BIOL 47/147 – Genomics: From Data to Analysis

Spring 2018
Classroom: LSC 105
Class Meeting times – MWF 10:10AM-11:15AM

Instructor: Olga Zhaxybayeva (aka Professor Z)
Office Location: LSC 333
Email: olgazh@dartmouth.edu
Office Phone: 603-646-8616
Office Hours: To be announced weekly via Canvas. Also by appointment, if needed
X-hour: See schedule on Canvas for details

Course Goals
With advances in sequencing technology, analyses of DNA, RNA and protein data have become central in many biological and medical research projects. Through lectures, discussion sessions, hands-on in silico exercises and in-class problem-solving, we will explore how genomic data analysis advances biological knowledge.

By the end of this course, you will be:

- Familiar with algorithms of nucleotide and amino acid sequence data analysis and able to apply them using specific software and online tools.
- Able to explain reasoning behind the widely-used bioinformatic algorithms and methods.
- Able to analyze and evaluate authentic genomic datasets.

Pre-Requisites
BIOL 13 OR BIOL 15 OR Permission of Instructor

Reading Materials
Our textbook is "Concepts in Bioinformatics and Genomics" by Nomand and McCurdy, Oxford University Press, 2016. It is available on reserve in the Dana library. To supplement the textbook, review articles and online tutorials will be occasionally assigned. There are also several additional textbooks on reserve in the Dana library, and selected sections from them will be assigned mostly as alternative optional reading. This syllabus, power point presentations, links to pre-lecture videos, links to library reserves and additional readings, and in-class assignments will be posted to Dartmouth’s Canvas site (http://canvas.dartmouth.edu). To access the site, use your NetID and password.

Teaching Approach
Class periods will be a combination of mini-lectures, discussions of the assigned reading, software demonstrations, and computer-based exercises completed either individually or in groups. The exercises will give you ample opportunities to master analytic and problem-solving skills using real genomic data sets.
Grading

Exam #1 (30%) - April 18, 2018 (6-8PM), LSC 201
Exam #2 (30%) - May 16, 2018 (6-8PM), LSC 201

Independent Project (35%):
  - annotation report (7%) - due May 25, 2018;
  - presentation (12%) - in class, tentatively May 25 and 30, 2018;
  - final paper (16%) - due June 1, 2018;

In-class participation (5%)
  - active engagement during the class periods,
  - pre-lecture quizzes (will be posted via Canvas)

On Independent Project

Our research theme this term is "Genomic determinants of temperature adaptation". In the first week of the course I will introduce you to the study system (a bacterium that is capable of living at an astonishingly wide temperature range of 20-80 degrees Celsius), data (the bacterium's genome and transcriptome) and research questions. In my laboratory, we have identified a list of genes that may be involved in the ability of this bacterium to respond to temperature fluctuations. However, more work is needed to understand the role of these genes (if any) in adaptation to growth at a specific temperature. For your project, you will select several putatively temperature-responsive genes and analyze them using tools and methods that we will be learning during the term. You will make a presentation of your findings during the final week of the class. You will also summarize the project in a research paper. Details will be discussed in class.

Additional Assignment for Graduate Students

Each graduate student will analyze an additional gene related to the study system/organism related to the student's thesis research. The student will pick the gene in consultation with the professor during the first week of classes. The student will summarize the research findings in a short scientific report that would count towards the "independent project" portion of the grade.

Expectations

Here is what I expect from you:

1. to carefully read the assigned material before class,
2. to enthusiastically participate in class discussions,
3. to diligently prepare for the in-class hands-on activities,
4. to thoughtfully compose discussion questions and answers in discussion forums.

Cell Phones: Cell phone use during the class is not allowed. Please be sure all cell phones are turned off or disabled before class starts.
Laptops: Please bring your laptop to the class periods, as we will be analyzing data in most class periods. However, multi-tasking with the laptop that distracts you and your classmates from participating fully during class (such as checking e-mail or Facebook, making online purchases, reading the newspaper, etc.) is strictly prohibited in all class periods.

You can expect me to:

(1) Bring expertise into the classroom.
(2) Stimulate interest in the course material.
(3) Provide consultations during the hands-on activities and be available to answer questions on lectures and hands-on exercises.
(4) Return graded assignments promptly.

