



Más bits al servicio del ADN

Bioinformáticos brasileños crean herramientas para estudiar genomas

Marcos Pivetta

PUBLICADO EN FEBRERO DE 2013

Hace poco más de una década casi no había genomas completos por analizarse. Ahora hacen falta programas y mano de obra especializada para dar cuenta de la cantidad de secuencias de ADN depositadas en bases de datos públicas que surgen diariamente de una nueva generación de secuenciadores. Extremadamente veloces, esas máquinas determinan los pares de bases del material genético, las denominadas letras químicas, a un precio miles de veces menor que al comienzo de los años 2000, cuando se completó la epopeya de secuenciar el primer genoma humano. De cara a ese desafío, el matemático João Meidanis, socio fundador de la empresa Scylla Bioinformática y docente de la Universidad Campinas (Unicamp), desarrolló una línea de investigación consistente en crear métodos más sencillos y eficientes para comparar dos o más genomas.

Junto a Pedro Feijão, su ex alumno de doctorado, estableció en 2009 la base teórica de una

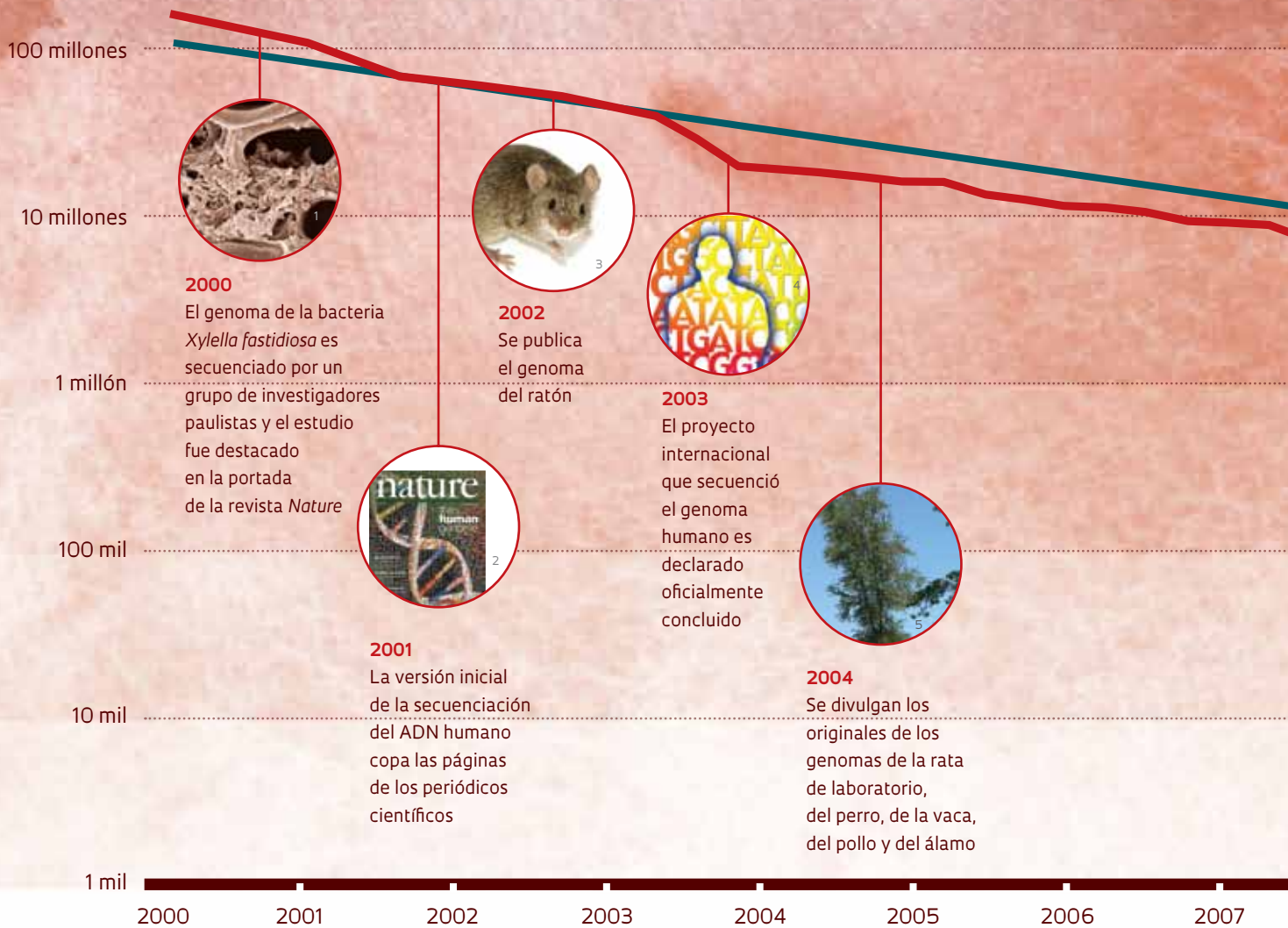
técnica destinada a comparar genomas completos denominada *single-cut-or-join* (SCJ) y, el año pasado, realizó prácticas con genomas de algunos organismos, tales como plantas y bacterias. “Con nuestro método podemos comparar fácilmente dos o más genomas sin que aumente exponencialmente la cantidad de cálculos, tal como ocurre con otras técnicas”, sostiene Meidanis. “Así podemos construir árboles genealógicos y comprobar cuáles son los genomas más próximos y los más distantes desde el punto de vista evolutivo”. El matemático fue uno de los coordinadores de bioinformática del proyecto que secuenció, en el año 2000, el genoma de la bacteria de la *Xylella fastidiosa*, causante del *amarelinho* [la clorosis variegada de los cítricos] de los naranjales. El trabajo fue por primera vez tema de tapa en la revista científica *Nature* para una investigación brasileña.

