
LA (R) EVOLUCIÓN INFORMÁTICA EN BIOLOGÍA: EL CASO DE LA GENÓMICA

The Informatics (R) Evolution in Biology: The Case of Genomics

LAYLA MICHÁN AGUIRRE¹, Ph. D; EDUARDO ALVAREZ¹,
LAURA ELIZABETH MONTOYA PÉREZ¹.

¹Departamento de Biología Comparada, Facultad de Ciencias,
Universidad Nacional Autónoma de México.
laylamichan@ciencias.unam.mx

Presentado 2 de febrero de 2011, aceptado 14 de abril de 2011, correcciones 1 de julio de 2011.

(R) EVOLUCIÓN INFORMÁTICA: LA GENÓMICA

No se nos ha escapado a nuestra atención que cuánto
más exploramos el genoma humano, más nos queda por explorar.
“No cesaremos de explorar. Pues al final de toda exploración llegaremos donde
empezamos, y conoceremos cuál es nuestro lugar por primera vez”
(Thomas Stearns, Eliot. 2001)

RESUMEN

La biología ha evolucionado rápidamente en las últimas décadas; uno de los indicadores de este fenómeno ha sido la introducción de nuevas disciplinas, tal es el caso de la genómica que se originó en la década de los años 80 y ha dado un giro inverosímil a la ciencia moderna, dicha disciplina hace referencia al estudio no solo de los genes, sino de sus funciones, relaciones entre sí y con el medio ambiente. Surgió con la consolidación del proyecto genoma humano en un periodo de transición donde el conocimiento genético específico se tornó crítico. Se diferencia de otros enfoques (por ejemplo: la ecología, la biología evolutiva) en el tipo de información que ofrece, las perspectivas de mejoras técnicas e intelectuales en la obtención y explotación de datos de todo el genoma (Murray, 2000).

Presentamos un breve análisis bibliométrico en el que se muestra el desarrollo y las principales tendencias de esta disciplina biológica en boga.

Palabras clave: biología, genómica, bibliometría, proyecto genoma humano, informática.

ABSTRACT

Biology has been revolutionized by the introduction of new disciplines, such is the case of genomics that introduced in the 80s has turned unlikely to modern science, the discipline refers to the study not only of genes but their roles, relations among themselves and with the environment. Arises, with the consolidation of the Human Genome Project and introduces us to a period of transition in which specific genetic

knowledge becomes critical. Differs from other approaches in the type of information provided, the prospects for technical and intellectual improvements in the collection and use of data from the whole genome. A brief bibliometric analysis which shows the development and the main trends of this biological discipline in vogue.

Key words: biology, genomics, bibliometrics, human genome project, informatics.

INTRODUCCIÓN

A partir de 1950 las tecnologías de la información y comunicación (TICs) se empezaron a desarrollar de manera notable y esto ha repercutido indiscretamente en las relaciones sociales, económicas, políticas y culturales del mundo como consecuencia del incremento en las capacidades para generar, sistematizar, compartir, transmitir, analizar y difundir información (Morales, 2003).

Esta (r)evolución informática, una revolución cultural resultado de la evolución de la tecnología, se ha caracterizado por la adopción del formato digital, la masificación, democratización, la actualización e inmediatez, influye y es influida por el progreso científico y tecnológico del siglo XX. Referirse a información en pleno siglo XXI implica la mención de términos, métodos, teorías novedosas e innovadoras como: sociedad del conocimiento, sociedad de la información, globalización, infodiversidad, acceso a la información, e-ciencia, e-investigación, redes de colaboración, ciberinfraestructura, laboratorios, conocimiento basado en la literatura, minería de textos (*text mining*), *web* semántica, índice de impacto, co-citación, *web* 2.0 y 3.0, redes sociales, plagio, acceso libre, computación en nube (*cloud computer*) u ontologías, por mencionar las más frecuentes (Hine, 2007; Michán, 2010).

Dos cambios importantes han surgido en los últimos treinta años: 1) la producción acelerada de una gran variedad de programas, aplicaciones, herramientas, utilidades recursos y servicios electrónicos para la práctica científica, muchos de ellos disponibles a través de internet; y 2) la adopción del formato electrónico. En la *web* se publica constantemente información digital diversa y colosal en forma de páginas electrónicas, *blogs*, *wikis*, libros y artículos que contienen texto, imágenes y hasta sonidos. La información se sistematiza en bases de datos, puede consultarse de forma libre o restringida. Esta trata sobre seres vivos o sus partes, los fenómenos o sus explicaciones. En ella encontramos información sobre publicaciones, investigadores, proyectos, grupos, líneas de investigación, convenios, subsidios, producción científica, instituciones de investigación o de enseñanza, colecciones biológicas y sociedades científicas. Estos cambios sin duda han repercutido en la transformación de la práctica científica en varios niveles como el objeto de estudio, la dinámica de la producción científica, las relaciones entre científicos, los métodos y los formatos de información (Hine, 2002; Hine, 2005).

