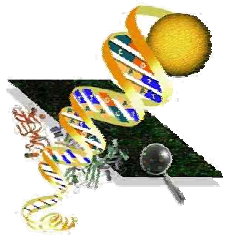
CS 177 Introduction to Bioinformatics



Dr. Tom Wilke

Department of Microbiology and Tropical Medicine

e-mail: mtmtxw@gwumc.edu

Tel.: 202 994 3635

Prof. Rahul Simha

School of Engineering and Applied Science (SEAS)

e-mail: simha@seas.gwu.edu

Prof. Robert Donaldson

Chair Department of Biological Sciences

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

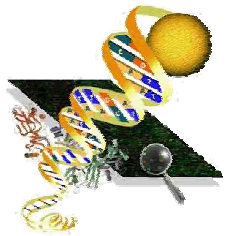
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

The Poliovirus Problem



Science magazine

VOL 297, 9 August 2002

Cello, J; Paul, A.V. & Wimmer, E.:

Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template

- they generated about 7.7 kilobases of single-stranded RNA genome based on the know genetic map
- DNA fragments were synthesized from purified oligo-nucleotides (average length 69: bases)
- the cDNA was then transcribed into highly infectious RNA

http://www.sciencemag.org/cgi/reprint/297/5583/1016.pdf

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

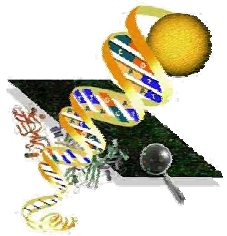
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

The Poliovirus Problem



washingtonpost.com

17 July 2002

Weiss, R.:

Mail-Order Molecules Brew a Terrorism Debate

- mail-order oligonucleotides can be used to manufacture a deadly virus

- because they are so small, most oligos lack a "fingerprint"

- call for more control and/or institutional oversight

- method could be used to manufacture other deadly viruses

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

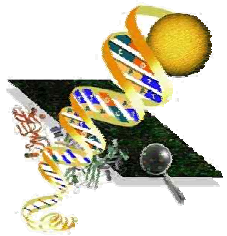
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

The Poliovirus Problem



Bioinformatics

- played a crucial role in the manufacturing of the poliovirus

- it could also play a critical role to control and prevent misuse of science

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

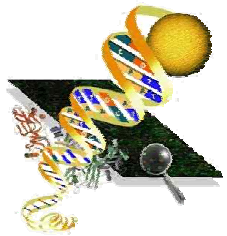
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



Bioinformatics.org

The scientific field of bioinformatics involves the use of information systems to analyze large biological data sets, often DNA and protein sequences. A subdiscipline of computational biology, it is relatively new, having been derived from individual efforts in the statistical analysis of sequences. The first reference to the word "bioinformatics" in the scientific literature was in 1991.

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

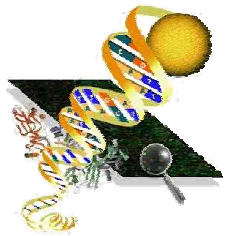
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



The European Bioinformatics Institute (EBI)

The EBI is a center for research and services in bioinformatics. The Institute manages databases of biological data including nucleic acid, protein sequences and macromolecular structures

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

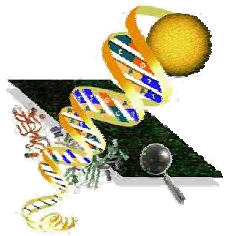
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



Bioinformatics (Journal)

The journal aims to publish high quality, peer-reviewed, original scientific papers and excellent review articles in the fields of computational molecular biology, biological databases and genome bioinformatics.

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

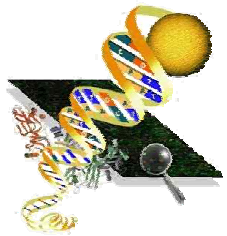
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



http://www.netsci.org/Science/Bioinform

The first level … can be defined as the design and application of methods for the collection, organization, indexing, storage, and analysis of biological sequences (both nucleic acids [DNA and RNA] and proteins). The next stage of bioinformatics is the derivation of knowledge concerning the pathways, functions, and interactions of these genes (functional genomics) and proteins (proteomics).

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

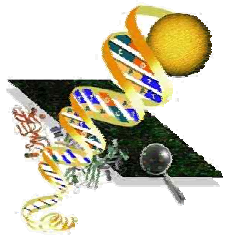
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



UCLA Bioinformatics Institute

Bioinformatics is the study of the inherent structure of biological information and biological systems. It brings together the avalanche of systematic biological data (e.g. genomes) with the analytic theory and practical tools of computer science and mathematics.

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

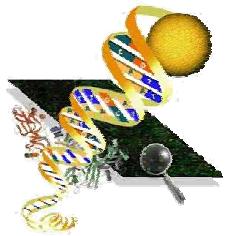
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



UPenn Center for Bioinformatics

… these disciplines deal with the management, analysis, and visualization of the flood of information generated in molecular biology, genomics, and other areas of biology and biomedicine.

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

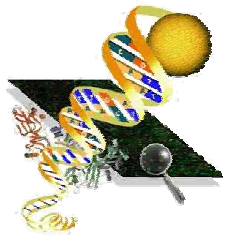
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



Bioinformatics (S.M. Brown, 2000)

Bioinformatics can be defined as the use of computers for the acquisition, management, and analysis of biological information. It exists at the intersection of molecular biology, computational biology, clinical medicine, database computing, the Internet, and sequence analysis.

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

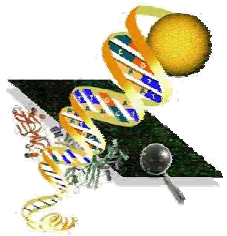
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



Weizmann Institute of Science

… although the term 'Bioinformatics' is not really well-defined, you could say that this scientific field deals with the computational management of all kinds of biological information, whether it may be about genes and their products, whole organisms or even ecological systems.

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

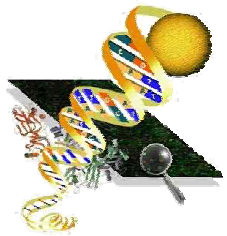
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



National Institutes of Health (NIH)

Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

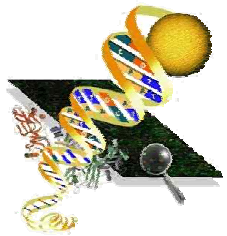
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



sensu stricto

An interdisciplinary field involving biology, computer science, mathematics, and statistics to analyze biological sequence data, genome content, and arrangement, and to predict the function and structure of macromolecules (D.W. Mount, 2001).

sensu lato

Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data (NIH).

