

BIOL 4540/6540, Bioinformatics I, Sequence Analysis. FALL 2011

This course covers the theory and practice of biological sequence analysis, including algorithms for alignment, database searching, and phylogenetics.

Place: Jonsson-Rowland Science Center Rm 2C13 **Time:** Monday, Thursday 2:00-3:50

Instructor: Chris Bystroff **office:** J-Rowl 3C07 **phone:**x3185 **email:** bystrc@rpi.edu

Office Hours: tbd.

Web sites: <http://www.bioinfo.rpi.edu/bystrc/courses/biol4540.html>

REQUIRED TEXT: [Z&B]

Marketa Zvelebil & Jeremy Baum. "Understanding Bioinformatics"

Garland Science Textbooks, 2007. ISBN: 0-8153-4024-9

Other required readings will be posted on the website.

REQUIRED SOFTWARE: UGENE

Download UGENE (ugene.unipro.ru) and install. This is free software!

GRADING: Homework 30%[‡] Midterm exam 20%[•]

Term project 20%[∞] Final exam 20%[•]

In-class exercises, participation, attendance: 10%^{*}

• Both midterm and final have a written and a practical (computer) part. Exams are closed-book. A cheat-sheet is allowed (midterm 1p., final 2pp). Missed exams can be made up only if the absence is excused (see below). Exam grades may be contested in writing up to one week after receiving the graded exam.

**You are allowed one unexcused absence. After that, every unexcused absence will result in a grade deduction of 3 percentage points, to a maximum of 9 percentage points. New RPI attendance policy: excuses must be validated through the Student Experience Office (4th fl Academy Hall x8022, se@rpi.edu).*

‡Please turn in homework on paper at the beginning of class on the day due unless otherwise specified. Late homework will be accepted with a 10% penalty for each weekday late for up to 5 weekdays late. Thereafter, late homework is accepted until Nov 30 of class with a 50% penalty. Homework grades can be contested in writing up to one week after receiving the graded homework, but only if homework was turned in on time.

∞Term projects are usually done in teams of two. Outside reading and programming are required. Term projects consist of a written report and an oral presentation.

ACADEMIC DISHONESTY: See Academic Dishonesty in [The Rensselaer Handbook of Student Rights and Responsibilities](#). Any student committing an act of plagiarism will automatically receive an F for the course and the violation will be reported to the Dean of Students Office.

Learning objectives

1. Students will be able to critically assess the quality and significance of a protein or nucleic acid sequence alignment as demonstrated by answering questions on an exam.
2. Students will demonstrate a thorough understanding of algorithms for sequence alignment, sequence modeling and phylogenetic analysis by completing homework assignments.
3. Students will be able to communicate in the language of bioinformatics on the subject of molecular evolution as demonstrated by a written and oral presentation.
4. Students will be able to navigate databases and understand the input and output data of key bioinformatics algorithms by completing in-class exercises and by answering questions on an exam.

BIOL 4540/6410, Bioinformatics I, Sequence Analysis. FALL 2011

	<i>Date</i>	<i>Topic</i>	<i>Reading [Z&B]*</i>	<i>HW</i>
1	Mon 8/29	Thinking algorithmically.	1, 5.5	HW 1
2	Thu 9/1	Sequence databases. Sequencing. Dotplots. Motifs. Scoring.	4.1-4.3, 4.8-4.9 Ugene podcasts	
3	Thu 9/8	Pairwise sequence alignment with gaps. Dynamic programming.	4.4, 5.2	HW 1 due
4	Mon 9/12	Local versus global alignment. Multiple sequence alignment.	4.5	HW 2
5	Thu 9/15	Alignment scoring. Jukes-Cantor. BLOSUM, PAM.	5.1, 8.1	
6	Mon 9/19	Database searching. Indexing. FASTA and BLAST.	4.6-4.7, 5.3	HW 2 due
7	Thu 9/22	Statistical Significance. E-values.	5.4	HW 3
8	Mon 9/26	Multiple sequence alignment. Clustal, Muscle.	6.4-6.5	
9	Thu 9/29	Editing multiple sequence alignments. bAliBase.	14.2	HW 3 due
10	Mon 10/3	Profiles. Sequence weighting.	6.1, 6.6	
11	Thu 10/6	review		
12	Tues 10/11	MIDTERM		
13	Thu 10/13	special event		HW 4
14	Mon 10/17	Phylogenetic analysis. Orthologs and paralogs.	7.1-7.3	
15	Thu 10/20	Parsimony. Maximum likelihood.	8.2-8.4	HW 4 due
16	Mon 10/24	The Tree of life. Monophyly and Paraphyly.	8.5	HW 5
17	Thu 10/27	Markov chains. HMMs. Profile HMMs.	6.2-6.3	
18	Mon 10/31	Footprints. Microrepeats, Transposons. Language models.	6.6 , Box 9.5	HW 5 due
19	Thu 11/3	Secondary structure prediction. Neural nets. Support vector machines.	Ch. 11	HW 6
20	Mon 11/7	Gene design. Back translation.	11.9	
21	Thu 11/10	RNA secondary structure prediction and DNA hybridization. (special guest Michael Zuker)	3.1-3.4, 9.7	HW 6 due
22	Mon 11/14	Gene finding. Splicing.	9.1-9.5, 10.4-10.8	
23	Thu 11/17	Virtual cloning & other topics.	Ugene podcast #34	
24	Mon 11/21	Gene ontology. Sequence annotation.	9.6-9.7, p.54	
25	Mon 11/28	Single nucleotide polymorphisms.		all late HW due
26	Thu 12/1	Review. Work on projects.		
27	Mon 12/5	Student Presentations		
28	Thu 12/8	Student Presentations		
29	TBA	Final Exam		