

Genómica y genómica funcional en microbiología

Sergio Encarnación-Guevara*

Vol. 48, No. 2
Abril - Junio, 2006
pp. 131 - 145

RESUMEN. La genómica funcional esta cambiando nuestra comprensión de la biología y cambiando nuestra manera de realizar la investigación en microbiología. Esta nueva área de la investigación concerta el uso de herramientas de genética de alta capacidad de procesamiento con el análisis de transcripción de RNAm, proteínas, y metabolitos para contestar a la última pregunta que se obtiene de todos los proyectos de secuenciación del genoma y que es: ¿cuál es la función biológica de cada gene? La genómica funcional esta produciendo un cambio en el paradigma de la investigación más allá del análisis de solo los genes, proteínas, o metabolitos hacia el análisis de cada uno de estos parámetros en una escala global. Identificando y midiendo varios, sino el grupo completo, de los actores moleculares que participan en un proceso biológico dado, la genómica funcional ofrece un panorama completo para obtener una representación verdaderamente aproximada de la vida. Los métodos de la genómica funcional no son necesariamente hipótesis-dependiente, sin embargo, ofrecen información acerca de expresión de RNAm, regulación genética, expresión de proteínas, localización de la proteína, e interacciones proteicas. En su principio, la biología inicio al observar a la naturaleza y al experimentar con sus piezas aisladas. La investigación en Genómica ahora genera nuevos tipos de observación compleja de la naturaleza. Este capitulo describe herramientas como (proteómica, transcriptómica), genómica comparativa y bioinformática, que están actualmente siendo utilizadas en el trabajo de genómica estructural y genómica funcional y considera además el impacto de esta nueva disciplina en la investigación en microbiología.

Palabras clave: Genómica estructural, genómica funcional.

INTRODUCCIÓN

Si la genómica estructural es la rama de la genómica orientada a la caracterización y localización de las secuencias que conforman el ADN de los genes, la genómica funcional consiste en la recolección sistemática de información sobre la función de los genes, mediante la aplicación de aproximaciones experimentales globales que evalúen la función de los genes haciendo uso de la información y elementos de la genómica estructural. Se caracteriza por la combinación de metodologías experimentales a gran escala con estudios computacionales de los resultados y también de la información obtenida me-

ABSTRACT. Functional genomics is changing our understanding of biology and changing our approach to biological research. It brings about concerted, high-throughput genetics with analyses of gene transcripts, proteins, and metabolites to answer the ultimate question posed by all genome-sequencing projects: what is the biological function of each and every gene? Functional genomics is stimulating a change in the research paradigm away from the analysis of single genes, proteins, or metabolites towards the analysis of each of these parameters on a global scale. By identifying and measuring several, if not the entire, molecular group of actors that take part in a given biological process, functional genomics offers the panorama of obtaining a truly holistic representation of life. Functional genomics methods are defined by high-throughput methods which are, not necessarily hypothesis-dependent. They offer insights into mRNA expression, protein expression, protein localization, and protein interactions and may cast light on the flow of information within signaling pathways. At its beginning, biology involved observing nature and experimenting on its isolated parts. Genomic research now generates new types of complex observational data derived from nature. This review describes the tools that are currently being used for functional genomics work and considers the impact that this new discipline on microbiology research.

Key words: Structural genomics, functional genomics.

dante la secuenciación de genomas (genómica comparativa y bioinformática).

Con la genómica funcional el objetivo es llenar el hueco existente entre el conocimiento de las secuencias de un gen y su función, para de esta manera desvelar el comportamiento de los sistemas biológicos. Se trata de expandir el alcance de la investigación biológica desde el estudio de genes individuales al estudio de todos los genes de una célula al mismo tiempo, en un momento determinado.

Genómica funcional, el papel de los estudios de la expresión génica

Estamos siendo testigos de un cambio notable en la escala de la investigación en microbiológica molecular y estamos entrando en la era de la gran ciencia. En la última década nos hemos movido del tiempo cuando los artículos de investigación eran basados totalmente en el análisis de un solo gene y su producto u operon para cambiar a artículos

* Programa de Genómica Funcional de Procaríotes. Centro de Ciencias Genómicas, Av. Universidad s/n., Cuernavaca Mor. E-mail: encarnac@cifn.unam.mx

que describen la secuencia de todo un genoma. La secuenciación de genomas microbianos avanza rápidamente, sobre todo si consideramos que varios cientos de ellos están en vías de secuenciación ([HTTP://www.tigr.org/tdb/mdb/mdbcomplete.html](http://www.tigr.org/tdb/mdb/mdbcomplete.html)). La disponibilidad de este nivel de información genética ha creado los términos genómica funcional, genómica comparativa, transcriptómica, proteómica, así como los que describen la aplicación de la mutagénesis masiva o en gran escala, el análisis de regulación mediante bioinformática etc. Sin embargo, esto también ha producido una preocupación de que este tipo de acercamientos al entendimiento de la función o regulación de los productos del genoma se mueven hacia análisis no conducidos por una hipótesis o lo que ha sido llamado peyorativamente ciencia conducida o dirigida por los datos o resultados. Es claro que la inferencia científica utiliza una combinación de razonamiento deductivo e inductivo. La genómica funcional nos permite hacer ahora nuevos e inesperados acoplamientos entre la función de genes sin relación y genes hasta ahora no caracterizados, y sugerir hipótesis las cuales pueden ser probadas posteriormente por métodos tradicionales de genética molecular y bioquímica (Hughes 2000). La investigación en genómica funcional puede ser llamada ciencia libre de hipótesis, pero también debe ser vista positivamente ya que revelara la función de muchos genes que no han podido ser definidos por análisis convencionales. La necesidad de este tipo de aproximaciones es justificable ya que por ejemplo la secuencia del genoma de *E. coli*, contiene aún 1632 (38%) genes de función desconocida los cuales permanecen en espera de ser funcionalmente caracterizados (Hinton 1997). La función de estos genes podría no ser definida sin el uso de aplicaciones de la tecnología de genómica funcional combinada con experimentos creativos.

¿Cuál es la importancia de la bioinformática en la genómica funcional y la genómica comparativa?

Integración es la palabra clave para entender la importancia de la bioinformática, ya que a través de herramientas y utilizando la información ya depositada en bases de datos alrededor del mundo estamos comenzando a descubrir relaciones no triviales escondidas en el código de la vida.

La bioinformática ha empezado a ocupar un papel central como punto de unión que une a diversas áreas de la ciencia. Actualmente, se necesitan herramientas para gestionar información genética en paralelo. Para ello se emplean nuevas tecnologías para extracción de conocimiento, minería de datos y visualización. Se aplican técnicas de descubrimiento de conocimiento a problemas biológicos como análisis de datos del genoma, proteoma o transcriptoma. La bioinformática, en este sentido, ofrece la capacidad de comparar y relacionar y aún más generar información genética con una finalidad deductiva, siendo capaz de ofrecer respuestas que no parecen obvias a la vista de los resultados de los experimentos. Todas estas tecnologías de genómica funcional vienen justificadas por la necesidad de tratar información masiva, no individual, sino desde enfoques celulares integrados (genómica funcional, proteómica, expresión multigénica).

REFERENCIAS

1. Hinton, J. C. (1997). The *Escherichia coli* genome sequence: the end of an era or the start of the FUN?. *Mol Microbiol*, 3, 417-422.
2. Hughes, D. (2000). Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. *Genome Biol*, 1, REVIEWS0006.

Diversidad genómica de *Rhizobium etli*: el genoma completo y su variación intraespecífica

Víctor González,* Rosa I. Santamaría, Patricia Bustos, José L. Acosta, José L. Fernández, Ismael Hernández-González, Luis Lozano, Santiago Castillo, Guillermo Dávila

* Centro de Ciencias Genómicas, UNAM. Av. Universidad N/C Col. Chamilpa, Cuernavaca, Mor., 62210.

Apdo. Postal 565-A. vgonzal@ccg.unam.mx

INTRODUCCIÓN

Las bacterias capaces de establecer simbiosis fijadoras de nitrógeno con leguminosas pertenecen en su

mayoría a las α -proteobacterias, y a algunas especies de las β -proteobacterias (Moulin et al., 2001). Recientemente nuestro grupo publicó la secuencia completa de *Rhizobium etli* CFN42, una bacteria simbiótica fijadora de

nitrógeno asociada con el frijol (González et al., 2006). El genoma completo de otros microorganismos simbióticos fijadores de nitrógeno como *Mesorhizobium loti*, *Sinorhizobium meliloti*, *Bradyrhizobium japonicum* y *Rhizobium leguminosarum* también ha sido determinado (Galibert et al., 2001; Goodner et al., 2001; Kaneko et al., 2000; Kaneko et al., 2002; Wood et al., 2001; Young et al., 2006). En esta comunicación revisamos las principales características del genoma de *R. etli* y presentamos resultados preliminares de la comparación del genoma de *R. etli* CFN42 con secuencias genómicas parciales de 6 cepas de *R. etli* "bonafide" y la estimación de la variación genómica entre ellas.

