BIOL 47/147 – Genomics: From Data to Analysis

Fall 2019
Classroom: LSC 105
Class Meeting times: Tuesdays and Thursdays 2:25PM-4:15PM
X-hour: Wednesdays 4:35-5:25PM

Instructor:	Olga Zhaxybayeva (aka Professor Z)
Office Location:	LSC 333
Email:	olgazh@dartmouth.edu
Office Phone:	603-646-8616
Office Hours:	Most X-hours; Additional times will be announced weekly via Canvas; By appointment, if needed
X-hour:	Office hours and exam-related activities (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

Undergraduate Students: BIOL 13 *OR* BIOL 15 *OR* Permission of Instructor

Reading Materials

Most of the reading for this course will come from the "Concepts in Bioinformatics and Genomics" by Nomand and McCurdy, Oxford University Press, 2016. The book is available on reserve in the Dana library, so you do not need to buy a copy. This is not a perfect textbook and genomics is a quickly changing field of study. Hence, there are also several additional books on reserve in the Dana library, and selected sections from them will be assigned either as the main, alternative or optional additional reading. Review articles and online tutorials will also be occasionally assigned, and will be made available via *Canvas*. This syllabus, power point presentations, links to pre-lecture videos, links to library reserves and additional readings, and assignments will be posted to Dartmouth's *Canvas* site (http://canvas.dartmouth.edu). To access the site, use your NetID, password and Duo authentication.

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.

To make time for interactive in-class exercises, some lecture materials will be provided in the form of pre-recorded videos, which you will need to watch *before* coming to class. You will also be assigned "Muddy Points" quizzes to reflect on the pre-lecture video and reading materials. These quizzes will not be graded for content, but they will count toward your participation grade. I will read your muddy points reports before coming to class and will address them during the class periods.

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 two 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 last week of the term. You will also summarize the project in a research paper. Details will be discussed in class.

Additional Assignment for Graduate Students only

You will select a gene related to your own research project and will create and perform an analysis of the gene using tools and methods that we will be learning during the term. You will report your findings using a Canvas quiz. Details will be discussed in class.

Grading

Undergraduate Students:

Exam #1 (30%) – October 10 (in class; 1hr50m)

Exam #2 (30%) – November 7 (in class; 1hr50m)

Independent Project (35%):

- progress quizzes (5%) – due throughout the term;
- annotation report (5%) - due November 22, 10PM;
- presentation (10%) - in class, November 14 and 19;
- final paper (15%) - due November 22, 10PM;

In-class participation (5%)

- active engagement during the class periods;

  - pre-lecture quizzes (will be posted via *Canvas*)

Graduate students:

Exam #1 (30%) - October 10 (in class; 1hr50m)

Exam #2 (30%) - November 7 (in class; 1hr50m)

Independent Project (35%):

  - progress quizzes (4%) – due throughout the term;
  - annotation report (4%) - due November 22, 10PM;
  - presentation (9%) - in class, November 14 and 19;
  - final paper (13%) - due November 22, 10PM;
  - “your own gene” analysis (5%) – due October 30, 11:59PM;

In-class participation (5%)

  - active engagement during the class periods;
  - pre-lecture quizzes (will be posted via *Canvas*)

Expectations

Here is what I expect from you:

(1) to carefully read and watch 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 all 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, messaging, browsing social media feeds, 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 collaborate with others while designing analyses, running analyses, and drawing conclusions. Since you all will be using the same software for your exercises and projects, I also encourage you to consult with your classmates if you run into technical difficulties with running programs or accessing databases.

(The detailed description of the Dartmouth Honor Principle is available at https://students.dartmouth.edu/judicial-affairs/policy/academic-honor-principle)

Your Needs and Wellness

If you have questions about whether you need to receive academic adjustments and services, please contact the Student Accessibility Services (SAS) office (Carson Hall, Suite 125, 646-9900; all inquiries and discussions are confidential) as early in the term as possible. Once SAS has authorized services, please come to see me privately and bring with you the originally signed SAS paperwork.

If you have a religious observance that conflicts with your participation in the course, please meet with me before the end of the second week of the term to discuss appropriate accommodations.

If you encounter financial challenges related to this class, please let me know.

The academic environment at Dartmouth is challenging, our terms are intensive, and classes are not the only demanding part of your life. There are a number of resources available to you on campus 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.

Overview of the 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 that find and compare similar sequences and learn to interpret the results

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 a PHI-BLAST search
- Understand the concept of a position specific scoring matrix and a profile
- Master running PSI-BLAST and RPS-BLAST (CDD) searches

Accounting for Insertion and Deletion of Genetic Material Over Time: Multiple Sequence Alignments. Measuring Genetic Change over Time
Learning Objectives:

- Describe algorithms for aligning more than two sequences
- Master using 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

Inference and Visualization of Evolutionary Histories of Genes. 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 programs for estimation of selection

How do we determine DNA sequence of an organism? What genes are actually used by an organism? DNA Sequencing and Assembly. Transcriptomics
Learning Objectives:

- Learn basic principles of currently used "next-gen" DNA sequencing technologies
- Be able to calculate how much DNA needs to be sequenced for a genomic project

- Learn basic principles of post-sequencing genome assembly
- Be able to determine how genome "completeness" by examination of how many scaffolds and contigs it contains
- Learn basic principles of the RNA-Seq analysis
- Be able to interpret results of an RNA-Seq analysis

Genetic Make-up of Complex Microbial Communities: Metagenomics.
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

Human Genomics. Linking Genes to Traits: Genome-Wide Association Studies.
Learning Objectives:

- Describe genetic variation in a population
- 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