

## Fall 2023 - Practical Genomics INTEGBI 134L (4 units) - SYLLABUS

To be offered every: **Fall**

Unit Value: **4**

IB Group: **A (Evolution and Genetics)**

Instructor: **Peter Sudmant & Becca Tarvin**

Office: **5002 VLSB / 3120 VLSB**

Course Fee: **\$25 (starting in 2021)**

Email: [psudmant@berkeley.edu](mailto:psudmant@berkeley.edu), [rdtarvin@berkeley.edu](mailto:rdtarvin@berkeley.edu)

Lecture Info: In person, lecture T/Th 11:00am-12:30pm (Mulford 159); lab Th 2:00-5:00 Social Sciences 110, F 3:00-6:00 Social Sciences 122, Th 3:00-6:00 122 Social Sciences

### Office Hours

- PHS (4-5pm Mondays on zoom) (ending after midterm)
- RDT (4-5pm Wednesdays on zoom) (starting after midterm)

Instructional Format: **Lecture, Lab**

**Course Schedule:** Two lectures per week, 1.5 hours per lecture, 3-hour lab

Target enrollment: 1 lecture section, 3 lab sections (100 students, 1 GSI each lab)

Classroom: lecture hall for computer lab for both lecture and lab

**Prerequisites:** Bio1A and Bio1B

**Course description and aims.** Genome sequencing and analyses have transformed biology over the past two decades. This course provides a hands-on introduction to the world of computational biology and bioinformatics. Students will apply state-of-the-art techniques to analyze genome and microbiome data from the UC Berkeley campus fox squirrels and local California ground squirrels each week. Students will master practical bioinformatics skills and then take on their own scientific research projects, all using genomic data collected specifically for this course. In addition to learning about methods and techniques, we will explore key advances in the field of genomics over the past two decades of both humans and non-model organisms that have driven the current revolution in genome sciences.

### Learning Objectives

- 1) Students will gain an understanding of the current state-of-the art in genomics technologies including sequencing, transcriptomic profiling, assembly and annotation, from scientific collection to sequence data files
- 2) Students will gain an understanding of fundamental concepts of bioinformatics including a complete quantitative understanding of several core topics including: sequence alignment, sequence assembly, sequence modelling, Hidden Markov Modelling of genome features, evolutionary modelling of substitutions, and phylogenetics.
- 3) Students will learn how to design and carry out a hypothesis-driven genomics-based field research program.

- 4) Students will design and complete a research project using actual data. They will carry out peer review assessment of each other's proposals and present to one another in a symposium style at the end of the course.

### **Assessment**

**Lab Assignments: 25%**

**Midterm: 20%**

**Final Project: 35%**

- Pre-proposal 5%
- peer review 5%
- final proposal 5%
- report 10%
- presentation 10%

**Final Exam (cumulative): 20%**

**Grading:** All grades and final grade assessment will be made on an absolute scale (no curving)

**Midterm and final exam formats:** short answer, closed book, testing information recall and reasoning based on covered concepts.

**Lab assignment formats:** Lab assignments are performed on the Data Hub following jupyter notebooks. Students will be expected to turn in a completed lab assignment (short answer or multiple-choice) each week prior to the following lab section.

**Late assignment policy:** Late assignments are docked 10% per day. Under extenuating circumstances (e.g., family emergencies), exceptions to this policy will be made on a case-by-case basis. We encourage students to obtain accommodations in advance from the Disabled Students' Program when needed.

**Attendance policy:** Attendance is highly encouraged to get the most of out this course. Up to 5% of each exam grade will be assigned from electronically assessed in-class participation and attendance using PollEverywhere or a similar platform.

**Final Project:** Students will perform a genomic analysis on data generated by the course or publicly available data. The final project consists of four graded components.

1. Each student will submit a project proposal which will be graded on the following components: background, hypothesis, methodological approach, and expected outcomes.
2. Students will each perform a peer review of three fellow student's proposals. Reviews consist of a short summary and suggested feedback.
3. Students will, either independently, or with a partner, carry out the proposed research project, with feedback both from peer review and instructors. Students will submit a final report consisting of the methods applied and the results including at least one figure and one table.
4. Students will present a 5-min final presentation and be graded on their communication of the motivation, approach, and results of their research project.

A detailed overview of the final project grading scheme follows on the next page.

## Practical Genomics Final Project (35% of final grade)

Students will perform a genomic analysis on data generated by the course or publicly available data. The final project consists of four graded components.

- 1.) Pre-Proposal (5%), Due October 27/28 at the start of lab. Each student will submit a project proposal which will be graded on the following components: background, hypothesis, methodological approach, and expected outcomes. Great project proposals will have a clear hypothesis that is grounded in background literature. Students will have identified what data will be used and which methods will be employed. They should envision the final figure and table that will be part of their final report.
- 2.) Peer Review (5%), Conducted October 27/28 during lab. Students will each perform a peer review of three fellow student's proposals. Reviews consist of a short summary and suggested feedback on how to improve the proposal. Great reviews provide constructive, thoughtful critique that could help improve the project.
- 3.) Final Proposal (5%), Due November 3 at the start of lab. Students will respond point-by-point to peer review and incorporate suggestions into a revised proposal. Great responses will be courteous and provide rationale as to why the comments were or were not incorporated into the proposal.
- 4.) Report (10%), Due Wednesday December 7 (RRR week). Students will, either independently, or with a partner, carry out one proposed research project, with feedback from instructors. Students will submit a final report consisting of the methods applied and the results including at least one figure and one table. If students work with partners, they must include a paragraph outlining the contributions of each partner to the final report. Students must both participate in the computational aspect of the project to receive credit. Great reports will build on final project proposals by adding a detailed methods and results section written in the same style as a scientific paper.
- 5.) Presentation (10%), Conducted on November 29 and December 1 during class. Students will present a 5-min final presentation and be graded on their communication of the motivation, approach, and results of their research project.