Esto ha reducido la energía, costo y tiempo requeridos para el análisis de información biológica (Wilson, 1994). Se han desarrollado nuevas técnicas analíticas, tecnologías de acceso y modelos de organización para explotar las colecciones digitales de manera innovadora como la minería de textos y la semántica utilizando algoritmos. Se diseñan a diario nuevas herramientas para realizar búsquedas más eficientes y precisas, que también permiten hacer análisis mejores y más extensos. Esto ha aumentado la eficacia

y la exactitud en la obtención de información y ha generado nuevas oportunidades de investigación al incrementar el espectro de opciones técnicas y análisis factibles. La automatización ha permitido a los investigadores desarrollar modelos y simulaciones de fenómenos lo que ha promovido la aceleración del conocimiento (NSF, 2005).

El acceso inmediato a colecciones de datos que contienen una inmensa variedad de temas y fuentes especializadas de diversas procedencias ha fomentado integración y multidisciplinariedad del conocimiento. La facilidad de compartir datos digitales ha aumentado la colaboración entre diferentes disciplinas y permitido que investigadores, estudiantes, educadores e instituciones localizadas en diversas áreas geográficas interrelacionen (NSF, 2005).

La repercusión ha sido tal que se han modificado notablemente las etapas en el ciclo de la información científica, ahora es posible conocer los contenidos de un artículo antes de su publicación (preprints), los formatos digitales y las herramientas de la *web* 2.0 y 3.0 permiten una consulta de la información más dinámica, completa y rápida; la *web* 2.0 permite que existan conexiones (colaboración y distribución) entre usuarios, mientras que la *web* 3.0 además de eso, desarrolla sistemas de interoperabilidad y estructura sistemáticamente la información en mapas cognitivos invisibles para el usuario que permiten la búsqueda de significados y no únicamente de términos, como en los orígenes de internet. En la actualidad se puede consultar instantáneamente un artículo seleccionado, reconocer a los autores, sus correos electrónicos, sus trabajos más citados y a las instituciones a las que ellos pertenecen. También es posible consultar los artículos más relevantes de manera inmediata, las publicaciones relacionadas, las más citadas y diversos indicadores bibliométricos como el factor de impacto, la obsolescencia y la vida media de una revista o un documento. Además, es fácil acceder directamente a las referencias, las citas y las imágenes relacionadas.

Los tiempos de publicación han disminuido notablemente, se han probado nuevas formas de certificar las publicaciones. Actualmente se pueden someter artículos a revisión en línea y el diseño de recursos públicos de libre acceso (*open access*) dan la oportunidad a todo el público a la consulta de sus colecciones de datos lo cual es un inicio hacia la democratización de la investigación y del conocimiento (NSF, 2005).

Se han acuñado nuevos términos para referirse a esta nueva forma de hacer ciencia: *e-ciencia* (*e-science*), *e-investigación* (*e-research*) y *ciberinfraestructura* (*ciberinfrastructure*; NSF, 2007). La realidad biológica ya no se explora únicamente a través de los métodos tradicionales como la observación, la comparación, los experimentos y los modelos *in vivo* e *in vitro* sino que actualmente muchos de los descubrimientos se hacen *in silico* procesando información digital con base en modelos reales; es decir, la forma de estudio en la biología, ha evolucionado (Fig. 1). También, han surgido nuevos campos del conocimiento como la informática biológica (*biological informatics*), ciencia que se enfoca al manejo y análisis de los diferentes tipos de información biológica (Heidorn, 2007) con ramas que incluyen bioinformática, neuroinformática, informática de la ecología, informática de la biodiversidad (Soberón y Peterson, 2004; Kostoff, 2007) y la *e-taxonomía* o *cibertaxonomía* (Hine, 2008), la genómica y la biología de sistemas, por mencionar las más comunes.

La ciberinfraestructura y las aplicaciones disponibles a través de la *web* han protagonizado transformaciones profundas en la práctica biológica en los últimos veinte años.

Historia Biología (Problemas)

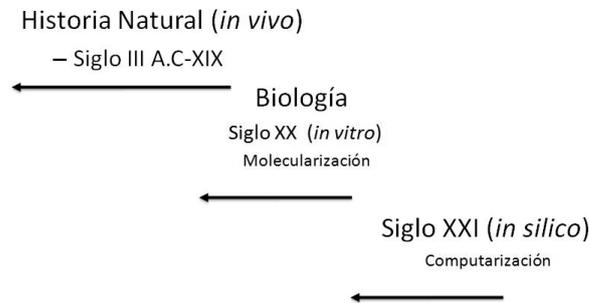


Figura 1. Evolución en el método de estudio de la Biología.

La primera se refiere a la infraestructura, programas y aplicaciones desarrolladas con tecnologías computacionales y formas complejas de comunicación que permiten opciones de seguridad y transferencia de datos, que han sido diseñadas por especialistas en cómputo Grid, han constituido las bases de la denominada e-ciencia. Esta tecnología permite analizar y compartir la información, facilitar el acceso inmediato a una variedad extensa de colecciones de datos especializadas de diversas procedencias y temas, es una de las características de la ciberinfraestructura que ha fomentado la colaboración en la investigación científica a gran escala (Hey y Trefethen, 2005), así como la integración y la interdisciplinariedad (NSF, 2005). Estos métodos facultan una reinención de la ciencia ya que permiten abordar el desafío que implica manejar conjuntos de datos grandes y complejos (Buetow, 2005).

Muchas de estas aplicaciones, utilidades, herramientas y servicios se encuentran disponibles en la *web*. Estos son de fácil acceso, eficientes y sencillos de usar, se convierten de gran utilidad para la investigación en ciencias biológicas ya que no demandan infraestructura mayor o asesoría personalizada de profesionales en bioinformática y/o ciencias de la computación. Una computadora personal y acceso a internet son los dos elementos que se requieren para poder explotar el uso de los recursos que la *web* ofrece (Truong *et al.*, 2006; Morales *et al.*, 2010; Michán *et al.*, 2010).

