Introduction to Bioinformatics
Biostatistics & Medical Informatics 576
Computer Sciences 576
Fall 2002

Mark Craven
Dept. of Biostatistics & Medical Informatics
Dept. of Computer Sciences
craven@biostat.wisc.edu
www.biostat.wisc.edu/bmi576/

BMI/CS 576: Bioinformatics

• Instructor: Prof. Mark Craven
  – craven@biostat.wisc.edu or
  – craven@cs.wisc.edu
• Office hours: 2:00-3:00 Wednesday, 2:30-3:30pm Thursday,
  or by appointment
  – room 6730, Medical Sciences Center
• Course home page: www.biostat.wisc.edu/bmi576/
• Course mailing list: TBA
Finding My Office

Course TA

- Wei Luo
  - luo@biostat.wisc.edu
  - 6749 Medical Sciences Center
    (across the hall from my office)
  - Office hours: 2:30-3:30 Tuesday, 1:00-2:00 Friday
Computing Resources for the Class

- UNIX workstations in Dept. of Biostatistics & Medical Informatics
  - no “lab”, must log in remotely
  - more details later
- CS department offers UNIX orientation sessions
  - 4:00pm in 1325 Computer Sciences
  - September 4, 5, 9, 10
- also can buy a copy of “CS 1000” (UNIX tutorial) at the DoIT Tech Store

Expected Background

- CS 367 (Intro to Data Structures) or equivalent
- statistics: good if you’ve had at least one course, but not required
- molecular biology: no knowledge assumed, but an interest in learning some basic molecular biology is mandatory
Course Emphases

• Understanding the types and sources of data available for computational biology.
• Understanding the important computational problems in molecular biology.
✓ Understanding the most significant & interesting algorithms.

Related Courses

• BMI/CS 776 (graduate level, 576 is a prerequisite)
• Biochemistry 711/712, “Sequence Analysis”, taught by Prof. Ann Palmenberg
• not-for-credit evening BioModules on “Sequence Analysis”, “Genetics Computing” and “Desktop Molecular Graphics” etc. (see Ann Palmenberg’s web pages)
• CS 731, “Advanced Artificial Intelligence with BiomedicalApplications”, taught by Prof. David Page
• Statistics 692, “Statistical Methods in Genomics”, taught by Prof. Nicole Perna and Dr. Bob Mau
Course Requirements

• homework assignments: ~40%
  – mostly programming (in Java)
  – computational experiments (e.g. measure the effect of varying parameter $x$ in algorithm $y$)
  – some written exercises
• midterm exam: ~25%
• final exam: ~35%

Course Readings

• articles from the primary literature (scientific journals, etc.)
Reading Assignment

• this week read:
  – Molecular Biology for Computer Scientists. L. Hunter
  – DOE Primer on Molecular Genetics
  – both are available from course web page
  – additionally the web page lists two recommended readings

What is Bioinformatics

• representation/storage/retrieval/analysis of biological data concerning
  – sequences
  – structures
  – functions
  – activity levels
  – networks of interactions
  of/among biomolecules
• sometimes used synonymously with computational biology
  or computational molecular biology
Topics to be Covered: Computational Problems in Molecular Biology

- pairwise sequence alignment
- sequence database searching
- multiple sequence alignment
- gene modeling and recognition
- “signal” modeling and recognition
- protein structure and function prediction
- gene expression analysis
- phylogenetic tree construction

Topics to be Covered: Computer Science Issues & Algorithms

- string algorithms
- dynamic programming
- machine learning
- Markov chain models
- hidden Markov models
- EM algorithms
- clustering
- tree algorithms
- and more…
What do two sequences/genomes have in common?

- string algorithms
- dynamic programming

Where are the genes in this genome?

- Markov chain models
- hidden Markov models
Can diseases be characterized by patterns of gene activity?

- clustering
- machine learning

What does the protein encoded by a given gene look like? What does it do?

- dynamic programming
- branch & bound
- hidden Markov models
- Tarot cards?
How are these species related?

- tree inference
- search methods

Student Survey

- tell me the following
  - name
  - taking course for credit or sitting in
  - grad/undergrad and year
  - major/home department
  - CS background
  - biology background
  - statistics background
  - why you are taking this course