

José L. Oliver
Grupo de Genómica Evolutiva y Bioinformática
Dpto. de Genética. Universidad de Granada

BIO INFORMÁTICA

If you can't do Bioinformatics,
you can't do Biology

J. D. Tisdall, 2001

Esta frase, del autor de uno de los textos más populares de bioinformática ('Beginning Perl for Bioinformatics'), resume bastante bien la posición que ocupa esta disciplina dentro de la Biología.

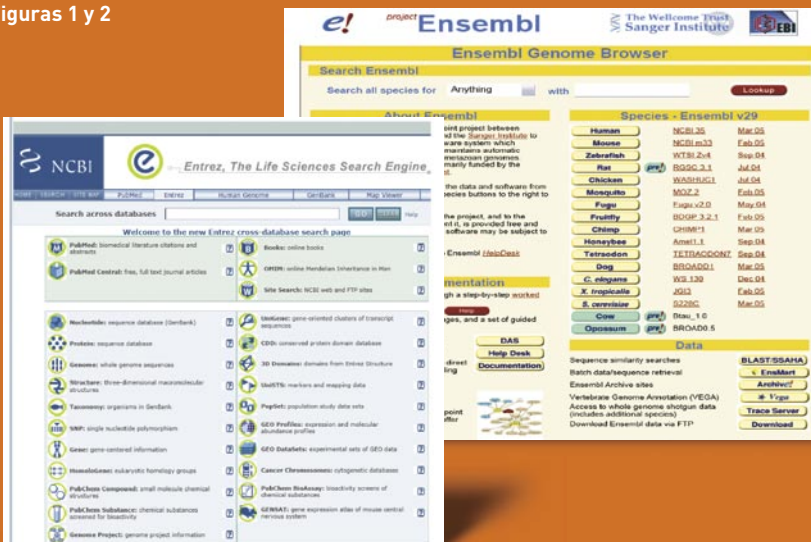
La variedad y complejidad de los datos moleculares a que hoy se enfrenta el biólogo, generados no solamente por la genómica, sino también por disciplinas más clásicas como la biología estructural, la fisiología o la biología de sistemas, hace de la bioinformática una herramienta que todo biólogo, tanto investigador como docente, necesita tener a mano.

En menos de una década hemos pasado de analizar el efecto de los genes tomados de uno en uno (ignorando

así al 99.99% restante...), a disponer de la colección completa de los genes de una especie. Antes se hacía una tesis con un gen, hoy cuesta conformarse con menos de un genoma. Alguien podría pensar que es solo una diferencia cuantitativa, pero el reto es mucho más importante que eso: los datos de genomas y proteomas completos están provocando un cambio profundo en toda la Biología.

Las preguntas que podemos plantearnos ahora son cualitativamente distintas a las de hace unos años. Ello se debe en buena medida a la facilidad de acceso a la información genómica, propiciada por herramientas bioinformáticas cada vez más potentes (figuras 1 y 2).

Figuras 1 y 2

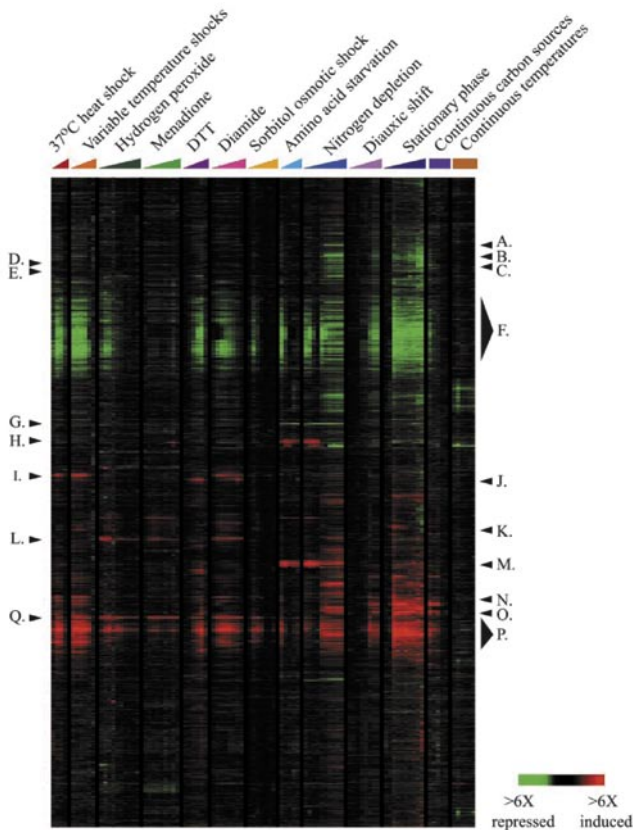


Navegador genómico Ensembl (<http://www.ensembl.org/>) que permite un acceso rápido e integrado a las bases de datos de ADN, proteínas y genomas completos. Ensembl trabaja con los datos de la EMBL Data Library, editada por el Instituto Europeo de Bioinformática (EBI, <http://www.ebi.ac.uk/>). Además del acceso a los datos, el EBI ofrece también acceso público a multitud de programas de bioinformática mediante interfaces web.