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

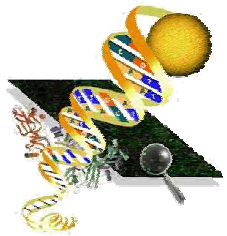
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

What is Bioinformatics?



Google search

	Sep 2002	Jan 2003
Proteomics	**158,000**	**261,000**
Genomics	**606,000**	**813,000**
Bioinformatics	**837,000**	**1070,000**
	1,601,000	**2,144,000**
Michael Jackson		**775,000**
George W. Bush		**1,700,000**

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

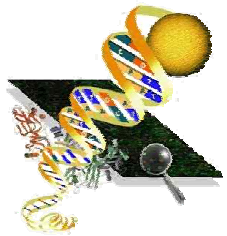
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

History of Bioinformatics



Gregor Mendel ("Father of Genetics") cross-fertilized different colors of the same species of flowers. In a journal, he kept careful records of the colors of flowers that he cross-fertilized and the colors of flowers they produced.	1865
Pauling and Corey propose the structure for the alpha-helix	1951
Watson and Crick propose the double helix model for DNA based on x-ray data obtained by Franklin and Wilkins	1953
The sequence of the first protein to be analyzed, bovine insulin, is announced by F. Sanger	1955
The first integrated circuit is constructed by Jack Kilby at Texas Instruments	1958
Margaret Dayhoff starts the Atlas of Protein Sequence and Structure	1965
The details of the Needleman-Wunsch algorithm for sequence comparison are published	1970
Protein Sequence Database (PSD) by Margaret Dayhoff	1972

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

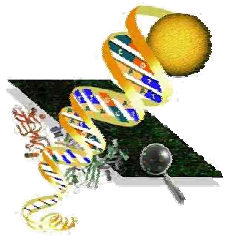
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

History of Bioinformatics



The first recombinant DNA molecule is created by Paul Berg and his group	1972
Stanley Cohen invented DNA cloning	1973
Sanger et al. invent cycle sequencing	1977
The first complete gene sequence for an organism (Bacteriophage FX174) is published. The gene consists of 5,386 base pairs which code nine proteins	1980
The Smith-Waterman algorithm for sequence alignment is published	1981
IBM introduces its Personal Computer to the market	1981
The PCR reaction is described by Kary Mullis and co-workers	1983
The FASTP algorithm is published by Lipman & Pearson	1985

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

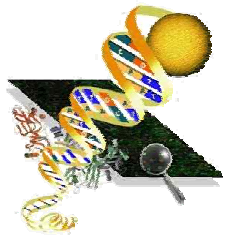
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

History of Bioinformatics



The term "Genomics" appeared for the first time. It was coined by Thomas Roderick as a name for the new journal	1986
The SWISS-PROT database is created (University of Geneva and the European Molecular Biology Laboratory)	1986
The Human Genome Initiative is announced by DOE	1986
Perl (Practical Extraction Report Language) is released by Larry Wall.	1987
The National Center for Biotechnology Information (NCBI) is established at the National Cancer Institute in Bethesda	1987
The physical map of *E. coli* is published	1988
The FASTA algorithm for sequence comparison is published by Pearson & Lupman	1988
The BLAST program (Altschul, et. al.) is implemented	1990
The first reference to the word "bioinformatics" in the scientific literature (source: Bioinformatics.org)	1991?

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

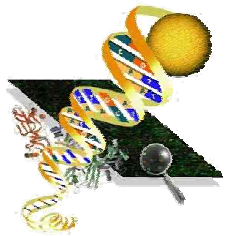
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

History of Bioinformatics



The research institute in Geneva (CERN) announces the creation of the protocols which make-up the World Wide Web	1991
Human Genome Systems, Gaithersburg Maryland, is formed by William Haseltine	1992
The Institute for Genomic Research (TIGR) is established by Craig Venter in Rockville	1992
The PRINTS database of protein motifs is published by Attwood and Beck	1994
Sun releases version 1.0 of Java. Sun and Netscape release version 1.0 of JavaScript	1995
The *Haemophilus influenzea* genome (1.8 Mb) is sequenced	1995
Affymetrix produces the first commercial DNA chips	1996
Craig Venter forms Celera in Rockville, Maryland	1998
The Swiss Institute of Bioinformatics is established in Geneva	1998
A draft of the human genome (3,000 Mbp) is published	2001

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

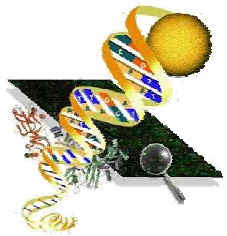
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

History of Bioinformatics



2003

Having 40 dedicated students interested in Bioinformatics

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

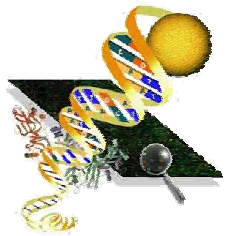
Course outline

Bioinformatics sources

Bioinformatics careers

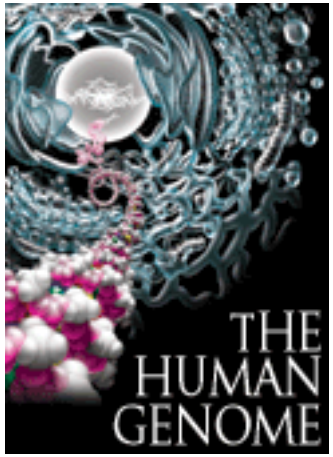
Examples

The Human Genome Project



The Human Genome Project is complete - almost

- publication of draft sequence on 15 Feb 2001 (based on data "frozen" on 7 Oct 2000)
- approximately 3.29 billion base pairs
- approximately 35,000 genes
- **draft** sequence means the DNA was sequenced on average four times
- there are still gaps, typos and assembly errors
- **finished** sequence will have eightfold coverage with an error of about 1 in 10,000 (2003)



Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

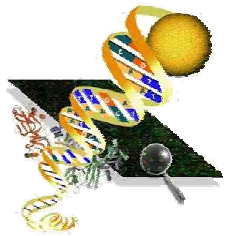
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

The Human Genome Project



Gene number estimates

Anticipated in 1998: 60,000-140,000

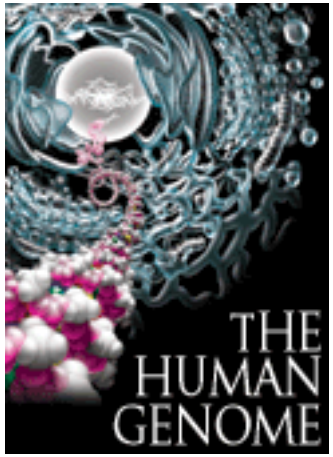
Celera: ~35,000

Affymetrix Gene Chip: ~60,000

GenBank: ~50,000 gene coding sequences

UniGene: ~96,000 clusters of unique human ESTs
(an expressed sequence tag is a small part of the active part of a gene, made from cDNA)