Para comparar todo el material genético de una especie con el de otra, los investigadores deben recurrir a simplificaciones. La principal consiste en

La depreciación del ADN

En una década, el costo de la secuenciación del genoma humano se redujo de 100 millones de dólares a menos de 10 mil dólares

* El eje y en escala logarítmica (en US\$)



considerar que los genes presentes en los genomas comparados son exactamente los mismos, aunque se encuentren ordenados en forma diferente en la secuencia específica de cada organismo. Partiendo de ese razonamiento, los métodos para comparar genomas cuentan el número de reordenamientos que han sido necesarios para que un genoma se transforme en otro. Esos reagrupamientos son consecuencia del movimiento de grandes segmentos de ADN ocurrido a lo largo del tiempo en la secuencia original. De este modo, cuanto menor fuera la cantidad de reordenamientos que separan a dos genomas, más cercanos se encuentran dentro del árbol evolutivo.

En su método, Meidanis y Feijão formularon una definición alternativa para el concepto de punto de ruptura (*breakpoint*), un parámetro importante para ubicar reordenamientos en una

secuencia y así calcular la cercanía entre dos genomas. El punto de ruptura es el sitio donde hay una interrupción en un largo segmento conservado en los genomas que se están comparando.

El año pasado, ambos también perfeccionaron otro método de comparación de genomas, más minucioso que el SJC. Esta segunda técnica, propuesta inicialmente en el año 2000, permitía confrontar solamente genomas circulares. Con su perfeccionamiento, también pasó a ser útil para comparar el material genético de cromosomas lineales. “Ésa era una de las limitaciones en la técnica original”, comenta Feijão, quien actualmente trabaja en Scylla. El nuevo método, basado en lo que los matemáticos denominan formalismo algebraico por adyacencias, todavía no fue probado en genomas reales. Por ahora existe como teoría.

LEY DE MOORE

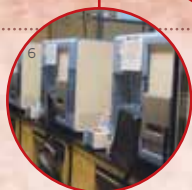
Hasta mediados de los años 2000, la disminución del costo de secuenciar siguió el abaratamiento tecnológico producto de la Ley de Moore. Después, el precio cayó abruptamente



2011 Científicos chinos del BGI anuncian un proyecto para secuenciar 3 millones de genomas de plantas, animales, seres humanos y microorganismos



2012 El genoma completo de 1.092 personas, de diferentes continentes, se publica en la revista *Nature*



2008

Las máquinas de segunda o última generación irrumpen en el mercado y derrumban el costo de la secuenciación de genomas

