

Bio-informática un Campo por conocer - Bio-computer science a Field to know

Víctor Hernán Arcila Quiceno

Médico Veterinario Zootecnista U de C.
Grupo de Investigación en Ciencias Animales Universidad Cooperativa
de Colombia

Contacto por e_mail: varcila23@hotmail.com

Resumen

El desarrollo de la Tecnología de la Información y la Comunicación ha generado nuevos campos de conocimiento los cuales manejan un lenguaje universal (lenguaje de las computadores u ordenadores) y a su vez permiten utilizar los datos que en muchos de los casos son gratuitos y se encuentran publicados en la web o internet pero que requiere una formación académica en biología y sistemas informáticos para dar un uso adecuado de ella. La interacción de diversas disciplinas permite tener una visión diferente de la ciencia y permite extraer mayor información de manera eficiente acelerando los procesos de investigación. En la actualidad constantemente se crean programas que refinan las búsquedas en bases de datos biológicas o que permiten identificar la estructura molecular de forma grafica a partir de la información presente en una base de datos de proteínas, carbohidratos o lípidos; o reconocer la composición de un gen en una especie animal o vegetal, todo esto es el aporte de la interacción entre el conocimiento de la biología y la informática reunidos en una disciplina conocida como **Bioinformática**.

Summary

The development of the Technology of the Information and the Communication has generated new fields of knowledge which manage an universal language (language of the computers) and in turn allow to use the data that are gratuitous in many of the cases and they are published in the web or internet but that it requires an academic formation in biology and computer systems to give an appropriate use of its. The interaction of diverse disciplines allows to have a vision different from the science and it allows to extract bigger information in an efficient way accelerating the investigation processes. At the present time constantly programs are believed that refine the searches in biological databases or that they allow to identify the molecular structure in a graphic way starting from the present information in a database of proteins, carbohydrates or lipids; or to recognize the composition of a gene in an animal or vegetable species, all this is the contribution of the interaction among the knowledge of the biology and the computer science gathered in a well-known discipline as **Bioinformatic**.

Con el creciente auge de la tecnología de la computación se han desarrollado líneas de especialización aplicadas a cada una de las ciencias de la vida y dentro de ellas la biología. Esto ha generado un **nuevo vocabulario y conocimiento** que no debería ser extraño a los profesionales que se relacionen con las ciencias biológicas incluyendo biólogos, bacteriólogos, microbiólogos, médicos, médicos veterinarios, entre otras profesiones. Cada vez es más frecuente encontrar grupos de investigación interdisciplinarios los cuales son constituidos considerando el aporte de cada individuo con puntos de vista y percepción diferente sobre un proceso determinado y particular común a todos.

En la labor diaria de un profesional el uso de un computador (ordenador) se ha convertido en una herramienta primaria para el desarrollo de su trabajo; se hace necesario el conocimiento básico para manejarlas y utilizar las aplicaciones conocidas como "software" y realmente lo mínimo que se espera es que se desempeñe en procesadores de texto y hojas de calculo y por que no, manejar adecuadamente algunas aplicaciones frecuentes. Usualmente no tiene importancia el considerar los software comerciales (que corren en Windows como OFFICE, LOTUS, etc) frente a los conocidos como Libres (Unix, LINUX), pero cada vez es mas importante el uso de software aplicado a nuestra área de desempeño es decir utilizar diferentes herramientas computacionales creadas para desarrollar nuestra labor profesional como Bases de Datos (DB del ingles Data Base) creadas con información de diversos tipos (desde el conocimiento general hasta el especializado) la cual puede ser consultada acorde a nuestro interés en servidores o equipos locales o a través de la "World Wide Web - www" para búsqueda de información general que es el uso que le da el común de la población.

Ahora existe un volumen importante de información biológica (y la cual crece día a día) de carácter gratuita disponible en la web y la cual requiere un conocimiento en la utilización de aplicaciones informáticas para sacar el máximo provecho de procesos que generados en el laboratorio gastarían mas tiempo y aun costo mayor por ejemplo como el desarrollo del Genoma no solo en el hombre sino en algunas especies de plantas y animales.