Desperate need for theoretical network



Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

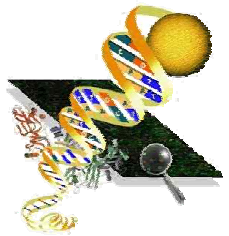
Course outline

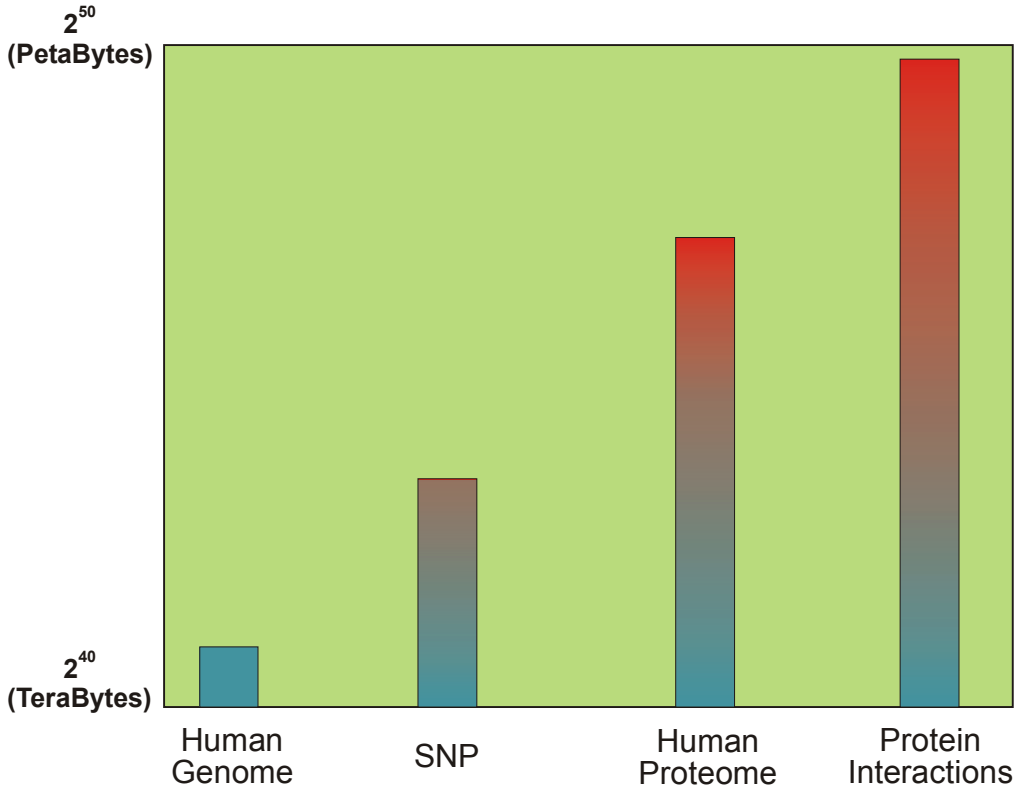
Bioinformatics sources

Bioinformatics careers

Examples

Information Complexity: Genomics vs. Proteomics





- **Problem:** bioinformatics is major bottle neck in many genomics/proteomics applications relative to data analysis, storage, management, search, and retrieval

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

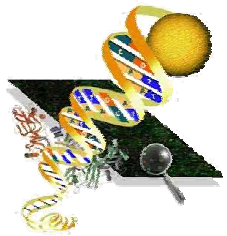
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Winston Churchill, 1942



"Now this is not the end.
It is not even the beginning of the end.
But it is, perhaps, the end of the beginning."

Introduction

The Polio problem

What is Bioinformatics?

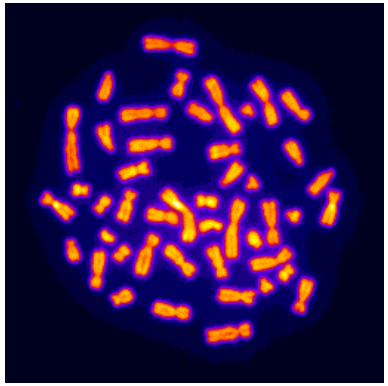
History of Bioinformatics

Course outline

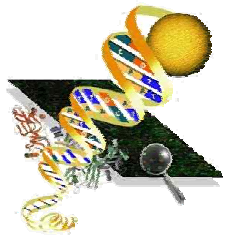
Bioinformatics sources

Bioinformatics careers

Examples



Course outline



Course objective

The general objective of the course is to provide a one semester introduction and overview to the fields of bioinformatics

The aim is to provide a practical description of the topics, tools, issues and current trends in the fields

As an introductory course, the focus will not be on the theoretical and computational aspects of the fields.

Students should become familiar with the terminology, principles, and strategies in bioinformatics

They will learn to use conventional software and web-based applications

Students should gain competence in the field of bioinformatics by using the approach of problem-based learning

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

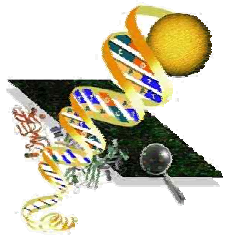
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Course outline



Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Course schedule

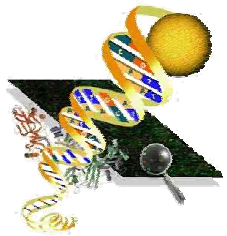
Lecture 2 (Jan 27): Cell Biology Overview
* Organisms, organs, tissues, cells.
* Cell components.
* Eurokaryotic vs. prokaryotic cells.
* Overview of diseases, immune system.
* Biomolecules, proteins

Lecture 3 (Feb 2): DNA/RNA overview
* DNA and its components
* RNA and its components
* PCR, sequencing
* Mutations

Lecture 4 (Feb 9): Nucleotide and protein databases
* Public sequence databases
* Sequence retrieval and examples
* Similarity searching
* Gene identification
* Genetic and physical map
* Protein databases
* Data exchange and management

Lecture 5 (Feb 16): Hands-on lab with databases
* Motivating problem: the poliovirus
* Review nucleotide and protein databases
* Sequence formats
* Lab exercises in using GenBank

Course outline



Course schedule (continued)

Lecture 6 (Feb 23): The Alignment problem
Part I:
* Pairwise alignment
* Dynamic programming algorithm
Part II:
* Multiple alignment
* Editing and formatting alignments

