

Bioinformática

Rodrigo Santamaría



Bioinformática

Contacto
Conocimientos
Introducción
Planificación
Temario
Evaluación
Bibliografía



Contacto

- ◆ Rodrigo Santamaría
 - ◆ Profesor Ayudante Doctor
 - ◆ Departamento de Informática y Automática
 - ◆ Dirección:
 - ◆ Edificio San Bartolomé 1-8, Pz. Fray Luis de León
 - ◆ Correo: rodri@usal.es
 - ◆ P. Web: <http://vis.usal.es/rodrigo>
 - ◆ Tutorías
 - ◆ Martes y Miércoles, de 16:00 a 19:00
 - ◆ A concertar por correo

Conocimientos requeridos

- ◆ Conocimientos de Informática
 - ◆ Bases de datos (SQL)
 - ◆ Sistemas operativos (UNIX)
 - ◆ Programación (fundamentos básicos)
- ◆ Conocimientos de otras asignaturas
 - ◆ **Genética molecular**
 - ◆ **Estadística**
 - ◆ Bioquímica
 - ◆ Biología Celular
 - ◆ Genética

Introducción

¿Qué es la bioinformática?

Cantidad de información

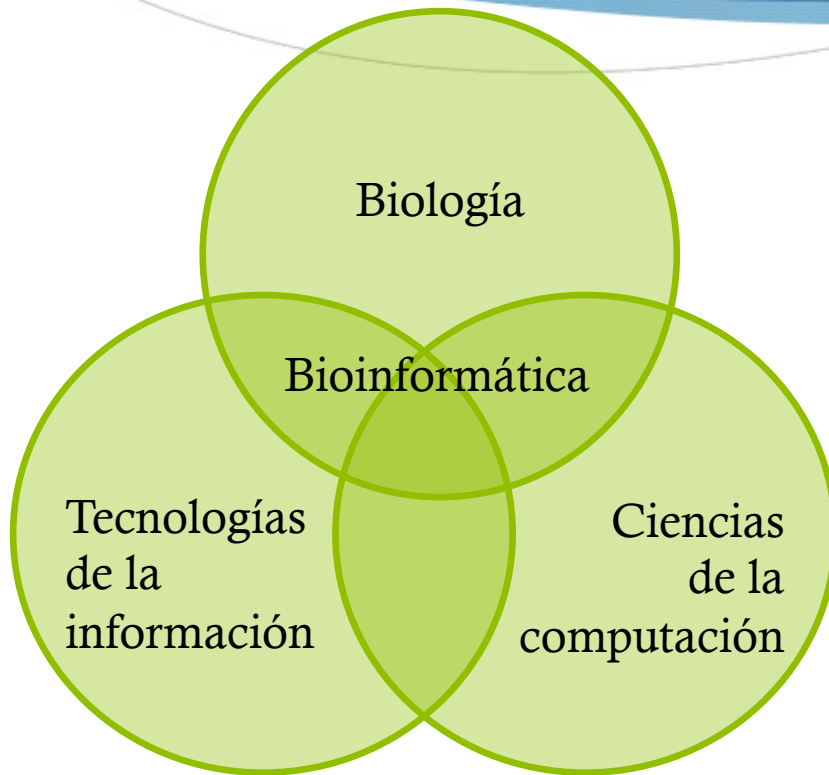
Tipos de experimentación

Algunas definiciones

Algunos ejemplos



¿Qué es la bioinformática?