EL GENOMA DE *R. etli* CFN42

El genoma de *R. etli* CFN42 consiste de un cromosoma circular y seis plásmidos (Tabla 1). El cromosoma contiene la mayoría de los genes esenciales para la vida celular, mientras que los plásmidos codifican funciones auxiliares en muchos casos desconocidas. El plásmido simbiótico o p42d contiene los genes más importantes para realizar la simbiosis. Hay un plásmido conjugativo (p42a) que contiene todos los genes *vir*, necesarios para transferir el T-DNA comúnmente encontrados en los plásmidos tumorigénicos de *Agrobacterium*, pero carece de T-DNA. Por lo tanto no es infectivo. En el resto de los plásmidos, excepto en el plásmido p42e que tiene los genes *minCDE* necesarios para la división celular, no hay genes absolutamente indispensables. Estos plásmidos codifican para diversas funciones como para la síntesis de tiamina y cobalamina, utilización de nitrito, degradación de catecol y protocatecuato, entre otras.

Como en otras Rhizobiales, el genoma de *R. etli* tiene un elevado número de transportadores de diversos metabolitos y reguladores transcripcionales, así como 23 subunidades sigma. Muchos de estos genes corresponden

a familias de parálogos de origen reciente o ancestral, que suman 27% del contenido total de genes de *R. etli* (6,034). Esto indica que *R. etli* tiene una gran redundancia genética, que a la vez se traduce en plasticidad metabólica. En *R. etli* hay una gran cantidad de genes que codifican isozimas, repartidos en el cromosoma y en los plásmidos.

EVOLUCIÓN GENÓMICA

El contenido promedio de GC del genoma de *R. etli* es de 61.5%, sin embargo los plásmidos p42a y p42d tienen un GC promedio de 58%. Estos plásmidos también son diferentes al resto del genoma en el uso de codones y en la firma de varios dinucleótidos. Además, tanto el p42a como el p42d pueden ser transferidos por conjugación. Así mismo, la presencia de un gran número de secuencias de inserción en estos plásmidos y en el cromosoma, pero no en el resto de los plásmidos, hace suponer que no han tenido una historia conjunta con el resto del genoma, y que son adquisiciones recientes del genoma.

La comparación del genoma de *R. etli* con otros genomas de Rhizobiales disponibles, revela una gran conservación entre los cromosomas pero no con los plásmidos. Las especies más relacionadas con *R. etli* son *S. meliloti* y *A. tumefaciens*, y en menor grado *M. loti* y otras bacterias no simbióticas como *Brucella*. Estas comparaciones muestran una estructura en mosaico del cromosoma de *R. etli*, con regiones altamente conservadas y regiones variables. En estas regiones variables se localizan genes de origen externo como ISs, algunos genes de fagos y de plásmidos. También las regiones variables contienen muchos de los genes relacionados con la síntesis de la envoltura externa de la bacteria, como son lipopolisacáridos y exopolisacáridos. Algunos de estos genes están relacionados con la nodulación (González et al., 2006).

Tabla 1. Características generales del genoma de *Rhizobium etli* CFN42. ¹ Número de secuencias de un promedio de 600 bases, empleadas para el ensamble final de cada molécula. ² El tamaño es indicado en pares de bases (pb). Datos de acuerdo a González, et. al. 2006.

Molécula	Secuencias ¹	Tamaño ²	Genes	Relevancia
Cromosoma	117,596	4,381,611	4,067	Funciones celulares esenciales
Plásmido F (p42f)	13,476	642,519	573	Transportadores tipo IV (<i>vir</i>), conjugación
Plásmido E (p42e)	8,737	505,334	459	Síntesis de tiamina y lipopolisacáridos
Plásmido D (p42d)	6,210	371,255	354	Transportadores ABC (azúcares y aminoácidos)
Plásmido C (p42c)	4,759	250,948	234	Simbiosis (nodulación y fijación de nitrógeno)
Plásmido B (p42b)	3,215	184,338	165	Síntesis de cobalamina, división celular
Plásmido A (p42a)	4,218	194,229	182	Fijación de nitrógeno, asimilación de nitritos
Totales	117,596	6,530,234	6,034	

Tabla 2. Variación nucleotídica de distintas cepas de *R. etli* comparadas con el genoma completo de *R. etli* CFN42.

Cepa	Origen	Secuencias	Sustituciones	Indeles	Cambios/kb
S20	España	8,436	8,264	464	14.5
brasil5	Brasil	7,654	15,460	655	19.9
kim5	USA	9,524	4,941	340	11.9
CIAT894	Colombia	8,436	8,575	439	14.3
GR56	España	12,628	5,904	532	10.5
IE4771	México	14,613	10,239	649	20.9

DIVERSIDAD GENÓMICA

Estudios pioneros mediante electroforesis de isozimas demostraron que *R. etli* esta formada por una población genéticamente muy diversa (Piñero et al., 1988). Para conocer los niveles de variación molecular en *R. etli* así como para determinar las tasas de evolución en este linaje, hemos comparado la secuencia completa de *R. etli* CFN42 con secuencias genómicas parciales de 6 cepas de *R. etli* de diferente origen geográfico (Tabla 2). Un 77% de las secuencias obtenidas resultó ser homóloga a *R. etli* CFN42, mientras que el resto representa secuencias huérfanas (10%) o secuencias con homólogos en el Genbank pero no con *R. etli* CFN42 (13%). Además, encontramos que la tasa de sustitución nucleotídica varía de 10 a 20 sustituciones por kilobase. Las sustituciones nucleotídicas se distribuyen al azar en el genoma, pero existen regiones altamente polimórficas que probablemente se originaron por recombinación.

CONCLUSIÓN

La gran redundancia genética y metabólica encontrada en *R. etli* correlaciona con su estilo de vida en el suelo o como simbiote. La simbiosis, de acuerdo a nuestros datos no es un carácter ancestral, sino adquirido. La conservación cromosomal entre especies relacionadas con *R. etli*, permite suponer un cromosoma ancestral común a todas ellas, pero no a los plásmidos. Estos pudieron ser generados independientemente a lo largo de la diferenciación de las especies. En la evolución de *R. etli*, las duplicaciones, las pérdidas y ganancias de genes han tenido un papel preponderante. El análisis de la variación genómica indica que probablemente la recombinación juegue un papel importante en la coherencia de *R. etli* como especie, contribuyendo a la vez a expandir la diversidad genética.

AGRADECIMIENTOS

Los autores agradecen la participación técnica de José Espíritu y Víctor del Moral. Este trabajo es apoyado por donativos del PAPIIT IN223005-3 y CONACYT U46333-Q.

REFERENCIAS

- Galibert, F., T.M. Finan, S.R. Long, A. Pühler, P. Abola, F. Ampe, et al. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science. 293:668-72.
- González, V., R.I. Santamaría, P. Bustos, I. Hernandez-González, A. Medrano-Soto, G. Moreno-Hagelsieb, et al. 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc Natl Acad Sci USA. 103:3834-9.
- Goodner, B., G. Hinkle, S. Gattung, N. Miller, M. Blanchard, B. Qurollo, et al. 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science. 294:2323-8.
- Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, et al. 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. 7:331-8.
- Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiyumi, S. Sasamoto, et al. 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110 (supplement). DNA Res. 9:225-56.
- Moulin, L., A. Munive, B. Dreyfus, and C. Boivin-Masson. 2001. Nodulation of legumes by members of the beta-subclass of Proteobacteria. Nature. 411:948-50.
- Piñero, D., E. Martínez, and R.K. Selander. 1988. Genetic diversity and relationships among isolates of *Rhizobium leguminosarum* biovar *phaseoli*. Appl Environ Microbiol. 54:2825-32.
- Wood, D.W., J.C. Setubal, R. Kaul, D.E. Monks, J.P. Kitajima, V.K. Okura, et al. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science. 294:2317-23.
- Young, J.P., L.C. Crossman, A.W. Johnston, N.R. Thomson, Z.F. Ghazoui, K.H. Hull, et al. 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol. 7:R34.

Genómica funcional de Rizobiaceas

Sergio Encarnación,* Emmanuel Salazar, Ángel Gabriel Martínez-Batallar, Magdalena Hernández, Agustín Reyes-Pérez, Sandra Contreras, María del Carmen Vargas, Rocío Domínguez-Vidaña, Claudia Gonzaga-Jáuregui, Yolanda Mora, Ma. Ricarda Rivero, Jaime Mora.

Programa de Genómica Funcional de Procaríotes. Centro de Ciencias Genómicas, Av. Universidad s/n., Cuernavaca Mor.
E-mail: encarnac@cifn.unam.mx.

INTRODUCCIÓN

Existen dos herramientas de investigación que permiten estudiar la expresión global del genoma, estas herramientas son el transcriptoma y el proteoma, estas metodologías estudian los productos del genoma que son primero, los RNAs mensajeros y posteriormente las proteínas lo que se ha llamado trasciptoma y proteoma respectivamente.

