BIOL 47/147 – Genomics: From Data to Analysis

Fall 2023
Meeting in-person in LSC 105
Class Meeting times: Tuesdays and Thursdays 2:25PM-4:15PM
X-hour: Wednesdays 5:30-6:20PM

Instructor: Olga Zhaxybayeva (aka Professor Z)
Email: olgazh@dartmouth.edu
Office Phone: 603-646-8616

Office Hours:

- X-hours; LSC 105
- Additional times on Zoom/In-Person; TBA weekly via Canvas
- By appointment, if needed
- I am also happy to stay immediately after class periods

X-hour: Office hours, Pre-exam Q&A, Post-exam Q&A; LSC 105

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

Overview of the Course Modules

Detailed schedule is on *Canvas*

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
- Be able to examine structures from RDP and AlphaFold databases

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

Reading and Video Materials

Genomics is a quickly changing field of study and textbooks get outdated very fast! Therefore, no single textbook purchase is required for this course. Instead, we will rely on several textbooks that are available in Baker-Berry library for a 2-hour loan, with some of the excerpts from these texts available as PDF files on *Canvas.* Additional readings will include review articles and online tutorials, links to which will be provided 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. You will need to use Dartmouth's VPN client to access some of the posted resources.

Class exercises will be available via *Google Docs* on *Dartmouth's Google Drive*. To access the Dartmouth's Google Suite, use your NetID, password and Duo authentication.

Meeting Places

The regular class periods will be in person, with Zoom as a back-up in case of illness or emergencies. All class periods will be recorded, and the videos will be made available via Panopto on *Canvas* (see **Consent to Recording** section below.)

The office hours will take place in person during X-hour in our regular classroom and in person (or over Zoom) during the times that will be communicated weekly via "Announcements" section of *Canvas*.

Dedicated Zoom link for the class is available via *Canvas*. Please use Dartmouth's Zoom account, as this meeting place is limited to Dartmouth users only.

Teaching Approach

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

To make time for interactive in-class exercises, most 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 quizzes to reflect on the pre-lecture video and reading materials, to be submitted *before* the class. These quizzes will not be graded for content, but they will count toward your participation grade. Each pre-lecture quiz will contain "What are your muddy points?" question, which will allow me to know what material in the pre-lecture video and reading assignments was unclear. I will read your muddy points reports before coming to class. I will try to address most muddy points during the class periods, but due to time limitations some questions may not be addressed (you are encouraged to follow up on your unanswered questions after class or during office hours).

On Independent Project

Our research theme this term is "Coronaviruses"! During the course I will introduce you to the genomes of coronaviruses, including the genome of SARS-CoV-2. For your project, you will be assigned one gene from a coronavirus genome and will work on it with several of your classmates. You will analyze the gene and its protein product using tools and methods that we will be learning during the term, following the provided instructions. You are expected to work on the project throughout the term, to keep "analyses notebook" and to report interim results in three progress quizzes. At the end of the term, your group will make a presentation of your findings, which will be reviewed and graded by me. You will also reflect on your classmates' presentations in a graded assignment. Finally, your group will summarize the project in a research paper and submit your "analyses notebook".

Details will be discussed in class and specific instructions provided via *Canvas* throughout the term.

Additional Assignment for Graduate Students only
You will come up with an interesting question about a gene or protein of your choice, ideally related to your own research. You 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.

Grading
Undergraduate Students:

Exam #1 (30%) – October 4, 6:30PM - October 5, 2:00PM (on Gradescope; 2hrs)

Exam #2 (30%) – November 1, 6:30PM - November 2, 2:00PM (on Gradescope; 2hrs)

Independent Project (35%):

- 3 progress quizzes (6%) – due September 29, October 13, November 6, 10PM;
- presentation (8%) - due November 14, in class;
- reflection on presentations of other students (1%) – due November 14, 11:59PM
- analyses notebook (5%) - due November 21, 11:59PM;
- final paper (15%) - due November 21, 11:59PM;

In-class participation (5%)

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

Grading Scale (%): 92-100 (A), 90-91 (A-), 88-89 (B+), 82-87 (B), 80-81 (B-), 78-79 (C+), 72-77 (C), 70-71 (C-), 60-69 (D), 59 and below (E). For description of grades, see https://www.dartmouth.edu/reg/transcript/grade_descriptions.html

Graduate students:

Exam #1 (30%) – October 4, 6:30PM - October 5, 2:00PM (on Gradescope; 2hrs)

Exam #2 (30%) – November 1, 6:30PM - November 2, 2:00PM (on Gradescope; 2hrs)

Independent Project (35%):

- 3 progress quizzes (5%) – due September 29, October 13, November 6, 10PM;
- presentation (7%) - due November 14, in class;
- reflection on presentations of other students (1%) – due November 14, 11:59PM
- analyses notebook (4%) - due November 21, 11:59PM;
- final paper (13%) - due November 21, 11:59PM;
- “your own gene” analysis (5%) – due October 27, 11:59PM;

In-class participation (5%)

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

Grading Scale (%): 95-100 (HP), 80-94 (P), 60-79 (LP), 59 and below (NC). For description of grades, see https://graduate.dartmouth.edu/policy/satisfactory-progress

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 and group work,
(3) to help your group mates with in-class hands-on exercises.