Lecture 7 (Mar 3): The new biology lab
* Old technology overview: microscopy …
* PCR, sequencing,
* Microarrays
* Crystallography
* Mass-spec

Lecture 8 (Mar 10): Proteins, part 1 (Structure-function relationships)
* Review of protein structures
* Experimental techniques to determine protein structures
* Protein databases

Lecture 9 (Mar 24): Proteins, part 2 (Computational modeling)
* Database similarity search
* Protein family analysis
* Structural analysis
* Three-dimensional comparative modeling
* Three-dimensional structural analysis in laboratory

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

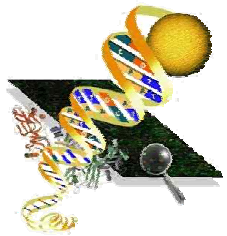
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Course outline



Course schedule (continued)

Lecture 10 (Mar 31): Phylogenetics I
* Evolution: overview
* Taxonomy and phylogenetics
* Phylogenetic trees
* Cladistic vs. phenetic analyses
* Models of sequence evolution

Lecture 11 (Apr 7): Phylogenetics II
* Phylogenetic trees and networks
* Cladistic and phenetic methods
* Computer software and demos

Lecture 12 (Apr 14): Field trip

Lecture 13 (Apr 21): Student presentations

Lecture 14 (Apr 29): Simulations (time permitting)
* Simulations
* Biological metaphors in computing

Final Examination (May 5)

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

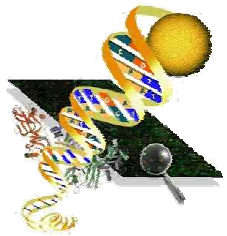
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Course outline



What is expected from students?

- Students should anticipate spending a minimum of ~ 3 hrs a week outside of class reading and studying the lecture notes and reading assignments and carrying out the assigned homework/exercises

- Students will need access to a computer with an internet connection and e-mail

- PC access is available on campus at Tomkins 405 and Himmelfarb library

- E-mail is necessary for submission of homework

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

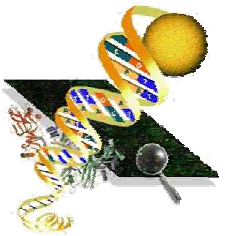
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Course outline



Introduction

The Polio problem

What is Bioinformatics?

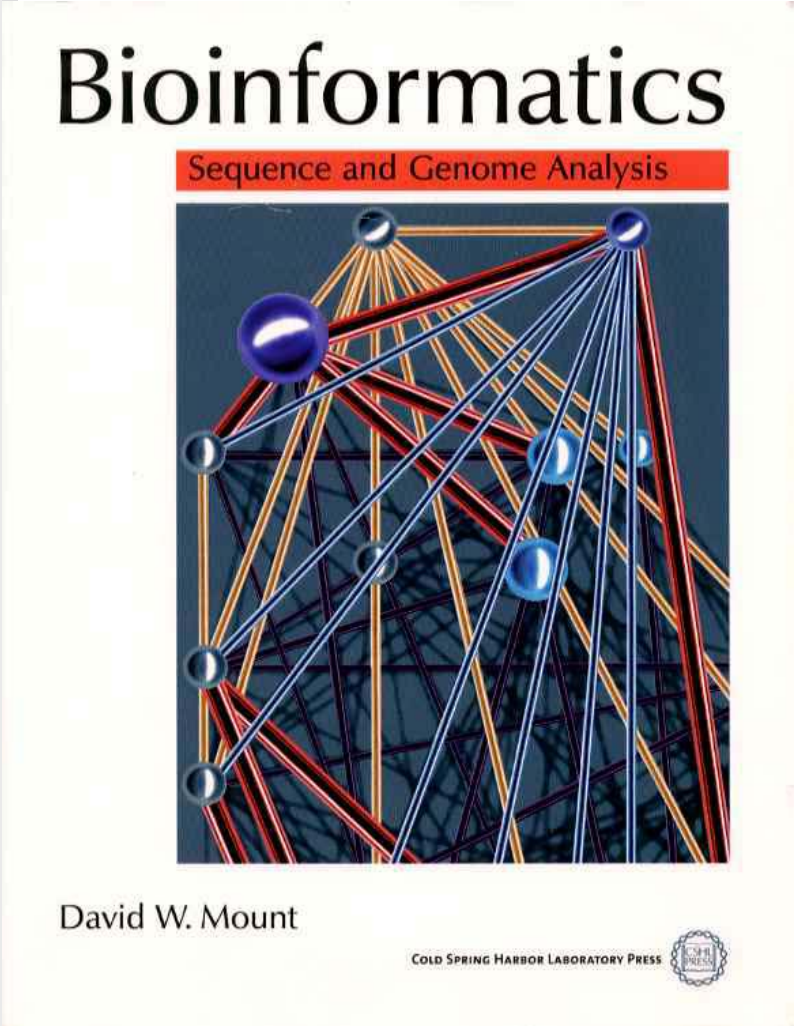
History of Bioinformatics

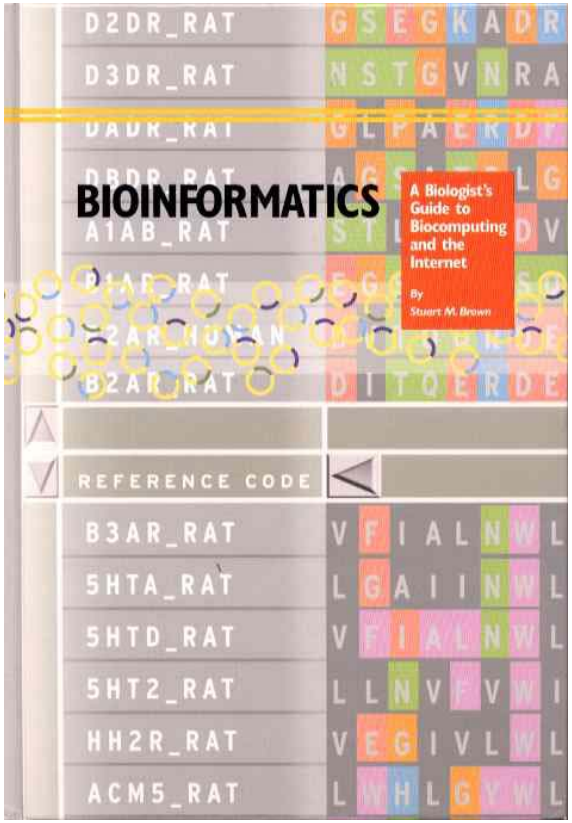
Course outline

Bioinformatics sources

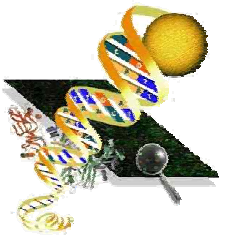
Bioinformatics careers

Examples





Bioinformatics sources



Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

Course outline

Bioinformatics sources

Bioinformatics careers

Examples

NCBI HomePage - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites History

Address http://www.ncbi.nlm.nih.gov Go Links

NCBI

National Center for Biotechnology Information

National Library of Medicine National Institutes of Health

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search Nucleotide for Go

SITE MAP
Guide to NCBI resources

About NCBI
The science behind our resources. An introduction for researchers, educators and the public.