Página de entrada a las bases de datos del NCBI (Centro Nacional de Información en Biotecnología) de los Estados Unidos (<http://www.ncbi.nlm.nih.gov/Entrez/>), donde se edita GenBank. Esta página permite buscar en muchas bases de datos simultáneamente: secuencias de nucleótidos y proteínas, estructuras 3D, expresión génica, etc.

La bioinformática ha puesto a nuestro alcance (‘a un clic de ratón’) cantidades ingentes de información relevante para el biólogo. Antes podíamos estudiar la filogenia de un grupo de especies utilizando como marcador un gen o una proteína concreta; hoy disponemos de herramientas para alinear y comparar genomas completos. Antes podíamos estudiar la regulación y el funcionamiento de un determinado gen o grupo de genes; hoy podemos analizar el funcionamiento simultáneo de todos los genes de un genoma, gracias a técnicas como los ‘microarrays’.

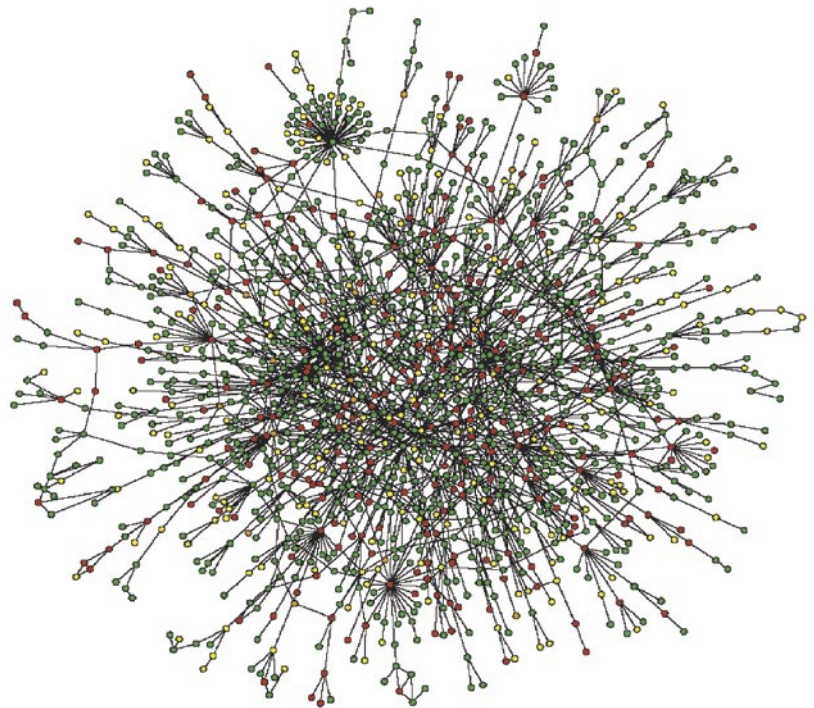
Figura 3



Los ‘microarrays’ permiten analizar la expresión génica de todos los genes de una especie simultáneamente. La figura muestra los patrones de expresión de los genes (filas) de la levadura identificados hasta el momento del experimento [6200] en respuesta a diferentes condiciones ambientales (columnas). [Gasch et al., Molecular Biology of the Cell 11, 4241–4257, 2000].

Antes estudiábamos la interacción de dos proteínas; hoy, gracias a técnicas como la del ‘doble híbrido’, podemos estudiar el ‘interactoma’ completo, es decir la totalidad de interacciones entre todas las proteínas de una determinada especie.

Figura 4



El interactoma de la levadura (*Saccharomyces cerevisiae*). La figura muestra la red de interacciones (representadas aquí por conexiones) entre las proteínas (nodos) de esta especie, deducidas mediante la técnica del ‘doble híbrido’. Puede observarse que la mayoría de las proteínas tienen pocas interacciones, mientras que algunas de ellas se comportan como ‘concentradores de red’, mostrando un gran número de interacciones con otras proteínas. Esta red muestra independencia de escala, lo que quiere decir que se vería igual a distintos niveles de resolución [Jeong H et al., Nature 411: 41-42, 2004].

