

**Universidad Católica “Nuestra Señora
de la Asunción”**

Facultad de Ciencias y Tecnología

Ingeniería Informática

TAI 2

Trabajo Práctico: Bioinformática

Martín Escobar

2009

Índice

Contenido

Contenido.....	2
Introducción.....	4
Definición.....	5
Involucra:.....	5
Función:.....	6
Campo de Investigación:.....	6
Términos a Diferenciar:.....	7
Historia y hechos más importantes que contribuyeron a la Bioinformática.....	10
Aporte Social de Bioinformática:.....	10
Últimos descubrimientos y proyectos que se están llevando a cabo a nivel mundial:.....	11
Proyectos actuales y cercanos a nuestra realidad en el área de Bioinformática.	11
Estado de la informática en el Área Salud.....	13
Problemas a solucionar:.....	13
Lo que propone BioLinux:.....	13
Últimos proyectos.....	14
EMBOSS v 2.9.0	18
GDE	19
RasMol	19
PyMOL.....	20
GROMACS	20
Phylip	20
Arka /Genpak	20
ClustalX /ClustalW	20
TCOFFEE	21
Seaview	21
Glimmer	21

Mummer	21
Garlic	22
Primer	22
XMolCalc.....	22
Babel V 1.6	22
NJPlot	22
Pratt	22
Críticas a la Biotecnología.....	24
En qué puede trabajar un Lic. en Bioinformática.....	25
Conclusión.....	26

Introducción

Estamos en la era post-genoma humano, la era de las ómicas: la Genómica, la Proteómica, la Metabolómica y la Citómica, que conllevan una gran cantidad de datos a analizar.

La aplicación de las nuevas técnicas en *Proteómica* (Es la ciencia que correlaciona las proteínas con sus genes, estudia el conjunto completo de proteínas que se pueden obtener de un genoma) y *Genómica Funcional* (es un campo de la biología molecular que se propone utilizar la vasta acumulación de datos producidos por los proyectos de genómica para describir las funciones e interacciones entre genes y proteínas) ha cambiado el escenario: de conocer una función, buscar el gen responsable y estudiar la proteína individual a tener una gran cantidad de genes a los cuales hay que buscar su función y la necesidad de estudiar las relaciones entre muchas proteínas.

El análisis de éstos datos entrega las enormes posibilidades en áreas como la Medicina, la Biología y la Biotecnología, desde los estudios evolutivos hasta la producción de nuevas medicinas capaces de interrumpir selectivamente puntos específicos en redes de interacción entre proteínas, por lo cual, la Bioinformática es crucial para lograr estos objetivos.

Todas las tecnologías tienen sus ciclos, cuando una era tecnológica empieza a decaer entonces otra comienza a subir; y la nueva era tecnológica que viene es la Biotecnología. Que es el uso de organismos biológicos para el beneficio humano.

A lo largo de este trabajo veremos los principales aspectos relacionados con la Bioinformática, como ser definiciones varias, avances, proyectos llevados a cabo, herramientas, aplicaciones y hasta Sistemas Operativos específicos para trabajar con esta fascinante área de las ciencias.

Definición.

La Bioinformática, llamada también Biología Molecular Computacional, corresponde como tal a una disciplina científica que utiliza tecnología de la información para organizar, analizar y distribuir información de Biomoléculas con la finalidad de responder preguntas complejas.

La Bioinformática abarca el uso de técnicas y herramientas utilizadas en tres disciplinas separadas; la Biología Molecular (donde se originan los datos a analizar), la Computación (que proporciona el hardware, las vías de comunicación de los resultados entre investigadores) y el análisis de datos mediante algoritmos (que entrega los programas y resultados a analizar).

Es la aplicación de la informática a la gestión y análisis de datos biológicos.

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 a 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”.

Involucra:

- Informática
- Matemática Aplicada
- Estadística
- Ciencias de la Computación
- Inteligencia Artificial
- Química
- Ciencias de la Salud
- Bioquímica
- Biología Molecular
- Medicina Molecular

Función:

- **Solucionar Problemas:** Los principales problemas están relacionados con la gran cantidad de datos que deben ser procesados y analizados. Específicamente se debe trabajar con una gran cantidad de secuencias de ADN y Proteínas que se encuentran almacenadas en bases de datos; esta información se puede utilizar para buscar las propiedades de una secuencia, comparándola con secuencias semejantes con propiedades conocidas o a través de reglas para predecir propiedades de secuencias nuevas.
- **Analizar Datos:** toda la información biológica y bioquímica debe ser combinada, se trabajan con miles de millones de datos; y la Bioinformática permite el análisis e interpretación de éstos para su posterior utilización. Sin ella sería imposible trabajar con Bases de Datos y toda la información disponible actualmente
- **Simular Sistemas y Mecanismos:** Por medio de la estadística, utilizando técnicas de data mining, inteligencia artificial, entre otros; se pueden crear diferentes modelos de predicción, simular comportamientos, inferir a partir de ciertas características las funciones que cumplen ciertas proteínas, etc.
- **Comparacion de Datos:** Un claro ejemplo sería las comparaciones de secuencias genéticas que se realizan, a partir de las cuales se pueden obtener características similar o diferencias que presentan diferentes organismos.
- **Modelamiento:** Por ejemplo hay proteína que nosotros no conocemos como es, y solo conocemos la secuencia genética; y principalmente no conocemos lo que hace. Entonces se empieza a comparar con proteínas conocidas y empezamos a modelar esta proteína hasta encontrarle una estructura; y según esta estructura va a ser la función de la proteína. Entonces con el computador se simula, modela la estructura y en base a esa estructura se sabe su función.

Campo de Investigación:

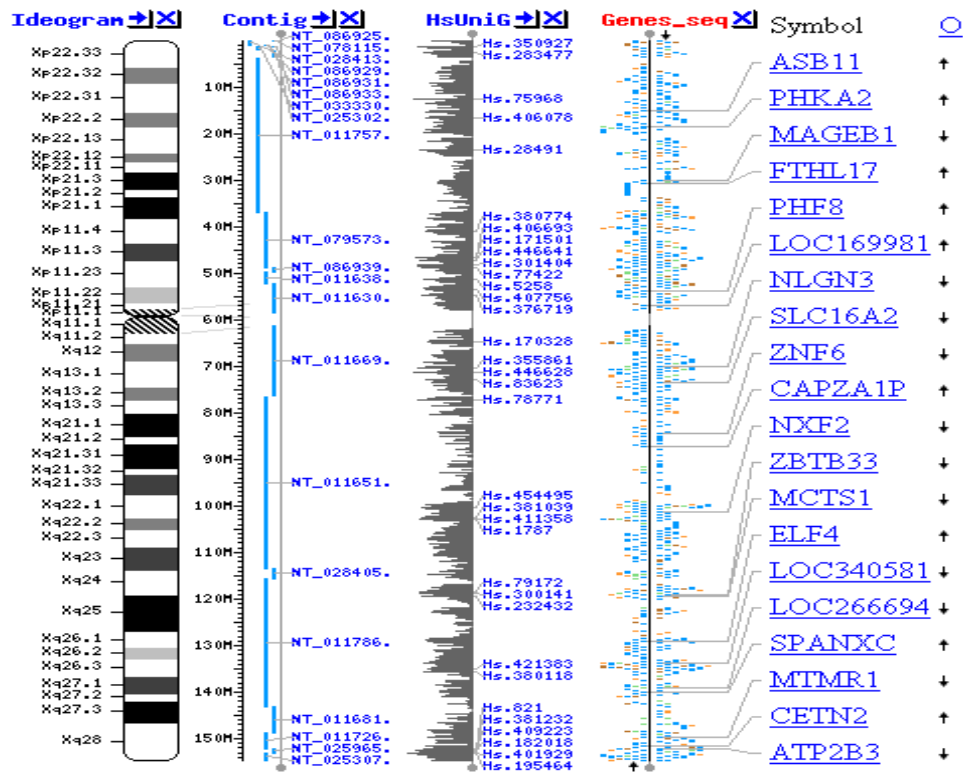
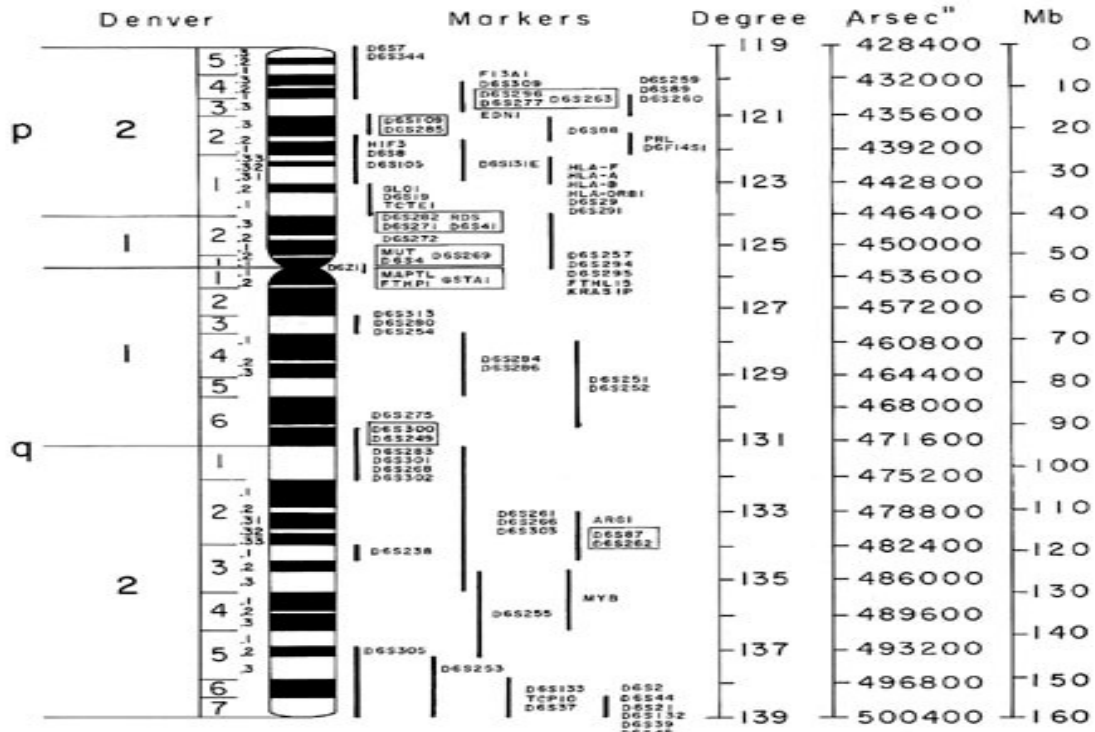
- Alineamiento de Secuencias.
- Predicción de Genes.
- Montaje del Genoma en la Secuenciación.
- Alineamiento estructural de Proteínas.
- Predicción de estructura de proteínas.
- Predicción de la expresión génica.
- Interacciones proteínas-proteínas.
- Modelado evolutivo.
- La obtención de secuencias de DNA o proteínas similares a la secuencia problema.
- La obtención de secuencias con palabras clave o información similar a la secuencia problema.

- La búsqueda de motivos funcionales o estructurales en la secuencia problema y obtención de secuencias que contienen los mismos motivos.
- El alineamiento múltiple de la secuencia problema con otras similares, y definición de regiones conservadas y variables.
- Ensamblaje de fragmentos de DNA y creación de mapas genómicos.
- Predicción de estructura y dinámica de macromoléculas.
- La reconstrucción de la filogenia a partir del alineamiento.
- La construcción de un motivo característico del alineamiento y el uso de éste para encontrar nuevas secuencias con motivos comunes con el alineamiento previamente realizado.
- Estudio de todos los genes y proteínas de un organismo: “Genómica y proteómica funcional”.
- La generación de partidores para diseñar PCR. (La reacción en cadena de la polimerasa es una técnica para amplificar una sola o unas pocas copias de un fragmento de ADN a través de varios órdenes de magnitud, generando miles de millones de copias de una secuencia de ADN en particular.)

Términos a Diferenciar:

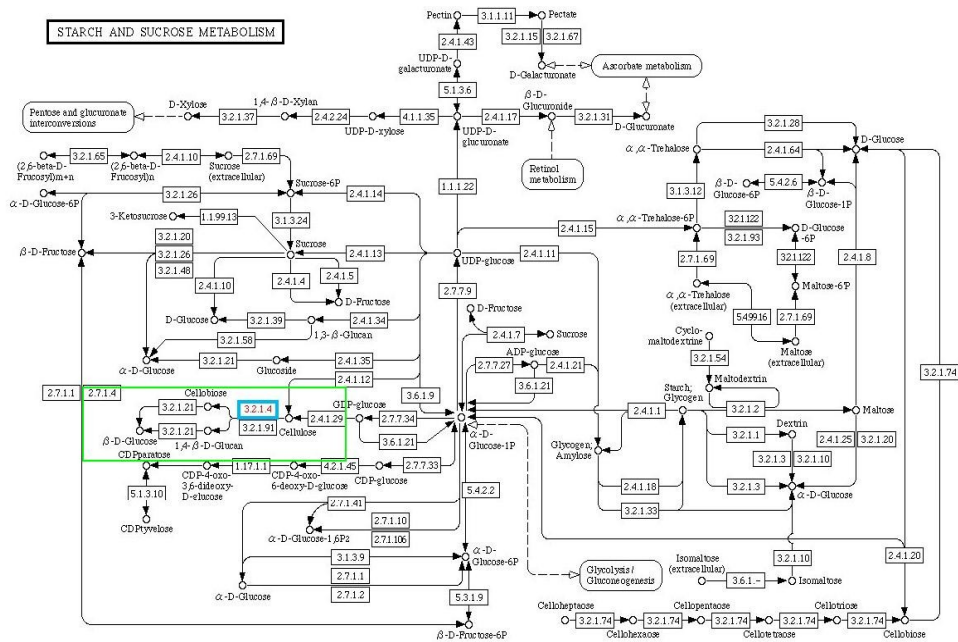
Bioinformática: Es la investigación, desarrollo o aplicación de herramientas computacionales y aproximaciones para la expansión del uso de datos biológicos, médicos, conductuales o de salud, incluyendo aquellas herramientas que sirvan para adquirir, almacenar, organizar, analizar o visualizar tales datos.

Chromosome 6



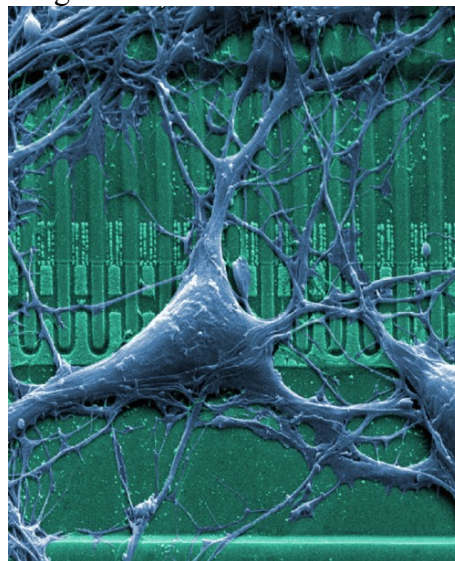
Estos son los datos de un cromosoma humano, y la Bioinformática se encarga de analizar estos datos.

Biología computacional: Tiende a solucionar problemas a nivel relaciones dentro de una célula en función de sus encimas, es decir, metabolómica; o también solucionar problemas a nivel de ecología.



Lo que se busca es ver cómo interactúan las encimas para poder lograr que el organismo funcione bajo ciertas condiciones.

Biocomputación: Es la construcción y uso de computadores que contienen componentes biológicos o funcionan como organismos.



Las neuronas se comunican a través de impulsos nerviosos, y se ha descubierto que las mismas pueden establecer conexiones con chips; por lo tanto pueden haber computadores utilizando las neuronas.

Las neuronas tienen una velocidad y capacidad de almacenaje de informaciones muy elevadas.

Historia y hechos más importantes que contribuyeron a la Bioinformática.

Desde finales de los sesenta, cuando las primeras secuencias de proteínas comenzaron a coleccionarse dando lugar a las primeras bases de datos en biología molecular, numerosos descubrimientos han acelerado el desarrollo científico. Un ejemplo claro es el Proyecto de Secuenciación del Genoma Humano, nuestra propia secuencia genética. Pero ninguno de estos avances sería posible sin el uso intensivo de los ordenadores.

Con la finalización de los primeros proyectos genoma y las grandes posibilidades que esta nueva información ofrece, numerosos campos científicos han experimentado un boom tanto en la concepción de nuevos experimentos, como en la obtención de resultados impactantes.

1. Watson y Crick proponen la estructura de doble hélice del ADN. (1953). Establecer que el material genético está almacenado en una secuencia de una molécula que tiene cuatro pares de bases A T C G. Y esa secuencia de miles de millones de clases es la que te entrega la característica genética.
2. Se secuencian la primera proteína.(insulina bovina) por F. Sanger.(1955)
3. Se publica el algoritmo Needleman – Wunsch para alineamiento de secuencias. (1970)
4. Se publica la primera secuencia completa de un organismo, el fago ϕ -X174 (5386 pares de bases que codifican 9 proteínas) .(1978)
5. Descubrimiento de la PCR (Polimetase Chain Reaction, reaction en cadena de la Polimerasa). Lleva a la multiplicación de muestras de ADN, lo que permitirá su análisis (1983)
6. Desarrollo de algoritmo de búsqueda en Base de Datos de secuencias. (1983)
7. Búsqueda rápida de similitudes entre secuencias. (1985)
8. Algoritmo FASTA para comparación entre secuencias. (1988)
9. Aparecen Bases de Datos biológicas (GenBank en 1982, Swiss-Prot en 1986)
10. Se consigue secuencia de genomas:
 - Haemophilus influenzae 1,8 Mbps 1995
 - Mycoplasma genitalium 0,58 Mbps 1995
 - Saccharomyces cerevisiae 12 Mbps 1996
 - Caenorhabditis elegans 97 Mbps 1998
 - Arabidopsis thaliana 100 Mbps 2000
 - Genoma hombre 3 Gbps 2003
 - Genoma mujer 3 Gbps 2003

Aporte Social de Bioinformática:

- Crea nuevas fuentes de trabajo para profesionales preparados y capacitados.
- Oportunidad para la optimización de la biodiversidad. Permitirá desarrollar usos y aplicaciones de la biodiversidad que abren la posibilidad de responder a retos como los biocombustibles, seguridad alimentaria, salud y el mercado en general.
- Sostenibilidad ambiental.
- Grandes beneficios en la salud pública y privada. Conocimiento biomédico que permitirá afrontar con éxito los retos de la medicina del futuro (medicina personalizada, medicina de calidad)

Últimos descubrimientos y proyectos que se están llevando a cabo a nivel mundial:

- **Drogas para prevenir los ataques cardíacos y coágulos de sangre.** Científicos de la CEH desarrollan una nueva droga de la saliva de la garrapata que controla el flujo de la sangre y previene la coagulación. Su descubrimiento es ahora objeto de un acuerdo internacional entre las organizaciones de investigación líderes en el Reino Unido, Singapur y Eslovaquia.
- **Proyecto Win – Win.** Estudian los beneficios ecológicos de la biodiversidad y la restauración del bienestar humano. La nueva investigación fue llevada a cabo por un equipo internacional de científicos de la Universidad de Alcalá, Centro para la Ecología y la Hidrología y la Universidad de Bournemouth.
- **El Proyecto Genoma Humano.** Se trata de un programa internacional que pretende construir mapas genéticos y físicos detallados del genoma humano, así como determinar su secuencia completa y localizar la posición de la totalidad de los genes que lo componen.
- **Métodos espectrales para agrupar secuencias de proteínas**
- **Cuantificación de la adaptación ambiental de las rutas metabólicas en la metagenómica**
- **Golden Rice.** Se trata de un ejemplo de planta transgénica obtenida en un programa con fondos públicos, cuyo objetivo es contribuir a mejorar las condiciones de nutrición de poblaciones del tercer mundo (principalmente en Asia), para reducir la incidencia de enfermedades ligadas al déficit de vitamina A.
- **Entre muchos otros.**

Proyectos actuales y cercanos a nuestra realidad en el área de Bioinformática.



