ENS 210 Fall 2017

Computational Biology

(December 16, 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 12:40-13:30 FENS L063
Computer lab: Fri 13:40-15:30 FASS G050

Course Description:

Are you curious whether your ancestors came form central Asia, the Arab peninsula, northern Africa or southern Europe? The company 23andMe provides an answer for $99 and a drop of your saliva. Miraculously, all the necessary information is in your DNA.
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:

Labs/pre-labs/quizzes:	25 %
Exam 1 (take-home): Nov 2 - 9	25 %
Exam 2 (take-home): Dec 14 - 22	25 %
Final exam (in-class): Dec 27	25 %

Python textbook:

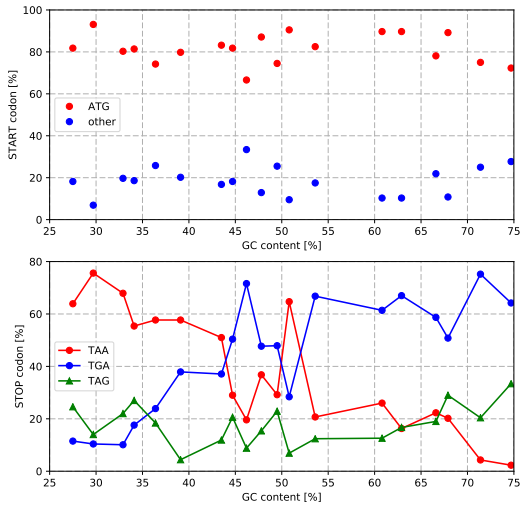
p4b: Martin Jones, *Python for Biologists*, `http://pythonforbiologists.com`

Supporting reading material:

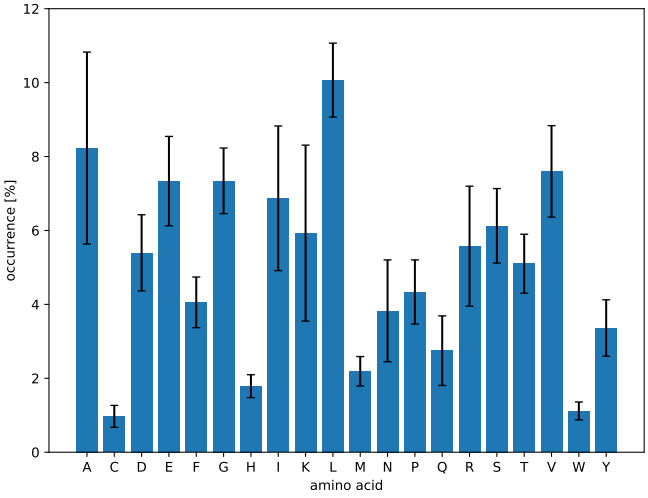
- Donald Forsdyke, *Evolutionary Bioinformatics*, 3rd ed., Springer 2015. (Chs: 2, 4, 7, and 12.)
- Stephen D. Bentley and Julian Parkhill, Comparative Genomic Structure of Prokaryotes, *Annual Reviews of Genetics*, **38**:771-791 (2004).
- John Lightfield, Noah R. Fram, Bert Ely, Across Bacterial Phyla, Distantly-Related Genomes with Similar Genomic GC Content Have Similar Patterns of Amino Acid Usage, *PLoS ONE*, **6**:e17677 (2011).
- Karen E. Nelson *et al.*, Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*, *Nature*, **399**: 323-329 (1999).
- Carl R. Woese and George E. Fox, Phylogenetic structure of the prokaryotic domain: The primary kingdoms, *Proc. Natl. Acad. Sci. USA*, **74**: 5088-5090 (1977).
- Nick Lane, *Power, Sex, Suicide: Mitochondria and the Meaning of Life*, Oxford University Press, 2005. (Introduction, Chs. 1 and 7.)
- Nick Lane, *The Vital Question: Energy, Evolution, and the Origins of Complex Life*, Norton & Company, 2015. (Introduction and beginning of Ch. 4.)

