

LA BIOLOGÍA GENÓMICA EN EL CENTRO DE INVESTIGACIÓN SOBRE FIJACIÓN DE NITRÓGENO

José Guillermo Dávila Ramos*

EL SISTEMA BIOLÓGICO

Una agronomía menos perjudicial para los sistemas ecológicos parece ser una contradicción de términos ante una humanidad que necesita más alimentos, producidos más rápido y a menores costos. Es decir: en tanto que un cultivo natural aprovecha la diversidad propia de una especie, el cultivo artificial –la agricultura– la elimina al emplear monocultivos (sin diversidad). Mientras que el primero contienda con las adversidades climáticas y bióticas, acelerando o alentando sus ciclos reproductivos, nosotros eliminamos toda competencia con herbicidas, matamos cualesquiera de sus enemigos con plaguicidas y aceleramos su crecimiento usando fertilizantes, con el único fin de incrementar el rendimiento.

El problema radica en que el uso indiscriminado de estos compuestos altera el equilibrio ecológico nativo, precio pagado por el medio ambiente. En tanto que el

problema es cada vez más evidente, hasta el momento no ha surgido una solución integral.

Si nos enfocamos en el uso de fertilizantes, tal vez podríamos buscar la manera de nutrir estos cultivos con una mayor eficacia, es decir, sin que falte ni se desperdicie el fertilizante. Afortunadamente existen asociaciones naturales entre algunas plantas y microorganismos del entorno. De entre estas asociaciones, una de las mejor estudiadas es la simbiosis que realizan las plantas leguminosas con bacterias del suelo. Los microsimbiontes, bacterias del grupo de los rizobiales, interaccionan con las raíces de la planta (el macrosimbionte) penetrándolas y estableciéndose en una estructura globular, elaborada por la planta, llamada nódulo. En el interior de estos nódulos ocurre un intercambio de nutrientes; la planta le suministra a las bacterias compuestos hidrocarbonatos ricos en energía elaborados en la fotosíntesis, en tanto que las bacterias retribuyen este presente de carbono con nitrógeno.

Estas bacterias son capaces de captar el nitrógeno atmosférico (N_2 o N_2aN), que es muy abundante pero

* Investigador del CINF

no asimilable por los vegetales, y reducirlo químicamente para formar amoníaco (NH_3), que sí es utilizado por las plantas. El proceso, conocido como "fijación biológica del nitrógeno" (FBN), sólo ha sido descrito en organismos procarióticos (bacterias y archaeobacterias). Como consecuencia de esta asociación simbiótica entre las plantas leguminosas y esta clase de bacterias, suelos vírgenes pobres en nitrógeno pueden ser colonizados. Pero también es claro que cultivos de estas bacterias (seleccionadas o mejoradas) pueden ser desarrollados para usarse como inoculantes o fertilizantes biológicos naturales.

Las limitaciones más importantes del sistema son que no toda planta de interés agrícola puede realizar esta simbiosis y que la eficiencia del proceso no produce altos rendimientos. Del intento de mejorar la eficiencia de la FBN del microsimbionte y de extender el rango de plantas con las que pueda asociarse, se desprenden las dos líneas generales de investigación en el área.

EL GENOMA DE LOS PROTAGONISTAS

Toda la información que define a un ser vivo, estipulando sus capacidades de desarrollo adaptación y reproducción, está contenida en su genoma. Cualquier especie del planeta, pasada o presente, se caracteriza por un genoma *sui generis* diferente en contenido, extensión, organización, historia y contexto.

El genoma es ADN. El ADN es una cadena (polímero) cuyos peldaños están formados por cuatro nucleótidos o bases, representados por las letras A, T, G y C. La sucesión de estos nucleótidos es precisa para cualquier región del ADN, determinando el código primario de la cadena. Obtener el genoma de un organismo implica establecer la secuencia de bases del total de su ADN. Por tamaño, el rango de los genomas varía desde poco más de medio millón de bases (0.5 mega bases, Mb) para algunas bacterias parásitas, hasta muchos miles de Mb (en algunas plantas coníferas).