Para quienes se forman en las ciencias de la biología los computadores son equipos que nos ayudan en nuestro trabajo (herramienta) y tratamos de acomodarnos a las diversas aplicaciones como fueron planteadas considerando las propuestas de los desarrolladores – usualmente ingenieros de sistemas (informáticos) – quienes se forman en las ciencias de la computación pero carecen de conocimientos en biología pero obviamente son los llamados a actuar como desarrolladores a partir de los requerimientos que a ellos se plantean y quienes buscan que las aplicaciones planteadas o desarrollados por mecanismos como los **algoritmos** sean las adecuadas para quien será el usuario final.

Con mayor frecuencia se plantea la necesidad de la interacción entre profesionales de las ciencias biológicas con desarrolladores de diversas aplicaciones (software) buscando solucionar las necesidades en manejo y análisis de la información biológica considerando las variantes que se puedan predecir; a partir de esto surgen ciencias como la bioinformática o Biología Molecular Computacional.

Se han planteado tres acepciones en las que se unen la biología y la informática, pero con objetivos y metodologías bien diferenciadas haciendo claridad que esto dependerá del autor que se consulte:

1- Bioinformática o Biología Molecular Computacional: investigación y desarrollo de la infraestructura y sistemas de información y comunicaciones que requiere la biología molecular y la genética (Redes y bases de datos para el genoma, análisis de microarrays, entre otros). (Informática aplicada a la biología molecular y la genética). La **Bioinformática** es el uso de las matemáticas y de las técnicas informáticas para resolver problemas biológicos, normalmente creando o usando programas informáticos, modelos matemáticos o ambos. Una de las principales aplicaciones de la bioinformática es la simulación, la minería de datos (data mining) y el análisis de los datos obtenidos en los proyectos genoma o el proteoma. Otras aplicaciones son el alineamiento de secuencias, la predicción de estructuras proteicas y las redes metabólicas.

2- Biología Computacional: computación que se aplica al entendimiento de cuestiones biológicas básicas, no necesariamente en el nivel molecular, mediante la modelización y simulación. (ecosistemas, modelos fisiológicos). (Informática y matemáticas aplicadas a la biología).

3- Biocomputación: desarrollo y utilización de sistemas computacionales basados en modelos y materiales biológicos. (Biochips, biosensores, computación basada en ADN, redes de neuronas, algoritmos genéticos).

Sin embargo para algunos autores o grupos que trabajan sobre este conocimiento consideran que no deberían plantearse diferencias y encuadrar este conocimiento en una ciencia definida como Bioinformática o Biología Computacional.

Dos "escuelas" la de EE.UU y la del Reino Unido con Australia le dan diferentes coberturas al término.

Para los estadounidenses la bioinformática se enmarca dentro de lo mencionado en los párrafos anteriores.

Para los británicos la bioinformática se extiende al estudio de cualquier tipo de dato obtenido de entidades biológicas independientemente de la escala de estudio, desde moléculas hasta paisaje, ecosistemas etc.

En la web es posible encontrar definiciones como las siguientes:

“Bioinformática es una disciplina científica emergente que utiliza tecnología de la información para organizar, analizar y distribuir información biológica con la finalidad de responder preguntas complejas en biología”.

“Bioinformática es un área de investigación multidisciplinaria, la cual puede ser ampliamente definida como la interfase entre dos ciencias: Biología y Computación y esta impulsada por la incógnita del genoma humano (y actualmente con genoma de diversas especies animales y plantas) y la promesa de una nueva era en la cual la investigación genómica puede ayudar dramáticamente a mejorar la condición y calidad de vida humana. Avances en la detección y tratamiento de enfermedades y la producción de alimentos genéticamente modificados son entre otros ejemplos de los beneficios mencionados más

frecuentemente. Involucra la solución de problemas complejos usando herramientas de sistemas y computación. También incluye la colección, organización, almacenamiento y recuperación de la información biológica que se encuentra en base de datos”.

La Bioinformática se ha desarrollado de la necesidad de entender el código de la vida, el ADN no solo en los humanos si no igualmente sobre todos los seres vivos. Proyectos de secuenciamiento de ADN masivos han evolucionado y adicionado en el crecimiento de la ciencia bioinformática. El ADN es la molécula la cual directamente controla los fundamentos biológicos de la vida. Estos codifican para genes los cuales codifican para proteínas las cuales determinan las características biológicas para cualquier organismo viviente.

