

Introduction to Bioinformatics
BiMM 140

Lec 1: Introduction to Bioinformatics

What is bioinformatics?

- **Bioinformatics is the use of databases and computers to ask and answer biological questions.**
- **Bioinformatics thus comprises two main activities:**
 - _ computerized annotation of genomic and biological information and data (databases).
 - _ Transformation and manipulation of the data (software tools).
- **Overall Aim of Bioinformatics:**
 - _ Provide biologically important predictions from annotated data and transformation / manipulation of these data.

Lec 1: Introduction to Bioinformatics

Need for and Use of Bioinformatics

- **Bioinformatics plays a key role in modern biology and is especially important in:**

- _Molecular biology
- _Genomics
- _Functional genomics
- _Systems biology
- _Protein design and engineering
- _Pharmaceutical development
- _Medicine
- _Ecology / population genetics

Lec 1: Introduction to Bioinformatics

Need for and Use of Bioinformatics

- **Keywords from Some Recent Review Articles:**

- _ Finding genes, locating coding regions, predicting function
- _ Function, Evolution, Sequence, Structure (FESS relationships)
- _ Metabolic genotype, phenotype, redundancy
- _ Genes to Pathways; Genes to Biological Knowledge
- _ Proteomics: ‘Proteome’ of an Organism
- _ Assigning Gene Sets to different Species: Homologs vs Paralogs
- _ Expression profiles, relation to Metabolic Pathways / Genetic Networks
 - Experimentally Analyse Thousands of Genes simultaneously
- _ Gene Synteny between Species: Gene Adjacency in Genomes
- _ Polymorphisms, Haplotypes, Propensity for Genetic Disease

Lec 1: Introduction to Bioinformatics

Bioinformatics is New, Hot, and Growing:

- **Chronology of Review Articles via PubMed Search**

- _ 731 for 'bioinformatics AND review'
- _ (91 for 'bioinformatics AND review AND 2003')
- _ 372 for 'bioinformatics AND review AND 2002'
- _ 257 for 'bioinformatics AND review AND 2001'
- _ 144 for 'bioinformatics AND review AND 2000'
- _ 67 for 'bioinformatics AND review AND 1999'
- _ 52 for 'bioinformatics AND review AND 1998'
- _ 18 for 'bioinformatics AND review AND 1997'
- _ 13 for 'bioinformatics AND review AND 1996'
- _ 6 for 'bioinformatics AND review AND 1995'
- _ 7 for 'bioinformatics AND review AND 1994'
- _ 1 for 'bioinformatics AND review AND 1993'
- _ 2 for 'bioinformatics AND review AND 1992'
- _ 0 for 'bioinformatics AND review AND 1991'

Lec 1: Introduction to Bioinformatics

Topics we will cover in this course

- **Information Resources**
- **Homology based prediction of function**
- **Sequence database searching**
- **Gene modeling**
- **RNA secondary structure prediction**
- **Phylogenetic trees**
- **Protein structure prediction**
- **Protein homology modeling**
- **Functional genomics & proteomics**

Lec 1: Introduction to Bioinformatics

Expectations and Level of Course

- **Focus of the course is on computational methods, how they work, and how one evaluates the results of various analyses.**
- **We do not deal with formal descriptions of algorithms, complexity analysis or programming – although qualitative descriptions of algorithms are a central part of the course.**
- **Some mathematical background is necessary, although this seldom if ever reaches beyond simple algebra.**
- **Basic understanding of molecular biology is assumed, e.g., the composition of DNA and proteins, replication, transcription, and translation, protein folding, etc.**

Lec 1: Introduction to Bioinformatics

Basic Course Information

- **Michael Gribskov**
156 Chemistry Research Building
gribskov@sdsc.edu 534.8312
Office hours by arrangement
- **Doug Smith**
5254 Muir Biology Building
dsmith@ucsd.edu
Office hours by arrangement
- **TA: Aaron Parker**
asparker@ucsd.edu
- **Lecture:**
2207 Warren Lecture Hall ... MWF 1:00 – 1:50 pm
- **Discussion:**
2207 Warren Lecture Hall ... F 10:00 - 10:50 am

Lec 1: Introduction to Bioinformatics

Basic Information

- Course Webpage:

<http://www.sdsc.edu/~gribskov/bimm140>

- Required Textbook

Bioinformatics, David W. Mount, Cold Spring Harbor Press

Available in bookstore

Two copies on reserve in Biomed Library

textbook website: *<http://www.bioinformaticsonline.org/>*

- Grading

Quizzes 40% of grade

4 such, in Discussion Section, each 10% of course, each covers 2 week period

Final 60% of grade

Covers entire course

Exams cover assigned readings, supplementary readings and lectures

Lec 1: Introduction to Bioinformatics

Other Books

• Non-mathematical

- _ "Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins." Ed., Andreas D. Baxevanis and B. F. Francis Ouellette. 2nd Edition. Wiley InterScience Press, 2001.
- _ "Bioinformatics: Sequence, Structure, and Databanks - A Practical Approach." Ed., Des Higgins and Willie Taylor. Oxford University Press. 2000.
- _ "Introduction to Computational Molecular Biology." Joao Meidanis and Joao Carlos Setubal. PWS Publishing, Boston, 1997.

• Moderately Mathematical

- _ "Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids." Richard Durbin, S. Eddy, A. Krogh, and G. Mitchison. Cambridge University Press. 1999.
- _ "Bioinformatics: The Machine Learning Approach." Pierre Baldi and Soren Brunak. MIT Press. 1998.
- _ "Statistical Methods in Bioinformatics: An Introduction." Warren Ewens and Gregory Grant. Springer Verlag. 2001.

• Highly Mathematical

- _ "Computational Molecular Biology: An Algorithmic Approach." Pavel Pevzner. MIT Press. 2000.
- _ "Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology." Dan Gusfield. Cambridge University Press. 1997.
- _ "Introduction to Computational Biology: Maps, Sequences and Genomes", Michael S. Waterman. Chapman and Hal/CRC.

• Programming oriented

- _ "Beginning Perl for Bioinformatics", James D. Tisdall, O'Reilly and Associates, 2001.
- _ "Developing Bioinformatics Computer Skills." Cynthia Cibas and Per Jambeck. O'Reilly and Associates. 2001.

Lec 1: Introduction to Bioinformatics

Tools sites

- **textbook website - <http://www.bioinformaticsonline.org/>**
 - _ Has useful links to material discussed in text
 - _ Must purchase textbook to get password

- **As we discuss methods in the course, we suggest that you try some of them. We will point these out as we go and make suggestions for where you can try some of the applications. The following are large sites with many links to servers and software:**
 - _ CMS Molecular Biology Resource - <http://restools.sdsc.edu/>
 - _ Baylor College of Medicine - <http://searchlauncher.bcm.tmc.edu/>
 - _ European Bioinformatics Institute - <http://www.ebi.ac.uk/>
 - _ National Center for Biotechnology Information – <http://ncbi.nlm.nih.gov>

Lec 1: Introduction to Bioinformatics

2 April 2002: Resources, Sequence Databases, Finding Information I

- **Database Issues**
- **Kinds of Information Resources**
 - _ NCBI
 - _ TIGR
 - _ PDB
- **Basic Database Principles**