El primer genoma secuenciado fue el de una bacteria patógena (*Haemophilus influenzae*).¹ Desde entonces, cerca de 200 secuencias genómicas han sido publicadas, incluyendo la del genoma humano, de más de tres mil Mb.² Dada la impresionante cantidad de información que se genera por la secuencia misma, su

catálogo, caracterización y clasificación son almacenados en bases de datos. Asimismo, los sistemas de análisis globales se realizan con programas de cómputo, generados recientemente, que son actualizados y mejorados, lo que da lugar a una nueva disciplina de la computación conocida como bioinformática genómica.

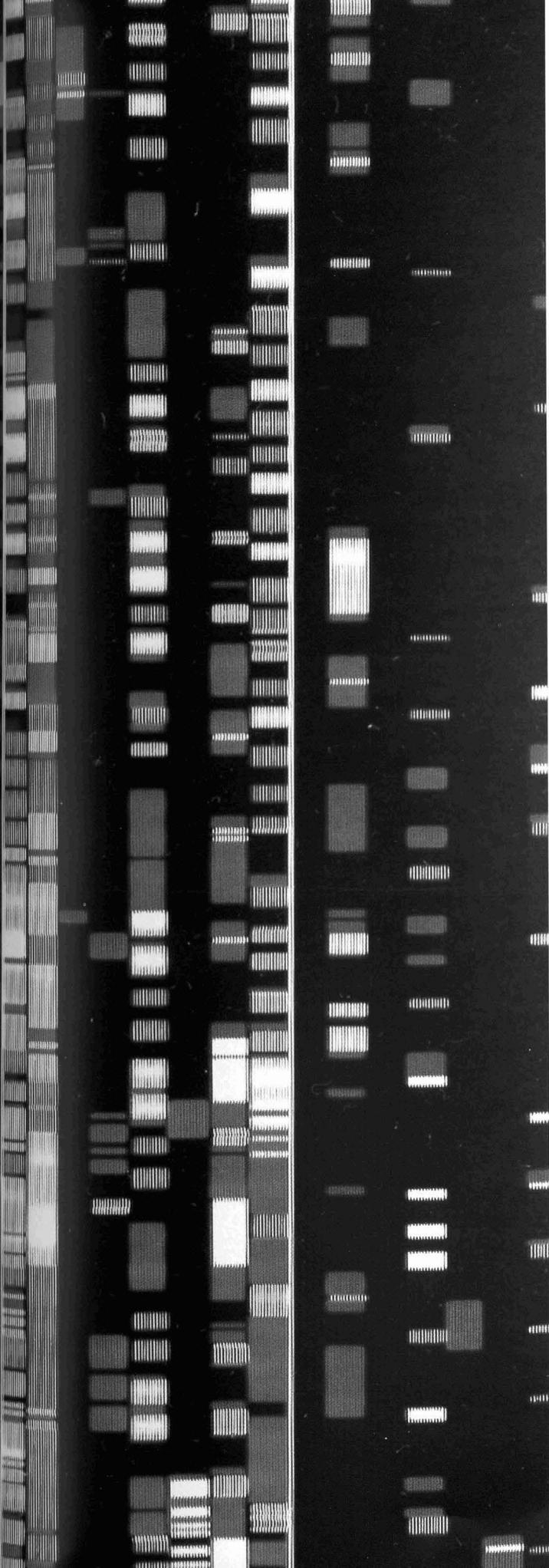
En nuestro país, la investigación sobre genomas se ha concentrado principalmente en la fase de análisis postsecuencia, en tanto que los proyectos de secuenciación genómica son emergentes. El proyecto "El desarrollo de las ciencias genómicas en México" propuesto por CIFN y apoyado por el Conacyt y la UNAM es, a mi entender, el primer proyecto nacional que pretende cubrir todos los aspectos de las ciencias genómicas, desde la obtención de la secuencia hasta su análisis funcional y evolutivo. El modelo biológico seleccionado fue el de la bacteria *Rhizobium etli*. Esta bacteria produce nódulos fijadores de nitrógeno que se asocian específicamente con las raíces del frijol (*Phaseolus vulgaris* L.). La planta es originaria de América, donde se diversificó y domesticó, por lo que es de esperarse que nuestros suelos cuenten con la población de *Rhizobium etli* con la mayor poza genética.