La PROTEÓMICA es el estudio de los proteomas. Esta configura una disciplina fundamental de la era post-genómica que trata de descubrir el grupo de proteínas que dan a las células su estructura y función.

Distintas tecnologías permiten obtener y comparar “instantáneas” de las proteínas que se están expresando en un momento determinado en una célula (electroforesis 2D, espectrometría de masas, chips, bioinformática). Este análisis puede incluir a todas las proteínas que se expresan en una célula en un determinado momento, ya sean las isoformas como las proteínas modificadas e incluso las interacciones proteicas.

Hace algunos años el protocolo básico para la electroforesis en geles de doble dimensión (2-DE) fue descrito (O'Farrel, 1975). Este protocolo es aún válido actualmente. Nosotros usamos una combinación de 2-DE con espectrometría de masas (MS) para revelar los patrones de síntesis proteica perteneciente a estímulos o regulones en *R. etli* (Encarnación *et al.*, 2002; 2003).

Los estímulos son estudiados comparando el perfil de expresión en una condición control contra una condición problema, mientras que los regulones son usualmente identificados usando cepas en las cuales el gene que codifica para la proteína regulatoria está mutado y es examinada en condiciones específicas, comparando el perfil de expresión proteica con el de la cepa silvestre (Encarnación *et al.*, 2005).

El protocolo utilizado para 2-DE, y la metodología de MS, así como las estrategias relacionadas con la preparación de la muestra, electroforesis y visualización proteica (aparatos y reactivos), identificación mediante espectrometría de masas (procedimiento de digestión en gel) y su posterior identificación mediante MALDI-TOF-MS y/o

LC-ESI-MS/MS, son mostrados en las Figuras 1 y 2 y discutidos ampliamente en Encarnación *et al.*, 2005.

Identificación de proteínas

Las proteínas seleccionadas fueron sujetas a digestión enzimática (Encarnación *et al.*, 2005) y los fragmentos péptidos resultantes fueron analizados usando los métodos complementarios de espectrometría de masas MALDI-TOF-MS y LC-ESI-MS/MS. Cada uno de estos métodos pueden ser usados independientemente para el análisis de proteínas, pero cuando un método masivo de análisis como el MALDI-TOF-MS es utilizado en combinación con LC-ESI-MS/MS se facilita enormemente la identificación de proteínas incluso las no reportadas en las bases de datos. (Encarnación *et al.*, 2005).

El primer paso para la identificación proteica es utilizando MALDI-TOF-MS, un espectro representativo del tipo de resultados que se pueden obtener, es mostrado en la Figura 3. La lista de masas generadas en el MALDI-TOF-MS (Autoflex, Bruker Daltonics, Billerica MA) fueron usadas para identificar las proteínas en las bases de datos existentes. En este ejemplo la proteína fue identificada exitosamente como GroEL.

Cuando la información obtenida mediante MALDI-TOF-MS es insuficiente para la identificación de la proteína, la obtención de la secuencia de un péptido de esta se vuelve necesario (Encarnación *et al.*, 2005). Este es el caso del ejemplo de la proteína 2, la cual demostró un buen espectro de masas mediante MALDI-TOF-MS (Fig. 4). Sin embargo la proteína no pudo ser identificada y el LC-ESI-MS/MS fue utilizado. En este proceso dos péptidos fueron seleccionados mediante cromatografía líquida (Fig. 5A) y posteriormente éstos fueron fragmentados en una cámara de colisión mediante disociación activa. Entonces las proteínas pueden ser identificadas ya que la secuencia de un péptido desconocido puede ser determinada de los resultados obtenidos de los patrones de fragmentación. Un ejemplo representativo es el presente en las Figuras 5B y 5C en las cuales se muestran los datos obtenidos

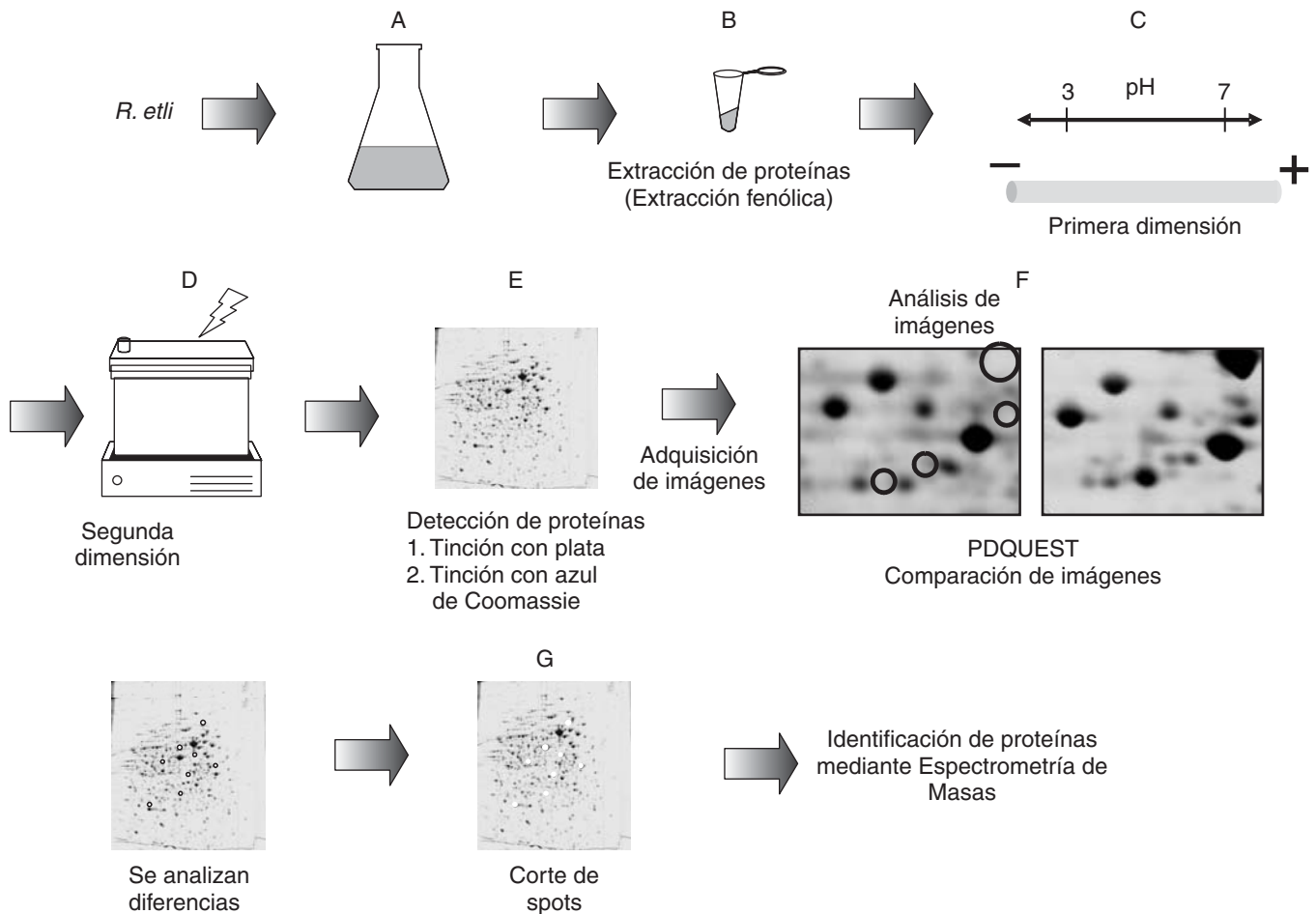


Figura 1. Estrategia proteómica sugerida en este capítulo para identificar proteínas. Se cultivan las cepas bacterianas a ser comparadas (A), después de colectar las células se extraen las proteínas (B). El extracto de proteínas es aplicado a la primera dimensión en un gel en tubo que separa las proteínas con base a su punto isoeléctrico (C). Después de este paso, el gel es aplicado a un gel plano para la segunda dimensión donde las proteínas son desnaturadas y separadas con base a su tamaño (D). Los gels son fijados y teñidos con azul de coomassie o tinción con plata (E). El mapa de proteínas resultante es adquirido y cuantificado (F). Las proteínas de interés son cortadas del gel (G) y sujetas a identificación mediante espectrometría de masas.

mediante el análisis utilizado LC-ESI-MS/MS, en donde se observan las secuencias logradas con los patrones de fragmentación usados para identificar esta proteína exitosamente.

Por estas razones la espectrometría de masas ha sido ampliamente reconocida como la piedra angular de la investigación en proteómica. Ya que es altamente sensible (a nivel femtomolar) y puede ser utilizada para hacer un análisis masivo de muestras. Otro aspecto clave es la compatibilidad que existe entre la separación proteica utilizado electroforesis en gels de doble dimensión y estas técnicas, adicionalmente, la espectrometría de masas puede ser usada en estudios que analizan complejos proteicos y modificaciones postraduccionales. (Porubleva *et al.*, 2001).

