Aplicaciones bioinformáticas

Origen de la bioinformática

Bioinformatics applications

Origin of bioinformatics

José González Cabeza[1]

Recibido: 17 de agosto de 2017
Aceptado: 24 de agosto de 2017

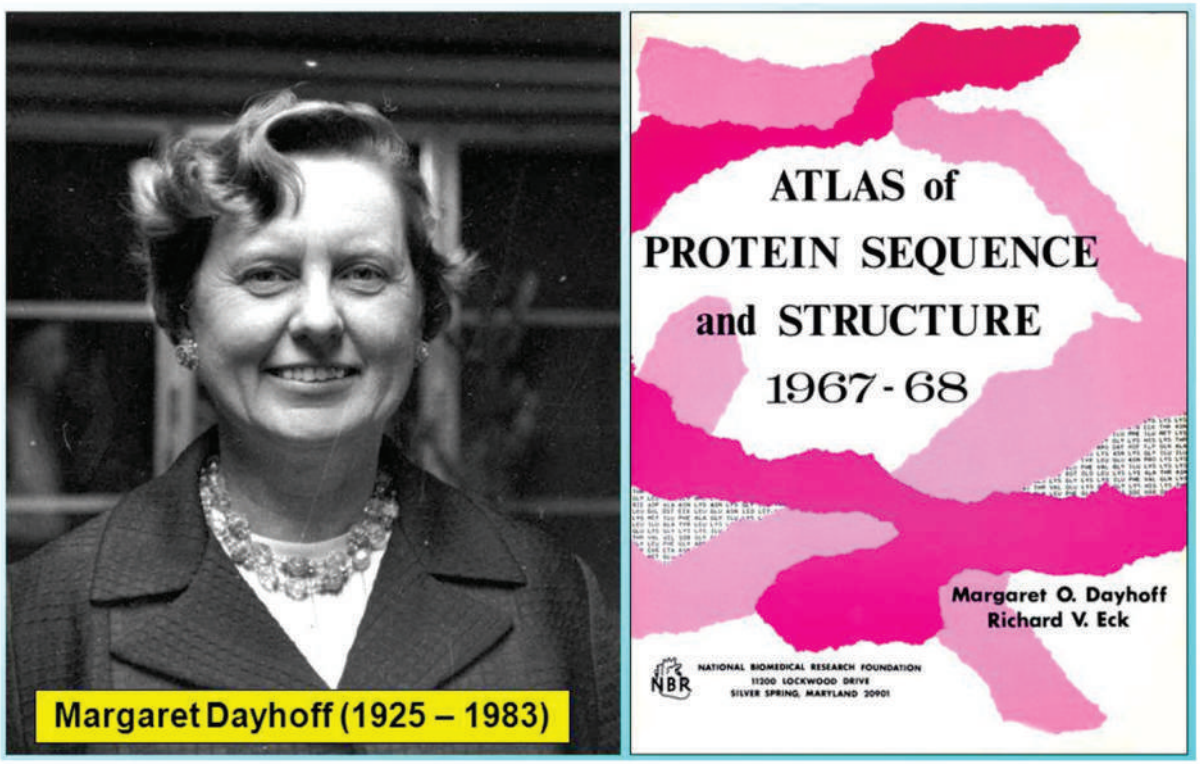
La ciencia de la bioinformática emerge en la actualidad, dentro de esta era post-genómica; sin embargo, de ninguna manera, es una ciencia nueva. Los trabajos pioneros se remontan a la década de los 60, con las investigaciones efectuadas por Margaret Dayhoff, Richard Eck y Robert Ledley, quienes a través de su experiencia y entrenamiento en informática (computación) potenciaron el análisis de datos de secuencias aminoacídicas de proteínas y evolución de proteínas asistido por computador, trabajos que pueden considerarse los pioneros dentro del mundo de la bioinformática.

Es así que en 1965, Dayhoff, Eck y otros investigadores, lograron compilar el primer *Atlas de estructura y secuencia de proteínas,* en el cual se presentaban aproximadamente unas 50 secuencias conocidas hasta ese momento. El segundo volumen de esta obra fue publicado en 1966, en que se reportó poco más de 100 secuencias. Todo ello resulta importante, dado que representan los estudios predecesores de las actuales bases de datos de genes y proteínas que constituyen la columna vertebral de la bioinformática. En los años posteriores, este atlas creció en tamaño y popularidad bajo el liderazgo de Dayhoff, el cual se convirtió en *The Protein Information Resource* (PIR), ahora bajo la administración de la Universidad de Georgetown.