Detailed Course Content:

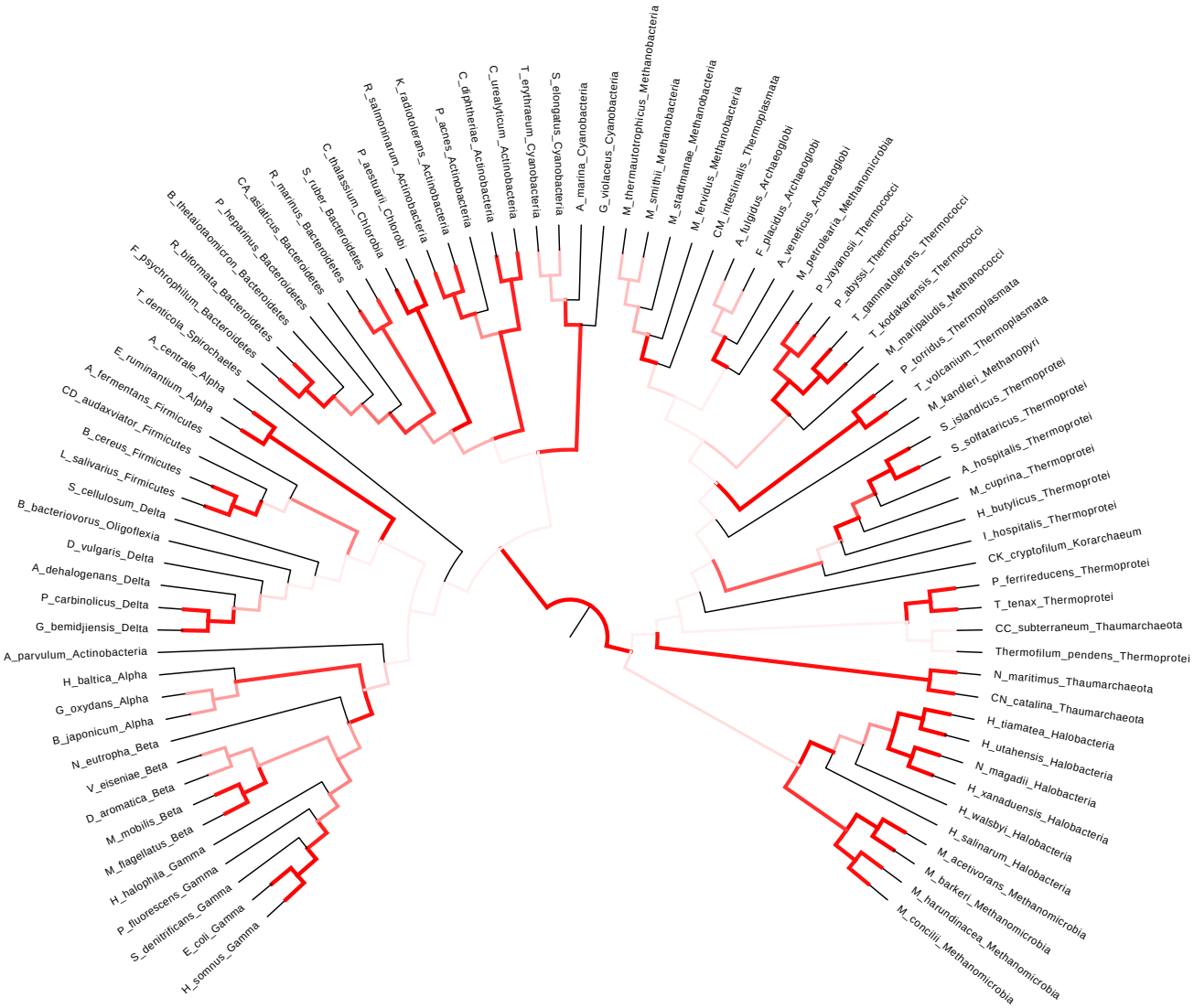
week	date	Lecture, Computer lab, `Python` (**p4b** Ch.)
		I. DNA and protein sequences
1	Sep 21	General information about the course. Sizes of cells and genomes.
	Sep 22	How many Bytes is a genome?
	Sep 22	Lab1: Nucleotide content of 8 prokaryotic genomes. `Strings` (Ch.2). `Reading text from a file` (Ch.3a).
2	Sep 28	Chargaff's parity rule and DNA structure. Chargaff's second parity rule.
	Sep 29	Duplets, triplets, and probabilities.
	Sep 29	Lab2: Content of nucleotide duplets in genomes ("genomic signature"). `Lists and loops` (Ch.4). `Functions` (Ch.5). `Dictionaries` (Ch.8).
3	Oct 5	Random numbers with Python. Probabilities.
	Oct 6	Conditional probabilities.
	Oct 6	Lab3: Generating random DNA with given dinucleotide composition. `Conditional tests` (Ch.6). `Writing to a file` (Ch.3b).
4	Oct 12	Dinucleotide correlations over long separations. The genetic code.
	Oct 13	How long is a "typical" protein?
	Oct 13	Lab4: Finding candidate protein-coding genes in a genome. `Making plots with matplotlib.pyplot`.
5	Oct 19	Transcription and translation. NCBI database of genomic sequences.
	Oct 20	Processing a GenBank file.
	Oct 20	Lab5: START and STOP codon frequencies in genomes of 19 prokaryotes. `Regular expressions` (Ch.7).
6	Oct 26	Protein structures and functions. Amino acid physical properties.
	Oct 27	Mutations and the genetic code.
	Oct 27	Lab6: Frequencies of amino acids in genomes of 40 prokaryotes. `Using Python's operating system (os) module`.
7	**Nov 2**	**Midterm Exam 1** (take-home) Nov 2- Nov 9
	Nov 3	No lecture this Friday.