El desarrollo simultáneo de las ciencias de la información (CI), las tecnologías la información y la comunicación (TICS) en particular las bases de datos e internet, han producido formas sistémicas de analizar información en cantidades colosales (*terabites*) (Neufeld y Cornog, 1986; Saracevic, 1999). Este fenómeno ha repercutido de manera importante en la forma de hacer estudios y análisis sobre ciencia, tarea que ha sido tradicionalmente el objeto de investigación de filósofos, historiadores y sociólogos de la ciencia. Recientemente han aparecido nuevas (inter)disciplinas que aplican métodos innovadores, técnicas novedosas y herramientas integradoras para sistematizar, acceder y analizar la información generada por los científicos de acuerdo con el progreso de sus propias disciplinas (Schoepflin y Glanzel, 2001; Schubert, 2002) Esto permite estudiar a la ciencia con un sustento empírico, cuantificable y comparativo (Bradford, 1948; Leydesdorff, 2001; Garfield, Pudovkin *et al.*, 2002), lo cual aporta elementos interesantes para el análisis e interpretación de la dinámica científica (Wilson, 1994; Bar-Ilan 2008; Zitt y Bassecoulard, 2008).

La e-ciencia resulta del uso y aplicación de la ciberinfraestructura en la práctica científica, la inter y multidisciplinariedad, la colaboración, la participación de un gran número de investigadores (en algunos casos cientos) localizados en diversas regiones y con diferentes especialidades que forman grupos trabajo (Hey y Trefethen, 2005). Dentro de la ciberinfraestructura resaltan las grids y el uso de colecciones de información digitales en bases de datos.

Uno de los indicadores de este fenómeno ha sido el amplio uso de computadoras, la masificación de la *web*, la adopción del formato digital que implica bajo costo y poco espacio, la explosión de la información que se registra y publica en colecciones de datos inmensas (Fig. 2; Fig. 3), el procesamiento de miles y millones de datos simultáneamente (meta-análisis), la adopción y popularización del cómputo para la práctica científica, la adopción de nuevos modelos biológicos (Michán *et al.*, 2010) y la generación de nuevas disciplinas biológicas, tal es el caso de la genómica.

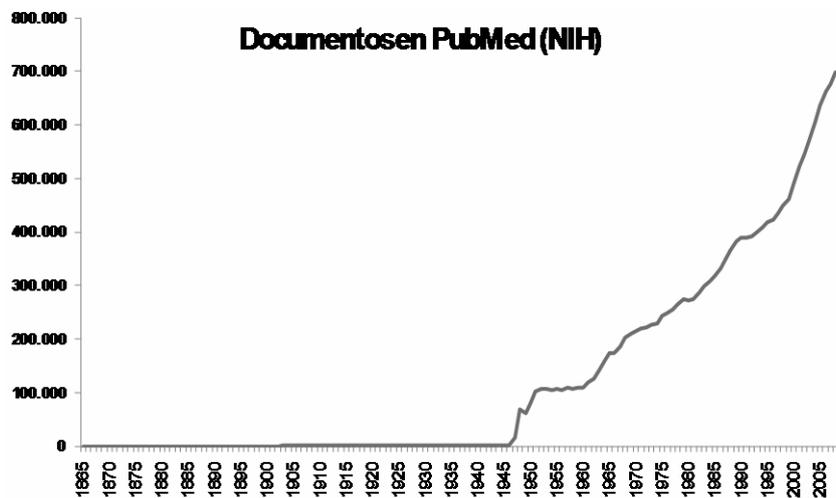


Figura 2. Explosión de la información que se registra y publica en colecciones de datos inmensas.

RESULTADOS

Uno de los primeros proyectos de e-ciencia más conocido es el proyecto genoma humano (PGH), cuyos primeros resultados se publicaron en el año 2001 con un día de diferencia en las revistas *Nature* y *Science*. El primero "*Initial sequencing and analysis of the human genome*" fue escrito por 79 autores adscritos a 48 instituciones diferentes de países y contiene 181 referencias bibliográficas. El segundo, "*The Sequence of the Human Genome*", firmado por 276 autores, de 14 instituciones diferentes y con 452 citas (*The International Human Genome Sequencing Consortium*, 2001; *The Sequence of the Human Genome*, 2001). El resultado de este gran proyecto fue sin lugar a dudas, la consolidación de la genómica, estudio sistemático de la documentación completa de las secuencias de ADN de los organismos (MeSH, 2010), dicho estudio ha ayudado a identificar polimorfismos dentro de especies, la interacción de proteínas y la evolución (Brent, 2000). Incluyetodos