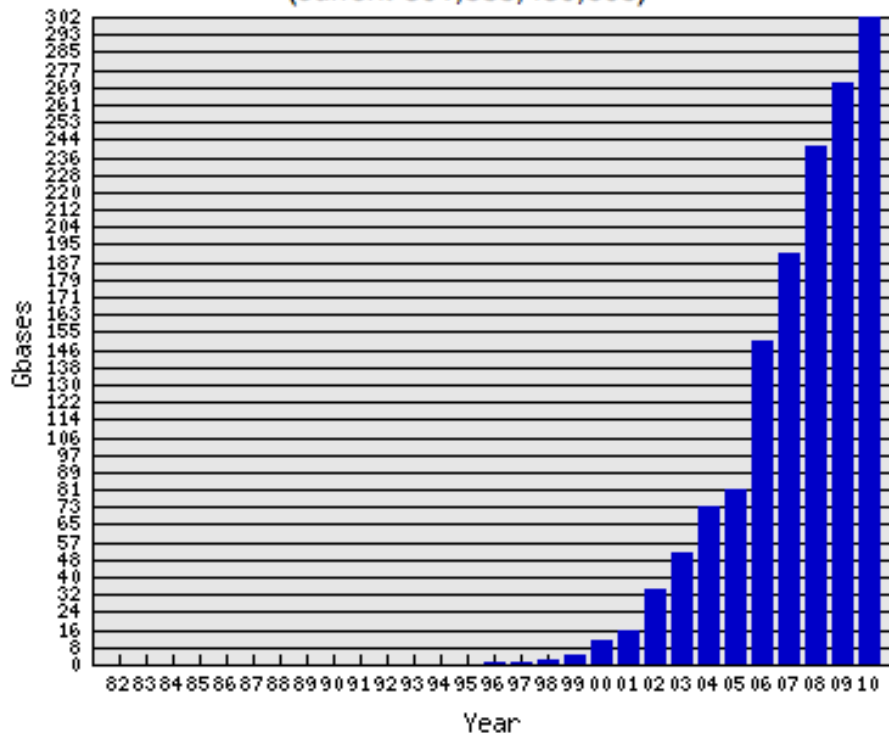


- ◆ Necesidad de la informática para resolver problemas biológicos
- ◆ Ingente cantidad de información debido a las revoluciones tecnológicas en biología
 - ◆ PCR (Polimerase Chain R.)
 - ◆ Human Genome Project
 - ◆ 1000 Genomes project
 - ◆ ...

Cantidad de información

EMBL Database Growth

Total nucleotides
(current 301,588,430,608)



- Despunta con el comienzo del siglo XXI
- Primer borrador del genoma humano en 2000
- El crecimiento es exponencial

Número de nucleótidos en la base de datos de secuencias de EMBL
(<http://www.ebi.ac.uk/embl/Services/DBStats>)

Cantidad de información

- ◆ Toda esta información debe manejarse a nivel informático
 - ◆ Gestión
 - ◆ Almacenamiento y consulta en bases de datos
 - ◆ Envío a través de redes
 - ◆ Entrada y salida de datos biológicos
 - ◆ Análisis
 - ◆ Detección de marcadores
 - ◆ Búsqueda de patrones
 - ◆ Predicción de comportamientos

Tipos de experimentación

- ◆ Tipos de experimentación biológica
 - ◆ Con un ser vivo (*in vivo*)
 - ◆ En un entorno artificial (*in vitro*)
 - ◆ En un entorno informático (*in silico*)
- ◆ No es una modalidad marginal o pasajera
 - ◆ Es parte básica de la mayoría de los avances recientes en biología
 - ◆ Genómica, metagenómica, biología de sistemas, tratamientos de cáncer, medicina personalizada, técnicas forenses, etc.

Algunas definiciones

◆ Algoritmo

- ◆ Lista bien definida, ordenada y finita de operaciones que permite hallar la solución a un problema
 - ◆ entrada → algoritmo → salida

◆ Programa

- ◆ Implementación en un ordenador de un algoritmo, mediante el uso de un determinado lenguaje de programación
 - ◆ Los lenguajes de programación más usados en bioinformática son Perl, R y Matlab

Algunas definiciones

- ◆ Minería de datos (Data mining)
 - ◆ Extracción no trivial de información que reside de manera implícita en los datos
 - ◆ Los datos no son lo relevante, si no los patrones, relaciones y estructura que encierran
 - ◆ Sus bases se encuentran en
 - ◆ El análisis estadístico
 - ◆ La inteligencia artificial



Perspectivas

- ◆ Molécula
 - ◆ Dogma central de la biología molecular (genómica)
 - ◆ DNA (genoma) → RNA (transcriptoma) → proteína (proteoma) → fenotipo celular
- ◆ Organismo
 - ◆ Distintas partes del cuerpo
 - ◆ Distintos momentos de su desarrollo
 - ◆ Distintos estados fisiológicos o patológicos
- ◆ Organismos
 - ◆ Evolución
 - ◆ Interacción entre organismos

Perspectivas

- ◆ Molecular:
 - ◆ Obtención y comparación de secuencias
 - ◆ Alineamiento de secuencias
 - ◆ Secuencias y evolución
- ◆ Organismo
 - ◆ Análisis de datos de microarray
 - ◆ Genómica funcional
 - ◆ Estructura de proteínas

Temario

Teoría

1. Bases de datos.
2. Estadística de secuencias genómicas
3. Alineamientos de pares de secuencias
4. BLAST
5. Alineamiento de múltiples secuencias
6. Filogenética
7. Análisis de microarray
8. Secuenciación de nueva generación

Práctica

1. Uso de BBDD principales
2. -
3. Alineamientos del NCBI
4. Algoritmos BLAST NCBI
5. Algoritmos y herramientas MSA
6. Algoritmos y herramientas filogenéticas
7. Análisis en R de microarrays
8. -