Resultados obtenidos mediante proteómica

En este proyecto hemos podido identificar proteínas ("spots") detectables en gels de doble dimensión (2-D) en condiciones de metabolismo aeróbico fermentativo y simbiosis entre los rangos de pH 3–10. Con esta estrategia hemos logrado identificar más de 1,000 proteínas llegando a conclusiones importantes acerca de estos metabolismos en *R. etli* y *S. meliloti*.

Usando la combinación mencionada, de electroforesis en gels de doble dimensión, espectrometría de masas y bioinformática, nuestra meta fue identificar:

(i) proteínas específicas, expresadas durante el metabolismo aeróbico, fermentativo y en la simbiosis *R. etli*-*Phaseolus vulgaris* (frijol).

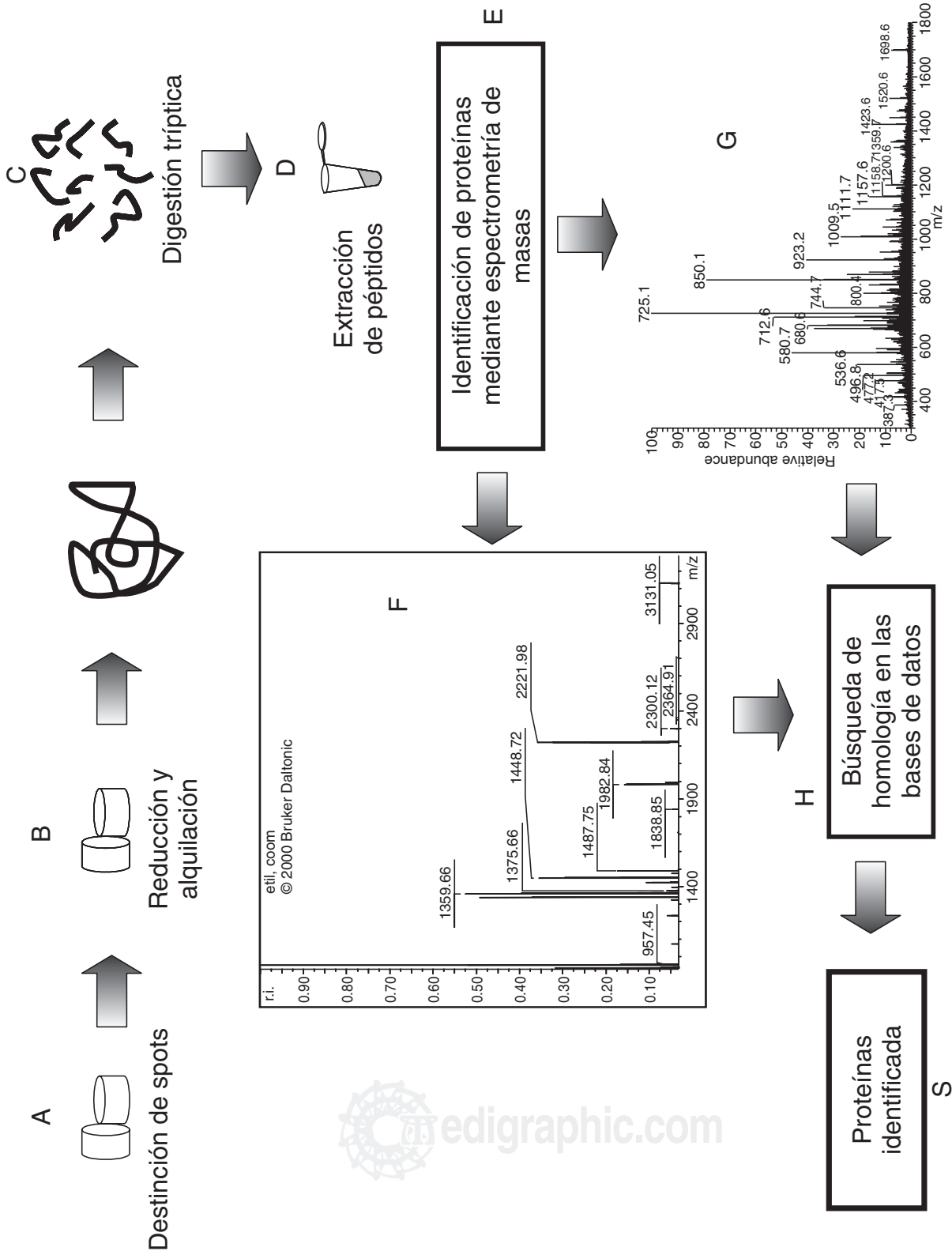


Figura 2. Diagrama esquemático de la identificación de proteínas mediante espectrometría de masas. Después de la separación bidimensional, las proteínas son recortadas y sujetas a digestión (A), reducción y alquilación (B), previo a la digestión con tripsina (C). El volumen fue ajustado a ~20 µl (D), en este punto la muestra está lista para su identificación mediante espectrometría de masas (E). Una alícuota del sobrenadante conteniendo los péptidos tripticos es analizada mediante MALDI-TOF-MS, del cual resulta en mapa peptídico de la proteína; el panel muestra los picos del espectro de masas (F). El remanente del sobrenadante es analizado mediante ESI-MS. La figura muestra la secuencia derivada de la fragmentación del mismo péptido (G).

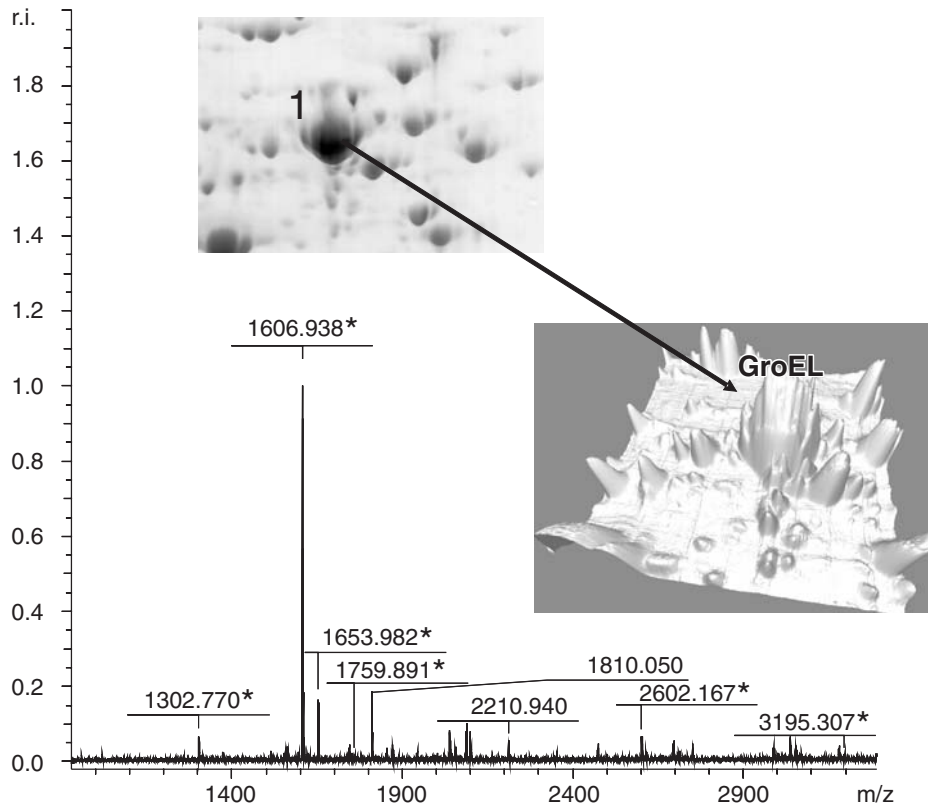


Figura 3. Espectro de masas de la digestión de la proteína no. 1, obtenido mediante MALDI-TOF. Los picos que dan homología en la base de datos con GroEL están marcados con asteriscos. El recuadro muestra la proteína analizada.