Academic Honor
The Dartmouth Honor Principle applies to all work you submit for a grade in this course.

During the in-class activities, however, you are encouraged to consult with others while designing analyses, running analyses, and drawing conclusions.

Student Needs and Wellness
Students who may need disability-related classroom accommodations are encouraged to make an appointment to see me before the end of the second week of the term. All discussions will remain confidential, although the Student Accessibility Services office may be consulted to discuss appropriate implementation of any accommodation requested. Student Accessibility Services (http://www.dartmouth.edu/~accessibility/facstaff/).

I recognize that the academic environment at Dartmouth is challenging, that our terms are intensive, and that classes are not the only demanding part of your life. A number of on-campus resources is available to you to support your wellness, including your undergraduate dean (http://www.dartmouth.edu/~upperde/), Counseling and Human Development (http://www.dartmouth.edu/~chd/), and the Student Wellness Center (http://www.dartmouth.edu/~healthed/). I encourage you to use these resources to take care of yourself throughout the term.

Additional Support for your Learning
- Academic Skills Center (http://www.dartmouth.edu/~acskills/)
- The Research Center for Writing, and Information Technology (RWiT) (http://www.dartmouth.edu/~rwit/)
Overview of Course Modules

What is in a Genome?
Learning Objectives:

- Be able to identify components and features of a genome
- Be able to discuss genomic architecture and signatures of prokaryotes and eukaryotes
- Be able to use computational tools to predict genes in prokaryotes and eukaryotes
- Be able to examine genomic features of completed genomes using browser-based software and genomic databases

Essentials of DNA and Amino Acid Sequence Comparisons
Learning Objectives:

- Understand homology and its relationship to sequence similarity
- Understand local alignment algorithms and assessment of significance of sequence similarity
- Master running programs from the FASTA and BLAST packages and interpret the results

Transcriptomics
Learning Objectives:

- Understand basic principles of the RNA-Seq analysis
- Be able to interpret RNA-Seq output

Elements of Protein Structure and its Evolution
Learning Objectives:

- Understand the concept of protein motifs and domains
- Learn how to interpret sequence logos
- Learn how to represent motifs as regular expressions and how to run PHI-BLAST search
- Understand the concept of a position specific scoring matrix and a profile
- Master running PSI-BLAST and RPS-BLAST (CDD) searches

Multiple Sequence Alignments. Measuring Genetic Change over Time
Learning Objectives:

- Describe algorithms for aligning more than two sequences
- Master using the ClustalX and MUSCLE multiple sequence alignment programs
- Describe alignment program benchmarking techniques
- Be able to assess alignment quality
• Be able to calculate genetic distances from nucleotide and amino acid sequences
• Know the difference between observed and expected actual number of substitutions
• Discuss why it is important to select appropriate substitution models
• Learn how to model DNA (protein) substitutions

Phylogenetic Inference and Visualization. Inference of Natural Selection at the Molecular Level

Learning Objectives:

• Learn major classes of phylogenetic reconstruction methods
• Learn bootstrapping technique for statistical assessment of tree reconstruction
• Master visualization and Interpretation of produced phylogenetic trees
• Understand how to measure selection in protein-coding genes
• Learn models and statistical tests for detecting selection in molecular data.
• Master interpreting results of the PAML program for estimation of selection.


Learning Objectives:

• Describe genetic variation in a population using terms like allele, locus, haplotype, SNP
• Understand concept of genetic association
• Understand concept of multiple test correction in genome wide association studies
• Know how to perform allelic trend test
• Learn to navigate the human population genomics data via Ensembl portal
• Learn to link phenotypic and genomic data via OMIM database
• Understand the concept of linkage disequilibrium and how it is used to detect genomic regions under selection

Metagenomics. Human Microbiome

Learning Objectives:

• Understand the rationale behind going "metagenomic" in microbial community analysis
• Understand methods of taxonomic classification of microbial communities
• Master using RDP tools to classify unknown 16S rRNA sequences
• Be able to calculate metrics that quantify population diversity
• Explore recent advances in our understanding of the human microbiome