GenBank
Sequence submission support and software

Molecular databases
Sequences, structures and taxonomy

Literature databases
PubMed, OMIM, Books and PubMed Central

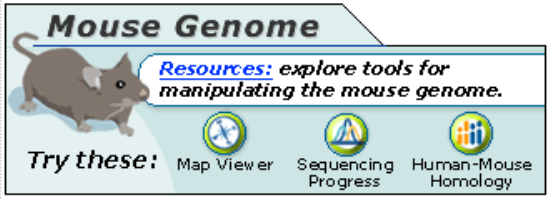
Genomic biology
The human genome, whole genomes and related resources

Tools
for data mining

Research at

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...



BLAST the zebrafish genome!

NCBI introduces a new Zebrafish Genome BLAST page. BLAST Zebrafish mRNA, EST, HTGS, Trace (WGS and EST), and reference sequences. More...

NCBI in the News

NCBI now offers quick links to online resources through LinkOut. This new feature of

Hot Spots

- Cancer genome anatomy project
- Clusters of orthologous groups
- Coffee Break
- Electronic PCR
- Gene expression omnibus
- Genes and disease
- Human genome resources
- Human map viewer
- Human/mouse homology maps
- LocusLink
- Malaria genetics & genomics
- Mouse genome resources
- ORF finder
- Reference

Internet

European Bioinformatics Institute

Research at the European Bioinformatics Institute

Bioinformatics Products and Services

FLASH Intro

European Bioinformatics Institute
a part of the European Molecular Biology Laboratory

the path to knowledge

© Copyright European Bioinformatics Institute 2002. All Rights and Trademarks Reserved.

Bioinformatics.Org: The Open Lab - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
Address http://bioinformatics.org/projects/
Go Links

Bioinformatics.Org

BioBanner Visit MolBiol.Net BioBanner

Software/Group Search

Not Logged In
- Login
- Join BiO

The Organization
- Home Page
- About Us
- Contact Us
- Organizational Members
- Events

The Development Laboratory

Project Hosting
- Home Page
- Project List
- Software Map
- FTP Repository
- CVS Repository
- Mailing Lists
- Documentation

The Research Laboratory

New! Online Databases
- Immigrant Genes

Online Analysis Tools
- COMBOSA3D
- PeCoP
- SeWeR

News

Submit News | Archive | Subscribe

- **Systems down today for software upgrades**
  J.W. Bizzaro - 01/04/02 14:18 (1 comment)

  The systems may be down for a couple hours today (January 4) for some software upgrades, starting around 22:00 GMT. Please do not use the systems for any important work at that time.

- **FTP temporarily down due to system upgrade**
  J.W. Bizzaro - 11/09/01 17:21

  FTP will be unavailable for an hour or so after receipt of this notice. We are expanding the hard drive capacity of the projects server (currently dubbed "depot") by about 62 GB. The new space will eventually be used for project shell, Web and CVS space, in addition to FTP and backups.

- **File release function being worked on -- temporarily off line**
  J.W. Bizzaro - 10/13/01 02:10

  For project administrators, the file release function is being worked on this weekend. Some improvements are planned, such as placing the files on FTP rather than on the Web. Please do not use this feature until further notice. If you need to make a file available, please use your FTP directory.

- **Connection Problems NOT QUITE Fixed**

Statistics

Hosted Projects: 77
Members: **4007**

Hosted Projects

- A Database of Leukemia Genes
- A Database of Pancreatic Cancer Genes
- A Smith Waterman algorithm on GreenTea
- AcE
- AnnHyb
- ANTHEDNA
- Biochemical Network Visual Designer
- Bioinformatics Benchmarking System
- Bioinformatics Package Repository
- biolib fortran library
- BioQuery
- Cardiovascular Gene Database
- CGInformatics
- COMBOSA3D
- DeMasker
- DEODAS
- Distributed Computing Power Project
- dnacgr
- E-CELL Simulation Environment Version 1
- E-CELL Simulation Environment Version 2
- E-CELL Simulation Environment Version 3
- EnzymeLab
- FishermansFriend:The Old Man And The Sea
- FLIMS

Internet

The Institute for Genomic Research - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Netsite: http://www.tigr.org/ What's Related

TIGR THE INSTITUTE FOR GENOMIC RESEARCH

Home

Genome Databases
Comprehensive Microbial Resource
Unfinished Genomes
Eukaryotic Projects
Gene Indices
more >>

Microarray Projects
PFGRC
more >>

Software
Glimmer
MUMmer
more >>

Conferences
14th International GSAC
more >>

Scientific Publications
Faculty
Education & Training
Genome News Network
Related Links

News

09.04.2002 Genomics Course for Educators to Start in October
TIGR's genomics course for educators, which is designed to give high school biology teachers a thorough understanding of genomics, will be offered on Saturdays this fall, starting on October 19th.
Press Release

08.26.2002 TIGR Completes Whole Genome Sequence of an Emerging Human Pathogen
An analysis of the whole genome sequence of Streptococcus agalactiae serotype V provides valuable insights into the virulence mechanism of the pathogen, which is a leading cause of pneumonia and meningitis in newborns and the source of life-threatening illnesses in a growing number of adults with deficient immune systems.
Press Release

08.15.2002 TIGR, IBEA, and TCAG to Create New High-Throughput Genomic Sequencing Facility
The Institute for Genomic Research (TIGR), Institute for Biological Energy Alternatives (IBEA) and The Center for the Advancement of Genomics (TCAG), all not-for-profit organizations supported by the J. Craig Venter Science Foundation, announced today their plan to create a next generation, high-throughput DNA sequencing facility in Rockville, Maryland.
Press Release

More TIGR News >>

Genome News Network

August 30 to September 12, 2002

Putting their best paw forward

Start The Instit... Microsoft ... - Netscape Buy a car ... The Insti... - Netscape 3:25 PM

Welcome to Bioplanet

the bioinformatics homepage

BioArticles
PlanetJobs
Employers
PlanetChat
BioLinks
Education
Recruiters

Latest Bioinformatics topics from the Planet Forums :