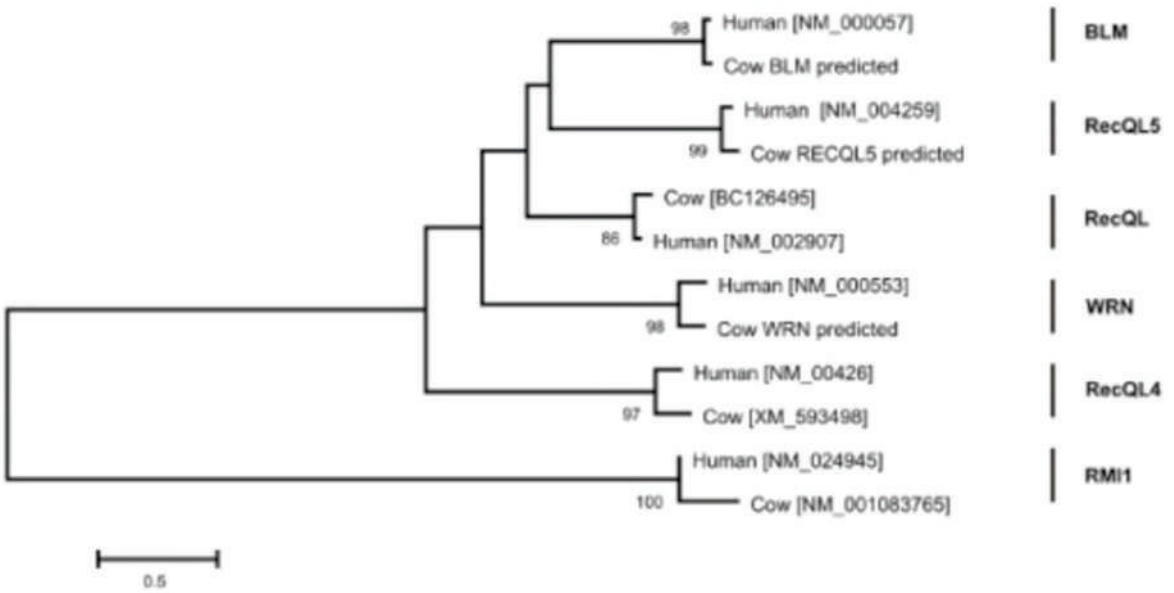
Margaret Belle Dayhoff, nació el 11 de marzo de 1925 en Filadelfia, y falleció el 5 de febrero de 1983; fue una fisicoquímica, profesora del Centro Médico Universitario de la Universidad de Georgetown, y una notable investigadora en bioquímica de la National Biomedical Reserche Foundation de los EE.UU. Se doctoró en el Departamento de Química de la Universidad de Columbia, donde diseño métodos computacionales para calcular energías de resonancia molecular de varios compuestos orgánicos. Realizó estudios postdoctorales en el Instituto Rockefeller (hoy Universidad Rockefeller) de la Universidad de Maryland, y se afilió en 1959 a la por aquel entonces recientemente creada *National Biomedical Research Foundation.* Fue la primera mujer en ocupar un cargo en la *Biophysical Society,* primero como secretaria para terminar como presidenta.

Consecuencia de su sólida formación en matemáticas, química y computación, direccionó todo ello para resolver problemas biológicos, particularmente en química de proteínas, y fue la pionera en la aplicación de las matemáticas y métodos computacionales a la bioquímica. Una de sus más importantes contribuciones fue desarrollar conjuntamente con Richard Eck, el código de una sola letra para los aminoácidos, utilizado por todas las herramientas bioinformáticas de análisis de proteínas. Todo ello, fiel reflejo de su intención de reducir el tamaño de los archivos empleados para describir las secuencias de aminoácidos en la era de la computación mediante tarjetas perforadas.

[1] Doctor en Biología, jefe del Laboratorio de Microbiología Molecular y Biotecnología, UPAO. jgonzalezc1@upao.edu.pe



Margaret Dayhoff, pionera en el campo de la bioinformática, que conjuntamente con Richard Eck publicaron el primer Atlas de secuencia de proteínas y estructura, en el año 1965.