FUENTE: NHGRI GENOME SEQUENCING PROGRAM

2008 2009 2010 2011 2012

METAGENÓMICA

Meidanis, lógicamente, no es el único en comprobar los efectos de la nueva realidad en su campo de actuación. De regreso en Brasil desde mediados de 2011, luego de ocho años en el Virginia Bioinformatics Institute, Estados Unidos, João Carlos Setubal, hoy en día profesor titular del Instituto de Química de la Universidad de São Paulo (USP), percibe que la demanda por servicios e investigación en su área ha crecido y se ha sofisticado en los últimos tiempos. Desde que retornó a São Paulo, ha recibido, por ejemplo, 16 propuestas para colaborar en iniciativas de otros investigadores. “Los secuenciadores de última generación aportan una cantidad astronómica de datos genómicos, de proteómica y del metabolismo de los organismos”, afirma Setubal,

quien también fue uno de los coordinadores de bioinformática del proyecto de la *Xylella*. “Con el abaratamiento de la tecnología, actualmente cualquier proyecto de investigación con un mínimo de recursos puede secuenciar el genoma de un organismo”.

En la última década, surgió un nuevo campo para los biólogos y bioinformáticos que es el de la metagenómica, donde se estudia la microbiota de un nicho ecológico. El proyecto principal de Setubal, un temático de la FAPESP sobre microorganismos presentes en el Zoológico de São Paulo, se incluye en dicha área. Mediante ese abordaje, en lugar de aislar y cultivar los microorganismos con el fin de extraer por separado el ADN de cada especie, el investigador toma una muestra directamente del ambiente que ha de estudiarse. En esa muestra, el ADN de varias especies se presenta “mezclado” y le cabe al bioinformático hallar técnicas para separar y caracterizar el material genético de cada una de ellas. “Estamos estudiando tres microbiomas en el Zoológico: el compost elaborado por los empleados del parque, el agua de los lagos y las heces de los monos aulladores [carayás]”, añade Setubal.

La metagenómica también es una forma de hurgar organismos desconocidos en un hábitat específico. El equipo de Ana Tereza Ribeiro de Vasconcelos, coordinadora del centro de bioinformática del Laboratorio Nacional de Computación Científica (LNCC), en Petrópolis, participó en el descubrimiento de bacterias magnéticas halladas en la laguna de Araruama, en la costa de Río de Janeiro, una de las más salinas del mundo. Una de esas bacterias encontradas fue la *Candidatus magnetoglobus multicellularis*. Esta bacteria, identificada por Ulysses Lins, de la Universidad Federal de Río de Janeiro (UFRJ), es difícil de aislar en el ambiente para mantenerla en un medio de cultivo. “Actualmente trabajamos en una decena de proyectos de metagenómica”, dice Ribeiro, quien posee tres secuenciadores en su laboratorio y un equipo formado por unas 25 personas.

La escala de tiempo y dinero involucrados en los proyectos dedicados a analizar el ADN de organismos se modificó radicalmente durante la última década. En los primeros años de la era genómica, tan sólo las grandes empresas osaban aventurarse en ese nuevo terreno. En abril de 2003, cuando el consorcio público internacional que secuenció por primera vez un genoma humano concluyó oficialmente su trabajo, la megainiciativa había insumido 13 años de trabajo de centenares de científicos de al menos 18 países (Brasil inclusive) y unos 2.700 millones de dólares. En una proporción mucho menor, pero no por ello irrelevante, la secuenciación de la *Xylella* le costó 12 millones de dólares a la FAPESP y requirió el trabajo de 192 científicos durante tres años.

Pero ahora, la secuenciación de genomas se ha convertido en una tarea entre 10 mil y 20 mil veces más barata de lo que era hace poco más de una década, según datos del National Human Genome Research Institute (NHGRI) de Estados Unidos. El arribo en masa al mercado, a comienzos de 2008, de los secuenciadores de segunda generación, que emplean una tecnología distinta a la de los primeros aparatos del tipo Sanger, ha llevado a que el costo de la secuenciación disminuya a un ritmo acelerado que supera largamente las mejoras en el desempeño producto de la Ley de Moore en la informática. Actualmente, en dos o tres días, con un costo de unos pocos miles de dólares, pueden determinarse las 3 mil millones de letras químicas del ADN de una persona. “La bioinformática es una nueva herramienta, una lupa, para comprender mejor el fenómeno biológico, que no se ha modificado, pero ahora puede observárselo de otra forma”, dice Gonçalo Pereira, del Instituto de Biología (IB) de la Unicamp.

