

CAP 5510: Introduction to Bioinformatics
CGS 5166: Bioinformatics Tools

Giri Narasimhan

ECS 389; Phone: x3748

giri@cis.fiu.edu

www.cis.fiu.edu/~giri/teach/BioinfS06.html

Overview of Courses

- Sequence Alignment: Multiple Sequence Alignment
- Sequence Analysis
- Sequencing and Mapping
- Phylogenetic Analysis
- Gene prediction techniques
- Pattern discovery techniques
- Protein structure alignment and analysis
- Genomics, Functional Genomics, Proteomics
- Gene Expression Data Analysis
- RNA Secondary structure
- RNA interference and small RNA
- Databases & Software Packages
- Statistics for Bioinformatics
- Computational Learning & Predictive Methods
- Biomedical Image Analysis
- Emerging Biotechnologies

Software Packages

- Databases (*GenBank*, *SwissPROT*)
- Programming Environments (*BioPerl*)
- Sequence Alignment (*BLAST*, *CLUSTALW*)
- Phylogenetic Analysis (*CLUSTALW*, *Phylip*, *PAML*)
- Learning Methods (*HMMPro*, *GeneCluster*, *ASOM*)
- Pattern Discovery Techniques (*GYM*, *TEIRESIAS*, *APRIORI*)
- Molecular Structure Analysis (*DALI*, *RASMOL*, *SPDBV*)
- Microarray Analysis (*CLUSTER*, *GeneCluster*, *TreeView*)
- Statistical Software Packages (*SAS*, *R*)

Evaluation

- Semester Project (50 %)
- Homework Assignments (20 %)
- Exams (25 %)
- Class Participation (5 %)

Course Homepage

www.cis.fiu.edu/~giri/teach/BioinfS05.html

- Lecture notes, required reading material, homework, announcements, etc.

1/10/06 CAP5510/CGS5166 4

Course Schedules

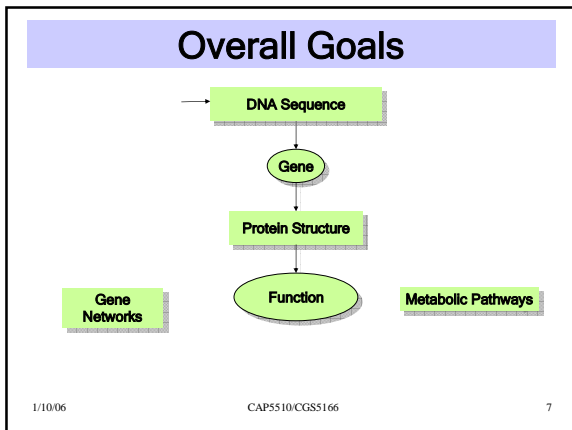
- CAP 5510 (3 credits) and CGS 5166 (2 credits) will meet every Tue from 2 PM to 4:45 PM. The first half of the lecture will be meant for both classes. The second half will be more focused on the CS majors
- Different exams and evaluation.
- Please attend all classes regardless of registered course.

1/10/06 CAP5510/CGS5166 5

Introduction

1. What is Bioinformatics?
 - Analysis of biological data with computing & statistical tools.
2. The different aspects of Informatics?
 - Data Management (Database Technology, Internet Programming)
 - Analysis/Interpretation of Data (Data Mining, Modeling, Statistical Tools)
 - Development of Algorithms/ Data Structures
 - Visualization and Interface Design (HCI, Graphics)
3. How to assist biological research?
 - propose new models or correlations based on data from experiments
 - verify a proposed model using known data
 - propose new experiments based on model or analysis
 - use predicted information to narrow down search in a biological investigation

1/10/06 CAP5510/CGS5166 6



General Information

- GenBank Release 144 (Oct 2004) contains over 38 million sequence entries totaling over 43 Gb from 2,645 organisms [http://www.ncbi.nlm.nih.gov] (Storage: 147 GB uncompressed)
- Human Genome has ~3 billion bp with 32,000+ genes
- 213 complete microbial genomes sequenced (274 more in progress)
- 1000 Viral genomes (300bp - 300Kb) (1st 1978: Simian virus; 5Kb).
- 14 complete eukaryotic genomes sequenced (46 more in progress):
 Caenorhabditis elegans, Arabidopsis thaliana, Drosophila melanogaster, Saccharomyces cerevisiae,
- 83 organisms have assemblies and chromosomal maps including:
 Mus musculus, Homo sapiens, Danio rerio, Oryza sativa
- Swiss-Prot Release 45.5 (Jan 2005): 167089 entries; 6 million amino acids.

Release 48.7 (Dec 05): 204086 entries, 7.4 million amino acids

1/10/06 CAP5510/CGS5166 8

Genome Sizes

Organism	Size	Date	Est. # genes
<i>HIV type 1</i>	9.2 Kb	1997	9
<i>H. influenzae</i>	1.8 Mb	1995	1,740
<i>M. genitalium</i>	0.58 Mb	1998	525
<i>E. coli</i>	4.7 Mb	1997	4,000
<i>S. cerevisiae</i>	12.1 Mb	1996	6,034
<i>C. elegans</i>	97 Mb	1998	19,099
<i>A. thaliana</i>	100 Mb	2000	25,000
<i>D. melanogaster</i>	180 Mb	2000	13,061
<i>M. musculus</i>	3 Gb	2002	~30,000
<i>H. sapiens</i>	3 Gb	2001	32,000+

1/10/06 CAP5510/CGS5166 9
