# Computational Methods for Studying Demographic History with Molecular Data

BIOL 1435

Spring 2023 Syllabus

## Instructor Information

**Course Instructor:** David Peede
**Email:** david_peede@brown.edu
**Office Location:** 164 Angell Street, 3rd Floor, Room 319.
**Office Hours:** Tuesday/Thursday 4:00 pm - 5:00 pm and by appointment.
**Message to Students:** You have worked hard to be here; you belong here! Congratulations on all your achievements thus far, and welcome to BIOL 1435! I'm excited you are here, and I hope it will be a great semester!

## Course Information

**Meeting Times:** Tuesday/Thursday 2:30 pm - 3:50 pm.
**Meeting Location:** 111 Thayer Street, Watson Institute, Room 116.
**Course Website:** https://github.com/David-Peede/BIOL1435 and on Canvas.
**Course JupyterHub:** https://biol1435.jupyter.brown.edu
**Prerequisites:** MATH 0100, BIOL 0470 or 0480; an introductory statistics course (BIOL 0495 or equivalent) will also be helpful. Instructor permission required. The course will assume that students are comfortable with concepts in genetics and evolution covered in the prerequisites, but there will be brief refreshers in this course. This course has limited enrollment.

## Course Description

**Overview:** This course broadly covers the field of population genetics and genomics, and focuses on how inferences about demographic history and population structure can be made from genetic variation observed across populations today. The main question we will endeavor to answer in class is "How can we infer demographic history in a population using next-generation sequencing data?" Students will also learn how to apply computational tools/methods to infer demographic history using both simulated and real DNA sequencing data.
**Learning Goals:** By the end of this course students will have (1) an improved understanding of major concepts in population genetics; (2) gained some level of fluency with the Python programming language; (3) learned how to model and simulate genetic variation within and between

population(s); (4) applied the course material to publicly available genetic data-sets to make inferences about the evolutionary history of humans and other taxa; (5) analyzed, critiqued, and presented primary journal articles to understand how said articles have advanced the field and/or instigated debates within the field.

**Learning Activities, Assessments, and Allocation:** Over 13 weeks, students will spend three hours per week in class (39 hours total), one hour per week in preparation (13 hours total), and four hours per week in reading primary literature (52 hours total). Students will complete five problem sets (estimated 20-40 hours total), write one reaction paper (5 hours) and prepare for leading a discussion (10-15). The final project requires research, a presentation and write-up. This will take approximately 20-40 hours. Thus over the course of the semester students will be expected to invest approximately 180 hours.

## Academic Support

**Accessibility and Accommodations Statement:** Brown University is committed to full inclusion of all students. Please inform me early in the term if you may require accommodations or modification of any of course procedures. You may speak with me after class, during office hours, or by appointment. If you need accommodations around online learning or in classroom accommodations, please be sure to reach out to Student Accessibility Services (SAS) for their assistance (sas@brown.edu, 401-863-9588). Undergraduates in need of short-term academic advice or support can contact an academic dean in the College by emailing college@brown.edu. Graduate students may contact one of the deans in the Graduate School by emailing graduate_school@brown.edu.

**Diversity and Inclusion:** It is my intention to create and foster a learning environment that is inclusive, supports a diversity of thoughts, perspectives, and experiences. Our classroom is a learning environment, and as such should be a safe, inclusive, and respectful place. Disrespecting fellow students as well as combative approaches, tones and/or actions are not acceptable. Please make me aware if there are classroom dynamics that impede your (or someone else's) full engagement.

## Course Materials

**Required Materials:** There is no required textbook for this course that students should buy. ***However, students need access to a laptop to successfully complete this course.*** Please see me if you don't have access to one.

**Readings on Canvas:** Despite having no required text book, ***this course has a heavy reading load.*** Reading for this course will be drawn from a number of journal articles and the occasional book chapter.

**JupyterHub and OSCAR:** This semester I hope to streamline coding issues that students have encountered in the past by setting up a JupyterHub for our course. This is a relatively new resource available to Brown University instructors so more information to come. Additionally, all students will have OSCAR accounts for the entire semester.