Podríamos seguir enumerando ejemplos, pero la conclusión siempre sería la misma: las preguntas que hoy se pueden plantear los biólogos son de una naturaleza y de un alcance muy distintos a las de hace tan sólo unos años, y la bioinformática ha tenido mucho que ver en ello.

La bioinformática utiliza una amplia variedad de métodos matemáticos, estadísticos y computacionales (algoritmos, programas, bases de datos...) para resolver problemas biológicos, utilizando para ello el ADN, las proteínas e información relacionada. Esta disciplina se ha convertido pues en una herramienta clave para generar nuevo conocimiento biológico, y para aplicar este conocimiento a áreas tan relevantes como la salud, la biotecnología, el medio ambiente o la evolución.

La información generada por los proyectos genoma se almacena en bases de datos, en su inmensa mayoría públicas, que son un ejemplo de empresa colectiva en ciencia. Todos los laboratorios del mundo contribuyen a nutrir con sus datos estas bases de datos, y cualquier investigador puede acceder en línea a su contenido. Estos inmensos bancos de datos se mantienen actualizados permanentemente en servidores públicos, de forma que la información que de ellos se obtiene es siempre de primera mano. Al mismo tiempo, dicha información es cada vez más fiable, porque es la comunidad científica en su conjunto, y los grupos líderes en cada campo concreto, quienes se encargan de actualizarla y depurarla.

Los programas como FASTA o BLAST permiten interrogar fácilmente las bases de datos de genes y proteínas. Ambos se ejecutan en servidores informáticos ubicados en grandes centros de investigación, como el Instituto Europeo de Bioinformática (EBI, Cambridge) o el Centro Nacional para Información en Biotecnología (NCBI, Bethesda, Maryland, USA). Sin embargo, los dos se pueden lanzar en remoto a través de un simple navegador web, sin requerir para ello ninguna otra infraestructura informática local. De esta forma, el biólogo tiene a su alcance una herramienta que le permite obtener rápidamente toda la información disponible acerca de un pequeño oligonucleótido o péptido problema: identificar la secuencia completa del gen o proteína a los que corresponde, identificar su papel metabólico, su actividad enzimática, sus cofactores o su perfil de expresión, averiguar en qué enfermedades genéticas puede estar implicado, explorar la estructura 3D de la proteína correspondiente, identificar genes homólogos en otras especies y aprovechar esta información para reconstruir la filogenia del grupo, etc.

La comparación de secuencias de genes o proteínas mediante alineamiento múltiple puede realizarse también en línea. Se pueden identificar así los aminoácidos conservados a lo largo de la evolución en una familia de proteínas, e incluso deducir qué aminoácido ocupaba una determinada posición en una especie extinguida hace millones de años.

Otro de los retos que tiene planteados actualmente la bioinformática es aumentar el poder de predicción de los programas para la búsqueda computacional de genes. El método de secuenciación que más se emplea actualmente ('por perdigonada' o shotgun) es muy rápido y eficiente, pero solo produce largas secuencias 'anónimas' que necesitan ser anotadas. Dicha

anotación (asignación de función) se hace en primera instancia por medios computacionales; solo una vez que pasan este filtro, los genes candidatos se someten a verificación experimental, mucho más costosa en tiempo y dinero.

Los programas para la búsqueda computacional de genes alcanzan una gran precisión (hasta el 95%) en los genomas de procariotas, pero no así en los eucariotas, donde diversos fenómenos biológicos (presencia de intrones, exones alternativos, etc.) complican la tarea, haciendo que la precisión sea considerablemente más baja. Ello explica el actual debate en torno al número de genes humanos. Según los programas, y los supuestos que estos adopten, así se puede predecir un número de genes u otro. Bien es cierto que las últimas estimaciones disponibles (en torno a 20.000-30.000 genes) son más coincidentes que las de hace unos años, pero queda aún mucho camino por recorrer antes de que la bioinformática pueda dar una respuesta definitiva a este problema. □



PARA SABER MÁS

Attwood, T.K. y D.J. Parry-Smith.
Introducción a la Bioinformática. Pearson Educación, 2002.

A Science Primer – Bioinformatics:
<http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html>

Bioinformatics Educational Resource:
<http://www.ebi.ac.uk/2can/index.html>

ENLACES DE INTERÉS

Grupo de Bioinformática de la Universidad de Granada
<http://bioinfo2.ugr.es/>

Red de Bioinformática de Andalucía
<http://bioinfo2.ugr.es/Rba/>

Red Nacional de Bioinformática
<http://www.pdg.cnb.uam.es/RedBio/>

Instituto Europeo de Bioinformática
<http://www.ebi.ac.uk/>

Master en Bioinformática, Universidad Internacional de Andalucía
http://www.unia.es/maestrias_unia/bioinformatica/