BioLinux

BioLinux es un Grupo de Profesionales que tiene como objetivos la difusión, investigación, desarrollo e implementación, de las tecnologías informáticas libres, en el ámbito de la salud.

BioLinux se funda en 1.997, en Buenos Aires, Argentina, por iniciativa de un pequeño grupo de profesionales del equipo de salud, con la motivación de investigar las nuevas tecnologías informáticas libres, para aplicarlas en entidades sanitarias y hospitalarias. Entre 1.998 y 2.000 se desarrolla el web site y la lista de discusión del grupo. A partir de esta etapa, BioLinux va afirmandose como grupo de referencia en Latinoamérica.

Logros:

- Año 2002:
 - Creamos **LinuxMed 2002**.
 - Primeras Jornadas de Software Libre aplicadas al Ámbito de la Salud. 1º al 3 de Julio, de 2002. Hospital Prof. Dr. Juan P. Garrahan.
 - Se crea el **Programa Salud Libre**, enfocado a la investigación, difusión, desarrollo e implementación de aplicaciones de software libre para entidades hospitalarias. (<http://saludlibre.org>)
- Año 2003:
 - Comienzo del trabajo de analisis del sistema informatico hospitalario Care2x (<http://care2x.org>), con el objetivo de personalizar su estructura como sistema base para hospitales de Latinoamérica.
- Año 2004:
 - Se librea el paquete **Care2x-Latino**, como sistema base para Latinoamérica. Es tomado por los grupos de desarrollo de varios paises, entre ellos: Brasil, Venezuela, Mexico, Perú y Colombia.
 - Nace **SaluX**, distribución oficial del grupo BioLinux, un sistema operativo para hospitales.
- Año 2005:
 - Comienza el trabajo de traducción y testeo de la plataforma de telemedicina **i-Path** (Internet Pathology Suite) en conjunto con la Univ. de Basilea, Suiza.
 - Se libera **Opencare**, como sistema de gestion de consultorios para Latinoamérica.
 - Nace la **Comunidad Salud Libre**, impulsada por BioLinux como fuente de cambio en la informatica médica regional.
 - Se desarrolla **LinuxMed 2005**. Primer Congreso Virtual Internacional de Free and Libre Opensource Software y su aplicación en el Area Salud. Septiembre-Noviembre, 2005 (<http://www.linuxmed.fac.org.ar>)
- Año 2006:
 - Se crea **TM-Sur**, un portal telemedicina opensource para la comunidad del equipo de salud de habla hispana.

- Se define **OpenHIS** como sistema informático hospitalario oficial desarrollado por el Grupo BioLinux.

Estado de la informática en el Área Salud

El grado de informatización hospitalaria es inadecuado e insatisfactorio a las demandas actuales de una sociedad moderna y de acuerdo a los problemas de salud locales y regionales.

La falta de existencia de estándares para el desarrollo de sistemas informáticos hospitalarios (SIH) es una constante en todos los establecimientos relevados. La ausencia de departamentos de informática y la no jerarquización de la especialidad (Informática en salud) son dos importantes problemas en varios países de la región.

Prácticamente más del 90% del parque informático de los establecimientos asistenciales se distribuye para áreas administrativas y/o contables (admisión, turnos, facturación) y se deja de lado los sistemas asistenciales, de docencia e investigación, como así también las áreas de toma de decisión (Dirección, Gerencias) y la Calidad de Atención.

Esto implica que los sistemas informáticos hospitalarios en la Latinoamérica ayudan poco al profesional en su tarea habitual sea esta asistencial o de gestión.

Problemas a solucionar:

- Bajo grado de informatización hospitalaria y escaso porcentaje de implementación de redes en hospitales
- No existe un único sistema operativo en la red hospitalaria ni regional, con la consecuente falta de estandarización de protocolos de comunicación, compartición de aplicaciones, herramientas, etc.
- No existe un sistema de Historia Clínica única en instituciones sanitarias, solo sistemas locales sin integración entre sí.
- No hay implementación unificada de sistemas informáticos hospitalarios con manejo de imágenes y telemedicina.
- Los sistemas de software son en su mayoría propietarios, habiendo sido estos desarrollados por programadores o empresas de software, con altos costos de desarrollo y licenciamiento.
- Además de esto, se evidencia una falta de presencia de sistemas operativos y herramientas informáticas bajo software libre.
- Se observan recursos limitados en cuanto al hardware disponible.
- Las principales áreas hospitalarias informatizadas corresponden a la administrativa, la contable y de facturación. No se pone énfasis en desarrollos orientados a la gestión y calidad de atención.

Lo que propone BioLinux:

El sistema Hospitalario es un conjunto de sub-sistemas en equilibrio dinámico:

- Sistema asistencial,
- Sistema administrativo,
- Contable,
- Técnico,
- Gerencial,
- Entre otros.

Para poder cubrir las necesidades básicas de un sistema hospitalario modelo, se deberían desarrollar cuatro sub-sistemas informáticos primarios:

- **Sistema Operativo base y redes informáticas.**
- **Sistema de Información Hospitalaria (SIH)**
- **Sistemas de Telemedicina (TM)**
- **Sistemas de Administración y**
- **Almacenamiento de Imágenes (PACS).**

Objetivo:

El Grupo BioLinux tiene como objetivo la investigación y desarrollo en estas cuatro líneas, necesarias para el normal funcionamiento de un sistema hospitalario, bajo los estándares del software libre, orientado a la gestión hospitalaria y la calidad de atención en efectores de salud. Esto hace la diferencia con los sistemas tradicionales.

Últimos proyectos.

- 1) SaluX



Constituye un Proyecto que se haya comprendido dentro del Programa Salud Libre (<http://saludlibre.org>) el cual el grupo Biolinux viene desarrollando desde agosto del año 2002. SaluX es la distribución oficial del Grupo BioLinux.