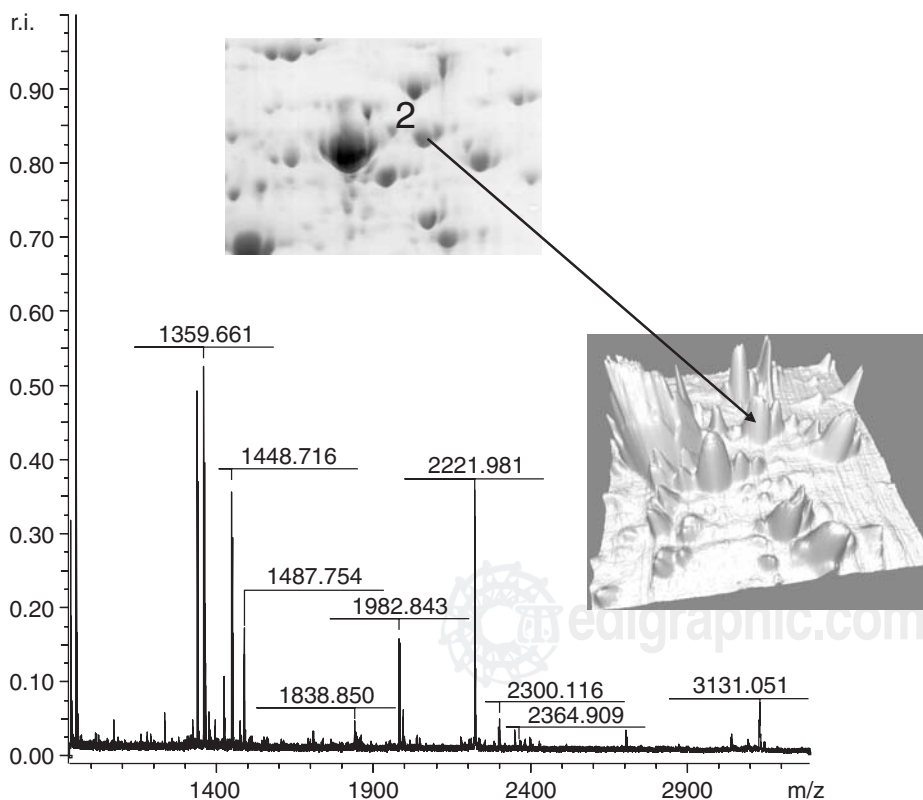


Figura 4. Espectro de masas de la proteína no. 2. En este caso la información obtenida mediante el espectrómetro de masas MALDI-TOF fue de buena calidad pero no suficiente para encontrar homólogos en las bases de datos por lo que no fue posible la identificación, fue necesaria un subsiguiente análisis mediante LC-ESI-MS/MS. El recuadro muestra la proteína analizada de un gel teñido con azul de Coomassie coloidal.

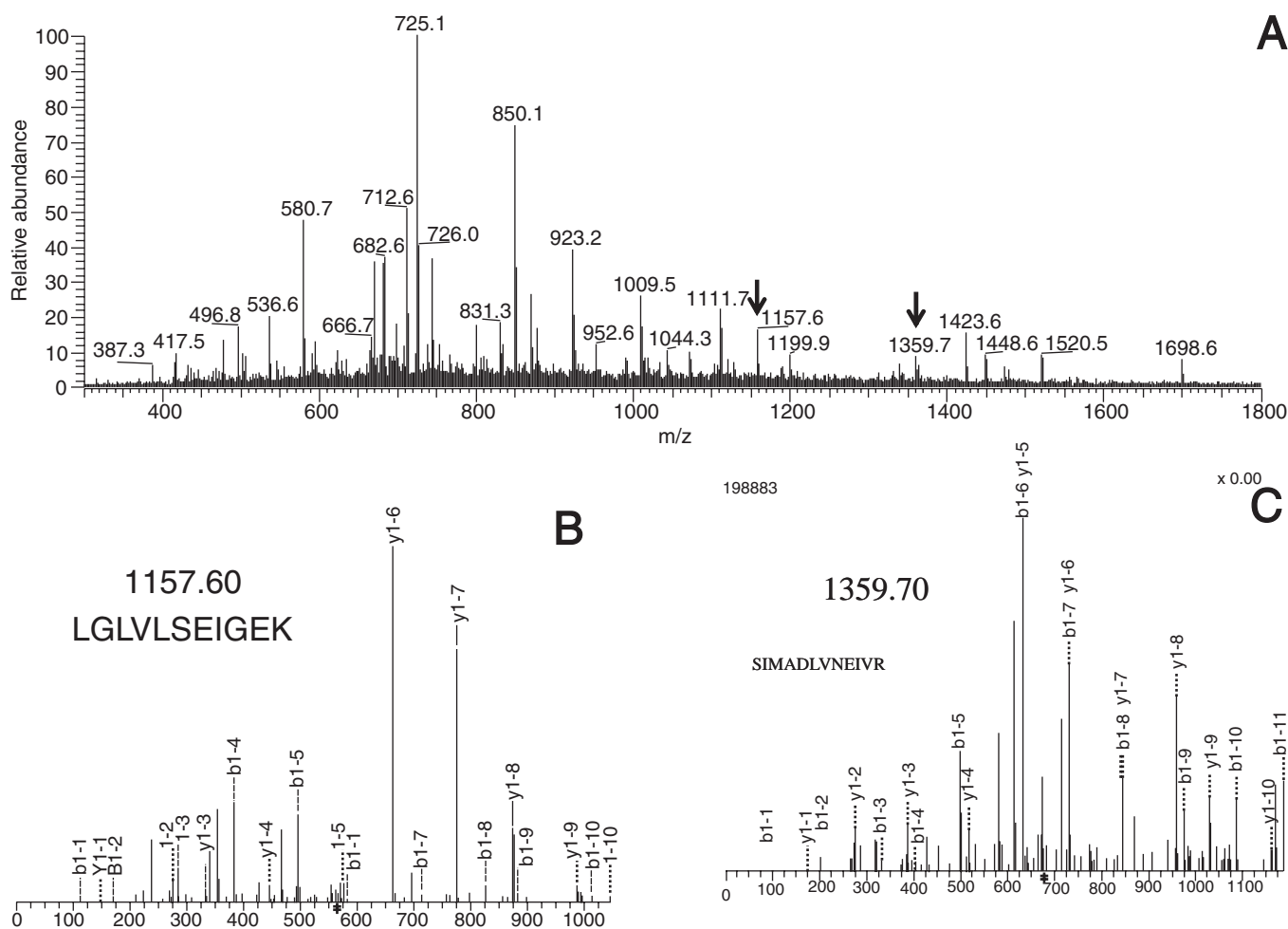


Figura 5. (A) Espectro completo de la proteína no. 2, mostrando los iones precursores analizados mediante LC-ESI-MS/MS. Dos picos fueron seleccionados para MS/MS para su posterior identificación vía colisión inducida. (B) y (C) El espectro de colisión inducida (ESI-IT-CID) de la misma proteína mostrando los iones correspondientes a la serie $-y$, y $-b$ y la secuencia del péptido. La búsqueda en las bases de datos con la información de esta fragmentación resulta en la identificación de la secuencia de los péptidos LGLVLSEIGEK y SIMADLVNEIVR como una proteína "trigger factor".

(ii) las rutas metabólicas activas durante las diferentes condiciones estudiadas.

Usando estas definiciones, 45 rutas metabólicas involucradas en procesos celulares anabólicos y catabólicos del metabolismo de pequeñas moléculas podrían estar consideradas a estar activas en ambas bacterias bajo las condiciones examinadas. Estas proteínas están presentes en:

(i) biosíntesis (de aminoácidos, cofactores, ácidos grasos, nucleótidos); (ii) funciones globales (ATP-síntesis) e intermediarios del metabolismo central (gluconeogenesis, vía de las pentosas, fijación de nitrógeno); (iii) transferencia de energía; (iv) transporte de electrones; (v) metabolismo energético; (vi) metabolismo de carbono (fermentación, respiración aeróbica, ciclo de ATC, glicólisis) y degradación (aminoácidos, compuestos de carbono).

Muchas vías esenciales fueron encontradas activas, incluyendo vías de síntesis de ácidos nucleicos, biosíntesis y catabolismo de aminoácidos, de síntesis de varias vitaminas (tiamina, riboflavina, B6, pantotenato, biotina, folato, y B12), de macromoléculas como peptidoglucanos, lipopolisacáridos, RNA, DNA, aminoacil-tRNAs, y proteínas ribosomales.

Siete enzimas del Ciclo de ATC fueron identificadas en vida libre y simbiosis. No es posible determinar si el ciclo está funcionando completamente, sin embargo la presencia de estas enzimas sugiere que ese es el caso.

Así mismo, la bacteria también expresa proteínas involucradas en la detoxificación de especies reactivas al oxígeno tales como SodB (superóxido dismutasa) y hp0014 (peroxiredoxina). Dado que la nitrogenasa es extremada-

mente sensible, la inactivación de especies reactivas del oxígeno es fundamental en la simbiosis.

Isoformas, modificaciones post-traduccionales

Aproximadamente el 28% de las proteínas presentes en 2D-geles fueron localizadas en al menos dos distintas posiciones en el gel, indicando variación en carga y en algunos casos en tamaño. Es común que muchas de estas proteínas sean modificadas por fosforilación lo que obliga al estudio del fosfoproteoma en estas bacterias. (ej: de proteínas con más de una posición electroforética fueron; aconitasa, cadenas alfa y beta de ATP sintasa, cisteína sintasa, DnaK, factor de elongación P, Ts y Tu, enolasas, GroEL2 GroEL1 y GroEL5, proteína de enlace a GTP (TypA), isocitrato deshidrogenasa, lipoamido deshidrogenasa, proteína ribosomal S1, cadena alfa de la succinil-CoA sintetasa.