NEW	PlanetCV	:	Bioinformatics Programmer in Scotland	: by	johanvni
NEW	PlanetChat	:	Masters Bioinformatics textbooks	: by	Exploited
NEW	PlanetChat	:	"The role of foreign worker scientists i	: by	Cannibal
NEW	PlanetChat	:	INTERNSHIPS ????????????	: by	nambisan
NEW	PlanetChat	:	A sincere advice for young kid to choose	: by	ric

905 Members, **333** Bioinformatics Jobs, **158** Resumes, **1202** Total posts

33 people live in forums : no members and 33 guests

Latest Bioinformatics News:

Datamonitor launches Immune Disorders and Inflammation Review...
BioPortfolio Fri Sep 06 09:54:00 GMT-0400 (Eastern Daylight Time) 2002

Biometrics tested at airport...
ZDNet Fri Sep 06 09:40:00 GMT-0400 (Eastern Daylight Time) 2002

Chiron-Genentech case to jury...
BayArea.com Fri Sep 06 08:50:00 GMT-0400 (Eastern Daylight Time) 2002

Datamonitor launches eHealth Review...
BioPortfolio Fri Sep 06 07:19:00 GMT-0400 (Eastern Daylight Time) 2002

Datamonitor launches Licensing Review...
BioPortfolio Fri Sep 06 06:49:00 GMT-0400 (Eastern Daylight Time) 2002

Zimbabwe 'accepts GM food'...
BBC Fri Sep 06 03:52:00 GMT-0400 (Eastern Daylight Time) 2002

Zimbabwe accepts GM food...
ONE News Fri Sep 06 02:54:00 GMT-0400 (Eastern Daylight Time) 2002

Biotech needs to slow down...
Red Herring Fri Sep 06 01:35:00 GMT-0400 (Eastern Daylight Time) 2002

Genetically-Engineered Food is Safe, Assures Colin Powell...
VOA Thu Sep 05 23:07:00 GMT-0400 (Eastern Daylight Time) 2002

Biopure gets grant for Hemopure trial...
digitalMASS Thu Sep 05 16:12:00 GMT-0400 (Eastern Daylight Time) 2002

The Top Bioinformatics Resource

We have received a Golden Web Award for the new design :
We are the most popular bioinformatics jobs site on Google
Yahoo! have selected our site for their directory

What is bioinformatics?

What skills are key? Read all about it at the UPDATED BioArticles page - contains articles and FAQs concerning bioinformatics related issues.

Bioinformatics related articles and tutorials

Here you will find a collection of articles and tutorials, relating to bioinformatics and genomics.

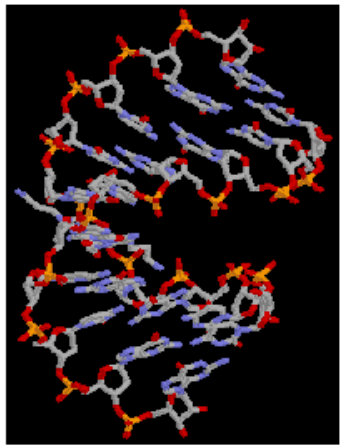
- What is bioinformatics??? **A brief description of this emerging field.**
- FAQs - Frequently asked questions on bioinformatics related topics Now you know what it is, find out what it takes to become a bioinformatician . Bioinformatics career questions answered.
- Human Genome Research A description of the Human Genome Project.
- Sequence analysis Keith Robison's guide to the exciting world of biosequence comparison! Useful background information on a variety of computational biology algorithms.
- Biocomputing For Everone an introduction to biocomputing for the layperson published by the VSNS biocomputing division.
- Biocomputing For Schools Another VSNS publication - this time aimed at highschool students, but fun to read for everyone. Includes articles on the application of bioinformatics to BSE research, and a ``Do-It-Yourself'' detailed example of a WWW search.
- Bioinformatics review articles published 1993-1996. A collection of off-line references to review papers, compiled by The Irish National Centre for BioInformatics. A great starting point for anyone wanting a general introduction to the field.
- Elementary Sequence Analysis - Database Searching by B. Golding Jan 1996. Fasta, blast, blitz, blaze, flash, blocks.
- How to become a bioinformatics expert, including a listing of ***European*** opportunities to study bioinformatics. Compiled by theVirtual School of Natural Sciences BioComputing Division.
- Pedestrian Guide to Analysing Sequence Databases
- Protein Sequence Alignment and Database Scanning Geoff Barton's review.
- A Practical Guide to protein sequence and structure analysis at UCL
- VHG Virtural HyperGlossary. Defines terms used in different subfields, currently Glycoscience, Protein Structure.
- A Bibliography on Computational Gene Recognition
- HGP Genomics Primer from the DOE
- Career Advice for Computational Biology By Amjad-Ali Khoja at U. Texas.

BIOINFORMATICS

Introduction

- Databases
- Finding Genes
- Protein Models

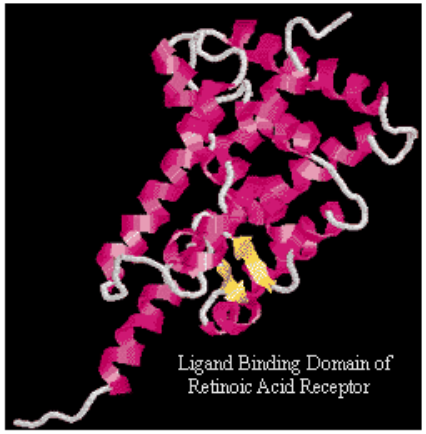
Courses
Grant Sources
Programs
Resources



In the last few decades, advances in molecular biology and the equipment available for research in this field have allowed the increasingly rapid sequencing of large portions of the genomes of several species. In fact, to date, several bacterial genomes, as well as those of some simple eukaryotes (e.g., *Saccharomyces cerevisiae*, or baker's yeast) have been sequenced in full. The Human Genome Project, designed to sequence all 24 of the human chromosomes, is also progressing. Popular sequence databases, such as GenBank and EMBL, have been growing at exponential rates. This deluge of information has necessitated the careful storage, organization and indexing of sequence information. Information science has been applied to biology to produce the field called **Bioinformatics**.

The simplest tasks used in bioinformatics concern the creation and maintenance of databases of biological information. Nucleic acid sequences (and the protein sequences derived from them) comprise the majority of such databases. While the storage and or ganization of millions of nucleotides is far from trivial, designing a database and developing an interface whereby researchers can both access existing information and submit new entries is only the beginning.