Según la definición del Centro Nacional para la Información Biotecnológica "National Center for Biotechnology Information" (NCBI por sus siglas en Inglés, 2001):

"Bioinformática es un campo de la ciencia en el cual confluyen varias disciplinas tales como: biología, computación y tecnología de la información. El fin último de este campo es facilitar el descubrimiento de nuevas ideas biológicas así como crear perspectivas globales a partir de las cuales se puedan discernir principios unificadores en biología. Al comienzo de la "revolución genómica", el concepto de bioinformática se refería sólo a la creación y mantenimiento de base de datos donde se almacena información biológica, tales como secuencias de nucleótidos y aminoácidos. El desarrollo de este tipo de base de datos no solamente significaba el diseño de la misma sino también el desarrollo de interfaces complejas donde los investigadores pudieran acceder los datos existentes y suministrar o revisar datos. Luego toda esa información debía ser combinada para formar una idea lógica de las actividades celulares normales, de tal manera que los investigadores pudieran estudiar cómo estas actividades se veían alteradas en estados de una enfermedad. De allí viene el surgimiento del campo de la bioinformática y ahora el campo más popular es el análisis e interpretación de varios tipos de datos, incluyendo secuencias de nucleótidos y aminoácidos, dominios de proteínas y estructura de proteínas a partir de aplicaciones de predicción.

Algunas subdisciplinas importantes dentro de la Bioinformática: Se basan en el desarrollo de nuevos algoritmos (fórmulas matemáticas o métodos de resolución de problemas que emplean una secuencia mecánica de pasos) y estadísticos con los cuales se pueda relacionar partes de un conjunto enorme de datos, como por ejemplo métodos para localizar un gen dentro de una secuencia, predecir estructura o función de proteínas y poder agrupar secuencias de proteínas en familias relacionadas que aseguren relaciones entre miembros de grandes bases de datos. Igualmente el análisis e interpretación de varios tipos de datos (nucleótidos, secuencias de a.a, dominios de proteínas, estructuras proteicas). Desarrollo e implementación de herramientas que capaciten el acceso eficiente y manejo de diferentes tipos de información. (www.ncbi.nlm.nih.gov/Education/index.html).

La información obtenida es almacenada en equipos pertenecientes a centros de investigación conocidos como servidores y ellos son las fuentes originales de información, mantienen las bases de datos de mayor importancia, y adicionalmente proporcionan acceso

a las mismas así como a un gran número de facilidades para procesar y digerir esta información.

La bioinformática se nutre especialmente de dos grandes áreas del conocimiento, las ciencias biológicas y las ciencias de la computación, dado este origen existen dos grandes líneas de trabajo: La primera en la cuál las ciencias de la computación utilizan modelos de las biológicas, ejemplo de ello lo constituyen las redes neurales, los algoritmos genéticos, computación con ADN, entre otras. La segunda en la cuál las ciencias biológicas utilizan modelos y herramientas de las ciencias de la computación, como se menciona anteriormente.

Tal vez una de las principales herramientas a utilizar son las Bases de Datos las cuales se han creado considerando la creciente información y estas pueden ser públicas o privadas (con acceso restringido o con un valor económico) manejadas por instituciones de diversos sectores de producción.

Las primeras bases de datos públicas de amplio interés biológico fueron las bases de datos de secuencias de ácidos nucleicos iniciadas en el **EMBL**, el **Laboratorio Europeo de Biología Molecular**. Esta iniciativa fué imitada enseguida en los Estados Unidos de Norteamérica con la base de datos GenBank, y se extendió pronto a otras bases de datos de macromoléculas como SwissProt, PIR (secuencias de proteínas) o PDB (Bases de Datos para Proteínas pero considerando las estructuras 3D). Las instituciones involucradas han procurado constantemente facilitar el acceso a las mismas con todos los medios a su alcance, incluyendo -por supuesto- la Web.

La Medicina molecular y la Biotecnología constituyen dos áreas prioritarias científico tecnológicas para el desarrollo e Innovación Tecnológica. El desarrollo en ambas áreas está estrechamente relacionado. En ambas áreas se pretende potenciar la investigación genómica y postgenómica así como de la bioinformática, herramienta imprescindible para el desarrollo de estas. Debido al extraordinario avance de la genética molecular y la genómica, la Medicina Molecular se constituye como arma estratégica del bienestar social del futuro inmediato. Se pretende potenciar la aplicación de las nuevas tecnologías y de los avances genéticos para el beneficio de la salud.