	Nov 3	No lab this week.
		II. Comparing two sequences
8	Nov 9	Trees and phylogeny. Hierarchical clustering.
	Nov 10	The UPGMA algorithm with NumPy.
	Nov 10	Lab7: A tree from amino acid content of 40 genomes. `Exposure to Python's numpy module`.
9	Nov 16	Horizontal gene transfer in *Thermotoga maritima*. Ribosomal RNAs.
	Nov 17	Edit distance and alignment of DNA sequences.
	Nov 17	Lab8: Is *T. maritima* a bacterium according to its 5S rRNA?
10	Nov 23	Endosymbiotic theory: eukaryotes, mitochondria and chloroplasts.
	Nov 24	5S ribosomal RNA is present in prokaryotes, eukaryotes and chloroplasts.
	Nov 24	Lab9: Tree of life according to 5S rRNA. Ancestors of chloroplasts.
11	Nov 30	A closer look at mitochondria and mitochondrial genomes.
	Dec 1	Aligning protein sequences (BLOSUM).
	Dec 1	Lab10: The sister group of mitochondria (according to cox1).
12	Dec 7	Semi-global alignment. Finding a gene in a mitochondrial genome.
	Dec 8	What are the midichlorians?
	Dec 8	Lab11: Mutation rate of mitochondrial *atp8* gene in the last 30 My's.
13	Dec 14	**Midterm Exam 2** (take-home) Dec 14 - Dec 22
	Dec 15	No lecture this Friday.
	Dec 15	No lab this week.
14	Dec 21	No lecture this Thursday.
	Dec 22	Solution of exam questions.
	Dec 27	**Final Exam** (in-class)



Usage of START and STOP codons in 19 prokaryotes.



Amino acid frequencies in 40 prokaryotes.



Bill Martin's disappearing tree.

A poor man's version of Fig. 6 in Sousa *et al.*, Early bioenergetic evolution, *Phil. Trans. R Soc. B*, **368**:20130088 (2013).

Bacterial ancestors of **chloroplasts** (top) and **mitochondria** (bottom).