Department: BME, CSE & MCB

Course No.: 1401

Credits: 3

Title: Honors Core: Computational Molecular Biology

Contact: David Knecht

Content Area: CA 3 Science and Technology

Catalog Copy: MCB 1401. Honors Core: Computational Molecular Biology (120). (Also offered as BME 1401 and CSE 1401 and PNB 1401). Three credits.
Introduction to research in computational biology through lectures, computer lab exercises, and mentored research projects. Topics include gene and genome structure, gene regulation, mechanisms of inheritance, biological databases, sequence alignment, motif finding, human genetics, forensic genetics, stem cell development, comparative genomics, early evolution, and modeling complex systems.
CA 3.

Justification: The above course was specifically designed in response to a curriculum development solicitation by the Honors program to address the need for early undergraduate research experiences. The course complements existing course offerings in computational biology and bioinformatics within the CSE and MCB departments (as well as the BME program) by providing a rigorous yet accessible introduction to the area of computational genomics. The course will be offered on the Storrs campus with the support of the Honors program. Two faculty (Mandoiu and Nelson) will each teach one section of the course per semester for the 2007-08 and 2008-09 academic years, in addition to their regular teaching load.

Course Information : This course is an introduction to computational genomics through lectures, computer lab exercises, and mentored research projects. Started in 1995 by the completion of the first genome sequence of a free-living organism, H. influenza, the genomic era has already led to hundreds of complete genome sequences deposited in public databases and many more genome projects at various stages of completion. The huge amounts of available genome data are revolutionizing biomedical research, but fully exploiting them requires powerful computational and statistical methods.

The main objective of the course is to provide students with a general understanding of the field of computational genomics, including current problems and research. Students will become familiar with fundamental molecular biology concepts and computational techniques, and will learn how to use the Matlab bioinformatics toolbox for solving problems in genomics. Grading is based on in-class quizzes (1/3), computer lab exercises (1/3), and team final projects (1/3).

a. Course goals and objectives: The main objective of the course is to provide students with a general understanding of the field of computational genomics, including current problems and research.

Students will become familiar with fundamental molecular biology concepts and computational techniques, and will learn how to use the Matlab bioinformatics toolbox for solving problems in genomics.

b. Course requirements: Grading is based on in-class quizzes, computer lab exercises, and team final projects, with each of the three components contributing equally to the final grade. Weekly in-class quizzes are used to test student understanding of the molecular biology and computational concepts covered in lectures and assigned reading material. Weekly discussion sections include hands-on computer activities requiring students to perform specific genomic data analyses using the Bioinformatics Toolbox of Matlab, with similar analyses assigned as take-home exercises.

Last five weeks of the semester are devoted to a final project done in teams of three students. For the project, students pick a computational genomics topic not covered in lectures or discussion sections and research it in more depth. Teams are required to give weekly progress reports, submit a written final report of 15-20 pages, and give a 15-minute presentation at the end of the semester.

c. List of topics

Molecular biology topics:

- Anatomy of a genome
- Structure and function of DNA, replication
- Transcription, translation, genetic code
- Anatomy of a gene: enhancers, promoters, UTR's, and ORF
- Homology, orthology, and paralogy
- Gene duplication and deletion
- Mutation & polymorphism
- Mitochondrial DNA & Human evolution
- Evolution and Natural Selection
- HIV & the immune system
- SARS and viral evolution
- Virus-host interactions
- Gene Expression
- Yeast cell cycle
- Circadian clock
- Mechanisms of gene regulation

Computer Science & Statistics Topics:

- Introduction to computer algorithms
- Probabilistic models and statistical sequence analysis
- Gene Finding
- Hypothesis Testing
- Sequence alignment

- Genetic distance, modeling sequence evolution
- Phylogenetic trees
- Quantifying natural selection
- Estimating Ka/Ks
- Structure and representation of phylogenetic trees
- Tree inference: distance matrices, neighbor joining
- Measuring gene expression with microarrays
- Data clustering
- Motif representation and scoring
- Motif finding

Meets Goals of Gen Ed:

The course addresses the following goals of General Education:

Become articulate: The course requires students complete a final project in which to research in more depth a computational genomics topic not covered in lectures. Teams are required to give weekly progress reports, submit a written final report of 15-20 pages, and give a 15-minute presentation to their colleagues at the end of the semester.

Acquire intellectual breadth and versatility: The course will help students develop a general understanding of the field of computational genomics. Students will become familiar with fundamental molecular biology and computational concepts.

Acquire awareness of their era and society: Students will become aware of ongoing research in the area of computational genomics, an area that is expected to significantly impact all life sciences in the 21st century.

Acquire a working understanding of the processes by which they can continue to acquire and use knowledge: The course will promote the development of lifelong learning skills by introducing students to genomic databases and other online educational resources.

CA3 Criteria:

1. Explore an area of science or technology by introducing students to a broad, coherent body of knowledge and contemporary scientific or technical methods. The course will introduce students to the fundamental molecular biology concepts and computational techniques used in the interdisciplinary area of computational genomics.

2. Promote an understanding of the nature of modern scientific inquiry, the process of investigation, and the interplay of data, hypotheses, and principles in the development and application of scientific knowledge. The course material is introduced using carefully selected case-studies relicating the main results from some classic research papers in the field of computational genomics. The case-studies approach provides a rigorous yet accessible introduction to the field for both biologically and computationally minded students. Furthermore, the case studies allow students to understand how abstract scientific ideas

are applied in concrete contexts, and help emphasizing the importance of interdisciplinary research in modern science.

3. Introduce students to unresolved questions in some area of science or technology and discuss how progress might be made in answering these questions. The objectives of the course include introducing students to is to ongoing research and open problems in computational genomics.

4. Promote interest, competence, and commitment to continued learning about contemporary science and technology and their impact upon the world and human society. The course will introduce students to an area of science that is undergoing rapid development and is likely to have great impact on many other disciplines. Students will become aware of the potential impacts of computational genomics, and will develop the technical vocabulary and background knowledge necessary to follow future developments in the field.

Supplementary Information: This course has been developed with support from curriculum development grant provided by the Honors program. Two faculty (Mandoiu and Nelson) will each teach one section of the course per semester for the 2007-08 and 2008-09 academic years, in addition to their regular teaching load.