Dentro de las actividades financiables, existen acciones estratégicas, de infraestructura, centros de competencia y grandes instalaciones científicas. En esta área, la dotación de infraestructura se plasmará en la creación y dotación de unidades de referencia tecnológica y centros de suministro común, como Centros de Bioinformática, que cubran las necesidades de la investigación en Medicina Molecular. En cuanto a centros de competencia, se crearán centros de investigación de excelencia en hospitales en los que se acercará la investigación básica a la clínica, así como centros distribuidos en red para el apoyo a la secuenciación, DNA microarrays y DNA chips, sumado, en coordinación con la red de centros de investigación genómica y proteómica que se proponen en el área de Biotecnología. En esta área la genómica y proteómica se fundamenta como acción estratégica o instrumento básico de focalización de las actuaciones futuras.

Las tecnologías de la información jugarán un papel fundamental en la aplicación de los desarrollos tecnológicos en el campo de la genética a la práctica médica como refleja la presencia de la Bioinformática médica y la Telemedicina dentro de las principales líneas en patología molecular. La aplicación de los conocimientos en genética molecular y las nuevas tecnologías son necesarios para el mantenimiento de la competitividad del sistema sanitario no sólo paliativo sino preventivo. La identificación de las causas moleculares de las enfermedades junto con el desarrollo de la industria biotecnológica en general y de la farmacéutica en particular permitirán el desarrollo de mejores métodos de diagnóstico, la identificación de dianas terapéuticas y desarrollo de fármacos personalizados y una mejor medicina preventiva.

El Centro Nacional para la Información en Biotecnología **NCBI** (del inglés **National Center for Biotechnology Information**) depende de la Librería Nacional de Medicina **NLM** (del inglés **National Library of Medicine**) y el Instituto Nacional de Salud **NIH** (del inglés **National Institutes of Health**) coordina junto al Instituto de Bioinformática Europeo **EBI** (del inglés **European Bioinformatics Institute**) y el Banco de Datos de DNA de Japón **DDBJ** (del inglés **DNA Data Bank of Japan**) el mantenimiento sincronizado de las principales bases de datos de interés biológico. Al igual que el EBI, el NCBI también dispone de un sistema de consulta de las bases de datos que mantiene, en este caso basado en Entrez (<http://www.ncbi.nlm.nih.gov/Entrez/>). Entrez permite consultar no solo bases de datos sino también la biblioteca digital del NCBI, que incluye un número creciente de libros entre los cuales figuran textos de referencia reconocidos internacionalmente.

Una de las áreas de mayor interés para las ciencias biológicas en general es la utilización de Bases de Datos las cuales reúnen información de diferentes tipos y son pilares para la construcción del conocimiento generado a nivel mundial.

Bases de datos

Una Base de Datos es una colección de datos lógicamente interrelacionados y los cuales se crean con la finalidad de facilitar el manejo de la información sobre áreas del conocimiento particular.

Existen **bases** de datos primarias, que contienen información directa de la secuencia, estructura o patrón de expresión de ADN o proteína, y secundarias, que contienen datos derivados del análisis de las bases de datos primarias, como mutaciones, relaciones evolutivas, agrupación por familias o funciones, implicación en enfermedades, etc.

Bases de Datos considerando nucleótidos

La colaboración de las tres bases de datos más importantes hace posible acceder a casi toda la información de secuencias de ADN desde cualquiera de sus tres sedes:

- **EMBL-BANK** en el Instituto europeo de Bioinformática (**EBI**) Enlace externo: EMBL-BANK (<http://www.ebi.ac.uk/embl/index.html>)

- **DNA Data Bank of Japan (DDBJ)** en el Centro de Información Biológica (**CIB**)
Enlace externo: DDBJ (<http://www.ddbj.nig.ac.jp/>)
- **GenBank** en el Centro Nacional de Información Biotecnológica (**NCBI**) Enlace externo: GenBank Entrez Nucleotide (<http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Nucleotide>)