EL GENOMA DEL MICROSIMBIONTE

Rhizobium etli es una bacteria gram-negativa del suelo, móvil con forma de bacilo; sin embargo, lo que permite distinguirla del resto de las bacterias del entorno es su capacidad de invadir las raíces del frijol para formar nódulos fijadores de nitrógeno.

El genoma de *Rhizobium etli* está constituido por varias moléculas de ADN circular (replicones) que se duplican (replican) dentro de la célula de manera independiente. En general el replicón de mayor tamaño es el cromosoma, definido por contener la información de los genes necesarios para el mantenimiento celular. Los replicones restantes, denominados plásmidos, varían en número, contenido y tamaño.

Rhizobium etli CFN42 es la cepa tipo de la especie; tiene un cromosoma de alrededor 4.3 Mb y seis plásmidos nombrados de p42a a p42f, con tamaños que varían entre 180 y 600 kilobases (kb). El plásmido p42d de 371 kb contiene la mayor parte de la información indispensable para nodular y fijar nitrógeno en frijol (genes *nod*, *nif* y *fix*) por lo que se conoce como



plásmido simbiótico o pSim. El p42d ha sido estudiado en nuestro laboratorio desde principios de los noventa; de él determinamos su estructura, mapa físico, caracterizado sus genes simbióticos y determinado su plasticidad.³ En 1996 iniciamos el proyecto para la obtención de la secuencia completa de su ADN, que fue hecha pública en el 2002.⁴

EL PROYECTO GENÓMICO DE *RHIZOBIUM ETLI*

En el 2000 adquirimos el primer secuenciador automático de ADN de tercera generación. Con él iniciamos el proyecto para la obtención del genoma completo de *Rhizobium etli* CFN42 por secuenciación de fragmentos al azar. Organizamos nuestro laboratorio para preparar las más de 50 mil muestras de secuencia necesarias para lograr una cobertura de ocho veces. Se elaboraron dos tipos distintos de librerías genómicas: la primera, donde se clonan fragmentos muy grandes de ADN (ente cien y 200 kb) utilizando un vehículo especial llamado BAC (por sus siglas en inglés: Bacterial Artificial Chromosomes), que nos permite mantener y propagar las clonas independientes de la librería en células de *E. coli*, y la segunda, también con fragmentados al azar de segmentos más pequeños, con un tamaño promedio de 3.5 kb. La primera librería nos sirve en la elaboración de mapas generales tanto del cromosoma como de los dos plásmidos de mayor tamaño (p42e y p42f); con tal propósito se secuencian las regiones de los extremos del inserto de cada BAC. La segunda librería se usó para el proyecto de secuencia propiamente dicho.

La adquisición por parte de nuestra universidad de un segundo secuenciador de alto rendimiento a principios del presente año, aunado a la optimización de nuestros protocolos de preparación de muestras y reacciones de secuencia y la experiencia adquirida –durante el proyecto del plásmido simbiótico– para el ensamble en regiones contiguas (CONTIGS) de las clonas secuenciadas, nos permitió terminar la secuencia de los cinco plásmidos restantes en junio y alcanzar las ocho veces de cobertura requerida para el resto del genoma en el mes de agosto. El 12 de septiembre del presente año concluimos el ensamble final del cromosoma de esta bacteria. Éste es, a mi entender, el primer genoma completo de un organismo realizado por entero en nuestro país.