The most pressing tasks in bioinformatics involve the analysis of sequence information. **Computational Biology** is the name given to this process, and it involves the following:



- Finding the genes in the DNA sequences of various organisms
- Developing methods to predict the structure and/or function of newly discovered proteins and structural RNA sequences.
- Clustering protein sequences into families of related sequences and the development of protein models.
- Aligning similar proteins and generating phylogenetic trees to examine evolutionary relationships.

The process of evolution has produced DNA sequences that encode proteins with very specific functions. It is possible to predict the three-dimensional structure of a protein using algorithms that have been derived from our knowledge of physics, chemistry and most importantly, from the analysis of other proteins with similar amino acid sequences. The diagram below summarizes the process by which DNA sequences are used to model protein structure. The processes

BioTech Logo

BIOINFORMATICS

Introduction

- Databases
- Finding Genes
- Protein Models

Courses
Grant Sources
Programs
Resources

Bioinformatics Resources

Primer on Molecular Genetics	A very good primer, kept up by the Department of Energy, on Molecular Genetics. This represents the molecular biology side of molecular bioinformatics.
JAM: the Biocomputing Survival Guide	This 'Survival' Guide is intended to serve as a cookbook for the casual user of computers in molecular biology. It is not intended to explain science or make the programs' algorithms more transparent.
A Bioinformatic Thesaurus	This basic guide discusses some of the more common terms used in bioinformatics.
Bioinformatics Resource List at Organelle Genome Megasequencing Program (Montreal)	This server contains lists of research groups, servers, and information sources.
Bioinformatics at Parallab	A brief paper discussing some basic tools of bioinformatics research.
Bioinformation Page at Duesseldorf	List of resources at other sites.
Computational Biology or Bioinformatics References at UCSC	A very thorough list of tools for computational biologists, some with a bit of brief annotation.
Internet Skills for Bioinformatics	A basic introduction on how to gear the Internet toward biological research and information exchange.
Links to bioinformatics servers on the Web at Bergen	A thorough list of servers, organized by category.

Computer Science Areas Related to Bioinformatics

Molecular Biology Information Servers/BioInformatics - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://www.nwfsc.noaa.gov/protocols/bioinformatics.html
What's Related

Fish Health / Microbiology

US Dept Commerce/NOAA/NMFS/NWFSC

Search the NWFSC Protocols:

Or go to Advanced Search

Protocols
Forums
Bioinformatics
Journals
Other Resources
MSDS Search

Fish Health - Microbiology Team

Northwest Fisheries Science Center

NMFS/NOAA

Last updated on 6/4/02

Bioinformatics: Biology and Molecular Biology Resources

▸ DNA and Protein Sequence Search, Comparison, Retrieval
▸ Genomics
▸ General Information and Resources

▸ DNA and Protein Sequence Search, Comparison, Retrieval

- National Center for Biotechnology Information (NCBI)
  - Entrez: entry point for all NCBI offerings
  - Searching
    - Nucleotides
    - Proteins
    - 3D structures
    - Genomes
    - Taxonomy
    - Literature - PubMed
  - BankIt: Sequence submission tool
  - Blast Homology Searches (Basic Local Alignment Search Tool)
- FHCRC BLOCKS WWW Server
- Protein Data Bank (PDB)
- Services at ExPASy (European Molecular Biology Server)
  - SWISS-PROT - Annotated protein sequence database
  - PROSITE - Dictionary of protein sites and patterns
  - SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis database
  - SWISS-3DIMAGE - 3D images of proteins and other biological macromolecules
  - CD40Lbase - The European CD40L Defect Database
  - ENZYME - Enzyme nomenclature database

Start The Institute for Genomic... Microsoft PowerPoint - [L... Molecular Biology Info... 3:36 PM

Bioinformatik - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://www.bioinformatik.de/
What's Related
bioinformatik.de
add url guestbook about bioinformatik.de email
TEXT-QUOTED
search
Software
Unix, Mac, Windows, Online Tools...
Companies
Bioinformatics & Software, Cloning & Expression...
Microarray Technology
Companies...
Literature
Books, Literature search, Protocols...
Miscellaneous
Fun...
International Trade Fairs
Institutes
Service Centers
Usenet
Societies
Research and Education
Online Courses and Tutorials...
Databases
Genome Projects, Genes, Proteins...
PCR
Primer Design...
Jobs
Career Tools...
Other Resources
Catalogs...
Meetings and Workshops
BioRegionen (in Germany)
Bioinformatics News
Evolution and Phylogeny
Origin of life...
NEW! Latest additions to bioinformatik.de
NEW! Bioinformatik - ein Leitfaden für Naturwissenschaftler
Start
The Institute for Genomic...
Microsoft PowerPoint - [L...
Bioinformatik - Netsca...
3:38 PM

Celera - Academic Products - Netscape

File Edit View Go Window Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Netsite: http://www.celera.com/genomics/academic/home.cfm What's Related

Weather GroupWise GWU GroupWise ANS Die Welt Online Tagesschau Hotmail Google

CELERA
an Applera Corporation Business

Discovery can't wait.™

Home Online Business Therapeutic Discovery Corporate Info Investor Press Careers -- Quick links --

Academic & Not-for-Profit Offerings

From Data to Discovery

- **Overview**
- Celera Discovery System™
- Ordering Information
- Publication Library
- Login
- Product & User Literature
- Contact & Request Info

Overview

The foundation of Celera's product offering is the Celera Discovery System™, an integrated web-based discovery platform. The unified platform provides online access to the most comprehensive and up-to-date set of genomic and biological data available, along with computational tools and super-computing power. Data integration and management with the Celera Discovery System saves time and reduces costs when trying to set up bench experiments or sifting through enormous amounts to data, thus allowing researchers to accelerate their publication process.

Using Celera's genomic expertise combined with integrated biological information and query and analysis tools, researchers can discover novel genes, predict their function and enhance overall understanding of important biological mechanisms and diseases. Celera's staff of experienced scientists has developed a comprehensive platform to allow researchers to extensively understand how genes and their related proteins function in biological context.