# Assessment

## Paper Presentation and Discussion

Throughout the semester students will be asked to present and lead discussions on primary articles assigned for that lecture—hereafter referred to as "discussion leaders." The number of discussion leaders per paper, and how often each student is a discussion leader over the course of the semester, will be determined by the size of the class, but every student will be a discussion leader at least once during the semester. The discussion leaders will be asked to outline the question being addressed in the paper, the main conclusions of the paper, the methodological approaches, the scientific advances as they see it, and critique any weaknesses that are perceived. They will be assessed by the instructor in how comprehensive their understanding of the papers' concepts and conclusions are, their ability to convey that information and their responses to questions regarding the paper from students in the class and the instructor. ***If an emergency arises that prevents you from leading the discussion on your assigned date, notify your co-leaders and the instructor as early as possible by email.*** The students leading discussion must meet with the instructor at least one lecture prior to their scheduled discussion. The purpose of this hour-long meeting is to go over the paper(s) being discussed, receive feedback on the paper presentation slides, and to discuss how to facilitate the upcoming class discussion. ***The exact time of the meeting must be determined at least one week prior to the scheduled discussion; this is best done by email, which should be initiated by the future discussion leaders.***

## Quizzes

There will be four quizzes throughout the semester after completing the first four modules. These quizzes will be taken during the first 15 minutes of class and are designed to help reinforce key concepts.

## Problem Sets

Five problem sets will be assigned. Problem sets will roughly be due every two weeks. Problem sets will be computational in nature—they will consist of running simulations, working with real data, gathering and summarizing results to compare in a quantitative way the effect of different demographic scenarios. There will be multiple coding demonstrations in class and help writing scripts will be provided during office hours to facilitate running simulations and data munging/plotting so students can focus on the scientific conclusions. Any command line interface knowledge needed to run the scripts will be taught in the class. Grading will be based on students successfully producing results and correctly interpreting the results.

## Class Participation

Participation in class and discussions will also be used in determining course grades—see grading breakdown, below. ***Attendance is expected, if you are unable to attend class, please let me know in advance.*** Discussion from other students can include request to discussion leaders to clarify difficult parts of the paper, responses to questions posed by the instructor, and volunteering their own observations regarding the paper being discussed. Therefore, all students in the class must be familiar with the paper that will be under review prior to the discussion sessions.

## Reaction Paper

One two-page response to any of the assigned readings must be turned in over the course of the semester. Students may choose when their reaction paper is turned in, but students must turn them in prior to the beginning of class with said assigned reading. The discussion leaders may write a reaction paper about the scientific paper for which they are discussion leaders.

## Final Project

Students will complete a final project addressing a specific question in evolutionary genetics chosen by the student with the instructor's help using simulations and/or empirical data. Students must submit a one-page final project proposal—due before the beginning of class on 2MAR23—that includes: (1) the motivation for the research question; (2) the approaches used to study the research question; (3) two potential outcomes and their interpretations; and (4) the broader significance of the research. During the reading period, students will give 10 minute presentations going over the research they conducted over the semester. Lastly, at the end of the semester students will submit a two-page report synthesizing their research—due by the end of the final exam period on 10MAY23—which must include: (1) an introduction of the research topic; (2) methods used in the final project; (3) the results including at least one figure and one table; (4) a discussion of the results and proposed follow-up analyses; and (5) fully annotated code that can be used to replicate all results.

## Late Work Policy

Unexcused late assignments will receive a 25% deduction for each day late—eg if a student turns in an unexcused assignment one day late with no incorrect answers, said student would receive a 75%. Requests for extensions must be made 48 hours in prior to the due date, by email to the instructor, and extensions will be granted by the instructor on a case by case basis.

## Grading Breakdown

Detailed rubrics of how I will evaluate all assessments will be available on Canvas, but here is the grading breakdown for final grades:

- Class Participation - 5%
- Reaction Paper - 5%
- Problem Sets - 50%
  - Five problem sets each worth 10%
- Quizzes - 10%
  - Four quizzes each worth 2.5%
- Paper Presentation and Discussion - 10%
- Final Project - 20%
  - Proposal - 5%
  - Presentation - 5%
  - Report - 10%

## Tentative Course Schedule as of 13JAN23

The instructor reserves the right to make changes to the course schedule throughout the semester. Students will always be immediately notified via Canvas of changes to the schedule and any change will always benefit the students—ie due dates may be extended, but will never be shortened. Additionally, I have intentionally not planned anything for 20APR23 and 25APR23 to give myself some wiggle room with the course material, but there will be class on those days.