Suggested final project categories:

- Compare mitochondrial genomes within or across species
- Compare repetitive sequence content between genomes of different species
- Compare assembly metrics of different published genomes
- Compare CpG island content across species
- Compare genetic diversity across sequences or species
- Compare gene expression across samples/tissues/species
- Blast search for horizontal gene transfer (or assembly errors)
- Phylogenetics of any gene or population
- Compare the microbiome of several samples or species
- Have another idea? Ask Dr. Tarvin and your GSIs

Time to work on projects will be given during the following class times (11.5hr total!):

- Lectures on November 15, 17, & 22 (tentative)
- Labs on November 3/4 & 17/18

## Tentative Course Schedule

<b>08/25/2023</b>	History of genomics and sequencing technologies (Sudmant, lecture)
LAB	Introduction to Jupyter notebooks, bash, and datatypes
<b>08/30/2023</b>	Genome Sequencing I
<b>09/01/2023</b>	Genome Sequencing II
LAB	Examining biological data using bash
<b>09/06/2023</b>	Genome Assembly I
<b>09/08/2023</b>	Pairwise Alignment
LAB	Mapping reads and pairwise alignment
<b>09/13/2023</b>	Genome Assembly II
<b>09/15/2023</b>	Sequence modeling, prediction, and annotation I
LAB	Genome annotation
<b>09/20/2023</b>	Sequence modeling, prediction, and annotation II
<b>09/22/2023</b>	Sequence modeling, prediction, and annotation III
LAB	GWAS
<b>09/27/2023</b>	Population Genetics I
<b>09/29/2023</b>	Population Genetics II
LAB	Population genetics in Structure/ Project guidelines
<b>10/04/2023</b>	GWAS I
<b>10/06/2023</b>	GWAS II
LAB	Phylogenies and ancestral state reconstruction
<b>10/11/2023</b>	Review
<b>10/13/2023</b>	Midterm
LAB	No Lab
<b>10/18/2023</b>	Project Overview; Scientific Collection
<b>10/20/2023</b>	Whole-genome and reduced-representation sequencing
LAB	Field collection
<b>10/25/2023</b>	Metagenomics
<b>10/27/2023</b>	Transcriptomics
LAB	Project proposals due. Peer review of project proposal during lab
<b>11/01/2023</b>	Phylogenetics I
<b>11/03/2023</b>	Phylogenetics II
LAB	Students work on projects
<b>11/08/2023</b>	Study Design
<b>11/10/2023</b>	Molecular Evolution / Future of Genomics
LAB	No lab (Veteran's Day)
<b>11/15/2023</b>	Open coding hour / Special topic / Lecture Overflow
<b>11/17/2023</b>	Open coding hour / Special topic
LAB	Student open-hours
<b>11/22/2023</b>	Open coding hour / Special topic
<b>11/24/2023</b>	Thanksgiving
LAB	No lab (Thanksgiving)
<b>11/29/2023</b>	Student presentations
<b>12/01/2023</b>	Student presentations
LAB	Student presentations (tentative)
Dec 13	Final Exam to take place during scheduled time (8-11am)

## Relevant Readings (although we will not be following these in class)

- Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
- Consortium (IWGSC), T. I. W. G. S. *et al.* Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, eaar7191 (2018).
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016).
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860 (2001).
- Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology* **36**, 338–345 (2018).
- Kunte, K. *et al.* *doublesex* is a mimicry supergene. *Nature* **507**, 229–232 (2014).
- Mudge, J. M. & Harrow, J. The state of play in higher eukaryote gene annotation. *Nature Reviews Genetics* **17**, 758–772 (2016).
- Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
- Prum, R. O. *et al.* A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526, 569 (2015).
- Rogers, R. L. *et al.* Genomic takeover by transposable elements in the Strawberry Poison Frog. *Mol Biol Evol* 35, 2913–2927 (2018).
- Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
- Shendure, J. *et al.* DNA sequencing at 40: past, present and future. *Nature* **550**, 345–353 (2017).
- Shulse, C. N. *et al.* High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* 27, 2241–2247 (2019).
- Xue, Y. *et al.* Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**, 242–245 (2015).

## General Support Availability Announcement

We pledge to make your experience this semester worth your effort.

**COVID-19 resources and support.** We care about your health and safety. Please check:

<https://coronavirus.berkeley.edu>

<https://uhs.berkeley.edu>

This includes support for mental health issues at Counseling and Psychological Services (CAPS)

<https://uhs.berkeley.edu/caps>

and Crisis Counseling for Urgent Concerns

<https://uhs.berkeley.edu/counseling/urgent>

**The Division of Equity & Inclusion** is providing support virtually for:

### Basic needs – Food and Housing

[https://docs.google.com/document/d/1WwPF-Q3Z8EXBfM-Wf\\_WwBzdTU39hfz85JL2F8Z5IfDE/edit](https://docs.google.com/document/d/1WwPF-Q3Z8EXBfM-Wf_WwBzdTU39hfz85JL2F8Z5IfDE/edit)

<https://sa.berkeley.edu/covid19>

### Disability Access & Compliance

<https://dac.berkeley.edu/message-campus-disability-community-re-covid-19>

**Health and Wellness**

<https://diversity.berkeley.edu/health-and-wellness>

**Student Technology Equity Program (STEP)**

Resource for connecting laptops, Wi-fi hotspots, and other technology you might need.

<https://technology.berkeley.edu/STEP>

**Academic Accommodations Hub**

This site provides support resources and academic accommodations to ensure all students have a fair chance at academic success. <https://evcp.berkeley.edu/programs-resources/academic-accommodations-hub>