- Fine Mapping & Genotyping
  Celera SNP Reference Database
- Protein Classification
  Celera's Proprietary Protein Families and Subfamilies using Panther™
- Genome Navigation
  Celera's Assembled and Annotated Human and Mouse Genomes

Document: Done

BioinformaticsOnline.org > The official Web site of Bioinformatics: Sequence and Genome Analysis, by - Netscape

File Edit View Go Window Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: http://www.bioinformaticsonline.org/ What's Related

Weather GroupWise GWU GroupWise ANS Die Welt Online Tagesschau Hotmail Google

Bioinformatics

Sequence and Genome Analysis

NOT A MEMBER? TRY THESE LINKS:

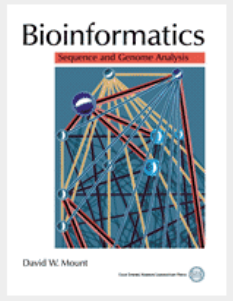
Registration
Privacy Policy
Chapters
Purchase the book

MEMBERS LOG IN HERE:

E-mail Password

CLICK HERE TO LOG IN

BOOK COVER



CLICK TO ENLARGE

Overview

The application of computational methods to DNA and protein science is the newest and most exciting development in biology. *Bioinformatics: Sequence and Genome Analysis* by David W. Mount is a comprehensive functional and theoretical introduction to this new discipline. Sequence alignment, structure prediction, phylogenetics and gene prediction, database searching, and genome analysis are amply explained and illustrated in the printed text. Based on the author's extensive teaching experience at the University of Arizona, this is a uniquely educational book, ideal for investigators, graduate students, and biology undergraduates studying this fast-changing discipline.

This Web site provides a link to on-line resources for the book. Tables explaining numerous Web sites for software and programming are presented here in hyperlinked form. Additional material not in the printed text is provided, and problems for classroom use are also given. Further or new material will be added as the Web site develops and the field progresses.

About the Author

David W. Mount is Professor of Molecular and Cellular Biology, Professor of Biochemistry, and Professor of Ecology and Evolutionary Biology at the University of Arizona in Tucson. His broad-ranging research interests include sequence analysis and genomic databases, mechanisms of genetic recombination and radiation resistance in plants, as well as conservation genetics, molecular ecology, and molecular evolution. His book *Bioinformatics: Sequence and Genome Analysis* is based on a course he teaches at the University of Arizona.

About This Web Site

The Bioinformatics Web site augments the content of *Bioinformatics: Sequence and Genome Analysis*. Purchasers of the book can:

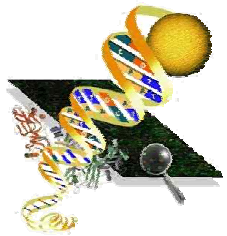
- access Web sites mentioned in the text from hyperlinked versions of the relevant tables
- access problem sets for classroom use and other useful material not included in the print edition
- access additional material to complement the print edition
- receive announcements about updates and additions
- receive E-mail alerts about peer-reviewed, new, and updated material that extends the scope and content of the book

© 2001 by Cold Spring Harbor Laboratory Press. All rights reserved.

No part of these pages, either text or image, may be used for any purpose other than personal use. Therefore, reproduction, modification, storage in a retrieval system, or retransmission, in any form or by any means, electronic, mechanical, or otherwise, for reasons other than personal use, is strictly prohibited without prior written permission.

Document: Done

Bioinformatics sources



Bioinformatics glossaries

College of Staten Island
http://www.library.csi.cuny.edu/~davis/Bioinfo_326/bioinfo_glossary.html

Incyte Genomics
http://www.incyte.com/glossary/index.shtml

www.SequenceAnalysis.com
http://www.sequenceanalysis.com/glossary.html

BIOINFORMATICS TEACHING & LEARNING
http://www.bscbioinformatics.com/Stu/Glo/glossary.html

CHI (Cambridge Healthtech Institute)
http://www.genomicglossaries.com/content/Bioinformatics_gloss.asp

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

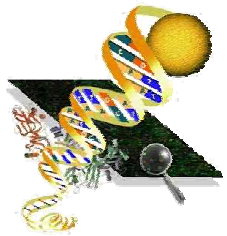
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Bioinformatics careers



Science careers @

August 2002 (476 positions advertised)

Virology	12
Pharmacology	43
Genetics	81
Biochemistry	103
Proteomics	38
Genomics	64
Bioinformatics	30

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Bioinformatics careers

Science: 5 June 2002

"Career opportunities in bioinformatics are very, very good … it seems that every time you turn around a company has decided to set up a bioinformatics group …"

M. Greene, Gene Logic Inc., Gaithersburg

Companies will … look for individuals who first and foremost are biologists but have key computational skills.

Those skills are:
- knowledge of UNIX and relational databases, skill with Structured Query Language (SQL) and programming skills (C, Perl, Java)
- Expert knowledge of sequence-analysis programs like BLAST and FASTA
- Web skills, e.g. Hypertext Markup Language (HTML)

"Recruiters get excited over applicants who have applied computational skills in a practical way"

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

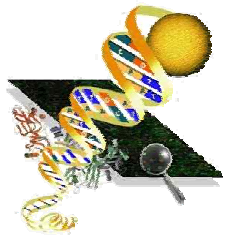
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Bioinformatics careers



What types of jobs are available in bioinformatics?

- Jobs are available from programmers and data analysts to senior level scientists and research directors
- Employment is available with private and public industries, research institutions, government institutions, and universities around the globe.

Online bioinformatic employment resources:

- Science magazine career section
- Science jobs in Nature magazine
- SmithKline Beecham - a leading employer of bioinformatic professionals
- Bioinformatics jobs listed by PlanetJobs
- Genome Jobs: resource for employment in genomics, bioinformatics, biotechnology and biocomputing.
- BiotechFind.com - a directory of international links covering the fields of Biotechnologies
- BioSpace Career Center

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

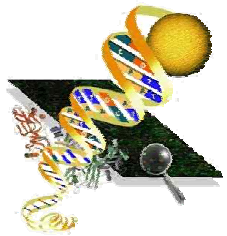
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

Bioinformatics careers



Planning a career

- Define your goals and objectives, start planning your career today!

- Talk to us about your career plans

- Get in contact with guest lecturers

- watch the market requirements

- assess your strengths and weaknesses

- take additional courses if necessary

- publish a paper(s)

- refine your collaborative skills, establish study groups

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

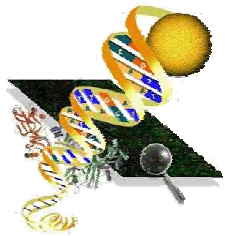
Course outline

Bioinformatics sources

Bioinformatics careers

Examples

The future of Bioinformatics



- in the future, bioinformatics is likely to become more central to the way biology is done

- as we enter the "post-genomic era"', information about gene expression, protein structure and function, data from DNA array technology, as well as epidemiological and disease susceptibility data, are all being integrated with genome sequence information.

- "When graduate students approach me these days about what is an interesting area to go into if you want to make a major contribution to biomedical research, the first thing out of my mouth is bioinformatics … we are woefully short in terms of having a critical mass of people who understand both biology and computational approaches"

Francis Collins, director of the National Human Genome Research Institute

Introduction

The Polio problem

What is Bioinformatics?

History of Bioinformatics

Course outline

Bioinformatics sources

Bioinformatics careers

Examples