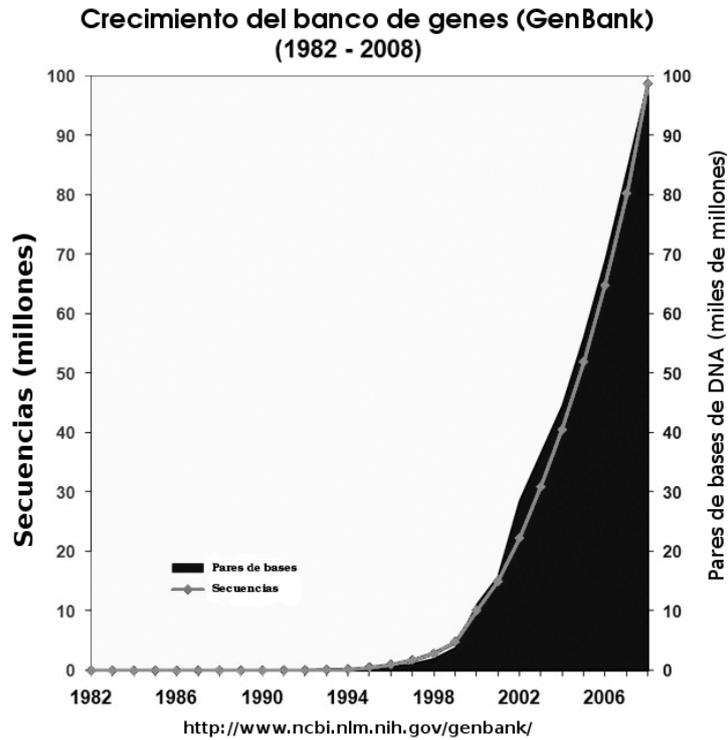


Figura 3. Producción acelerada sobre secuencias de DNA.

los métodos que recopilan y analizan datos completos acerca de los genes, como las secuencias, la abundancia de ácidos nucleicos y las propiedades de las proteínas que codifican (Murray, 2000). Nuestra capacidad para estudiar la función génica está aumentando en la especificidad gracias a esta nueva disciplina (Collins *et al.*, 2003) que promete acelerar el descubrimiento científico en todos los ámbitos de la ciencia biológica (Burley, 2000).

Los eventos históricos relevantes que marcaron la genómica son, por mencionar algunos y en orden cronológico: Gregorio Mendel descubre las leyes de la genética en 1865, re-descubrimiento de las leyes de la genética en 1900, James Watson y Francis Crick describen la estructura de la doble-hélice de ADN, Marshall Nirenberg, Har Gobind Khorana y Robert Holley determinan el código genético en 1966, en 1985 se propuso un proyecto con la finalidad de secuenciar los nucleótidos del genoma humano; en 1990 el PGH se inició oficialmente en Estados Unidos bajo la dirección de los institutos nacionales de salud (NIH) y el Departamento de Energía de Estados Unidos (DOE) con un plan de 15 años y 3 billones de dólares para completar la secuencia del genoma. En 1997 se forma el *Joint Genome Institute* y en el mismo año se inicia *The National Human Genome Research Institute* (NHGRI; Collins, *et al.*, 2003). El 14 de abril de 2003 el Consorcio Internacional de Secuenciación del Genoma Humano (NHGRI) y el DOE, anunciaron la terminación exitosa del PGH con dos años de anticipación (Cervantes, 2005).

El surgimiento del término ‘genómica’ se documentó como sigue: T. H. Roderick del laboratorio Jackson de Bar Harbor, Maine, propuso el término para la nueva disciplina que nace de la alianza de la biología molecular y celular con la genética clásica, y es a su vez impulsado por las ciencias computacionales. Sabemos también que la palabra ‘genoma’ es un híbrido de genes y cromosoma, término al que se suma el sufijo ‘ics’ (en inglés) para sugerir un método de ataque (McKusick y Ruddle, 1987). De manera formal el término ‘genómica’ apareció por primera vez en la literatura publicado en la revista *Pathology and Immunopathology Research* en 1988; en el artículo “*The Genomics of Human Homeobox-containing Loci*”, escrito por CA Ferguson-Smith y FH Ruddle del Departamento de Biología, en la Universidad de Yale, New Haven, Connecticut (Ferguson y Ruddle, 1988). Ya que el acceso a dicho artículo se encuentra restringido, desconocemos con exactitud la definición que proporcionaron al término, pero podemos estar seguros que hoy el término tiene una connotación mucho más amplia debido a las innumerables innovaciones que día a día se incorporan a este enfoque.

Fue a partir del año 1990 que comenzó la aparición de artículos en diferentes revistas científicas sobre genómica, mostrando un notable incremento en número en el año 1997, se presenta un periodo de estabilidad en los años 2004-2006, para después incrementar exponencialmente la producción hasta la fecha, actualizando a su vez el conocimiento sobre el tópico, no resulta raro que la influyente revista *Science* encabece el número de publicaciones sobre el tema (Fig. 4).

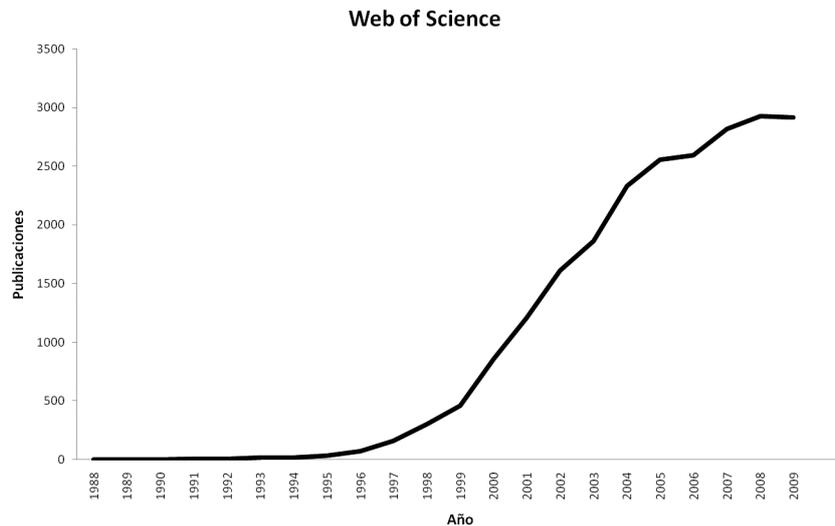


Figura 4. Aparición de artículos en diferentes revistas científicas sobre Genómica (por año), analizado en *Web of Science*.

