Topics in Computational Biology and Biological Modeling

Course number: BIOL 571  
Schedule: Fall 2019, Wed 2:00pm  
Instructor: Junhyong Kim ([junhyong@sas.upenn.edu](mailto:junhyong@sas.upenn.edu); Lynch 304G)  
Office Hours: By appointment

Intended Audience: First year PHD students in computational biology

Pre-requisites: Calculus, Part I (Math 104), Statistics (BIOL 446 or equivalent), Intermediate level Biology (BIOL 121 + 200 level course);

Recommended: Calculus, Part II and III, knowledge of programming

Synopsys:

Biology is flooded with data that cannot be understood without computational analysis and modeling. For example, new instruments are sequencing the DNA and RNA at such a rate that less than 0.001% of the data will be ever seen by human eyes. Computational Biology is a growing field that develops algorithms, statistical analysis methods, and ultimately biological models for these kinds of data. That is, Computational Biology is a subfield of natural science where quantitative approaches are used to discover and understand biological and medical phenomena.

The goal of this course is to develop a deeper understanding of techniques and concepts used in Computational Biology. The course will strive to focus on a small set of approaches to gain both theoretical and practical understanding of the methods. We will aim to cover practical issues such as programming and the use of programs, as well as theoretical issues such as algorithm design, statistical data analysis, theory of algorithms and statistics. This course WILL NOT provide a broad survey of the field nor teach specific tools but focus on a deep understanding of a small set of topics.

Course Structure: The course will meet once a week. Readings from a draft textbook will be given out each week with reading assignments. Each week, we will discuss the reading materials and application of the concepts. Students will rotate presenting the materials covered in the textbook each week. In the last four weeks of the course, we will examine real research problems and how to approach the problems with computational solutions.

Topics to be discussed (not in order):

Grades:

Grades will be compiled from weekly presentations (30%) and a final project (70%).

Recommended Reading

There are no required textbooks for this course. You must read all posted material that is marked for reading. The following is a list of books you may use to complement the lecture notes. Lecture notes will be posted weekly.

An introduction to bioinformatics algorithms (Computational Molecular Biology)
N.C. Jones and P.A. Pevzner

A Primer of Genome Science
Greg Gibson and Spencer Muse, North Carolina State University

Other books of interest:

Machine learning approaches to bioinformatics, Zheng-rong Yang, World Scientific (a fairly comprehensive survey of machine learning approaches)

Algorithms on strings, trees, and sequences: computer science and computational biology, Dan Gusfield, Cambridge Univ Press. (A very comprehensive algorithms book with easy to read style)

Computational Molecular Biology: An Algorithmic Approach
Pavel A Pevzner (Somewhat hard core computational biology)

Molecular Biology of the Cell, Alberts et al. (Again standard textbook—long)

Recombinant DNA, Watson et al., W.H. Freeman. (Quite outdated but still a great introduction to molecular biology)

Statistical methods in bioinformatics: An introduction, Grant and Ewens, Springer-Verlag. (This is one of the few statistically focused books on bioinformatics. And, you can directly talk to the author!)