

UMBC UGC New Course Request: BIOL 295 - Introduction to Bioinformatics and Computational Biology

Date Submitted: 10/08/2009

Proposed Effective Date: Spring 2010

	name	email	phone	dept
dept chair	Lasse Lindahl	lindahl@umbc.edu	X52996	Biology
contact	Ivan Erill	erill@umbc.edu	X52470	Biology

COURSE INFORMATION:

course number(s)	BIOL 295
formal title	Introduction to Bioinformatics and Computational Biology
transcript title (≤24c)	Bioinformatics Intro
prerequisite	MATH 151 pre-requisite
credits	3
max. repeat credits	0
grading method(s)	<input checked="" type="checkbox"/> Reg (A-F) <input type="checkbox"/> Audit <input type="checkbox"/> Pass-Fail

PROPOSED CATALOG DESCRIPTION:

This is an introductory course to the field of Bioinformatics and Computational Biology, an emerging interdisciplinary field dealing with the application of computing methods in biology. The course will provide a brief introduction to contemporary molecular biology and evolutionary theory. It will introduce the concept of bioinformatics, its applications and its career opportunities. Applying a question-answer approach, it will then explore some of the main problems faced by biology in the last decades, such as genome sequencing and annotation, sequence-based search, structure or function prediction, and some of the computational methods and answers provided by the emerging field of Bioinformatics. Lecture content will be complemented with paper discussion sessions and home-based lab assignments. No prior programming/biology skills required. The course is intended for a multi-disciplinary audience with an emphasis on BINF, CMSC, BIOL, MATH, STAT and IS majors. Pre-requisite: MATH 151 or equivalent.

RATIONALE FOR NEW COURSE:

- Why is there a need for this course at this time?** The reasons are manifold. Currently, Bioinformatics (BINF) majors do not get exposure to their subject field until their senior year (BIOL 495), thus preventing them to make clear connections among the different subjects taken in the intervening years. The proposed course will bridge this gap, exposing BINF students to bioinformatics in their early years, and will be a core course in the remodeled BINF curriculum. Basic skills in bioinformatics are also becoming increasingly required of graduating Biology (BIOL) majors. Even though some bioinformatics tools are already broached in different BIOL courses, BIOL students lack a course providing them a comprehensive view of bioinformatics, as well as exposure to web-based bioinformatics resources and their applications. Finally, the current BINF major draws mostly from BIOL majors, creating a marked imbalance with respect to similar undergraduate programs in other universities. The introduction of a new 200-level introductory course would allow prospective CMSC, IS, MATH or STAT majors to get acquainted with this emerging field, potentially leading to an increase of non-BIOL undergraduates in the BINF major.
- How often is the course likely to be taught?** The course will be taught in spring, yearly.
- How does this course fit in with your department's curriculum?** The course fits naturally within the remodeled BINF and BIOL curriculums (see above).

- d) **What primary student population will the course serve?** The primary target are BINF majors, since the course will become part of their core course list. However, the course intends to reach a campus-wide audience, focusing especially on BIOL, CMSC, IS, MATH and STAT freshman and sophomore students.
- e) **Why is the course offered at the level chosen?** The course is offered at the 200 level to allow reaching students from different majors and with different backgrounds. This is an introductory course that will provide basic bioinformatics notions to non-BINF majors and will serve as a stepping stone for prospective BINF majors.
- f) **Explain the appropriateness of the prerequisites.** Even though the course assumes no prior knowledge in CMSC or BIOL, students need to command significant quantitative and analytical skills in order to grasp the main algorithmic concepts in bioinformatics.
- g) **Explain the appropriateness of the P/F or regular grading method.** This is an introductory course and also a core course for BINF majors. It covers fundamental concepts in Biology, Computer Sciences and Bioinformatics. It is mostly lecture based and the regular grading method is appropriate.
- h) **Provide a justification for the repeatability of the course:** N/A

ATTACH COURSE OUTLINE:

Course outline

The course will cover 4 main topics, with Topic III taking up a third of the course. **Topic I** will introduce students to the main contemporary concepts behind molecular biology and evolution, in order to provide a common ground and lexicon that will allow students from highly different backgrounds to follow the remainder of the course. Students will be introduced to the molecular building blocks of life (DNA, RNA and proteins) and their interplay through the Genetic Code in the Central Dogma of molecular biology. Students will also learn the basics of evolution by natural selection, the interlinked concepts of genotype and phenotype and the dynamic concept of an adaptive landscape. Through the concept of orthology, dot-plots and other classical methods, students will be made aware of the relevance of evolution to Bioinformatics. **Topic II** will address the advent and coming of age of Bioinformatics and Computational Biology as independent research fields, as a consequence of the increasing amounts of data generated by modern high-throughput biology. It will analyze the challenges posed by high-throughput data, the reinforcement cycle of bioinformatics and biology and the applications of and career opportunities in Bioinformatics. **Topic III** will cover the foundational basis of Bioinformatics: sequence and genome analysis. It will slowly introduce students to the concept of sequence comparison and alignment, analyzing in depth the method of dynamic programming and its computational complexity. Genome assembly will be analyzed as a foundational topic of modern bioinformatics before proceeding to explore the problem of genome annotation, subdivided into gene finding and functional annotation by comparison. The students will learn how gene-finding algorithms exploit genome-wide statistics and signal search to locate search, as well as the basics behind Bioinformatics most famous algorithm (BLAST), the difference between similarity and homology and the relevance of statistics in assessing the reliability of results. **Topic IV** will cover several common advanced topics in Bioinformatics, but focusing more on the nature of the problem and the solutions proposed, rather than immersing students in the algorithmic detail of the methods studied. In this section, students will be exposed to the holy grail of bioinformatics, protein folding, and the basic ideas and limitations of conventional approaches to structure prediction. They will also extend their previous knowledge on sequence alignment to multiple sequences and how computational progressive methods cope with exponential complexity. Finally, students will be introduced to the concept of molecular phylogeny, how it can address biologically and socially important questions and how the basic methods approach the task of tree reconstruction.

Course schedule

Topic I – Introduction to molecular biology (2.5 weeks)

- Life and information
 - Definitions of life. Information, autopoiesis and evolution. Life requirements.
- Introduction to molecular biology
 - Cells, organisms and molecules. Molecular biology. Central dogma. The genetic code. Life's alphabet. Transcription and translation. Regulation. Regulation of transcription.
- Evolution
 - Basic aspects and requirements. Genotype and phenotype. The adaptive landscape. Relevance to bioinformatics.

Topic II – Introduction to bioinformatics (2 weeks)

- Origins and etymology. Raisons d'être.
- High-throughput methods
 - Old and new biology. High-throughput technologies: sequencing, arrays & hybrids.
- Biological databases.
- Goals and challenges of bioinformatics. The in-silico, in-vitro, in-vivo cycle. The -omics.
- Careers and applications.

Topic III – Sequence and genome analysis (4.5 weeks)

- Sequence analysis
 - Sequence comparison and the alignment problem.
 - Local and global alignment. Dynamic programming. Pair-wise alignment.
- Genome sequencing and genome assembly
 - Conventional sequencing. Shotgun sequencing. Genome assembly.
- Genome analysis
 - Genes and genomes. Genome statistics. Gene finding algorithms.
 - Genome annotation
 - Databases. Search by content. Homology and similarity. BLAST. E-values.

Topic IV – Advanced topics in bioinformatics (3 weeks)

- Protein folding
 - Structure and function. Protein folding and structure prediction algorithms.
- Multiple sequence alignment
 - Alignment and complexity. Exact and progressive methods. Tradeoffs, greediness and errors.
- Phylogeny
 - Ancestry and relatedness. Trees and cladograms. Phylogeny and its applications. Alignment and circularity. Neighbor-joining and parsimony.

Learning objectives

At the end of the semester, students should have:

- Acquired basic notions and vocabulary of molecular biology and evolutionary theory that should enable them to interact with other class members regardless of their declared majors, read related bibliography and attend related courses in the BIOL, CMSC and BINF majors.
- Gained an understanding of the problems faced by 21st century biology and how biology has changed with the advent of new experimental techniques and the increasing reliance on data processing. Specifically, students should be familiar with:
 - Genome sequencing and other high-throughput techniques
 - Genomic annotation
 - Sequence comparison

- Understood the basic computational methods developed to tackle these biological problems and be aware of the availability of alternative methods to address those same problems. Specifically, students are expected to understand the main concepts behind:
 - Sequence assembly
 - Pair-wise and multiple sequence alignment
 - Signal and content-based genomic analysis
- Grasped the multi-disciplinary nature of Bioinformatics and Computational Biology and the importance of interdisciplinary collaborations to address most of biology current problems. The students should also have become aware of:
 - The possible contributions from each of the students' fields
 - The different career options available for Bioinformatics majors and other majors with mixed training or Bioinformatics M.S. degrees.
- Learned to read and interpret a popular science paper in the context of Bioinformatics. In doing this, the students should have learned:
 - To analyze the structure and content of a paper
 - How to interact with students from other majors in discussions within and between groups
 - How to summarize the findings and basic ideas of a paper
- Mastered basic computing skills to complete the lab assignments. In particular, the students should have a basic understanding of:
 - How to navigate through bioinformatics web resources
 - How to use some of the studied bioinformatics methods to solve biological problems

Assessment of student knowledge

Student learning will be assessed continuously through the course. Students will be given 2-3 exams prior to the final, providing them with feedback on their assimilation of the covered material. They will also be queried with short quizzes at the end of paper discussion sessions and upon completion of their lab assignments.