SaluX se constituye en un portador de sistemas informáticos de código abierto de avanzado desarrollo a nivel internacional en las áreas de telemedicina, medicina rural y sistema de registro médico electrónico. Como así también OpenHIS, el sistema informático hospitalario en desarrollo en el grupo Biolinux como parte integral del Proyecto Salud Libre.

Objetivos de SaluX:

1. Promover una paridad informática hospitalaria en Latino América.
2. Disponer de una distribución (sistema operativo libre) como elemento base informático para hospitales y entidades sanitarias.

3. Estimular las capacidades locales a través de un desarrollo informático de código abierto que pueda ser personalizable y escalable por terceros.
4. Fomentar el uso del software libre en el área de salud en Latino América.
5. Mejorar las tecnologías de comunicación, incentivando el conocimiento y la coparticipación en efectores de salud urbana, sub urbana y rural.

Características Generales de SaluX:

Kernel de la serie 2.6.

Es un sistema operativo basado en Debian

Instalación grafica simplificada

Idioma: Español

Paquetes desarrollados por BioLinux para la Gestión Hospitalaria.

Conectividad con redes cableadas, WiFi, Bluetooth, IrDA

Integración con otros sistemas existentes en red como las plataformas MSWindows, Novell y Unix

Acceso sencillo a internet e intranets.

Capacidades de Terminal Server para thin clients.

2) OpenHIS



OpenHIS es el nuevo emprendimiento del Grupo BioLinux para cubrir la necesidad de un sistema informático hospitalario, integral, modular y escalable.

OpenHIS es un desarrollo creado totalmente desde cero, con nuevo diseño, y con el respaldo que el Grupo BioLinux puede ofrecer en su larga trayectoria en análisis, diseño y desarrollo de sistemas informáticos para el ámbito de la salud. OpenHIS significa Entorno Abierto de Información para Salud (Open Health Information System Environment) y está diseñado con HOME (Hospital Object Model Environment), un nuevo modelo de objetos para el análisis y creación de sistemas informáticos hospitalarios, desarrollado por el Grupo BioLinux.

OpenHIS es una plataforma de información hospitalaria abierta, desarrollada de acuerdo a la realidad asistencial y gerencial hospitalaria de la región. OpenHIS, está pensado como contenedor de información (CMS) y pretende ser la base para el desarrollo futuro de sistemas de información hospitalarios, para Latinoamérica.

3) Telemedicina. i-Path

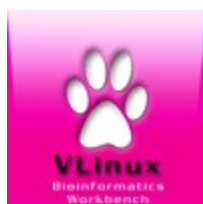
Los sistemas de telemedicina(TM) son altamente complejos y costosos para implementarse masivamente en los efectores públicos. Pero sin embargo, son un elemento de importancia en la calidad de atención hospitalaria, brindando la posibilidad de crear redes de consulta regionales e inter-regionales. El software libre, permite crear sistemas de TM con bajo costo para estos fines. Desde el año 2005, el Grupo BioLinux, se asocia con la Universidad de Basilea, Suiza, para colaborar en el desarrollo de i-Path: Internet Pathology Suite, un sistema abierto de TM que puede correr sobre intranets o internet con bajo costo de implementación.

Como resultado de esta colaboración se prepara i-Path para implementarse en hospitales de Argentina y se desarrolla Telemedicina Sur como portal telemédico para Latinoamérica.

Distribuciones y Programas para Bioinformática.

Tres de las mejores distribuciones **Linux Live** que incluyen aplicaciones de la bioinformática son:

🐧 **VLinux**



Desarrollador: V. Vimalkumar

URL: bioinformatics.org/vlinux/

Este es un buen sistema para que un usuario Nuevo pueda comenzar con él. Posee una fácil interfaz y la inclusión del sistema de categorías EMBOSS menú gráfico es una bendición.



VLinux banco de trabajo de bioinformática es una distribución de Linux para la bioinformática. Es fácil de usar, no requiere instalación, basada en CD de distribución basada en Knoppix 3.3. Incluye una variedad de secuencias y packages. Es un producto de código abierto publicado bajo la licencia GNU GPL.

Los siguientes paquetes están incluidos en la versión 1.0 VLinux.

EMBOSS, GDE, RasMol, PyMOL, GROMACS, Phylip , Arka / Genpak, ClustalW / ClustalX, Toffee, Seaview, Glimmer, Mummer, Garlic, Ramachandran Plot viewer, Stride, Primer, XMolCalc, Babel, NJPlot, HMMER, Pratt

Vigyaan



Desarrollador: PK Agarwal, el Laboratorio Nacional Oakridge

URL: www.vigyaancd.org

En general, este sistema vale la pena probarlo y es adecuado especialmente para los nuevos usuarios que tienen un interés en la química y la estructura además de programas de análisis de la secuencia.



Vigyaan es un banco de trabajos electrónicos para la Bioinformática, la biología computacional y la química computacional. Ha sido diseñado para satisfacer las necesidades tanto de principiantes y expertos. VigyaanCD es un Live CD de Linux que contiene todo el software necesario para arrancar el equipo con listas para usar software de modelado. VigyaanCD v1.0 se basa en KNOPPIX v3.7.

Lista de aplicaciones con las que cuenta Vigyaan:

Arka / GP, Artemisa, Bioperl, BLAST (NCBI-herramientas), ClustalW / ClustalX, Cn3D, herramientas EMBOSS, ajo, Mica, GROMACS, Ghemical, GNU R, Gnuplot, el GIMP, ImageMagick, Jmol, MPQC, MUMer, NJPlot, Open Babel, Octave, PSI3, PyMOL, visor de representación de Ramachandran, RasMol, Raster3D, Seaview, TINKER, XDrawChem,

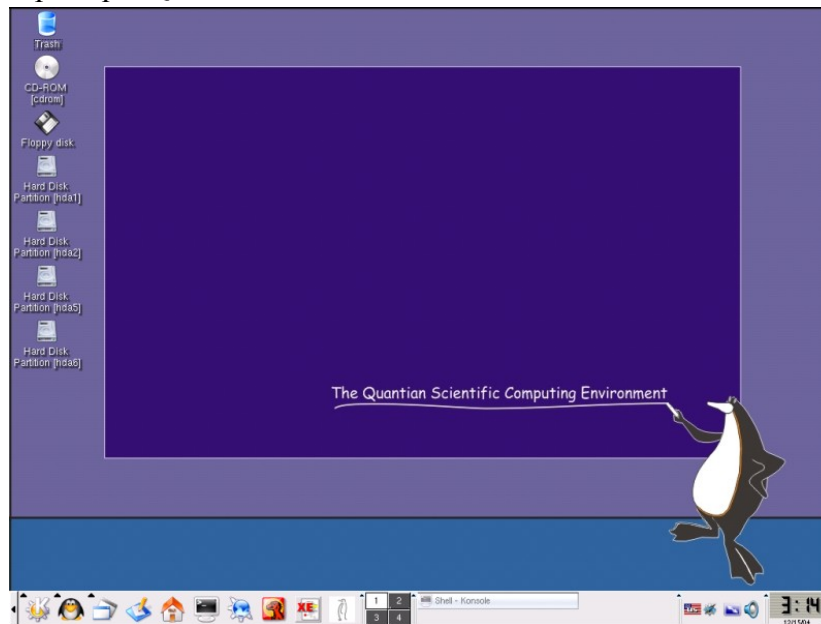
Xmgr y Xfig. GNU C / C + + / compiladores de Fortran y herramientas adicionales de Linux (como ps2pdf) también están disponibles.