Es la revista *Genomics* la primera publicación que en el nombre incluyó el término estudiado, publicada en el año 1987 ha evolucionado como el campo que lleva su nombre, cuenta con V. A. McKusick y F. H. Ruddle como sus editores fundadores y con J. Quackenbush del Dana-Farber Cancer Institute, Boston, MA, Estados Unidos como redactor en jefe (McKusick y Ruddle, 1987).

Dentro de los países que más publican artículos relacionados a genómica se encuentran en primer lugar Estados Unidos con el casi 50% de la producción, seguido por Alemania; gráfica en la cual América Latina no aportar ni el 0,1% (Fig. 5). La institución que más ha publicado en el tema de la genómica es la Universidad Harvard seguida por la Universidad de Washington (Fig. 6). El artículo más citado con la palabra genómica dentro del texto y palabras clave es el de Kumar *et al.*, 2004, con el título: “MEGA3: El *software* integrado para el análisis molecular y genética evolutiva alineamiento de secuencias citado 6.848 veces”.

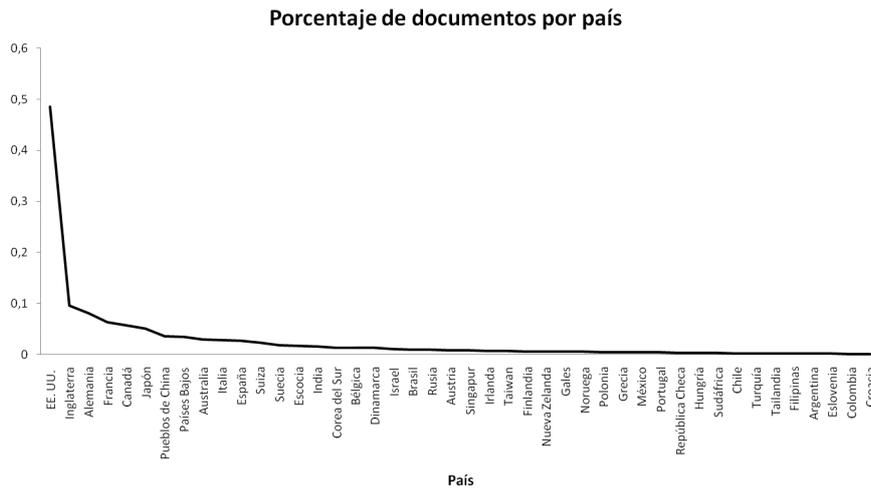


Figura 5. Porcentaje de publicación de artículos en diferentes revistas científicas sobre Genómica (por país), analizado en *Web of Science*.

Numerosas revisiones han sido realizadas respecto al tema tal es el caso de “Los avances en análisis genómico de trazos ¿Qué hemos aprendido y hacia dónde vamos? (Lanktree, 2010), donde se resaltan los avances tecnológicos del análisis genómico en tópicos de estudio como bioquímica y biología molecular, herencia y genética, biotecnología y microbiología aplicada, farmacología y farmacia, microbiología, biología celular, biofísica, oncología, inmunología, biología evolutiva, ecología, entre otras, áreas en las cuales la genómica ha innovado la manera de abordar la temática, haciendo tangibles aspectos que en el pasado pudieron ser un tanto subjetivos.

A partir del surgimiento de esta disciplina, se ha avanzado a pasos agigantados, muestra de nuestra aseveración como lo muestran los trabajos del científico estadounidense Craig Venter los cuales marcan un momento estelar en el desarrollo de la ciencia moderna. El pasado 20 de mayo Craig Venter, Daniel Gibson y 22 científicos más del Instituto J. Craig Venter de los Estados Unidos, anunciaron una de las noticias más sorprendentes en el mundo científico, noticia que con seguridad cambiará el rumbo de la biología como ciencia; la obtención sintética del genoma de un ser vivo, la bacteria *Mycoplasma genitalium*; es decir, “el primer paso para crear vida en el laboratorio, lo que provocará, no solo importantes avances científicos sino también grandes discusiones éticas y filosóficas” (Gibson *et al.*, 2010).

