

BIOL 2xxx: Computational Project Lab

Provides hands-on experience applying common computational techniques to actual sequence datasets generated by research labs in the Biology Department or in public databases. You'll learn a range of computational skills and practices as you design and carry out your own research project, conveying results through a presentation and a short report formatted as a research paper. You'll also develop "best practices" and learn the value of collaboration in computational work by sharing your ideas, obstacles, and results with peers. Topics include read mapping, counts normalization, differential expression analysis, and phylogenetic analysis and the tools, workflows, and basic statistics involved in these approaches.

Prerequisites: BIOL 2xxx (Genome analysis skills workshop)

Learning objectives. Upon completion of the course, you will be able to:

- Implement basic workflows for processing computational data
- Submit and track large jobs on Northeastern's high-performance computing cluster
- Design and carry out a computational project
- Apply appropriate statistics in analysis of data
- Collaborate with other researchers to troubleshoot and exchange feedback
- Present research findings to a group
- Write a report describing the project and results
- Practice "best practices" in computational research

Format. The course will be run as a computational project lab, involving mostly hands-on exercises working with various tools, with limited lecture/discussion to introduce research data and concepts as needed. As part of the course, you'll participate in at least one meeting of CompCore (see below).

CompCore: Serves as a resource for undergraduate students, graduate students, and postdocs involved in computational research. CompCore has open weekly meetings where people share resources, help troubleshoot, develop projects, share results and latest findings in the field, and provide overall moral support.

AI use: Personally, I'm tremendously excited about the possibilities offered by AI, while also wary of the downsides, and I look forward to experimenting with its use in this class, with you! We will have frequent conversations during the term about ways in which AI is useful for our work, and ways in which it can lead us astray. For purposes of this experiment, please document your uses of AI and share them in class. The objective is to use AI to make us smarter and not dumber!

Grading. You'll be assessed as weighted in the following categories, which are described on the Assignments page:

Basic skills review	25 pts
Project design and implementation	20 pts
Research report	20 pts

Presentations	15 pts
Worksheets	20 pts

Grades are determined based on the following scale of percentages:

A[90+]	B+[87-89]	B[83-86]	B-[80-82]	C+[77-79]	C[73-76]
C-[70-72]	D+[67-69]	D[63-66]	D-[60-62]	F[<60]	

Unit I: Skill set review. You'll review the basic skills required for the course and demonstrate mastery through a series of short tutorials.

- Unix commands tutorials
- Explorer high-performance computing cluster tutorials
- R tutorials

Unit II. Analyses. You'll familiarize yourself with the contents and structure of various datasets, and review common types of analyses performed on them, including:

- Differential gene expression analysis
- Phylogenetic analysis in R
- Finding and analyzing variants
- Assessing transposable elements

Unit III. Computational lab. You'll carry out your project employing a repeated process of designing a workflow, troubleshooting, and refining the analysis, while tracking your progress in a lab notebook. Weekly, students will "debrief" in collaboration with one another, sharing obstacles, insights and feedback.

Unit IV. Preparation and presentation of project results. You'll present your findings to CompCore or class, and write a short research report outlining your project goals, results, discussion, conclusions, methods, and references.

Proposed attributes: NUpath Analyzing/Using Data, NUpath Creative Express/Innov, **Writing Intensive**

#	Date	Topics	Activities
1	Thu Jan 8	Course introduction; Command line setup, best practices	
2	Mon Jan 12	UNIX review: tips and troubleshooting	
3	Thu Jan 15	UNIX review: scripts	Complete tutorial 1
4	Mon Jan 19	NO CLASS (Holiday)	
5	Thu Jan 22	High-performance computing review	Complete tutorial 2
6	Mon Jan 26	RStudio review: working with genomic data	
7	Thu Jan 29	RStudio review: visualizing data	Complete tutorial 3
8	Mon Feb 2	Workflows and lab notebooks	

#	Date	Topics	Activities
9	Thu Feb 5	Overview of analyses: differential gene expression	
10	Mon Feb 9	Overview of analyses: variant analysis	Worksheet due: DGE analysis
11	Thu Feb 12	Overview of analyses: phylogenetic analysis	Worksheet due: Variant calling
12	Mon Feb 16	NO CLASS (Holiday)	
13	Thu Feb 19	Overview of analyses: transposable elements	Worksheet due: MSA to tree
14	Mon Feb 23	Lab notebooks: best practices	Worksheet due: TE analysis
15	Thu Feb 26	Project brainstorming session!	
16	Mon Mar 2	NO CLASS (Spring Break)	
17	Thu Mar 5	NO CLASS (Spring Break)	
18	Mon Mar 9	Develop workflow; provide to classmate	
19	Thu Mar 12	Present classmate's workflow to class	
20	Mon Mar 16	"Groundbreaking" ceremony: begin project!	Due: workflow, description of dataset
21	Thu Mar 19	Troubleshooting round robin	
22	Mon Mar 23	Report on progress	Progress report 1 due
23	Thu Mar 26	Background, references; writing troubleshooting	Background, references due
24	Mon Mar 30	Continue project; troubleshooting	Progress report 2 due
25	Thu Apr 2	Results and methods writing	
26	Mon Apr 6	Informal presentation of methods, results	Methods, results due
27	Thu Apr 9	Informal presentation of	Lab notebook due
28	Mon Apr 13	Presentations	
29	Thu Apr 16	Presentations	