Bioinformatics Algorithms
CS 579, 3 Credits, Fall 2015
Tues, Thurs 12:00 - 1:15 P.M., Lawson 1106
Prof. Alex Pothen
Dept. of Computer Science
apothen at purdue dot edu

This course is intended to be a first course in Bioinformatics algorithms for graduate students in computer science, computer engineering, mathematics, statistics, computational life sciences, and related fields. This is suitable as a foundational course for students planning to do research in bioinformatics with a focus on algorithms; it is also suitable for those who wish to acquire some breadth during graduate study and learn about important recent developments in bioinformatics. Computer science students will be able to take a qualifying exam in Bioinformatics in this course.

New high through-put technologies make it possible to study the genome (all the genes of an organism) and the proteome (all of the proteins in an organism), making biology an information-rich science. New technologies such as RNA-Seq measure quantitatively gene expression, and metagenomics is able to identify genes from multiple organisms present in an environment. Computational algorithms and software tools are critical to process the massive, error-plagued, data to understand the processes of life at the molecular level. This increased understanding advances life sciences, and has practical benefits too: personalized medicine (the design of drugs sensitive to one’s genetic make-up), the engineering of microorganisms for industrial uses, an understanding of how all organisms are related, etc. Bioinformatics or computational biology is the moniker for this area of research, and it involves the development of string algorithms, combinatorial algorithms, databases, statistical methods, high performance computing, and software.

A recent National Science Foundation study on Simulations based Engineering and Science (SBES) states that the shortage of trained students is the biggest bottleneck to progress in bioinformatics and computational biology. Students who have a working knowledge of bioinformatics concepts should find that their employment prospects are improved by their skills in bioinformatics. There are many excellent opportunities in academia, industry, and the national labs available to students who brave the challenge.

Students from computer science, the engineering disciplines, and mathematical sciences may be concerned about how much previous exposure to biological sciences they need to study bioinformatics. For this introductory course the answer is, not much—students do not need any previous knowledge of biology or bioinformatics. We will begin with a quick review of some of the biological concepts needed for bioinformatics, and will learn more as needed. Pre-requisites include at least an undergraduate course in algorithms and some programming experience.

What about students from the biological sciences? The same pre-requisites apply; hence you would benefit from this course if you have the computational and mathematical knowledge corresponding to an undergraduate algorithms course. You have the option of being graded in the course based on home work and a project. In case of questions, please talk to me before registering for the course.
The course will provide an introduction to the basic techniques and algorithms in bioinformatics. The topics discussed will include:

- **Biological Sequences:** pairwise global alignments of genes and proteins; pairwise local alignment of genes and proteins; scoring matrices; multiple alignment of proteins
- **Database search for sequences**
  BLAST and variants
- **Phylogenetic Trees and networks:**
  distance based, parsimony, and maximum likelihood methods.
- **Whole Genome Alignment:** suffix trees and algorithms
- **Overview of Systems Biology and Biological Networks**
  Clustering and module discovery in biological networks. Network alignment
- **Algorithms for analyzing RNA-Seq data**
- **Metagenomics**

Other topics could be included based on the interests of participating students, but you need to let me know within the first four weeks of the course if you wish to see some other topics addressed.

There are two ways to take this course. Students who wish to take a qualifying exam (Qual-1) in Bioinformatics, will be graded on regular homework problems, programming assignments, and two in-class exams: a midterm exam and a final exam. Students who do not wish to take a Qual-1 exam will be graded on the basis of homework and a class project, for which they would need to write a report and and also present the results in class.

Students will find the first of the books listed below accessible for the material in the first half of the course. The first book focuses more on algorithms, but the writing style (especially the proofs) can be hard to follow in places. The second book is the basis for a Bioinformatics Algorithms course taught through Coursera by Compeau and Pevzner. This is a good source for self-study and for open-ended projects which introduce research problems. The third book is focused on students in the life sciences, and emphasizes the use of bioinformatics databases and software tools to solve problems and leads to projects.

Other lectures will be based on material from some of the books listed below and from the research literature. The last book is a brief introduction to the molecular biology background needed for a mathematical or computational scientist who wishes to read the research literature in the biological sciences.