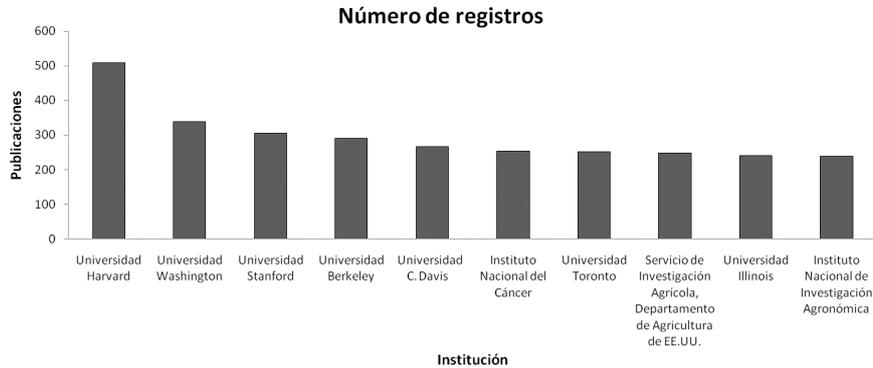


Figura 6. Publicación de artículos en diferentes revistas científicas sobre Genómica (por institución), analizado en *Web of Science*.

DISCUSIÓN

Causa y efecto de las TICs en biología, como mencionamos anteriormente, son las grandes colecciones de información digital que permiten procesar cantidades inmensas de información. El GenBank es una colección anotada de todas las secuencias de nucleótidos a disposición del público y su traducción de proteínas a cargo del *National Center for Biotechnology Information* (NCBI), *European Molecular Biology Laboratory* (EMBL) de datos de bibliotecas del *European Bioinformatics Institute* (EBI) y *DNA Data Bank* de Japón (DDBJ). En estas colecciones se registran las secuencias producidas en laboratorios de todo el mundo de más de 100.000 organismos distintos y crece a un ritmo exponencial duplicándose cada 18 meses (Fig. 2). STS's, un *software* producido en febrero de 2003, el cual contenía más de 29.300 millones de bases nucleotídicas en más de 23,0 millones de secuencias. Se construye mediante el envío directo de los distintos laboratorios y de los centros de secuenciación a gran escala (Fig. 7; Olson *et al.*, 1989).

Ha sido tal el impacto de la genómica en la ciencia moderna, que en pocos años han aparecido gran cantidad de las denominadas ciencias 'ómicas' (*"omics"*, por su término en inglés) que están involucradas en el análisis de grandes cantidades de parámetros bioquímicos, en vez de solo unos cuantos, generalmente son proteínas "proteómica" (*proteomics*), lípidos "lipidómica" (*lipidomics*) y metabolitos "metabolómica" (*metabolomics*). El término *'omics'* deriva del sufijo griego *'ome'* que significa muchos o masa. Las mediciones se hacen con base en marcadores químicos que son indicativos de algún evento biológico (*biomarkers*). Los valores asociados con los parámetros medidos, son analizados con el fin de encontrar una correlación con fenómenos o enfermedades. Su meta es comprender comprensiva e integralmente procesos biológicos, mediante la identificación y correlación de varios 'elementos' (e.g. genes, RNA, proteínas, metabolitos) en vez de estudiar cada uno de ellos de manera individual.

Lamentablemente, un estudio sobre la probabilidad de reproducir la aplicación de la investigación en *genomics* (asociar grupos de genes con enfermedades, en el cual se compararon 370 estudios, reveló una alta frecuencia de investigaciones con resultados no reproducibles. Debido al enorme número de mediciones y el número limitado de

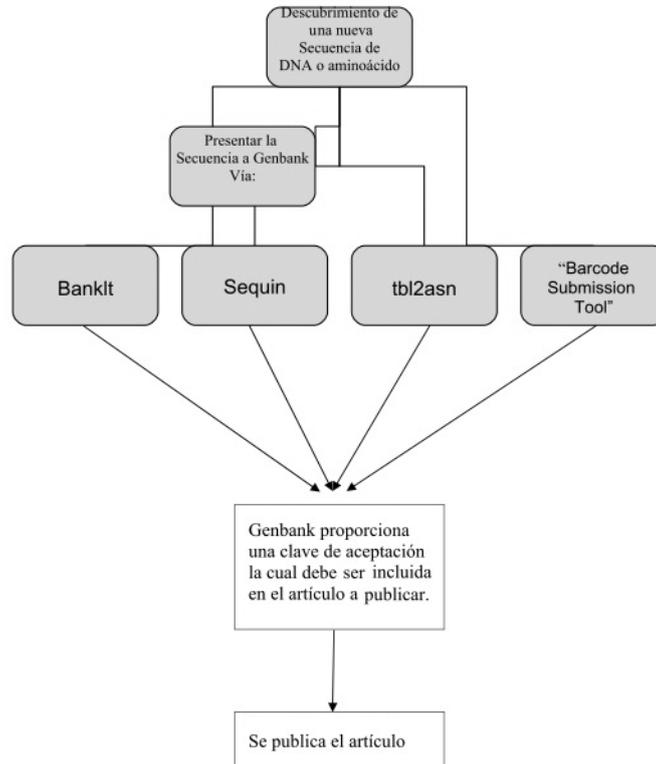


Figura 7. Construcción de GenBank mediante el envío directo de las secuencias desde distintos laboratorios y de los centros de secuenciación a gran escala.

muestras de investigación, surgen problemas relacionados a la estadística, el sesgo, la metodología y el uso inadecuado del método (Layjr *et al.*, 2006). La era de la genómica acaba de comenzar, nuestro genoma es “el hilo conductor, la fibra común que nos une a todos, nuestra herencia” (Sir John Sulston).

