevar a cabo múltiples transacciones metabólicas entre el mundo orgánico e inorgánico. Las plantas utilizan la energía solar para convertir el dióxido de carbono en azúcares, en polímeros de carbohidratos y en lípidos. También reducen los iones nitratos y sulfatos para sintetizar aminoácidos.

Conforme sea posible obtener la secuencia de los genomas de diferentes especies vegetales, podremos contar con bancos de datos que serán analizados para obtener información en cuanto a la historia evolutiva de una especie particular, o bien analizar los espectros de la diversidad entre ellas. Esto nos deja ante la posibilidad invaluable de contar con bancos de germoplasma y con información bien clasificada que nos permita explotar racionalmente nuestros recursos, así como planear el cuidado de los mismos.

LA GENÓMICA DEL FRIJOL EN LA UNAM

Para México debería ser de especial interés el desarrollo del proyecto genómico del frijol (*Phaseolus vulgaris*). El tamaño de su genoma (600 millones de pares de bases) es de los más pequeños de las plantas superiores de impor-

tancia socioeconómica, junto con el del arroz. El proyecto genómico del frijol representa para México una oportunidad única por varias razones fundamentales:

1) Es un cultivo estratégico, estrechamente ligado a nuestra cultura, ya que fue domesticado y se cultiva desde hace varios milenios en nuestro país.

2) México es centro de origen y de especiación importante, donde existe una gran diversidad de variedades, tanto cultivadas como silvestres, del género *Phaseolus*.

3) Desde hace varios años, existe en la UNAM un proyecto genómico pionero que estudia la bacteria que fija nitrógeno (*Rhizobium etli*) en simbiosis con el frijol.

4) Finalmente, existe la capacidad, tanto en la UNAM como en diversas instituciones de educación superior y organismos gubernamentales de nuestro país, de establecer grupos académicamente sólidos que estudian diversos aspectos tanto de la bioquímica, la fisiología, la genética y la biología molecular y celular del frijol.

La posibilidad de consolidar al frijol como un organismo modelo y de catalizar el estudio sistemático de la secuencia y expresión de su genoma tendrá un gran impacto en los programas de mejoramiento genético y de la preservación de la biodiversidad de este cultivo.

Éste es un proyecto detonador en el marco de las ciencias genómicas que se desarrolla en la UNAM.

EL FUTURO

El reto al que se enfrenta la comunidad científica en los próximos años es integrar la información generada de la secuencia del genoma, del transcriptoma, del proteoma, del metaboloma y fenoloma (el cambio fenotípico observado) de una célula, órgano, individuo en una condición patológica, estado fisiológico o del desarrollo particular. Ciertamente, estamos parados ante el umbral de un milenio donde la revolución de la era genómica sólo tiene la frontera de nuestra imaginación.

REFERENCIAS

- Aparicio, S., et al., *Whole-Genome Shotgun Assembly and Analysis of the Genome of Fugu rubripes*, en *Science* núm. 297, págs. 1301-1305, 2002.
- Baltimore, D., *Our Genome Unveiled*, en *Nature*, núm. 409, págs. 814-816, 2001.
- Feschotte, C., Jiang Ning, y Wessler, S., *Plant Transposable Elements: Where Genetics Meets Genomics*, en *Nature Review Genetics*, núm. 3, págs. 329-341, 2002.
- Goff, S.A. et al., *A draft sequence of the rice genome (Oryza sativa L. ssp japonica)*, en *Science* núm. 296, págs. 92-100, 2002.
- Hall, A.E., Fiebig, A., y Preuss, D., *Beyond the Arabidopsis Genome Opportunities for Comparative Genomics*, en *Plant Physiol*, núm. 129, págs. 1439-1447, 2002.
- Page, D. and Grossniklaus, U., *The Art and Design of Genetic Screens: Arabidopsis thaliana*, en *Nature Review Genetics*, núm. 3, págs. 124-136, 2002.
- The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana*, en *Nature*, núm. 408, págs. 796-815, 2002.
- Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G., y Tanksley, S., *Deductions about the Number, Organization, and Evolution of Genes in the Tomato Genome Based on Analysis of a Large Expressed Sequence Tag Collection and Selective Genomic Sequencing*, en *Plant Cell*, núm. 14, págs. 1441-1456, 2002.
- Yu, J. et al., *A draft sequence of the rice genome (Oryza sativa L. ssp indica)*, en *Science*, núm. 296, págs. 80-92, 2002.