Bases de Datos considerando proteínas

Bases de datos de secuencias de aminoácidos

- **Swissprot** contiene secuencias anotadas o comentadas, es decir, cada secuencia ha sido revisada, documentada y enlazada a otras bases de datos. Enlace externo: Swissprot en el EBI (<http://www.ebi.ac.uk/swissprot/access.html>), Swissprot en ExPasy (<http://us.expasy.org/sprot/>)
- **TrEMBL** por *Translation of EMBL Nucleotide Sequence Database* incluye la traducción de todas las secuencias codificantes derivadas del (EMBL-BANK) y que todavía no han podido ser anotadas en Swissprot. Enlace externo: TrEMBL (<http://www.ebi.ac.uk/trembl/>)
- **PIR** por *Protein Information Resource* está dividida en cuatro sub-bases que tienen un nivel de anotación decreciente. Enlace externo: PIR (<http://pir.georgetown.edu/>)
- **ENZYME** enlaza la clasificación de actividades enzimáticas completa a las secuencias de Swissprot. Enlace externo: ENZYME (<http://us.expasy.org/enzyme/>)
- **PROSITE** contiene información sobre la estructura secundaria de proteínas, familias, dominios, etc. Enlace externo: PROSITE (<http://us.expasy.org/prosite/>)
- **INTERPRO** integra la información de diversas bases de datos de estructura secundaria como PROSITE, proporcionando enlaces a otras bases de datos e información más extensa. Enlace externo: INTERPRO (<http://www.ebi.ac.uk/interpro/index.html>)
- **PDB** por *Protein Data Bank* es la base de datos de estructura terciaria 3-D de proteínas que han sido cristalizadas. Enlace externo: PDB (<http://www.rcsb.org/pdb/>)

Bases de Datos considerando genomas

- **Ensembl** integra genomas eucariotas grandes, por el momento contiene genoma humano, ratón, rata, fugu, zebrafish, mosquito, *Drosophila*, *C. elegans*, y *C. briggsae*. Enlace externo: Ensembl (<http://www.ebi.ac.uk/ensembl/index.html>)
- **Genomes server** y **TIGR** son portales con información y/o enlaces de todos los genomas secuenciados por el momento, desde virus a humanos. Enlace externo: Genome Server (<http://www.ebi.ac.uk/genomes/index.html>) Enlace externo: TIGR (<http://www.tigr.org/>)
- **Wormbase** es el portal del genoma de gusano *C. elegans*. Enlace externo: Wormbase (<http://www.wormbase.org/>)
- **Flybase** es el portal de la mosca del vinagre *Drosophila melanogaster*. Enlace externo: Flybase (<http://flybase.bio.indiana.edu/>)

Otras

- **Taxonomy** es el portal de clasificación taxonómica de organismos
- Enlace externo: Taxonomy Browser (<http://www.ncbi.nlm.nih.gov/Taxonomy/>)
- **Pubmed** da acceso gratuito al índice de publicaciones de la Biblioteca Nacional de Medicina (NLM), con enlaces a artículos completos. Enlace externo: PubMed (<http://www.ncbi.nlm.nih.gov/PubMed/>)
- **OMIM** por *Online Mendelian Inheritance in Man* es un catálogo de genes humanos relacionados con informaciones genéticas. Enlace externo: OMIM (<http://www.ncbi.nlm.nih.gov/omim/>)

Cabe citar que estas bases de datos recogen información biológica que incluye desde virus, bacterias hasta plantas y animales y por lo tanto son fuente importante de consulta para diversas disciplinas profesionales.

La bioinformática se nutre especialmente de dos grandes áreas del conocimiento, las ciencias biológicas y las ciencias de la computación, dado este origen existen dos grandes líneas de trabajo: La primera en la cuál las ciencias de la computación utilizan modelos de las biológicas, ejemplo de ello lo constituyen las redes neurales, los algoritmos genéticos, computación con ADN, entre otras. La segunda en la cuál las ciencias biológicas utilizan modelos y herramientas de las ciencias de la computación, como se menciona anteriormente.