Arriba, el equipo de Margaret Dayhoff (última a la derecha) con el ordenador. Abajo, un árbol filogenético. Fuentes: http://www.nlm.nih.gov/ y http://openi.nlm.nih.gov/

Richard Eck, estudió Ingeniería Química y Biología Vegetal. En 1961, Eck publicó un artículo en *Nature,* en el que comparó todas las secuencias de las variantes de hemoglobina y otras proteínas como la insulina, de diversas especies. De aquellas investigaciones dedujo que la información de dichas secuencias de aminoácidos podía organizarse de diferentes maneras y presentar patrones específicos. También identifico, numerosas sustituciones de aminoácidos en proteínas y que este patrón de sustituciones no era aleatorio. En una conferencia del año de 1964, presentó un método criptograma, método para rastrear la evolución de proteínas; él sugirió que empleando tales resultados uno puede calcular el grado de parentesco de una proteína con respecto a sus antepasados y dibujar un árbol genealógico, donde las distancias de las ramas exhibidas representan una medida cuantitativa del parentesco. Por tanto, así Eck estableció los cimientos de la reconstrucción de árboles filogenéticos.

Robert Ledley, estudió Física Teórica y Odontología, previo a los grandes aportes dentro de la aplicación de los ordenadores para el análisis de secuencias. Él sugirió que la cadena polipeptídica puede ser cortada en muchos fragmentos, los solapamientos de secuencias, las que podían ser analizadas por secuenciación de péptidos y darían la secuencia completa de la proteína, todo esto con auxilio de los ordenadores. Consecuencia de ello Ledley, sugiere que las computadoras pueden servir de soporte a los bioquímicos para establecer la secuencia de proteínas. Posteriormente, invitó a Dayhoff a laborar en la Oficina Nacional de Normas (NBRF), denominado más adelante Instituto Nacional de Estándares y Tecnología (NIST) en 1960, para continuar investigando en esta área. Dayhoff y Ledley escribieron programas FORTRAN, los cuales podían servir para el ensamblaje parcial de secuencias de péptidos parciales de forma correcta en menos de 5 minutos.

Robert Steven Ledley (1926 – 2012)

Dayhoff y Eck se involucraron en estudios evolutivos de proteínas, mientras que Ledley continuó con su interés en el empleo de computadoras dentro del campo biológico. Dayhoff, publicó por primera vez, la reconstrucción de un árbol filogenético basado en el método de máximo parsimonia; asimismo, ella desarrolló la primera matriz de sustitución de aminoácidos para estudiar la evolución de las proteínas, llamada matriz PAM, que significa mutación aceptada en un punto (también referida como el porcentaje de mutación aceptada), porque representa la mutación puntual aceptada por cada 100 residuos de aminoácidos. Una publicación de Dayhoff en la revista *Scientific American,* titulada *Computer Analysis of Protein Evolution,* puede considerarse como una de las más importantes publicaciones iniciales en bioinformática y filogenética molecular. Consecuencia de todo lo anterior, es legítimo considerarla fundadora de la bioinformática moderna.

DEFINICIÓN DE LA BIOINFORMÁTICA

El término "bioinformática" fue aplicado por Paulien Hogeweg y Ben Hesper en 1978. En un reciente artículo de revisión, donde se recapitula la historia de la bioinformática, Hogeweg declara que tanto él como Hesper, lo emplearon desde inicios de los años 70; no obstante, fue acuñado formalmente recién en 1978 en un artículo escrito en holandés. Desde un inicio el término se utilizó para denotar el estudio de procesos informáticos en sistemas biológicos. La bioinformática es básicamente, informática aplicada a la biología; es decir, el análisis asistido por computadora de bases de datos. Sin embargo, existen muchas definiciones y descripciones de la bioinformática; algunas de ellas, no hacen ninguna distinción entre la bioinformática y la biología computacional. Luscombe y col. definen la bioinformática como:

"La bioinformática, es conceptualizada biológicamente en términos de moléculas (en el sentido de su fisicoquímica), en la que se aplican técnicas de "informática" (derivadas de disciplinas de matemáticas aplicadas, y estadísticas), para comprender y organizar la información asociada a estas moléculas a una gran escala" (2001:347).

Asimismo Higgs y Attwood brindan dos definiciones de bioinformática, que en esencia son lo mismo, pero bajo diferentes ópticas: *1. La bioinformática es el desarrollo de métodos computacionales para estudiar la estructura, la función y la evolución de genes, proteínas y genomas completos. 2. La bioinformática es el desarrollo de métodos, para la gestión y el análisis de la información biológica generada por los amplios avances de la genómica.*

Por lo tanto, para los biólogos moleculares, la bioinformática es la disciplina del análisis asistido por computadora de la información relacionada con genes, genomas y sus productos. En otra palabras, para todos los propósitos prácticos, la bioinformática considerada como biología molecular computacional, que utiliza técnicas computacionales para estudiar la estructura, función, regulación y la intrincada red interactiva de genes y proteínas. El objetivo final es analizar y predecir la estructura, organización, función, regulación y dinámica de todo el genoma de un organismo.

