Genes and Blue Genes

I400: Introduction to Bioinformatics

Fall 2004
Basic Information

Class meets:
  Time: MW 11:15am – 12:45pm
  Place: Informatics 001

Instructor:
  Predrag Radivojac
  Office: Eigenmann 1005
  Email: predrag@indiana.edu
  Web: http://www.informatics.indiana.edu/predrag

Office Hours:
  Time: MW 1-2pm or by appointment
  Place: Informatics Building, Eigenmann 1005

Course Web Site:
What is a Gene?

- genes are material things that parents pass to offspring during reproduction; these things encode information essential for the construction and regulation of polypeptides, proteins and other molecules essential for the growth and functioning of the organism (Wickipedia)

- genes are sequences of nucleotides within nucleic acid molecules each of which determines the primary structure of some protein or polypeptide molecule (OED)
What is a Blue Gene?

• In December 1999, IBM announced the start of a five-year effort to build a massively parallel computer, to be applied to the study of biomolecular phenomena such as protein folding.

• The project has two main goals: to advance our understanding of the mechanisms behind protein folding via large-scale simulation, and to explore novel ideas in massively parallel machine architecture and software.

• Its processing power will be about 1 peta flops

• Blue Gene is due in 2006
Information

Probably first to use the word:
- 1387 - John de Trevisa
- 1390 - John Gower
- 1386 - Geoffrey Chaucer

Old meanings:
- Formation or moulding of the mind or character, training, instruction, teaching; communication of instructive knowledge
- An item of training; an instruction

Example of usage:
- “Whanne Melibee hadde herd the grete skiles and resons of Dame Prudence, and hire wise informacions and techynges”
- Divine instruction, inspiration

Modern meaning:
- Knowledge communicated concerning some particular fact, subject, or event; that of which one is apprised or told; intelligence, news. spec. contrasted with data.

(OED)
Informatics

• (1967) Informatics is the discipline of science which investigates the structure and properties (not specific content) of scientific information, as well as the regularities of scientific information activity, its theory, history, methodology and organization (OED)

• Example of usage:
  – 1973 Times Lit. Suppl. 28 Sept. 1133/1 The problem falls into two parts: the preparation of decisions, which is a matter of informatics, and the making of the decisions themselves, which is a matter of ‘politics’. (OED)

• We can also use this ‘definition’: Informatics is a scientific and practical discipline associated with the use of technology in knowledge extraction and solving real-life problems (in areas such as biology, chemistry, social sciences, etc.)

• Origin: information + “tic” which means theory in Greek
Bioinformatics

- The science of information and information flow in biological systems, esp. of the use of computational methods in genetics and genomics. (OED)

- First documented use of the word bioinformatics:
  - 1978 Paulien Hogeweg wrote about her field of study: “Since 1970 she has been a staff member at the Subfaculty of Biology of the University of Utrecht, with her main field of research in bioinformatics” (Simulation)

- **A simplified view:**
  - bioinformatics involves the use of technology in order to ask biological questions. We may say “biology”, but we actually mean “molecular biology”.

- **The ultimate goal of the field** is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discerned (NCBI)
Computational Biology, Biocomputing?

What is Computational Biology?
• The actual process of analyzing and interpreting data is referred to as computational biology (NCBI)

What is Biocomputing?
• The application of computing in biological research, esp. the analysis of statistical data and molecular structures, and the modeling of biological processes (1977 Science) (OED)

• For us, the terms computational biology and biocomputing will have identical meanings and will be more associated with algorithms used to discover information. Bioinformatics will be associated with knowledge extraction and interpretation of data.
Goal of the Course

- Bioinformatics is a practical scientific discipline

- This course is designed to introduce the fundamental concepts, research topics, and applications of bioinformatics

- Along with the core concepts, the course will provide information on types and sources of biological data and practical training with selected bioinformatics tools.

Some major applications

- Drug design
- Evolutionary studies
- Genome characterization
Overview of the Course

Textbook information:

Required:
• Fundamental Concepts of Bioinformatics - by Dan E. Krane and Michael L. Raymer

Recommended:
• Introduction to Protein Structure - by Carl-Ivar Branden and John Tooze
• Bioinformatics: Sequence and Genome Analysis - by David W. Mount

Supplementary material will be provided in class.
Overview of the Course

- fundamental biochemical principles and introduction to molecular biology
- biological sequences: proteins, RNA, DNA
- biological databases, overview and the importance
- pairwise and multiple sequence alignment
- searching biological databases
- phylogenetic analysis
- introduction to statistics and machine learning
- prediction of protein structure and function
- prediction of RNA structure
- genome analysis
- new directions of bioinformatics research
- applications of bioinformatics research (e.g. pharmacogenomics)
Guest Presentations and Research

- Scientists from Eli Lilly
  - Internship and employment information
- Scientists from Indiana University School of Medicine
- Scientists from Indiana University School of Informatics

- Research opportunities?
  - come see me
Grading Policy

• Midterm exam: 20%
• Final exam: 20% (will be a project!!!)
• Class presentations: 20%
• Homework assignments: 30%
• Class participation: 10%

• Midterm exam will be held in mid October
• Final exam will be as scheduled by the School of Informatics
Late Assignment Policy and Academic Honesty

• The homework assignments are due in class, on the specified due date

• No late assignments will be accepted unless there are legitimate circumstances

• All assignments are individual

• All the sources used for problem solution must be acknowledged (people, web sites, books, etc.)