Básicamente, los sistemas informáticos que se emplean en estos campos son:

- Bases de datos (Información Científica)
- Software para visualización de aplicaciones
- Programas para control de reactivos, geles y otros materiales
- Generación y ensamblaje de secuencias
- Programas para análisis de secuencias, Alineación de secuencias.
- Programas para predicción de estructura de proteínas
- Paquetes de integración y ensamblaje de mapas genéticos
- Software para clasificación y comparación
- Técnicas de Inteligencia Artificial
- Gestión de datos
- Bases de datos locales o accesibles mediante redes de comunicaciones.
- Literatura médica y científica unida a las secuencias.
- Distribución de datos
- Redes de comunicaciones
- Aplicaciones
- Gestión de datos en el laboratorio
- Automatización de experimentos
- Ensamblaje de secuencias contiguas
- Predicción de dominios funcionales en secuencias génicas
- Búsquedas en las bases de datos de estructuras
- Predicción de genes

- Evolución molecular. Árboles filogenéticos
- Documentos de difusión y apoyo a la Bioinformática

Está ocurriendo la transición de décadas de biología reduccionista, enfocadas a un gen o a una proteína, por utilización de las nuevas disciplinas "ómicas" que incluyen: Transcriptómica (espectro de moléculas de ARN en cada célula tipo), Metabólica (todos los químicos en la célula), Fisiológica (función de órganos, tejidos y ultimadamente el organismo completo).

Crítico para el éxito de esta revolución es la bioinformática y nuevas estrategias de computación, para modelar las propiedades de células, órganos y estados de enfermedad, produciendo órganos y pacientes virtuales (Goldman, M., Bio-It World. 2004).

Bibliografía

1. Claverie JM and Notredame C. Bioinformatics for Dummies. Wiley Publishing, Inc. 2003.
2. Gibas C and Jambeck Developing Bioinformatics Computer Skills. O'Reilly. 2001.
3. Lahoz – Beltra, R. Bioinformática. Simulación, Vida artificial e inteligencia artificial. 2004.
4. Valverde, J. Bioinformática en la WWW. Documento Maestría en Bioinformática Universidad Internacional de Andalucía. 2005.
5. <http://www.cecalc.ula.ve/bioinformatica/BIOTUTOR/bioinformatica.html>
6. <http://www.ebi.ac.uk/embl/index.html>
7. <http://www.ddbj.nig.ac.jp/>
8. <http://www.ebi.ac.uk/swissprot/access.html>
9. <http://www.ebi.ac.uk/interpro/index.html>
10. <http://www.ncbi.nlm.nih.gov/PubMed/>
11. <http://www.ncbi.nlm.nih.gov/Taxonomy/>
12. <http://www.rcsb.org/pdb/>
13. <http://us.expasy.org/enzyme/>
14. <http://us.expasy.org/prosite/>
15. <http://www.ncbi.nlm.nih.gov/Entrez/>
16. <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Nucleotide>
17. <http://www.iibce.edu.uy/2000-08/index.html>
18. <http://www.ugr.es/~eianez/Biotecnologia/forensetec.htm#3>
19. <http://www.ideal.es/waste/genomabiochip.htm>
20. <http://barrapunto.com/articles/100/02/10/2317259.shtml>
21. <http://www.bioinformacion.net/biochip.htm>
22. <http://www.inmuno.org/pdf/basesdedatos.pdf>
23. http://www.seis.es/i_s/i_s19/i_s19l.htm
24. http://www.genomicglossaries.com/content/Bioinformatics_gloss.asp
25. <http://www.biology.gatech.edu/bioinformatics/whatis.html>. What is Bioinformatics.
26. Zepeda García, O Maestría en Ingeniería de Sistemas Empresariales. Universidad Ibero Americana. 2003.

Trabajo recibido el 13/08/2006, nº de referencia 110630_RED.VET. Enviado por su autor. Publicado en [Revista Electrónica de Veterinaria REDVET](#), ISSN 1695-7504 el 01/07/06. [Veterinaria.org](#) - [Comunidad Virtual Veterinaria.org](#) - Veterinaria Organización S.L.® Se autoriza la difusión y reenvío de esta publicación electrónica en su totalidad o parcialmente, siempre que se cite la fuente, enlace con Veterinaria.org - <http://www.veterinaria.org/> y REDVET® <http://www.veterinaria.org/revistas/redvet> y se cumplan los requisitos indicados en [Copyright](#) 1006 2006