BIOINFORMÁTICA Y BIOLOGÍA COMPUTACIONAL

La biología computacional es un término general, que incluye cualquier subdisciplina en biología que use el análisis asistido por computadora, modelamiento y predicción. Algunos ejemplos incluyen el modelamiento de relaciones presa-predator en un ecosistema, modelamiento y predicción de poblaciones en un ecosistema, estructura cuantitativa, análisis de actividad y predicción de efectos biológicos por productos químicos, predicción del destino metabólico de productos químicos *in vivo,* y el modelado farmacocinético de fármacos y xenobióticos; en contraste, la bioinformática puede considerarse como biología molecular computacional, como se había señalado anteriormente.

Por lo tanto, de acuerdo con las definiciones anteriores, la biología computacional es mucho más amplia en su alcance, y la bioinformática es parte de ella. La bioinformática como otras áreas de la biología computacional, es esencialmente una ciencia multidisciplinaria, porque utiliza técnicas y conceptos de una serie de disciplinas, tales como la biología molecular y bioquímica, ciencias computacionales, estadística y matemáticas, e informática (ciencias informáticas).

OBJETIVOS DEL ANÁLISIS BIOINFORMÁTICO

El objetivo medular de la bioinformática es la capacidad de poder predecir procesos biológicos bajo condiciones de salud y enfermedad; para ello, es necesario tener la capacidad de comprender los procesos biológicos como tal, dado que resulta elemental para el análisis y la integración de la información obtenida a partir de los genes y las proteínas; a su vez, es tan necesario para poder desarrollar nuevas herramientas y mejorar el conjunto de las ya existentes para estos análisis.

Además de lo anterior, la bioinformática también tiene como objetivo, desarrollar herramientas que ayuden en la gestión y acceso a la información; involucrando una mejora de la investigación y la capacidad para recuperar datos genómicos a partir de múltiples bases de datos. Algunos ejemplos comunes de herramientas y análisis bioinformáticos que continuamente son mejorados y optimizados son: capacidad de búsqueda y almacenamiento de datos; utilización de las bases de datos; análisis de los datos; análisis de secuencias de ácidos nucleicos y de proteínas, conjuntamente con la anotación de secuencias; análisis estructural de proteínas y la predicción de la estructura de las proteínas, incluyendo estructura tridimensional (3D); predicción de dominios proteicos; predicción de genes; análisis de estudios funcionales; análisis de redes de genes y proteínas; y análisis filogenético.

Las herramientas analíticas en bioinformática, son algoritmos computacionales y estadísticos. Las mejoras en las capacidades existentes y el desarrollo de nuevas herramientas, están impulsados por la necesidad de nuevas interrogantes, poseer una mayor velocidad de análisis, así como la capacidad para poder administrar una cantidad cada vez una mayor de datos; sin embargo, el éxito y precisión en la predicción del análisis bioinformático, depende en última instancia de nuestro conocimiento que podamos tener sobre la biología de los organismos. Por lo tanto, a medida que se acumula una mayor información en las bases de datos, y exista una mayor disponibilidad de la información científica, esto marcará el progreso de esta ciencia, y su pronóstico estará dictado por el desarrollo de nuevas herramientas bioinformáticas.

BIOINFORMÁTICA COMO INSTRUMENTO TÉCNICO

El análisis bioinformático requiere de datos (como información de secuencias), bases de datos y herramientas de análisis. Las bases de datos se construyen a partir de datos obtenidos experimentalmente en el laboratorio; algunos de estás bases de datos para proteínas fueron creados hace más de 30 años atrás; hoy, las informaciones de estas bases de datos han sido curadas, resultando más refinadas y específicas para la investigación.

REFERENCIAS BIBLIOGRÁFICAS

Barreto Hernandez, E. 2002. Bioinformática: Historia y Perspectivas Futuras. Colombia Ciencia y Tecnología. Julio-Setiembre, Vol. 20, Número 3 COLCIENCIAS. Bogotá-Colombia pp.36-44.

Choudhuri, Supratim. 2014. Bioinformatics for Beginners. Genes, Genomes, Molecular Evolution, Databases and Analytical Tools. Edit. Elsevier INC. 226 p.p.

Luscombe, N.M. 2001. What is Bioinformatics? A Proposed Definition and Overview of the Field. Method Inform Med. 4:346-358.

Ramsden, Jeremy. 2009. *Bioinformatics. An Introduction.* 2°Edic. Edit. Springer.

Pevzner, P.; Shamir, R. 2011. *Bioinformatica for Biologists.* Printed in the United Kingdom at the University Press, Cambridge. 394.

Xiong, Jin. 2006. *Essentials Bioinformatics.* Published in the United States of America by Cambridge University Press, New York.