ENS 210 Fall 2017

Computational Biology

(October 14, 2017)

Instructor: Deniz Sezer
E-mail: dsezer@sabanciuniv.edu
Office: FENS G021
TA: Tandaç Furkan Güçlü

Lectures: Thu 8:40-10:30 FENS L056
Fri 12:40-1:30 FENS L063
Computer lab: Fri 1:40-3:30 FASS G050

Course Description:

This course offers a first exposure to DNA and protein sequences. Students will learn about aspects of the information contained in the genomes of organisms, will get familiar with databases making this genetic information available, and will develop programming and analytical skills to manipulate this information. (A detailed list of the course content is given on the next page.) Because the course is geared towards undergraduate students in their second year, previous exposure to biology is assumed to be on the level of NS 101-102. Writing computer code to perform the desired sequence analyses is going to be an essential part of the course. Therefore, **programming skills on the level of CS 201 are taken for granted**. Although we will use Python in the course, students are not expected to have used Python before.

Evaluation:

Recitations/homework/quizzes	25 %
Exam 1 (take-home): Nov 3 - 9	25 %
Exam 2 (take-home): Dec 22 - 28	25 %
Final exam (in-class): Dec ?	25 %

Textbooks:

p4b: Jones, *Python for Biologists: A programming course for complete beginners*
`http://pythonforbiologists.com`
EB: Forsdyke, *Evolutionary Bioinformatics*, 3rd ed., Springer 2015.

Detailed Course Content:
(On the next page.)

week	date	Lecture, Computer lab, `Python` (p4b Ch.)
		I. DNA and protein sequences
1	Sep 21	General information about the course. Sizes of cells and genomes.
	Sep 22	How many Bytes is a genome?
	Sep 22	Nucleotide content of 8 prokaryotic genomes. `Strings` (2). `Reading text from a file` (3a).
2	Sep 28	Chargaff's parity rule and DNA structure. Chargaff's second parity rule.
	Sep 29	Duplets, triplets, and probabilities (EB Chs. 2 & 4).
	Sep 29	Content of nucleotide duplets in genomes ("genomic signature"). `Lists and loops` (4). `Functions` (5). `Dictionaries` (8).
3	Oct 5	Random numbers with Python. Probabilities.
	Oct 6	Conditional probabilities.
	Oct 6	Generating random DNA with given dinucleotide composition. `Conditional tests` (6). `Writing to a file` (3b).
4	Oct 12	Dinucleotide correlations over long separations. The genetic code.
	Oct 13	How long is a "typical" protein?
	Oct 13	Finding candidate protein genes in a genome. Making plots using `matplotlib.pyplot`.
5	Oct 19	Databases of DNA and protein sequences.
	Oct 20	Processing a GenBank file.
	Oct 20	Ribosomal- and transfer-RNA genes in genomes. `Regular expressions` (7).
6	Oct 26	Querying a database.
	Oct 27	BioPython.
	Oct 27	Accessing ALL whole-genome sequences of prokaryotes
7	Nov 2	Codon usage bias
	Nov 3	**Midterm Exam 1** (take-home) Nov 3 - Nov 9
	Nov 3	`No lab this week`
		II. Comparing two sequences
8	Nov 9	Introduction to DNA sequence alignment
	Nov 10	
	Nov 10	?
9	Nov 16	Dynamic programming for sequence alignment
	Nov 17	Semiglobal and local alignment
		My first sequence alignment code
10	Nov 23	Matrix models of base substitution
	Nov 24	Phylogenetic distances and maximum likelihood
	Nov 24	Evolutionary distance between gorilla and orangutan
11	Nov 30	Aligning protein sequences
	Dec 1	Amino acid substitution matrices PAM, BLOSUM
	Dec 1	Evolutionary distances from protein sequences
12	Dec 7	Phylogenetic trees
	Dec 8	Constructing an UPGMA tree
	Dec 8	My first phylogenetic tree code
13	Dec 14	Neighbour joining
	Dec 15	The tree of life
	Dec 15	Phylogenetic trees with neighbour joining
14	Dec 21	Rates of synonymous and nonsynonymous substitutions
	Dec 22	**Midterm Exam 2** (take-home) Dec 22 - Dec 28
	Dec 22	`No lab this week`
	Dec ?	**Final Exam** (in-class)