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.

Late Assignments

Please speak with me at the beginning of the term if you anticipate circumstances that might affect your ability to get your work in on time and reach out if such a situation arises along the way. Reasonable allowances will be made as long as we connect about these matters *prior* to the assignment deadlines.

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://policies.dartmouth.edu/policy/academic-honor-principle-1

For the exams (graduate and undergraduate students), final paper (graduate and undergraduate students) and “your own gene” analysis (graduate students only) use of generative AI tools (such as ChatGPT) is prohibited (the reasons why will be discussed in class).

Public Health Information

For the health and safety of our class community, please **do not attend class when you have flu-like or other symptoms of respiratory illness (regardless if it is COVID-19 or not) or if you are tested positive for COVID.** You will be able to view class recordings in *Canvas* if you are unable to attend in person, and some of the office hours will be available via Zoom.

Per **college's policy regarding COVID-19**:

- If you feel sick or have COVID-19 symptoms, test immediately and wear a mask until you feel better.
- If you test positive for COVID-19, isolate for at least five days and wear a mask for at least 10 days. Wear a mask for longer if you continue to test positive or have new or worsening symptoms. Notify your close contacts.
- If you were exposed to COVID-19, wear a mask for at least 10 days and test on or after Day 6. Test sooner if you develop symptoms.

For the most up-to-date COVID-related health guidance, visit https://covid.dartmouth.edu/

Your Needs and Wellness

If you are requesting disability-related accommodations and services for this course, you are required to register with Student Accessibility Services (SAS; Getting Started with SAS webpage; student.accessibility.services@dartmouth.edu; 1-603-646-9900) and to request that an accommodation email be sent to me in advance of the need for an accommodation. Then, you should talk to me to determine relevant details such as what role SAS or its Testing Center may play in accommodation implementation. *This process works best for everyone when completed as early in the quarter as possible.* If you have questions about whether you are eligible for accommodations or have concerns about the implementation of accommodations, you should contact the SAS office. All inquiries and discussions will remain confidential.

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.

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, the Counseling Center, and the Student Wellness Center. I want you to be aware of these resources and encourage you to use them as needed.

Title IX

At Dartmouth, we value integrity, responsibility, and respect for the rights and interests of others, all central to our Principles of Community. We are dedicated to establishing and maintaining a safe and inclusive campus where all have equal access to the educational and employment opportunities Dartmouth offers. We strive to promote an environment of sexual respect, safety, and well-being. In its policies and standards, Dartmouth demonstrates unequivocally that sexual assault, gender-based harassment, domestic violence, dating violence, and stalking are not tolerated in our community.

The Sexual Respect Website provides a wealth of information on your rights with regard to sexual respect and resources that are available to all in our community. *Please note that, as faculty members, we are obligated to share disclosures regarding conduct under Title*

IX with Dartmouth's Title IX Coordinator. Confidential resources include licensed medical or counseling professionals (e.g., a licensed psychologist), staff members of organizations recognized as rape crisis centers under state law (such as WISE), and ordained clergy. Should you have any questions, please feel free to contact Dartmouth's Title IX Coordinator.

Consent to Recording

(1) Consent to recording of course

a) By enrolling in this course, you affirm your understanding that this course and associated **group** meetings involving students and the instructor, may be recorded within any digital platform used to offer instruction for this course;
b) You further affirm that the instructor owns the copyright to their instructional materials, of which these recordings constitute a part, and distribution of any of these recordings in whole or in part without prior written consent of the instructor may be subject to discipline by Dartmouth up to and including expulsion;
c) You authorize Dartmouth and anyone acting on behalf of Dartmouth to record your participation and appearance in any medium, and to use your name, likeness, and voice in connection with such recording; and
d) You authorize Dartmouth and anyone acting on behalf of Dartmouth to use, reproduce, or distribute such recording without restrictions or limitation for any educational purpose deemed appropriate by Dartmouth and anyone acting on behalf of Dartmouth.

(2) Requirement of consent to one-on-one recordings

By enrolling in this course, you affirm that you will not under any circumstance make a recording in any medium of any one-on-one meeting with the instructor without obtaining the prior written consent of all those participating, and you understand that if you violate this prohibition, you will be subject to discipline by Dartmouth up to and including expulsion, as well as any other civil or criminal penalties under applicable law.