👉 Quantian

Desarrollador: Dirk Eddelbuettel, Debian

URL: dirk.eddelbuettel.com/quantian.html

Cualquiera que busque un sistema portátil que les permita ejecutar análisis, escribir guiones, y en general a utilizar las herramientas de alto nivel, sería difícil encontrar un sistema más amplio que Quantian.



Principales Aplicaciones:

- **EMBOSS v 2.9.0**

EMBOSS es un nuevo paquete de software gratuito Open Source desarrollado especialmente para el análisis de las necesidades de la biología molecular (por ejemplo, EMBnet) comunidad de usuarios. El software automáticamente las arregla con los datos en una variedad de formatos e incluso permite la recuperación transparente de los datos de la secuencia de la web. Además, como bibliotecas extensas se proporcionan con el paquete, es una plataforma para permitir a otros científicos a desarrollar y lanzar el software en cierto espíritu de código abierto. EMBOSS también integra una gama de paquetes disponibles en la actualidad y herramientas para el análisis de la secuencia en un todo sin fisuras. EMBOSS rompe la tendencia histórica hacia paquetes de software comercial.

Dentro de EMBOSS encontrará más de 100 programas (aplicaciones). Estos son sólo algunos de los ámbitos cubiertos:

- Alineamiento de secuencias

- Búsqueda rápida en Base de datos de búsqueda patrones de secuencia
- Identificación de motivos de proteínas, incluyendo el análisis de dominio
- Análisis EST
- Nucleotide sequence pattern analysis, for example to identify CpG islands. Analizador de patrones de secuencias de nucleótidos, por ejemplo para identificar a las islas CpG.
- Identificación repetitiva de especies simples y específicas
- Análisis de uso de codones para pequeños genomas
- La identificación rápida de los patrones de secuencia en secuencia de conjuntos de gran escala.
- Herramientas de presentación para la publicación
- Y mucho más.

URL:<http://www.emboss.org>

- **GDE**

Facilita los análisis genéticos, es una herramienta rápida y de alto rendimiento, que combina la flexibilidad del entorno de datos genéticos (GDE) con el poder de Linux. Su principal ventaja es la capacidad de "envolver" en torno a una amplia gama de programas de bioinformática y mostrar la salida del programa. La incorporación de nuevas funciones es relativamente simple ya que el código para la edición de menú es similar al código utilizado para ejecutar la línea de comandos de software. El nuevo software está integrado por la edición de un archivo único que controla la apariencia del menú (. GDEmenus). El resultado final es una interfaz gráfica única que puede acceder a varios programas de bioinformática y puede ser adaptado a la búsqueda de ADN específicas locales y bases de datos de secuencia de amino ácidos.

URL:<http://www.bioafrica.net>

- **RasMol**

RasMol es un programa de gráficos moleculares previsto para la visualización de proteínas, ácidos nucleicos y moléculas pequeñas. El programa está dirigido a mostrar, enseñanza y la generar imágenes con calidad de publicación. RasMol lee archivos de cooperación molecular ordenados en una serie de formatos, y de forma interactiva muestra la molécula en la pantalla en una variedad de combinaciones de colores y representaciones. Actualmente soporta formatos de archivo de entrada como Brookhaven Protein Databank (PDB), Alchemy Tripos y Sybyl formatos Mol2, Molecular Design Limited (MDL's), formato de archivo de Mol, Minnesota Supercomputer Center (MSC) XMol formato XYZ, CHARMM formato, el formato MOPAC, CIF formato y los archivos de formato mmCIF. Si la información de conectividad y / o información de la estructura secundaria no está contenida en el archivo de esta se calcula de forma automática. La molécula de carga puede ser mostrada como alambre, cilindro (driending), reastro de alfa-carbono rastro, cintas macromoleculares, puentes de hidrógeno y la superficie de puntos. Los átomos también pueden ser etiquetados con cadenas de texto arbitrario. Por último, la imagen renderizada se puede escribir en una variedad de formatos incluyendo tanto raster y vector PostScript, GIF, PPM, BMP, PICT, Rasterfile Sol o como un script de entrada MolScript o Kinemage.

URL:<http://www.openrasmol.org>

- **PyMOL**

PyMOL es un sistema de gráficos moleculares con un intérprete de Python integrado diseñado para la visualización en tiempo real y generación rápida de alta calidad de imágenes moleculares de gráficos y animaciones. También puede realizar otras muchas tareas de gran valor (como la edición de archivos PDB) para ayudar en la investigación. PyMOL es un espectador molecular capaz, con soporte para animaciones de alta calidad de representación, cristalografía, y otras actividades moleculares comunes.

URL: <http://pymol.sourceforge.net> <http://pymol.sourceforge.net>

- **GROMACS**

GROMACS es un paquete de simulación de dinámica molecular y minimización de la energía.

URL:<http://www.gromacs.org>

- **Phylip**

PHYLIP, el paquete de inferencia Filogenia, es un paquete de programas para inferir filogenias (árboles evolutivos). Se ha distribuido desde 1980, y tiene más de 15.000 usuarios registrados, lo que es el paquete de más amplia distribución de los programas de la filogenia. PHYLIP está disponible como código fuente en C, y también como ejecutables para algunos sistemas informáticos comunes. Se puede inferir filogenias por la parsimonia, la compatibilidad, los métodos de matriz de distancia, y la probabilidad. También se puede calcular árboles de consenso, calcular distancias entre los árboles, sacar los árboles, los conjuntos de datos de arranque, editar los árboles, y calcular las matrices de distancia. Se puede manejar datos que son secuencias de nucleótidos, secuencias de proteínas, las frecuencias de genes, sitios de restricción, la restricción fragmentos, distancias, caracteres discretos, y los caracteres continuos.