Proteínas regulatorias

Contrariamente a la idea de que la electroforesis en geles de doble dimensión (2-DE) es una herramienta insuficiente para detectar proteínas de baja abundancia, en nuestras condiciones fue posible identificar como reguladores transcripcionales a 52 proteínas.

CONSTRUCCIÓN DEL MICROARREGLO DE ADN DEL GENOMA TOTAL DE *RHIZOBIUM ETLI* Y SU USO EN ESTUDIOS DE EXPRESIÓN GLOBAL EN VIDA LIBRE Y DURANTE LA SIMBIOSIS CON *PHASEOLUS VULGARIS*

A partir de la secuencia completa de *R. etli* (Gonzalez et al., 2006) trabajamos en el diseño de secuencias de oligonucleótidos de 70 bases que cubrieran los 6,034 marcos abiertos de lectura (ORF's) que contiene el genoma de *R. etli*, así como de sus respectivos controles. Las características requeridas de cada oligonucleótido fueron: tamaño (70 bases), temperatura de fusión ($TM = 80^{\circ} C \pm 5^{\circ} C$), hibridación cruzada (homología $\leq 70\%$), sobrelape de bases continuas entre los diferentes oligos (bases ≤ 20), sobrelape de bases entre el mismo oligonucleótido ("loops" ≤ 7), distancia después del 5' y antes del 3' (>50 bases). Para este "oligo set" diseñamos los controles negativos a partir de genomas de organismos no relacionados filogenéticamente como *Drosophila sp*, como positivos a genes que codifican para proteínas ribosomales y que se expresan "constitutivamente". Como números totales tenemos que de los 6,034 oligonucleótidos diseñados, 5,776 (95.73%) cumplen con todos los criterios solicitados, números similares son los que presenta el "oligo set" de *S. meliloti*, bacteria relacionada filogenéticamente a *R. etli*. Cada uno de los 6,034 oligonucleótidos además de los controles se en-

viaron a sintetizar a la compañía MWG Biotech finalmente cada oligonucleótido fue impreso en un portaobjetos ("slide"). Con esta herramienta nos planteamos realizar un análisis de expresión global del genoma de *R. etli* utilizando la metodología del transcriptoma en diferentes condiciones de cultivo.

La expresión del genoma de *R. etli* en 2 condiciones muy contrastantes (simbiosis y vida libre) mostró que 1,857 genes fueron expresados exclusivamente en medio rico durante la fase exponencial y 1255 genes fueron expresados solo durante la simbiosis entre *R. etli*-*P. vulgaris* a los 11 días posteriores a la inoculación (ddi) (manuscrito en preparación).

Adicionalmente, nuestro interés radico en conocer de manera global el papel de algunos reguladores transcripcionales de *R. etli*, para cumplir este objetivo utilizamos la secuencia del genoma completo de *R. etli*, en el cual mediante un análisis *in silico* buscamos las regiones teóricas consenso, de pegado para NifA y RpoN. Correlacionando estos datos obtenidos utilizando el microarreglo de *R. etli*. Hibridamos el cDNA sintetizado de la mutante NifA- a los 11 ddi y el cDNA proveniente de la cepa silvestre de *R. etli* a los 11 ddi. De este análisis logramos generar el regulón de RpoN-NifA el cual esta constituido por 66 genes de los cuales 54 pertenecen al plásmido simbiótico y 12 están repartidos en el resto del genoma (el detalle de los resultados de este análisis se encuentran en un manuscrito en preparación).

Proteoma y transcriptoma metodologías no excluyentes y si complementarias

Podemos afirmar que si bien hay un grupo de proteínas presentes cuya señal de mensajero no fue detectada, tenemos correlación de otro grupo cuyo transcrito (RNA mensajero) y su producto (proteína) fueron detectados en la misma condición, por ejemplo en simbiosis los transcritos de los genes *cpxP2*, *nifD*, *nifK*, *fixA*, entre otros, presentaron expresión en esta condición y su producto (proteína) también fue detectado sólo durante la simbiosis. Esta misma correlación la encontramos en algunos genes que detectamos su expresión en medio rico durante la fase exponencial a nivel de RNA mensajero y proteína.

Con los resultados obtenidos mediante de transcriptoma o proteoma podemos establecer diferencias o similitudes metabólicas presentadas por *R. etli* y *S. meliloti* en los diferentes metabolismos estudiados, así como en la simbiosis *R. etli*-*P. vulgaris*.

AGRADECIMIENTOS

Este trabajo fue parcialmente apoyado por el donativo DGAPA-UNAM 203003-3.

REFERENCIAS

1. Encarnación S, Guzmán Y, Dunn MF, Hernández, Vargas M del C, & Mora J. 2003. Proteome analysis of aerobic and fermentative metabolism in *Rhizobium etli* CE3. *Proteomics* 3:1077-1085.
2. Encarnación S., Hernández M., Contreras S., Martínez-Battallar G., Vargas M. del C. Mora J. 2005. Comparative proteomics using 2-D gel electrophoresis and mass spectrometry as tools to dissect stimulons and regulons in bacteria with sequenced or partially sequenced genomes. *Biological Procedures Online*. 7:117-135. 2005.
3. Encarnación S, Vargas M del C, Dunn MF, Dávalos A, Mendoza G, Mora Y, & Mora J. 2002. AniA regulates reserve polymer accumulation and global protein expression in *Rhizobium etli*. *J Bacteriol* 184: 2287-2295.
4. González, V., R.I. Santamaría, P. Bustos, I. Hernandez-González, A. Medrano-Soto, G. Moreno-Hagelsieb, *et al.* 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc Natl Acad Sci USA*. 103:3834-9.
5. Porubleva L, Vander VK, Kothari S, Livier DJ, & Chitnis PR. 2001. The proteome of maize: use of gene sequence and expressed sequence tag data for identification of proteins with mass fingerprints. *Electrophoresis* 22:1724-1738.

Genómica funcional del metabolismo actinobacteriano

Francisco Barona-Gómez*,**

* Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, UNAM, Av. Universidad 2001, Cuernavaca, Mor., 62210, México.

** Department of Biological Sciences, University of Warwick, Coventry, CV4 7AL, Reino Unido. barona@ibt.unam.mx

El principal reto al que se enfrenta la genómica microbiana como ciencia emergente es demostrar su valía en cuanto a la generación de información funcional biológicamente relevante. Dicho reto se pone de manifiesto en dos observaciones recurrentes; por un lado, hasta un 50% de los genes identificados por los proyectos de secuenciación genómica, y analizados con las técnicas bioinformáticas convencionales, carecen de una asignación de función; por el otro lado, el contenido genómico (genotipo) de un cierto microorganismo, y las funciones biológicas (fenotipo) esperadas en el mismo (muchas de estas funciones previamente descritas experimentalmente) no concuerdan, o bien, la relación gene-función no resulta obvia. El metabolismo bacteriano, incluidas las rutas metabólicas centrales "universales", no se escapa de estas discrepancias genómicas. Afortunadamente, los primeros éxitos del establecimiento inequívoco de la relación gene-función en discrepancias metabólicas de este tipo han comenzado a aparecer en la literatura (Osterman & Overbeek, 2003). Más aún, se propone que estos reportes sugieren la existencia de una relación íntima entre ambos tipos de discrepancias: es común encontrar situaciones en las que faltan genes metabólicos, los cuales se explican a través de la asignación de la actividad metabólica correspondiente, a genes con función desconocida (Morett *et al.*, 2003).

A estas alturas de la era genómica, cuando la información generada sobrepasa nuestra capacidad de análisis, resulta imperativo el tomar en cuenta la información tradicional disponible sobre fisiología y ecología microbiana, dentro de un contexto filogenéticamente delimitado, con la finalidad de optimizar el proceso de análisis. Este ejercicio se ve representado por la genómica comparativa, la

cual promete, como disciplina emergente, identificar rasgos genéticos que ayuden a entender las características morfológicas y fisiológicas de cierto organismo (o grupo de organismos), así como su evolución y ecología. Dado que la cobertura taxonómica por los genomas secuenciados cada vez es mayor, representando no tan sólo microorganismos patógenos, sino también microorganismos aislados del medio ambiente, las conclusiones que se derivan de dichas comparaciones genómicas resultan cada vez menos segadas y más atinadas. Un grupo de microorganismos que ejemplifican esta situación son las bacterias Gram-positivas con alto contenido de (G + C), también conocidas como actinobacterias o actinomicetos, dentro de las cuales se sitúa el patógeno *Mycobacterium tuberculosis*, así como especies productoras de antibióticos y de otros fármacos del género *Streptomyces*. Por ejemplo, a pesar de que un 25% de los cuarenta genomas actinobacterianos secuenciados a la fecha, o en vías de secuenciación, pertenecen a especies del género *Mycobacterium*, la diversidad de estos genomas incluye 21 géneros distintos, así como bacterias patógenas, comensales y de vida libre; mesófilos y termófilos; organismos acuáticos y terrestres; y con relevancia médica, agrícola e industrial.

