

Introduction to Bioinformatics

EBIO 4100

Summer 2024, 8/5-8/17 – Mountain Research Station

Instructor: Megan Frayer (megan.frayer@yale.edu)

Course Description

The ability to understand and process sequencing datasets is an essential skill for today's biologists. This course aims to provide a foundation for biologists seeking to process their own datasets, computer scientists seeking an introduction to genetic methods, or any students interested in exploring this growing field. Through this course, students will become familiar with different types of biological sequence datasets and how they can be used to generate insights about our natural world.

Learning Goals:

1. Understand the scope of bioinformatics.
 2. Understand various approaches to sequencing and the types of data they produce.
 3. Gain confidence using the command line.
 4. Understand workflows for processing genomic data.
 5. Gain familiarity with common databases (NCBI, Genome Browser, etc.).
 6. Learn best practices for reproducibility.
-

Course Materials

This course will require the use of a personal computer with internet access. All activities and datasets will be made available through a course website and GitHub page.

Course Structure

This course will be taught over a two-week period 8/5-8/17 (Monday-Saturday, no class Sun 8/11). Each day will consist of lectures in the morning and a hands-on activity in the afternoon. Daily comprehension quizzes will be available online following the lecture, and due by the start of the next lecture (the following day). Each hands-on activity will have an assignment that is due at the end of that session. At the end of the course, there will be a group project and a final exam.

Group Project

The group project will be completed during class on the second to last day. Students will apply their knowledge by assembling a project plan/proposal for a sequencing-based project.

Final Exam

The final exam will be drawn from the concepts covered during lectures.

Assignments/Grades

Lecture Comprehension Quizzes (10)	10 points each	20% of final grade
Hands On Activities (10)	15 points each	30% of final grade
Group Project	150 points	30% of final grade
Final Exam	100 points	20% of final grade
<i>Total</i>	<i>500 points</i>	<i>100%</i>

Letter Grade Assignments: 94-100 (A) ; 89-93 (A-) ; 86-89 (B+) ; 83-86 (B) ; 80-82 (B-) ; 77-79 (C+) ; 73-76 (C) ; 70-72 (C-) ; 67-69 (D+) ; 63-66 (D) ; 60-62 (D-) ; < 60 (F)

Attendance

Due to the accelerated nature of this course, attendance at all class sessions is essential. It will not be possible to make up the hands-on activities. Exceptions can only be made in the case of temporary medical conditions or injuries, as per university policy (<https://www.colorado.edu/disabilityservices/students/temporary-medical-conditions>).

Code of Conduct

All participants in this course are responsible for maintaining an acceptable learning environment. This is particularly important in an accelerated course, in which all disruptions can be detrimental to our learning goals. Please see university policies on Academic Integrity and the Code of Conduct: <https://www.colorado.edu/sccr/students/honor-code-and-student-code-conduct>

Accommodation Policy

If applicable, please submit a letter from Disability Services ahead of your arrival at MRS (<https://www.colorado.edu/disabilityservices>). If applicable, please notify me of any class conflicts due to religious observance prior to arrival at MRS (<https://www.colorado.edu/policies/observance-religious-holidays-absences-classes-or-exams>). Due to the length of this course, it will be important to prepare for any accommodation ahead of time.

Tentative Schedule:

		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Week 1	Morning Lecture	Lecture: What is bioinformatics?	Lecture: Central Dogma and sequencing techniques	Lecture: Sequencing techniques (Part II)	Lecture: Genetic variation	Lecture: Genome structure	Lecture: Molecular evolution
	Break						
	Mid-Morning Lecture	Lecture: Introduction to the command line and the dataset	Lecture: Common file formats	Lecture: Basic algorithms	Lecture: Variant calling	Lecture: Variant filtering and genome structure	Lecture: Phylogeny
	Lunch						
	Hands-on Activity	Hands-on: Computer set up and data skills	Hands-on: FASTA files and visual alignment with MEGA	Hands-on: Alignment of whole genome data	Hands-on: Variant calling	Hands-on: Variant filtering and genome structure	Hands-on: Phylogeny

		Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Week 2	Morning Lecture	Lecture: Gene ontology	Lecture: Large scale sequencing projects	Lecture: Reproducibility	Lecture: Beyond whole genome sequencing	Lecture: Review of topics with lowest comprehension scores	Final Exam
	Break						
	Mid-Morning Lecture	Lecture: Annotation	Lecture: Useful databases	Lecture: Reproducibility	Lecture: RNA-SEQ	Group Project	
	Lunch						
	Hands-on Activity	Hands-on: Annotation	Hands-on: BLAST and Genome Browser	Hands-on: Reproducibility	Hands-on: RNA-SEQ	Group Project	