URL:<http://evolution.gs.washington.edu/phylip.html>

- **Arka /Genpak**

Arka es un programa que sirve como una interfaz gráfica para los programas del paquete GP. El principal objetivo del programa es la manipulación y visualización de DNA / RNA / proteína secuencias

- **ClustalX /ClustalW**

Es una interfaz de Windows para la alineación de secuencias múltiples. Proporciona un entorno integrado para la realización de secuencias múltiples, alineamientos de perfil y el análisis de los resultados. El alineamiento de secuencias se muestra en una ventana en la pantalla. Un esquema de color versátil, se ha incorporado lo que le permite

destacar las características conservada en la alineación. Los menús desplegables en la parte superior de la ventana le permiten seleccionar todas las opciones necesarias para las secuencias múltiples tradicionales y la alineación de perfil.

URL:<http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX>

● **TCOFFEE**

T-Coffee es una herramienta versátil para hacer alineamientos múltiples de secuencias. Su objetivo es combinar fuentes heterogéneas de información. La aplicación actual ha sido especialmente diseñada para combinar las alineaciones locales y mundiales, respectivamente, de ClustalW y LALIGN.

URL:<http://mathbio.nimr.mrc.ac.uk/~cnotred/>

● **Seaview**

Seaview es un editor gráfico de alineamiento de múltiples secuencias. Es capaz de leer diferentes formatos de alineación (MSF, CLUSTAL, FASTA, PHYLIP, MASE). Permite editar manualmente la alineación, y también ejecutar DOT-PLOT o programas a nivel local CLUSTAL, para mejorar la alineación. Se maneja desde línea de comandos.

● **Glimmer**

Glimmer (Gene Locator and Interpolated Markov Modeler) es el principal buscador de genes microbianos en TIGR (The Institute for Genomic Research), y se ha utilizado para anotar los genomas completos de *B. burgdorferi*, *T. pallidum*, *maritima T.*, *D. radiodurans*, *M. tuberculosis*, etc. Consta de dos programas principales. El primero de ellos es el programa de formación, la acumulación de ICM. Este programa contiene un conjunto de secuencias de entrada y construye los resultados de la IMM para ellos. Estas secuencias de genes puede ser completo o sólo ORFS parcial. El segundo programa utiliza este IMM para identificar los genes putativos en un genoma entero.

URL:<http://www.tigr.org>

● **Mummer**

MUMmer es un software de código abierto para la alineación rápida de ADN's muy grandes y las secuencias de aminoácidos. Es un paquete modular y versátil que se basa en un sufijo de estructura de datos de árboles para la coincidencia de patrones eficientes. Utiliza árboles subfijos, que son aptos para grandes conjuntos de datos, ya que pueden ser contruidos y registrados en un tiempo lineal y espacial. Esto permite encontrar todos los partidos de máximo 20 pares de bases exacta entre 2 ~ 5 millones de pares base de genomas de bacterias en 20 segundos, con 90 MB de RAM, en un típico ordenador 1,7 GHz Linux de escritorio. Su tiempo de ejecución baja y los requisitos de memoria le permite ser utilizado en la mayoría de los equipos. MUMmer ha sido utilizado con éxito para alinear el genoma del ratón y del humano, mostrando que puede manejar la mayoría de las entradas disponibles.

URL:<http://www.tigr.org>

- **Garlic**

Garlic es un viewer y editor molecular libre, escrito para Unix y clones de Unix.

URL: <http://garlic.mefos.hr/garlic>

- **Primer**

Ayuda a diseñar las cartillas para las secuencias. Esto ayuda a asegurarse de que la secuencia no es un vector, ni quimérico y evitar elementos repetitivos.

URL:http://www-genome.wi.mit.edu/genome_software/other/primer3.html

- **XMolCalc**

Una pequeña utilidad que te permite calcular el peso molecular.

URL:<http://www.oci.unizh.ch/group.pages/hesse/xmolcalc>

- **Babel V 1.6**

Babel es un programa para convertir entre varios formatos de archivo.

URL:<http://openbabel.sourceforge.net>

- **NJPlot**

NJplot es un programa graficador de árboles, capaz de dibujar cualquier árbol binario expresado en el formato estándar de árbol filogenético (por ejemplo, el formato utilizado por el paquete de PHYLIP).

URL:ftp://pbil.univ-lyon1.fr/pub/mol_phylogeny/njplot/

- **Pratt**

Pratt le permite al usuario buscar patrones conservados en un conjunto de secuencias de proteínas. El usuario puede especificar qué tipo de patrones deben ser buscados, y cuantas secuencias deben coincidir con el patrón que se informa.

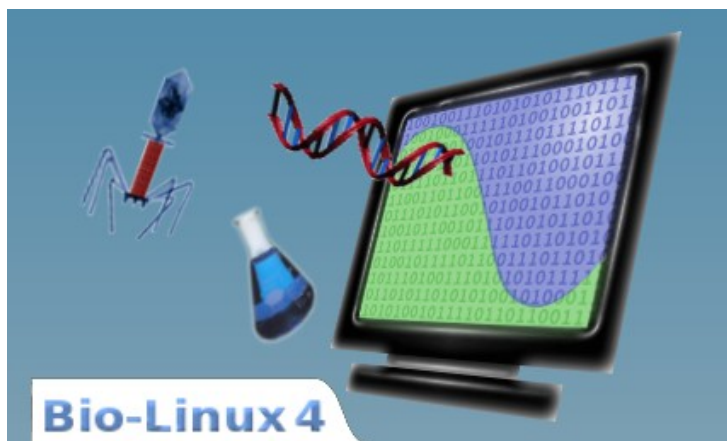
Sistemas Operativos Completos y grupos de trabajo.

Es poco probable que un CD o un DVD basado en la distribución sean suficientes para su sistema de Bioinformática primaria si desea realizar un trabajo serio y grande. Una instalación completa de Linux permite mucha más flexibilidad y poder. Sistemas a medida para el trabajo de Bioinformática se ofrecen gratuitamente por varios grupos.

A) **Bio-Linux**

Desarrollado por: The Environmental Genomics Thematic Programme Data Centre, Oxford

URL: envgen.nox.ac.uk/biolinux.html



Bio-Linux 5.0 es una estación de trabajo Bioinformática; completa, potente, configurable y fácil de mantener.

Bio-Linux ofrece más de 500 programas de Bioinformática en la base de un Linux Ubuntu. Hay un menú gráfico para los programas de Bioinformática, así como fácil acceso a la Bio-la documentación de Linux Bioinformática y los datos de la muestra útil para los programas de pruebas. Bio-Linux se basa en sistemas de código abierto y software, y así es libre de instalar y usar.

