

Bioinformatics Models and Algorithms

BGSU Math 4000 / Math 5820

Fall 2009 – August 24 - December 18 – 10:30-11:20 MWF

[Craig Zirbel](mailto:zirbel@bgsu.edu) - Mathematics and Statistics - Bowling Green State University - 419 372 7466 - zirbel@bgsu.edu

This is a new course designed primarily for students in mathematics and/or computer science who are interested in learning bioinformatics. Bioinformatics refers to using mathematics, statistics, and computer science to better understand biological data of a discrete nature (as opposed to using differential equations to model blood flow, for example, which would be called mathematical biology). Bioinformatics is a rapidly growing and constantly changing area. There is every reason to think that there will be good job opportunities for students who can combine expertise in mathematics, statistics, and computer science with biology and/or chemistry. Bioinformatics has strong connections with genetics, molecular biology, and medicine.

The main idea of the course is to introduce bioinformatics to an audience which has a solid math or computer science background but little to no chemistry or biology background. There is a place in every bioinformatics research group for people with a solid math and computer science background. For example, a few years ago, I started working with an undergraduate CS major / math minor who had very little biology or chemistry background, but who is now doing well in a Ph.D. program in biochemistry. The course should also be good for students with a strong biology or chemistry background who also have an appetite for mathematics. Biology and chemistry majors and graduate students are welcome!

The course will introduce several problems concerning biological sequences, including sequence alignment using dynamic programming, hidden Markov models, models of sequence evolution, inferring evolutionary relationships, genome searching, and the like. The emphasis will be on probabilistic models, optimization problems related to those models, and algorithms to solve those optimization problems, nearly all of which involve dynamic programming, which will be introduced as a topic of its own. We will work out dynamic programming algorithms, rather than simply turn to canned ones. This is not a course about how to use online databases, nor is it a course about physical modeling of molecules. It's more of a math course than anything. Students will code the algorithms we derive in the programming language of their choice.

I would be happy to have a wide variety of students in the course, and I'll work hard to accommodate students with a variety of backgrounds. Students should have solid background in at least one area. For example:

* mathematics, at least BGSU Math 3220 (Discrete Mathematics) or the equivalent, or

* computer science, at least three CS courses required for the CS major already completed. BGSU Math 2220 or 3220 (Discrete Mathematics) would also be good to have had, or

* biology or chemistry junior, senior, or graduate student, Calculus I completed, and an appetite for learning more mathematics

Please write to me if you're interested in the course but are not sure if you're prepared for it, or if you have any other questions.

You can preview the textbook, *An introduction to bioinformatics algorithms* by Jones and Pevzner, at books.google.com This book introduces bioinformatics optimization problems and dynamic programming. I will supplement with my own descriptions of probabilistic models for biological sequences, including HMM and SCFG, and for sequence evolution, including Markov processes. A good reference for this is the book *Biological Sequence Analysis* by Durbin et al, which can also be previewed at books.google.com.

The course will be taught in a distance-learning classroom at BGSU. We can explore the possibility of attending the class sessions in a similarly-equipped room at other schools.