AGRADECIMIENTOS

A Lyssania Macías y a Judith Aguirre por la lectura crítica del texto. A Adrián Echevarría Harris por la realización de la figura 7. A dos revisores anónimos por sus atinadas sugerencias para mejorar el texto. **Fuente de financiamiento:** agradecemos la subvención del proyecto DGAPA, UNAM PAPIME PE 201509 y CONACYT 13276 para la realización de este trabajo.

BIBLIOGRAFÍA

BAR-ILAN J. Informetrics at the Beginning of the 21st Century—A review. *J Informetr* [Serial online] 2008 January;2(1):1-52. Disponible en: <http://dx.doi.org/10.1016/j.joi.2007.11.001>

-
- BRADFORD SC. Documentation. London: Crosby Lockwood and Son, 1948, Ltd.
- BRENT R. Genomic Biology. Cell. [Serial online] 2000 January;100(1):169-183. Disponible en: [http://dx.doi.org/10.1016/S0092-8674\(00\)81693-1](http://dx.doi.org/10.1016/S0092-8674(00)81693-1).
- BURLEY SK. An Overview of Structural Genomics. Nat Struct Mol Biol. [Serial online] 2000;7S:932-934. Disponible en: <http://dx.doi.org/10.1038/80697>.
- BUETOW KH. Cyberinfrastructure: Empowering a "Third Way" in Biomedical Research. Science. [Serial online], 2005;308(5723):821-824. Disponible en: <http://dx.doi.org/10.1126/science.1112120>
- CERVANTES GARCÍA D. Proyecto genoma humano: situación actual y perspectivas. INVESTIGACIÓN Y CIENCIA de la Universidad Autónoma de Aguascalientes. [Serial online] 2005;(33). Disponible en: <http://www.uaa.mx/investigacion/revista/archivo/revista33/Articulo%208.pdf>.
- COLLINS FS, GREEN ED, GUTTMACHER AE, GUYER MS. A Vision for the Future of Genomics Research. Nature. [Serial online] 2003;422(6934):835-847. Disponible en: <http://dx.doi.org/10.1038/nature01626>.
- ELIOT TS. (Frase final del artículo de *Nature* de febrero 15 de 2001 en el que el Consorcio Internacional que secuenció el Genoma Humano presenta el primer análisis del genoma. Nótese que el inicio de la frase es el mismo que el de la frase final del artículo de *Nature* de 1953 de Watson y Crick, con lo que se quiere hacer patente el carácter de hito científico de este trabajo, el más destacado tras el descubrimiento de Watson y Crick.) Disponible en: <http://bioinformatica.uab.es/genetica/curso/frases.html>
- FERGUSON-SMITH AC, RUDDLE FH. The Genomics of Human Homeobox-containing Loci. Pathol Immunopathol Res. [Serial online] 1988;7(1-2):119-126. Disponible en: <http://view.ncbi.nlm.nih.gov/pubmed/2906120>
- GARFIELD E, PUDOVKIN AI, ISTOMIN VS. Algorithmic Citation-linked Historiography-mapping the Literature of Science. Proc Amer Soc Info Sci Tech. 2002;39(1):14-24. Disponible en: <http://www.garfield.library.upenn.edu/papers/asis2002/asis2002presentation.html>
- GIBSON DG, GLASS JI, LARTIGUE C, NOSKOV VN, CHUANG RY, ALGIRE MA, BENDERS GA, *et al*. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Science. 2010;329(5987):52-56.
- HEIDORN PB, PALMER CL, WRIGHT D. Biological Information Specialists for Biological Informatics. J Biomed Discov Collab. [Serial online] 2007; 2:1. Disponible en: <http://dx.doi.org/10.1186/1747-5333-2-1>.
- HEYT, TREFETHEN AE. Cyberinfrastructure for e-Science. Science. [Serial online] 2005;308(5723):817-821. Disponible en: <http://dx.doi.org/10.1126/science.1110410>.
- HINE C. Cyberscience and Social Boundaries. Sociological Research Online. 2002;7(2).
- HINE C. Internet Research and the Sociology of Cyber-Social-scientific Knowledge. The Information Society. 2005; 21(4):239-248.
- HINE C. Connective ethnography for the exploration of e-Science. Journal of Computer-Mediated Communication. 2007; 12(2):618-634.
- HINE C. Systematics as Cyberscience: Computers, Change, and Continuity in Science. Cambridge: The MIT Press. [Serial online] 2008. Disponible en: <http://www.worldcat.org/isbn/026208371X>

KOSTOFF R, BRIGGS M, RUSHENBERG R, BOWLES C, PECHT M, JOHNSON D, *et al.* Comparisons of the Structure and Infrastructure of Chinese and Indian Science and Technology. *Technol Forecast Soc.* [Serial online] 2007;74(9):1609-1630. Disponible en: <http://dx.doi.org/10.1016/j.techfore.2007.02.007>.

KUMAR S, TAMURA K, NEI M. MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Brief Bioinform.* [Serial online] 2004;5(2):150-163. Disponible en: <http://dx.doi.org/10.1093/bib/5.2.150>

LANKTREE-M B. Advances in Genomic Analysis of Stroke What Have We Learned and Where Are We Headed? 2010 Disponible en: http://apps.isiknowledge.com/full_record.do?product=WOS&search_mode=Refine&qid=2&SID=2A8NFO8Aglg2GEeCl@h&page=1&doc=1

LAYJR J, LIYANAGE R, BORGMANN S, WILKINS C. Problems with the "Omics". *TrAC Trends Analyt Chem.* [Serial online] 2006;25(11):1046-1056. Disponible en: [doi: 10.1016/j.trac.2006.10.007](https://doi.org/10.1016/j.trac.2006.10.007).

