Introduction to Bioinformatics and Molecular Evolution
BIOC3991C/BIOL3991C
Winter 2010, three credits
MW (lectures, DUNN 106) F (lab, DUNN102) 10:30-11:20
http://www.mta.ca/~ozhaxybayeva/bioinf2010/

Instructor: Olga Zhaxybayeva, olgazh@mta.ca, Flemington 122

Office Hours: M 11:20-12:20, Tue 16:00-17:00 or by appointment

Grading Criteria: 30% midterm (in-class, written, tentative date March 3rd), 30% final (written, during examination period), 30% homework + computer lab assignments, 10% independent mini-project

Textbook: Jin Xiong "Essential Bioinformatics", Cambridge University Press 2006.
Supplementary Textbook: Understanding Bioinformatics" by Marketa Zvelebil and Jeremy Baum, Garland Science, 2007 [available in the library on reserve].
Additional Reading Materials will be posted to Moodle.

Overview: The course is designed to introduce students to contemporary tools (both online and stand-alone) to analyze genetic and genomic data in an evolutionary context. The course will consist of lectures and hands-on computer labs. Upon completion of the course the student is expected to be able to make meaningful biological predictions about fragments of DNA obtained in the laboratory or from public databases, as well as have basic understanding of genome-wide and phylogenetic analyses.

Tentative Course Outline
Order of the topics may be altered if necessary; roughly one week per topic.

1. Overview. Databases at NCBI. Entrez. Other, specialized databases (e.g. RDP, Ensembl). Introduction to currently available biological sequence data (genes, genomes, ESTs, SNPs, environmental fragments, etc.). Computer Lab: Literature search, exploration of NCBI tools and other genomic resources.
2. Homology. Types of Homology. Gene Families. Sequence Similarity Statistics. BLAST. PRSS. Pairwise alignment. Computer Lab: BLAST
3. ORF finding, sequence features (introns, exons), homology-based functional annotations, functional databases (COG, PFAM). Computational predictions of protein properties. Computer Lab: annotation of a genomic fragment and examination of coding regions properties.
4. Multiple sequence alignments. Alignment Quality. Consensus sequence. Genome Alignments. Computer Lab: Alignments in CLUSTALX, DotPlots, MAUVE, SeqLOGO.
5. Protein Structure. Protein Domains. CDD. Structure Alignments. Protein Space. Computer Lab: PDB, protein structure prediction and exploration.
6. Profile Alignments. PSI-BLAST. HMMER. Computer Lab: PSI-BLAST, RPS-BLAST, TIGRFAMs
7. Phylogenetic Trees. Tree Space. Basics of sequence evolution modeling. Bootstrapping. Computer Lab: ClustalX, Tree-PUZZLE, PhyML, tree topology visualization.
8. Selection. Detection of positive and purifying selection. Computer Lab: MrBayes, PAML
9. Prokaryotic Genome Evolution: gene duplications, horizontal gene transfer, gene loss. Concept of pan-genome and gene core. Computer Lab: TaxTable, Comprehensive Microbial Resource tools at JCVI, DotLet.
10. Metagenomics. Fragment Recruitment. Assembly. Computer Lab: CAMERA database, MUMmer.

Independent Mini-Project throughout the course (introduced after second week): Annotate a metagenomic fragments (from Global Ocean Survey Environmental Database) using tools as we learn them along the class. The project will be a part of larger Annotathon project (http://annotathon.univ-mrs.fr/): Bioinformatics Training Through Metagenomic Sequence Annotation.