B) **BioBrew**

Desarrollado por: Glen Otero, Callident

URL: bioinformatics.org/biobrew



BioBrew es una colección de aplicaciones de código abierto para científicos y proyectos de Bioinformatics.Org. The BioBrew Roll for Rocks se pueden utilizar para crear Rocks / BioBrew Linux, una distribución personalizada tanto para clúster de computación y la Bioinformática: automatiza la instalación del clúster, incluye todo el software que necesita un clúster de HPC de fanáticos, y contiene aplicaciones Bioinformática popular.

Se utiliza en muchos sistemas de cluster de hoy! A la derecha, un Rocks / cluster BioBrew utilizado en la Conferencia de Supercomputación, noviembre de 2003. Creado por Glen Otero, el grupo que se muestra fue parte de un sistema que contiene 192 nodos. Se registró en TFLOPS en más de 1 (un billón de operaciones de punto flotante por segundo) de rendimiento, que se han convertido en una de las 500 computadoras más rápidas del mundo.



c) **Bio-Land**

Desarrollado por: Centro de Bioinformática, la Universidad de Pekín

URL: bioland.cbi.pku.edu.cn



Críticas a la Biotecnología.

Obs: Hablamos de las críticas a la Biotecnología, pues muchos proyectos Bioinformáticos contribuyen en este campo.

- Hay quienes acusan a la biotecnología de contubernio entre ciencia reduccionista e industria multinacional monopolizada, cuya visión utilitaria de la naturaleza --¿o deberíamos decir "ceguera"?--- favorece el intercambio de beneficios económicos por daños ambientales irreparables cual si se tratara de simples títulos accionarios. ¿Cómo ha llegado la arrogancia científica a concebir el proyecto de patentar la totalidad de la constitución genética humana y convertirla en propiedad privada?
- El gran problema es la incertidumbre en torno a las eventuales consecuencias de la agricultura transgénica. Con respecto al daño ambiental, la inevitable "polución genética" (liberación de organismos transgénicos al medioambiente) contaminaría a las variedades que no han sufrido manipulación y tendría efectos impredecibles. Se teme que esta polinización cruzada de lugar a supermalezas y superpestes que arrasen con otros cultivos y alteren la biodiversidad de una región. En cuanto a la salud humana, se desconocen los efectos de estos "alimentos Frankenstein" sobre el organismo, pero parece que los genes resistentes a los antibióticos podrían ingresar en nuestro patrón genético a través de la cadena alimentaria e inmunizarnos a nosotros también, lo que originaría, por ejemplo, pandemias sin fin.
- La biotecnología aplicada al agro, tiende a acentuar los efectos de la globalización mediante las prácticas monopólicas de las empresas transnacionales.
- Genera dependencia tecnológica y financiera en los países de menor desarrollo.
- Acentúa la degradación de los suelos y la pérdida de biodiversidad.
- Contribuye a la contaminación del agua y el aire.
- Se juega a ser dioses. Se modifican aspectos y características que solo Dios puede determinarlas en su perfecto plan.

- El alto precio que se debe pagar para probar las nuevas drogas, “curas” encontradas genéticamente. Este precio se refiere, a las muertes ocasionadas en animales y personas por probar la medicina genómica o genética.
- Entre otros.

En qué puede trabajar un Lic. en Bioinformática.

En el campo científico y empresario:

- Desarrollo de medicamentos y vacunas
- Modelización de especies
- Desarrollo de estudios para análisis del genoma
- Desarrollo estudios en modelización de los mecanismos de regulación de la expresión génica.

En el campo de la Salud Pública:

- Aplicación de métodos computacionales y matemáticos en inmunología y virología
- Colaboración en estudios de biodiversidad
- Elaboración de planes y políticas de salud
- Participación en emprendimientos biotecnológicos

Conclusión

Con el presente trabajo se introdujo a una nueva área de las ciencias, la Bioinformática, se presentaron definiciones, materias relacionadas a la misma, su función, una breve reseña histórica y otros datos de suma importancia para entender la idea de esta nueva tecnología. Este material de propósito no presenta con lujo de detalles explicaciones a nivel biológico; ni tampoco habla de las distintas técnicas de data mining o inteligencia artificial que se están utilizando en la actualidad; como podría ser el uso de micro arrays.

Como segunda parte del trabajo, se presentan proyectos a nivel mundial como “el arroz dorado”, pero también proyectos más cercanos a nuestra realidad como ser el de Biolinux, una comunidad de científicos radicados en argentina que están trabajando en pro de la salud, y con el claro objetivo de llegar a la salud personalizada.

Por último abordamos en tema de software específico, realizado para el estudio de la Bioinformática. Se citan diferentes distribuciones Live como distribuciones más potentes que pueden ser utilizadas a la hora de realizar trabajos genéticos, proteicos u otros. Como también se da una breve explicación de que hacen las principales aplicaciones orientadas al uso de las técnicas Bioinformáticas.

Como ya se menciona a lo largo del trabajo, esta tecnología no solo es una promesa, sino que ya es una realidad entre nosotros; los logros que se están alcanzando son increíbles y todo parece indicar que apenas es el inicio de una gran ola que se viene. Este es un buen momento, si nos interesa la investigación y la biología, para que apliquemos nuestros conocimientos informáticos en esta área.

Anexo

Videos Generales, biotecnología y alimentos.

<http://www.youtube.com/watch?v=heXzu1Gkdhw&feature=related>

<http://www.youtube.com/watch?v=4MkQOPQ9Uw0&feature=related>

<http://www.youtube.com/watch?v=1-8Rqc4xr7o&feature=fvw>

Programa de Discovery Channel.

<http://www.youtube.com/watch?v=8HXWDziT2w0&feature=related>

http://www.youtube.com/watch?v=_mjtziTx7pE&feature=related

<http://www.youtube.com/watch?v=GcH8a4EfQVk&feature=related>

Bibliografía

- <http://www.biolinux.org/>
- http://www.gchelpdesk.ualberta.ca/news/03mar05/cbhd_news_03mar05.php
- **Paper. Función e Importancia de la Bioinformática en el Desarrollo de las Ciencias, Especialmente en Biotecnología y Medicina Molecular. Autor Jorge Martínez Hormazabal**
- bioinformatics.org/vlinux/
- <http://www.bioingenieria.edu.ar/>
- <http://www.mkm-pi.com/mkmpi.php?rubrique500>
- <http://www.mkm-pi.com/mkmpi.php?article201>
- <http://www.cs.rhul.ac.uk/~alberto/>
- <http://www.almargen.com.ar/sitio/seccion/tecnologia/biotec/index.html>