De hecho, la genómica comparativa de actinomicetos ha dado buenos resultados en cuanto a la reconstrucción de rutas metabólicas en este grupo de bacterias. Un ejemplo en cuestión es la biosíntesis del amino ácido esencial L-triptofano, cuya biología molecular ha sido extensivamente estudiada en los organismos modelo *Escherichia coli* y *Bacillus subtilis* por más de cincuenta años (Yanofsky, 2001). Sorprendentemente, y a pesar de este gran cuerpo de información, la secuenciación de genomas actino-

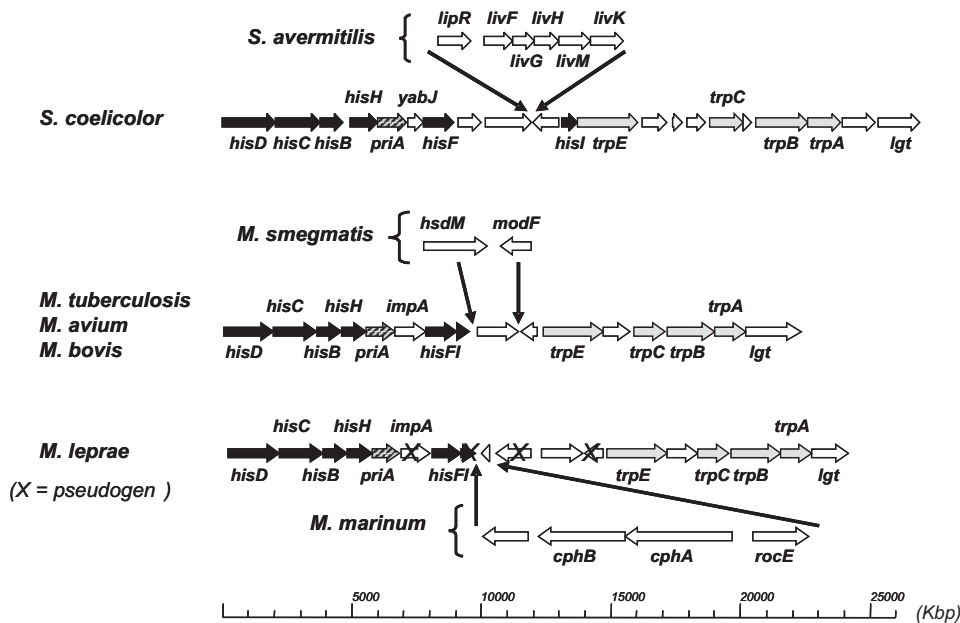


Figura 1. Regiones genómicas de actinomicetos selectos (especies del género *Streptomyces* y *Mycobacterium*) que muestran la conservación de genes *his/trp* en torno a *priA*. Los genes *his* y *trp* se muestran sombreados en negro y gris, respectivamente, y *priA* con trama a rayas. En *M. leprae* se incluyen aquellos genes que han deteriorado hacia pseudogenes. Interesantemente, *priA* se mantiene intacto, lo que manifiesta la esencialidad del producto de este gen en este patógeno.

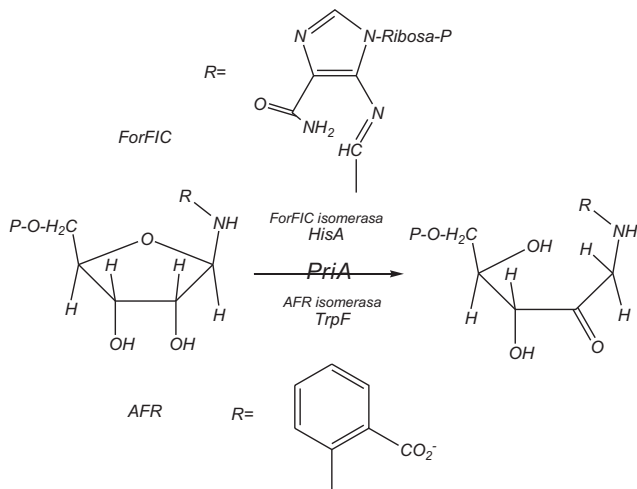


Figura 2. Los sustratos antranilato *N*'-5'-fosforribosil (AFR) y el ribonucleótido *N*'-[5'-fosforribosil formimino]-5-imidazolamina-4-carboxiamino (ForFIC), intermediarios en la biosíntesis de los amino ácidos triptofano e histidina, son análogos y su conversión conlleva un rearrreglo del tipo Amadori catalizado por TrpF y HisA, respectivamente, o PriA, para ambos amino ácidos. P representa grupos fosfato.

bacterianos fracasó sistemáticamente en cuanto a la identificación del gene biosintético *trpF*, esto a pesar de que la evidencia experimental daba cuenta de esta actividad enzimática en extractos celulares. Cabe mencionarse que una excepción a la falta del gen *trpF* en actinomicetos se presenta en los genomas de algunas especies del género *Cory-*

nebacterium, aunque en estos casos existe evidencia de que la presencia de este gen es el resultado de un evento de transferencia horizontal genética (Xie *et al.*, 2003). En el resto de los casos, haciendo uso de comparaciones genómicas, se notó que los genes biosintéticos *his*, involucrados en la biosíntesis del amino ácido L-histidina, se agrupaban con la mayoría de los genes biosintéticos *trp*, formando lo que se denominó el cluster *his/trp* (Fig. 1). La relevancia de esta observación radica en la premisa de que la conservación genómica, o sintenia, sugiere una asociación funcional de los genes en cuestión.

Esta información genómica, en combinación con un entendimiento del mecanismo de acción de la enzima codificada por el gen *trpF*, es decir, la antranilato *N*'-5'-fosforribosil (AFR) isomerasa (TrpF, EC 5.3.1.24), y de la estructura tridimensional de esta enzima de otras bacterias, dio lugar a la hipótesis correcta de que el producto del gen biosintético *hisA*, anotado originalmente como la enzima *N*'-[5'-fosforribosil formimino]-5-imidazolamina-4-carboxiamino (ForFIC) isomerasa (HisA, EC 5.3.1.16), participaba tanto en la biosíntesis de la histidina como en la del triptofano, esto a través de una especificidad de sustrato relegada que le permite al producto de este gen aceptar como sustrato a los intermediarios AFR y ForFIC (Fig. 2). Con base en estos resultados, se propuso renombrar a los homólogos de *hisA* de origen actinobacteriano, y con especificidad de sustrato dual, como *priA* (del inglés *phosphoribosyl isomerase A*) con la finalidad de reflejar la actividad enzimática de esta isomerasa sobre diversos sustratos fosforribosilados (Barona-Gómez & Hodgson, 2003). Este ejemplo demuestra el

potencial de la genómica comparativa para la reconstrucción de rutas metabólicas en actinomicetos, en aquellos casos en los que hagan falta genes.

REFERENCIAS

- Osterman, A. & Overbeek, R. (2003) Missing genes in metabolic pathways: a comparative genomics approach. *Curr. Opin. Chem. Biol.*, 7, 238-251.
- Morett, E., Korbel, J.O., Rajan, E., Saab-Rincón, G., Olvera, M., Schmidt, S., Snel, B. & Bork, P. (2003) Systematic discovery of analogous enzymes in thiamine biosynthesis. *Nat. Biotechnol.* 7, 790-795.
- Yanofsky, C. (2001) Advancing our knowledge in biochemistry, genetics, and microbiology through studies on tryptophan metabolism. *Annu. Rev. Biochem.* 70, 1-37.
- Xie G., Bonner C.A., Song J., Keyhani N.O. & Jensen R.A. (2003) Intergenomic displacement via lateral gene transfer of bacterial *trp* operons in an overall context of vertical genealogy. *BMC Biol.* 2:15, doi:10.1186/1741-7007-2-15.
- Barona-Gómez, F. & Hodgson, D.A. (2003) Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.*, 4, 296-300.

Identificación de riboswitches bacterianos mediante genómica comparativa

Abreu-Goodger, C. y Merino E

Departamento de Microbiología Molecular. Instituto de Biotecnología, UNAM, Av. Universidad 2001, Cuernavaca, Mor., 62210, México.