Prácticas obligatorias

Práctica

1. Uso de BBDD principales
2. -
3. Alineamientos del NCBI
4. Algoritmos BLAST NCBI
5. Algoritmos y herramientas MSA
6. Algoritmos y herramientas filogenéticas
7. Análisis en R de microarrays
8. -

práctica I: caracterización de una secuencia

práctica II: análisis de un microarray

Objetivos

- ◆ Familiarizarse la información biológica disponible
 - ◆ Y desarrollar las destrezas para consultarla e interrelacionarla
- ◆ Conocer las aplicaciones de la bioinformática en distintos campos
 - ◆ Biología molecular, medicina clínica, farmacología, biotecnología
- ◆ Conocer las principales técnicas y herramientas bioinformáticas utilizadas hoy en día para el análisis de datos biológicos
 - ◆ Comprender sus fundamentos algorítmicos
 - ◆ Desarrollar las destrezas para utilizarlas y aplicarlas
- ◆ Familiarizarse con problemas biológicos reales que se plantean hoy en día en un laboratorio y sólo pueden abordarse mediante la bioinformática

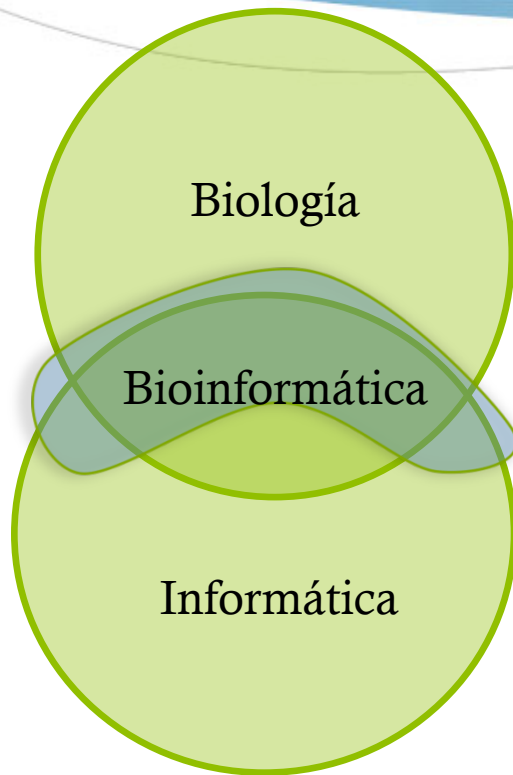
Objetivos

- ◆ En esencia, la bioinformática trata de **comparar**
 - ◆ Dos secuencias de nucleótidos o proteínas
 - ◆ Dos experimentos de expresión génica
 - ◆ Dos árboles filogenéticos ...
- ◆ Aprender a comparar
 - ◆ ¿Qué? Sabiendo obtener y discriminar los datos a comparar
 - ◆ ¿Cómo? Entendiendo los métodos de comparación
 - ◆ ¿Qué obtengo? Analizando de manera crítica los resultados de la comparación

Destrezas a adquirir

- ◆ Conocer, distinguir y consultar bases de datos de secuencias, proteínas, genomas, microarrays, etc.
- ◆ Conocer distintas herramientas para el análisis de secuencias de nucleótidos y proteínas
 - ◆ A nivel de usuario final
 - ◆ Pero comprendiendo los algoritmos subyacentes
 - ◆ Sabiendo interpretar los resultados
- ◆ Conocer los resultados de la experimentación con microarrays y saber analizarlos

Destrezas a adquirir



- ◆ Perfil de bioinformático a nivel de usuario
 - ◆ Familiaridad con las bases de datos y herramientas de análisis
- ◆ Pero con:
 - ◆ Conocimiento profundo de los fundamentos algorítmicos
 - ◆ Capacidad de análisis crítico de los datos y resultados

Perfil del bioinformático

Informático



Biólogo



Bioinformático



Relevancia del bioinformático

- ◆ *Ask 10 investigators in human genetics what resources they need most and it is highly likely that **computational skills and tools** will be at the top of the list.*
 - ◆ David Valle, M.D.
- ◆ *Scientists who enter genomics with traditional life science training must pick up several informatic skills along the way. They need a fundamental understanding of programming, bioinformatics and data mining. **It's critical to be able to mine your data and to be conversant in the language of programming***
 - ◆ Janet Warrington, Affymetrix

Relevancia del bioinformático

- ◆ *We prefer candidates who are comfortable working with databases, writing code to conduct statistical analyses and interpreting analysis results in light of recent biological research publications.*
 - ◆ Jason Johnson, Rosetta
- ◆ *Individuals who have a lot of biology experience and computed experience are **unique**. It adds to their value to have both sides*
 - ◆ Shawn Burgess, NHGRI
- ◆ [Genomics: Bridging Research Areas, Peter Gwynne, Science Career Magazine, May 2003](#)