PERSPECTIVAS GENÓMICAS POSTSECUENCIA

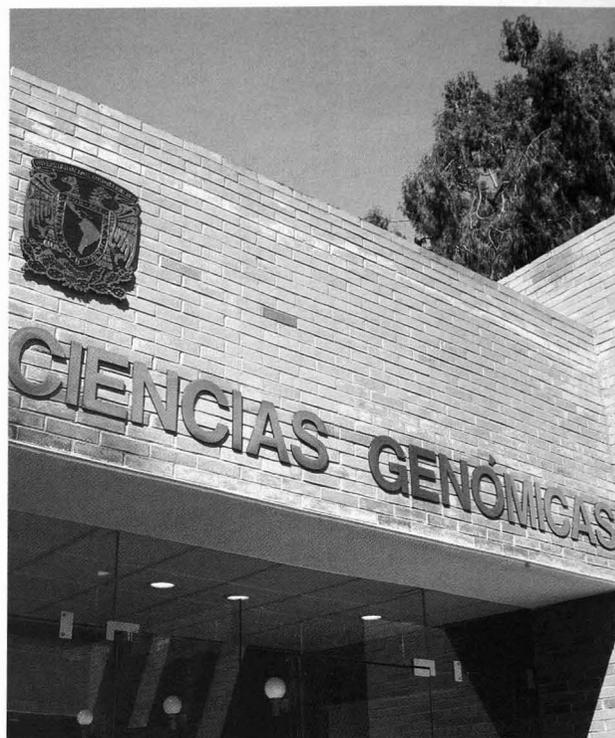
Una vez conseguido un ensamble final, es decir, un **CONTIG** cerrado para cada uno de los siete replicones del genoma de la cepa **CFN42**, se da inicio a las fases de refinamiento y análisis de la secuencia obtenida.

El refinamiento es el proceso que nos permite elevar la calidad y confiabilidad de la secuencia obtenida. A este respecto, los estándares internacionalmente aceptados demandan que no exista más de un error por cada diez mil bases de secuencia reportada, para así lograr resultados más precisos de los programas de búsqueda de regiones codificantes (**ORF's**), o de los utilizados para comparar nuestras secuencias. La asignación de genes y funciones para cada **ORF** da inicio al proceso de anotación del genoma con el que se pretende elaborar un listado de todo gen, real o putativo, de su tamaño, composición y función, y de todo sitio regulatorio relevante (promotores, terminadores, operadores, etcétera). Para cada elemento del listado se estipula una localización precisa de las coordenadas de la primera y última base del elemento. Por su misma naturaleza, el proceso de anotación tiene un principio pero no un fin, ya que es frecuente su actualización y corrección.

Un genoma anotado nos permite hacer un análisis de la expresión de la totalidad de sus genes, es decir, cuáles, cuántos y a qué nivel son copiados (transcritos) en moléculas de **ARN mensajero**; la metodología se conoce con el nombre de transcriptoma.

De igual forma, si lo que necesitamos es determinar la expresión global de las proteínas de un organismo, el genoma anotado nos permite predecir el tamaño, su secuencia de aminoácidos, el peso molecular exacto del polipéptido y su carga eléctrica (punto isoeléctrico). Con estos datos establecemos qué gen codifica para qué polipéptido y asignarle una posible función. En su conjunto, estas técnicas son conocidas como proteoma.

Por último, el genoma anotado nos permite hacer clasificaciones precisas del organismo y proponer hipótesis sobre su origen y tipo de evolución. ☉



NOTAS

- ¹ R. D. Fleishman *et al.*, "Whole genome random sequencing and assembly of *Haemophilus influenzae*", *Rd. Science*, núm. 269, 1995, págs. 496-512.
- ² E. S. Lander *et al.*, "Initial sequencing and analysis of the human genome", *Nature*, núm. 15, 2001, págs. 860-921; J. C. Venter *et al.*, "The sequence of the human genome", *Science*, núm. 291, 2001, págs. 1304-1351.
- ³ M. L. Girard *et al.*, "Structural complexity of the symbiotic plasmid of *Rhizobium leguminosarum* bv. *phaseoli*", *J. Bacteriol.*, núm. 173, 1991, págs. 2411-2419; D. Romero *et al.*, "Amplification and deletion of a *nod-nif* region in the symbiotic plasmid of *R. phaseoli*", *J. Bacteriol.*, núm. 173, 1991, págs. 2435-2441; A. García-de los Santos, S. Brom, S. y D. Romero, "Rhizobium plasmids in bacteria-legume interactions", *World J. Microbiol. Biotechnol.*, núm. 12, 1996, págs. 119-125.
- ⁴ V. González *et al.*, "The mosaic structure of the symbiotic plasmid of *Rhizobium etli* **CFN42** and its relation to other symbiotic genome compartments", *Genome Biol.*, núm. 4, 2003, rev. 36.