Los seres vivos se caracterizan, entre otras cosas, por su capacidad de responder a estímulos de su exterior. Desde el organismo unicelular más primitivo, hasta el más complejo de los animales, son capaces de percibir cambios en su medio y reaccionar de una manera apropiada. Muchas de estas respuestas, en un momento dado, requerirán de cambios en el estado de expresión del genoma: genes que estaban prendidos necesitarán ser apagados y genes latentes necesitarán ser activados. En organismos bacterianos, dicha regulación de la expresión genética se lleva primordialmente al inicio de la transcripción y está mediada comúnmente por proteínas reguladoras, ya sea para inhibirla o favorecerla, de acuerdo a las necesidades metabólicas de la célula. Sin embargo, estudios recientes de genómica comparativa han revelado la existencia de un novedoso sistema de regulación mediante elementos llamados riboswitches [Winkler. *et al.*, 2002]. Los riboswitches se encuentran en la región 5' no traducida de los mRNA, y pueden plegarse en el espacio formando complejas estructuras tridimensionales de RNA capaces de reconocer, con alta afinidad y especificidad, a pequeños metabolitos como vitaminas o nucleótidos, en total ausencia de proteínas [Batey, *et al.*, 2004]. El mecanismo por el cual la expresión genética es regulada involucra la formación de estructuras alternativas (dependiendo de la presencia o ausencia del metabolito) que causan la terminación prematura de la transcripción o el bloqueo del inicio de la traducción. Los riboswitches se han encontrado casi exclusivamente en organismos bacterianos y regulan muchos procesos incluyendo la biosíntesis de vitaminas como la riboflavina, tiamina y

cobalamina, así como el metabolismo de la metionina, lisina, glicina y purinas. La diversidad taxonómica de genomas que contienen riboswitches en adición al hecho de que la interacción directa de los riboswitches con sus ligandos no requiere de factores adicionales, sugiere que los riboswitches representan uno de los sistemas reguladores más antiguos. Al inicio de nuestra investigación siete familias de riboswitches habían sido identificadas mediante un proceso arduo de experimentación y búsqueda de motivos conservados en las regiones de regulación de los genes conocidos de las vías metabólicas de estudio. En nuestro grupo nos hemos dado a la tarea de desarrollar un método computacional capaz de identificar nuevos tipos de riboswitches de manera automática [Abreu-Goodger *et al.*, 2004] explotando la actual disponibilidad de cientos de secuencias de genomas totalmente secuenciados, con la ventaja adicional de prescindir de la información bioquímica o genética de las diferentes vías metabólicas.

El método desarrollado está basada en la hipótesis de que algunos genes ortólogos (genes homólogos funcionalmente equivalentes) a pesar de encontrarse en genomas filogenéticamente distantes, pueden compartir un mismo sistema de regulación. En general, la anterior premisa no es cierta cuando la regulación de los sistemas se da por proteínas que se unen al DNA, sin embargo, para los riboswitches conocidos esta hipótesis parece cumplirse.

Como grupos iniciales de genes ortólogos se tomó la base de datos COG (Clusters of Grupos de genes Ortólogos, por sus siglas en inglés) [Tatusov *et al.*, 1997], misma

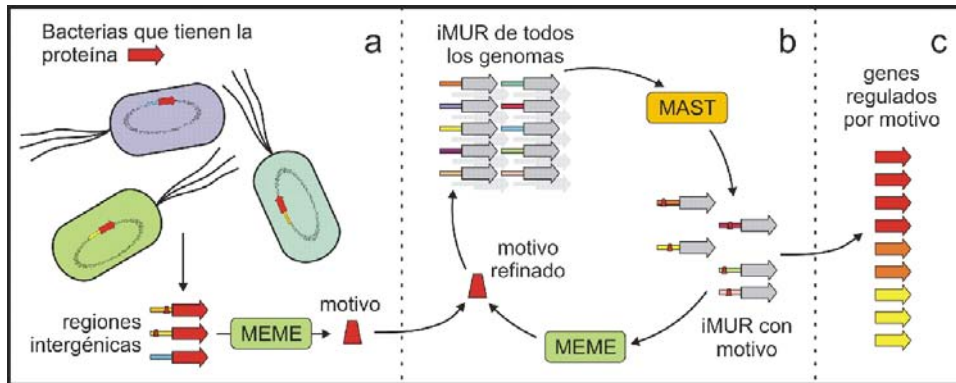


Figura 1. Resumen del método. a) Se toman todas las regiones intergénicas de las proteínas de un mismo COG (flecha roja) y con MEME se encuentran motivos sobre-representados. b) Cada motivo es refinado al ciclar entre buscarlo con MAST en todas las regiones intergénicas de genomas completamente secuenciados y redefinir el motivo con MEME a partir de las nuevas secuencias encontradas. c) Al final, se obtiene una lista de genes regulados por el motivo refinado que incluye miembros de otros COGs además del inicial. Se presume que todos éstos pertenecen a un mismo regulón.

que fue actualizada para incluir a 164 genomas bacterianos secuenciados en su totalidad al momento del análisis. Los operones de cada genoma fueron predichas de acuerdo a sus distancias intergénicas [Moreno-Hagelsieb. and Collado-Vides, 2002] y agrupadas en base a la relación de ortología de los genes de sus correspondientes unidades transcripcionales. Con el objeto de eliminar la sobre-representación de secuencias similares de organismos filogenéticamente cercanos, se tomó un solo representante de cada grupo de genomas redundantes. Al final, se obtuvieron cerca de 4,000 grupos de regiones de regulación ortólogas, cada una perteneciente a un COG diferente. En cada uno de estos grupos se buscaron motivos de secuencias cuya frecuencia de aparición fuera significativamente mayor a la esperada al azar, utilizando el programa MEME [Bailey and Elkan, 1994]. Los motivos así encontrados, fueron utilizados por el programa MAST [Bailey and Elkan, 1994] como “semilla” para identificar a nuevos miembros del regulón que compartieran la misma señal reguladora. Como resultado de este proceso, nuevos miembros fueron agregados a cada grupo, mientras que otros fueron eliminados. A partir de estos nuevos conjuntos de secuencias, nuevos y más representativos motivos fueron obtenidos, utilizando MEME, seguido por una nueva búsqueda con MAST. Este proceso fue repetido iterativamente hasta que ningún nuevo elemento fuera agregado o excluido del regulón putativo (ver Fig. 1 para un esquema del método).

El método de búsqueda desarrollado identificó todos y cada uno de los riboswitches previamente reportados en adición a un conjunto significativo de más de cien motivos altamente conservados en diferentes organismos. Con el objetivo de evaluar el significado biológico de las nuevas firmas encontradas, se analizó el contexto genómico de los genes que potencialmente

tuvieran elementos de regulación en común, encontrándose que en la mayoría de los casos, dichos genes pertenecían a vías metabólicas en común o bien estaban involucrados en procesos celulares similares. Una lista completa de los motivos conservados y genes potencialmente regulados por los mismos, puede encontrarse la página de Internet <http://www.ibt.unam.mx/biocomputo/ribex.html>. Este servidor además permite buscar sobre cualquier secuencia todos los motivos que han sido predichos. Es importante señalar que debido a que el método está basado exclusivamente en la conservación de secuencia en las regiones de regulación, no es posible diferenciar cuáles de los motivos encontrados corresponden a nuevos riboswitches y cuáles otros a secuencias de unión a proteínas reguladoras altamente conservadas.

Consideramos que nuestras predicciones computacionales de riboswitches y elementos de regulación conservadas es de gran valor para dirigir trabajos experimentales que caractericen los diferentes mecanismos de regulación. Interesantemente, este fue el caso de nuestra identificación *in silico* del riboswitch de glicina, cuya existencia fue paralelamente comprobada experimentalmente [Barrick, *et al.*, 2004]. Sin duda alguna, el potencial de metodologías computacionales para la identificación de elementos de regulación, interacciones proteína-proteína, redes metabólicas y muchos otros componentes de los sistemas celulares, serán grandemente favorecidos con los proyectos de secuenciación de miles de genomas que actualmente están en curso. Una nueva era en la genómica comparativa está por nacer.

Este trabajo estuvo parcialmente apoyado Consejo Nacional de Ciencia y Tecnología (CONACyT 44213-Q) y el Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT-UNAM IN203705).

REFERENCIAS

1. Abreu-Goodger, C., Ontiveros-Palacios, N., Ciria, R. and Merino, E. 2004. Conserved regulatory motifs in bacteria: riboswitches and beyond. *Trends in Genetics* 20:475-479.
2. Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., Wickiser, J.K. y Breaker, R.R. 2004. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. U. S. A.* 101: 6421-6426
3. Batey, R.T., Gilbert, S.D. and Montange, R.K. 2004. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*. 432: 411-415.
4. Bailey, T.L. and Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers Proceedings of the 2nd International Conference on ISMB, AAAI Presspp. 28–36.
5. Moreno-Hagelsieb, G. and Collado-Vides, J. 2002. A powerful nonhomology method for the prediction of operons in prokaryotes. *Bioinformatics* 18, 329–336.
6. Tatusov, R.L., Koonin, E.V. and Lipman, D.J. 1997. A genomic perspective on protein families. *Science*. 278: 631-637.
7. Winkler, W, Nahvi A and Breaker R.R. 2002. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419, 952–956.

Correspondencia:

Sergio Encarnación-Guevara.
Programa de Genómica Funcional de Procaríotes.
Centro de Ciencias Genómicas
Av. Universidad s/n., Cuernavaca Mor.
E-mail: *encarnac@cifn.unam.mx.