LEYDESDORFF L. *The Challenge of Scientometrics: The Development, Measurement, and Self-Organization of Scientific Communications.* Universal Publishers, 2nd ed; 2001.

MCKUSICK VA, RUDDLE FH. A New Discipline, a New Name, a New Journal. *Genomics.* [Serial online]1987;71:1-2. [doi: 10.1016/0888-7543\(87\)90098-X](https://doi.org/10.1016/0888-7543(87)90098-X).

MeSH [Serial online] 2010. Biblioteca Nacional de Medicina - Medical Subject Headings. Disponible en: http://www.nlm.nih.gov/cgi/mesh/2010/MB_cgi.

MORALES EC. Infodiversidad, globalización y derecho a la información. *Soc Invest Bibliot.* [Serial online] Buenos Aires; 2003. Disponible en: [http://sbdigital.ues.edu.sv/asp/getFicha.asp?glx=215692.glx&skin=ISBD&recnum=66&maxrecnum=96&searchString=\(@buscable%20S\)\%20and%20\(@encabezamiento%20INFORMACION%20and%20LEMB\)\&orderBy=&pg=3&biblioteca](http://sbdigital.ues.edu.sv/asp/getFicha.asp?glx=215692.glx&skin=ISBD&recnum=66&maxrecnum=96&searchString=(@buscable%20S)\%20and%20(@encabezamiento%20INFORMACION%20and%20LEMB)\&orderBy=&pg=3&biblioteca)

MURRAY AW. Whither Genomics? *Gen Biol.* [Serial online] 2000. 1 (1). Disponible en: <http://dx.doi.org/10.1186/gb-2000-1-1-comment003>

NSF NATIONAL SCIENCE FOUNDATION. Where Discoveries Begin. Press Release 05-143 \$150 Million TeraGrid Award Heralds New Era for Scientific Computing. August 17, 2005. Disponible en: http://www.nsf.gov/news/news_summ.jsp?cntn_id=104248

NSF NATIONAL SCIENCE FOUNDATION. Where Discoveries Begin. Press Release 06-007 2005: Year in Review. January 11, 2006. Disponible en: http://www.nsf.gov/news/news_summ.jsp?cntn_id=105693

NEUFELD ML, D CORNOG M. Database History: From Dinosaurs to Compact Discs. *Journal of the American Society for Information Science.* 1986; 37(4):183-190.

OLSON M, HOOD L, CANTOR C, BOTSTEIN D. A Common Language for Physical Mapping of the Human Genome. *Science.* [Serial online] 1989; 245(4925): 1434-1435. Disponible en: <http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubmed&Cmd=ShowDetailView&TermToSearch=2781285>

SARACEVIC T. The Stratified Model of Information Retrieval Interaction: Extension and Applications. *Proceedings of the American Society for Information Science.* 1997;34,313-327.

SCHOEPFLIN U, GLÄNZEL W. Two Decades of "Scientometrics". An Interdisciplinary Field Represented by its Leading Journal. *Scientometrics.* [Serial online]

2001 February; 50(2):301-312. Disponible en: <http://dx.doi.org/10.1023/A:1010577824449>.

SCHUBERT A. The Web of Scientometrics. *Scientometrics*, [Version online] 2002;53(1):3-20. Disponible en: [L http://dx.doi.org/10.1023/A:1014886202425](http://dx.doi.org/10.1023/A:1014886202425)

SOBERÓN J, PETERSON T. Biodiversity Informatics: Managing and Applying Primary Biodiversity Data. *Philos Trans Royal Soc Lond Series B, Biol Sci.* [Serial online] 2004;359(1444):689-698. Disponible en: <http://dx.doi.org/10.1098/rstb.2003.1439>.

THE INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Initial Sequencing and Analysis of the Human Genome. *Nature.* [Serial online] 2001;409(6822):860-921. Disponible en <http://dx.doi.org/10.1038/35057062>

THE SEQUENCE OF THE HUMAN GENOME. *SCIENCE.* [Serial online] 2001;291(5507):1304-1351. [DOI:10.1126/science.1058040]. Disponible en: <http://www.sciencemag.org/citmgr?gca=sci;291/5507/1304>

TRUONG T N, NAYAK M, HUYNH HH, COOK T, MAHAJAN P, TRAN LT, *et al.* Computational Science and Engineering Online (CSE-online): A Cyber-Infrastructure for Scientific Computing. *J Chem Inf Model* [versión online] 2006;46(3):971-984. Disponible en: <http://dx.doi.org/10.1021/ci0503917>

Wilson TD. Information needs and uses: fifty years of progress. *Information seeking behaviour, Behavior, Information needs, Information use*, 1994. p. 15-51. London: Aslib. Disponible en: <http://informationr.net/tdw/publ/papers/1994FiftyYears.html>

ZITT M, BASSECOULARD E. Quelques Defis Des Indicateurs Scientometriques: “déménagement” des données; mesure des flux de connaissance ; questions de diversité. *La revue MODULAD.* [Serial online] 2008;38. Available from: <http://www-rocq.inria.fr/axis/modulad/archives/numero-38/Zittetal-38/38-res-zitt.htm>.