Pero secuenciar es una cosa, y extraer información útil de los miles de millones de datos que las computadoras depositan cotidianamente en manos de los científicos es otra, bastante más compleja. “La secuenciación en sí misma ahora es barata, se transformó en una *commodity*, pero el análisis de los datos es caro”, dice el experto en informática João Paulo Kitajima, de Mendelics, una empresa creada recientemente que

trabaja con diagnósticos genómicos personalizados. “La demanda por el trabajo de bioinformática ha crecido exponencialmente y hay un *gap* entre la oferta y la demanda de expertos en Brasil y en el exterior”.

Resulta difícil estimar con precisión el tamaño de la comunidad de bioinformáticos en el país. Según Guilherme Oliveira, presidente de la Asociación Brasileña de Bioinformática y Biología Computacional (AB3C), unas 300 personas, entre docentes, alumnos e investigadores, mantienen vínculos con la entidad. “Tiempo atrás, el bioinformático era autodidacta”, dice Oliveira, coordinador del centro de bioinformática de la Fiocruz de Minas Gerais. “Actualmente, buena parte emerge de las carreras de posgrado y en cada estado hay algún especialista del sector. Una novedad es que ahora las empresas también trabajan en el área”. Grandes universidades del país, tales como la USP, la UFRJ y la Universidad Federal de Minas Gerais (UFMG), y la Fiocruz, cuentan con programas de posgrado específicos en bioinformática. Otras universidades trabajan en el tema como una línea de investigación en el ámbito de posgrado de un área más amplia, tal como biología o computación.

El abaratamiento de la secuenciación tornó accesibles las técnicas genómicas para proyectos con cualquier presupuesto

China posee el mayor centro de secuenciación

En menos de 15 años, un centro chino de bioinformática se ha transformado, desde su condición de socio menor del consorcio internacional que mapeó el primer genoma humano, en la mayor potencia global en términos de secuenciación de ADN. Fundado en 1999, el Beijing Genomics Institute (BGI, según su sigla en inglés), actualmente es dueño de 180 máquinas secuenciadoras, la mayor parte de ellas de última generación, que pueden generar diariamente 6 terabytes de datos, el equivalente a los genomas completos de 2 mil individuos. Posee 4 mil empleados y filiales en Estados Unidos, Europa y Japón. El trabajo de los chinos, realizado a una escala gigantesca, genera la expectativa de que el costo de secuenciar un genoma humano disminuya pronto hasta los 1.000 dólares, y los ubica en un sitio protagónico en proyectos de punta, que

van mucho más allá de la determinación de la secuencia genética de un símbolo nacional, el panda gigante, un logro alcanzado hace tres años.

En 2010, por ejemplo, el BGI secuenció el primer genoma completo de un ancestro del ser humano, el ADN de un esquimal que vivió hace 4 mil años. En 2012 aportó el ADN de 100 chinos para un esfuerzo internacional que estudia el genoma de alrededor de mil individuos de distintas regiones del planeta. También durante el año pasado, el centro se propuso secuenciar en los próximos años 3 millones de genomas, provenientes de seres humanos, plantas, animales y microorganismos.

La política de los chinos es agresiva en todos los sentidos, no sólo en el campo científico, sino también desde el punto de vista comercial. Además de vender sus servicios en bioinformática,

el BGI intenta asegurarse el acceso a los más recientes avances en el sector. Al comienzo de este año, el centro asiático obtuvo luz verde de los estadounidenses para adquirir por 177 millones de dólares una compañía en California, Complete Genomics, que desarrolló una nueva tecnología de secuenciación, cuyos resultados serán más precisos que los obtenidos mediante los métodos actualmente en uso en todo el mundo.



Panda gigante: uno de los genomas que los chinos del Beijing Genomics Institute (BGI) secuenciaron por completo



Aguas hipersalobres de la laguna de Araruama, en Río de Janeiro: estudios de metagenómica hallaron bacterias magnéticas en el lugar

Los trabajos de secuenciación y análisis del genoma del *Schistosoma mansoni*, el parásito causante de la esquistosomiasis, constituyen el proyecto con mayor visibilidad al que el centro de bioinformática de la unidad de Minas Gerais de la Fiocruz se ha dedicado durante los últimos años. Pero las seis máquinas secuenciadoras y los 15 especialistas en bioinformática del sector encabezado por Oliveira participaron en alrededor de 60 proyectos distintos, que incluyen estudios sobre el genoma del cáncer, de agentes infecciosos, de razas bovinas y trabajos en metagenómica. Ahora, el centro incluso genera y analiza datos para la Red de Investigación Molecular de la Biodiversidad Brasileña (Br-BoL), coordinada por Cláudio Oliveira, del Instituto de Biociencias de la Universidade Estadual Paulista (Unesp) de Botucatu, que catalogará 120 mil ejemplares de 24 mil especies de la naturaleza durante cuatro años. La BR-BoL es el brazo brasileño de un proyecto internacional, el Barcode of Life (Código de Barras de la Vida), que procura identificar especies mediante la caracterización de su ADN.