Horario

- ◆ Teoría (Aula C1)
 - ◆ Lunes 10:00 – 11:00
 - ◆ Miércoles 11:00 – 12:00
- ◆ Prácticas (Aula de Informática 2)
 - ◆ Miércoles 12:00 – 14:00

Material

- ◆ Moodle: <http://studium.usal.es>
 - ◆ Nombre: *Bioinformática (Lic. en Biotecnología)*
 - ◆ Transparencias
 - ◆ Foros
 - ◆ Ejercicios
 - ◆ Noticias
 - ◆ ...
- ◆ Página personal: <http://vis.usal.es/rodrigo>
 - ◆ Algún contenido adicional

Evaluación

◆ Evaluación

- ◆ Ejercicios prácticos entregados a lo largo del curso (50%)
 - ◆ Ejercicios opcionales semanales
 - ◆ Evaluación continua y ayuda para los ejercicios obligatorios
 - ◆ Dos ejercicios obligatorios finales de carácter general
 - ◆ Necesarios y suficientes para aprobar la parte práctica
- ◆ Examen final con preguntas principalmente teóricas (50%)
 - ◆ 27 enero 2012
 - ◆ 9 septiembre 2012

Metodología de trabajo

◆ Teoría

- ◆ Alumno: traer el tema para esa semana leído

- ◆ Clases:

- ◆ Se pasará lista

- ◆ Se repasará el tema, no de manera exhaustiva, si no explicando los conceptos más relevantes y atendiendo las dudas que les hayan surgido a los alumnos durante su lectura

- ◆ Se abrirá un debate en clase sobre cuestiones relacionadas con el tema

Metodología de trabajo

◆ Práctica

◆ Alumno:

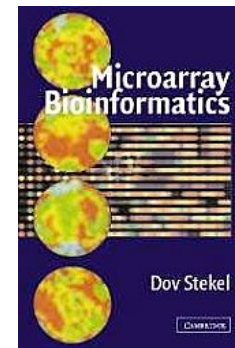
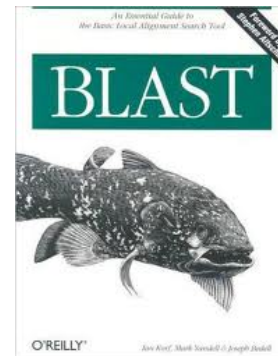
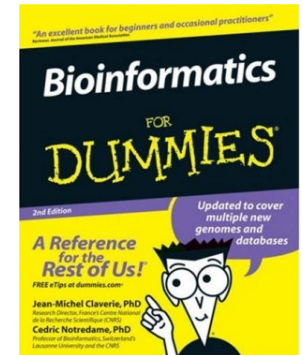
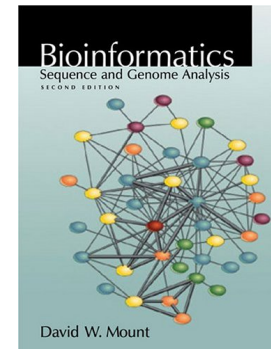
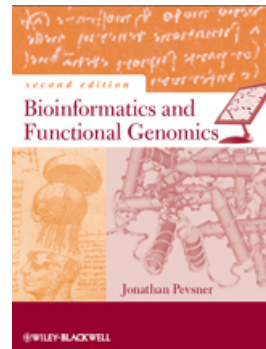
- ◆ Haber entregado los ejercicios de la sesión anterior (si había)
- ◆ Traer repasada la teoría del tema correspondiente (que ya ha sido leído y visto en clase)

◆ Sesiones:

- ◆ Se pasará lista
- ◆ Se atenderán dudas sobre los ejercicios de la sesión anterior (si había).
- ◆ Aleatoriamente, se preguntará sobre los ejercicios entregados.
- ◆ Se comenzará la realización de los ejercicios de esta sesión.

Bibliografía

- ◆ **Bioinformatics and Functional Genomics.** Jonathan Pevsner, 2nd Ed. Wiley Publishing 2009
- ◆ **Bioinformatics: Sequence and Genome Analysis.** Nello Cristianini and Mathew W. Hahn. Cambridge Univ. Press 2006
- ◆ **Bioinformatics for Dummies.** Jean-Michel Cleverie & Cédric Notredame, 2nd Rev. Ed. Wiley Publishing 2006
- ◆ **BLAST.** Ian Korf, Mark Yandell and Joseph Bedell. O'Reilly 2003.
- ◆ **Microarray Bioinformatics.** Dov Stekel. Cambridge University Press 2003





DNA Art es una empresa que, a partir de una muestra de saliva, analiza tu ADN y lo convierte en un cuadro

www.dna11.com