La bioinformática se ha difundido por el país y se ha asentado en centros alejados de las grandes capitales del sudeste. En la Universidad Federal de Pará (UFPA), Artur Silva realiza investigaciones en el sector y colabora con grupos de São Paulo. Desde mayo del año pasado, Sandro de Souza, quien durante años dirigió esa área en el Instituto Ludwig de Investigaciones sobre el Cáncer de São Paulo, trabaja en el Instituto del Cerebro de la Universidad Federal de Rio Grande do Norte (UFRN). Souza no cuenta con un secuenciador en dicha capital nordestina. Pero eso no le preocupa. “Puede hacerse secuenciación hasta en la nube de internet si uno quiere”, afirma Souza. “Estoy comenzando con mis trabajos en neurociencias sin ningún problema”.

Sucede que Souza aún tiene acceso a todas las máquinas del Ludwig – cuya sede administrativa

se encuentra actualmente en el hospital Sirio-Libanés de São Paulo – que fueron transferidas a la Facultad de Medicina de Ribeirão Preto (FMRP) de la USP, donde comenzó a funcionar el año pasado el Centro de Medicina Genómica. “Las técnicas de genómica y bioinformática revolucionarán la práctica médica, tal como lo hizo la medicina basada en imágenes”, sostiene Wilson Araújo da Silva Júnior, uno de los responsables por el nuevo centro de la FMRP.

Para tornar más accesibles los servicios de secuenciación y análisis de ADN y ARN, la Unicamp inaugura el 1º de marzo el Laboratorio Central de Tecnologías de Alto Desempeño (LacTAD), que actuará en los campos de la genómica, proteómica, biología celular y bioinformática propiamente dicha. Entre el instrumental del laboratorio hay dos secuenciadores de nueva generación de la empresa Illumina, de aquéllos que en pocos días son capaces de secuenciar por completo un ADN humano, y un tercero para secuenciar regiones específicas de genomas. En rigor, los aparatos del centro ya se encuentran en uso desde el año pasado, cuando arribaron a la universidad, pero distribuidos en distintas unidades. A partir del próximo mes, comenzarán a operar en el edificio de 2 mil metros cuadrados construido para el LacTAD.

“Creemos que hay una demanda latente por ese tipo de servicio, y la bioinformática se ha convertido en un problema para muchas investigaciones en las áreas biológicas”, sostiene el químico Ronaldo Pilli, rector de Investigación de la Unicamp, quien se encuentra al frente del proyecto del nuevo laboratorio. “Mantenemos la tendencia mundial de ofrecer ese tipo de servicio en forma centralizada. Así resulta más fácil adquirir, operar y mantener actualizados los equipamientos”. El instrumental del LacTAD costó unos 5,5 millones de reales y fue adquirido por medio del programa Multiusuarios de la FAPESP. El edificio, presupuestado en 4 millones de reales, fue financiado por la universidad.

El LacTAD prestará servicios para investigadores de la Unicamp y también de otras universidades y empresas. En el sitio *web* del laboratorio hay un formulario para los investigadores interesados en cotizar los precios por sus servicios. “Realizaremos trabajos que pueden costar desde 100 hasta 100 mil reales”, comenta Pilli. Se trata de la democratización de la bioinformática. ■

Proyectos

1. Estudios de la diversidad microbiana en el Parque Zoológico de São Paulo – n° 11/ 50870-6; **Modalidades** Programa Biota – Proyecto Temático; **Coordinadores** João Setubal – USP; **Inversión** R\$ 1.711.698,25 (FAPESP); 2. EMU: Laboratorio Central de Tecnologías de Alto Desempeño – n° 09/ 54129-9; **Modalidades** Programa de Equipamientos Multiusuarios; **Coordinadores** Fernando Ferreira Costa – Unicamp; **Inversión** R\$ 6.034.431,00 (FAPESP).