Key: Problem Set; Quiz; Final Project

### Module 1: Introduction to BIOL 1435 & Evolutionary Genetics

- **26JAN23:** Course overview and evolution refresher.
  - Welcome to your first day, no reading necessary!
- **31JAN23:** Describing variation and patterns of diversity.
  - Assigned readings
    - 1000 Genomes Project Consortium. "A global reference for human genetic variation." Nature 526.7571 (2015): 68.
    - Korunes, Katharine L., and Kieran Samuk. "`pixy`: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data." Molecular ecology resources 21.4 (2021): 1359-1368.
- **2FEB23:** Dave's example paper presentation and discussion.
  - Assigned reading
    - Konopiński, Maciej K. "Average weighted nucleotide diversity is more precise than `pixy` in estimating the true value of $\pi$ from sequence sets containing missing data." Molecular Ecology Resources (2022).
  - Problem set 1 assigned (due by 11:59 pm on 16FEB23).

### Module 2: Introduction to Coalescent Theory & Tree-Thinking

- **7FEB23:** The Wright-Fisher model and the standard coalescent.
  - Assigned readings
    - Rosenberg, Noah A., and Magnus Nordborg. "Genealogical trees, coalescent theory and the analysis of genetic polymorphisms." Nature Reviews Genetics 3.5 (2002): 380-390.
    - Cann, Rebecca L., Mark Stoneking, and Allan C. Wilson. "Mitochondrial DNA and human evolution." Nature 325.6099 (1987): 31-36.
    - Fu, Qiaomei, et al. "A revised timescale for human evolution based on ancient mitochondrial genomes." Current biology 23.7 (2013): 553-559.
  - Module 1 quiz.
- **9FEB23:** Measures of genetic distance.
  - Assigned readings

    * Rosenzweig, Benjamin K., et al. ”Powerful methods for detecting introgressed regions from population genomic data.” Molecular ecology 25.11 (2016): 2387-2397.

- **14FEB23:** Building trees.
  - Assigned readings
    * Cahpter 3 from Hein, Jotun, Mikkel Schierup, and Carsten Wiuf. Gene genealogies, variation and evolution: a primer in coalescent theory. Oxford University Press, USA, 2004.
    * Gascuel, Olivier, and Mike Steel. ”Neighbor-joining revealed.” Molecular biology and evolution 23.11 (2006): 1997-2000.
- **16FEB23:** Student-led paper presentation and discussion.
  - Assigned reading
    * Patterson, Nick, et al. ”Genetic evidence for complex speciation of humans and chimpanzees.” Nature 441.7097 (2006): 1103-1108.
  - Problem set 1 due by 11:59 pm.
  - Problem set 2 assigned (due by 11:59 pm on 2MAR23).

### Module 3: Population Differentiation & Demography

- **21FEB23: NO CLASS!**
- **23FEB23:** Population structure.
  - Assigned readings
    * Patterson, Nick, Alkes L. Price, and David Reich. ”Population structure and eigenanalysis.” PLoS genetics 2.12 (2006): e190.
    * Bhatia, Gaurav, et al. ”Estimating and interpreting $F_{ST}$: the impact of rare variants.” Genome research 23.9 (2013): 1514-1521.
  - Module 2 quiz.
- **28FEB23:** Demographic inference.
  - Assigned readings
    * Mather, Niklas, Samuel M. Traves, and Simon YW Ho. ”A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data.” Ecology and evolution 10.1 (2020): 579-589.
    * Kamm, Jack, et al. ”Efficiently inferring the demographic history of many populations with allele count data.” Journal of the American Statistical Association 115.531 (2020): 1472-1487.
- **2MAR23:** Admixture and introgression.
  - Assigned readings
    * Durand, Eric Y., et al. ”Testing for ancient admixture between closely related populations.” Molecular biology and evolution 28.8 (2011): 2239-2252.

    - Dagilis, Andrius J., et al. "A need for standardized reporting of introgression: Insights from studies across eukaryotes." Evolution Letters 6.5 (2022): 344-357.
  - Final project proposals due **before** class—ie by 2:29 pm.
  - Problem set 2 due by 11:59 pm.
  - Problem set 3 assigned (due by 11:59 pm on 16MAR23).
- **7MAR23:** Student-led paper presentation and discussion.
  - Assigned reading
    - Coughlan, Jenn M., et al. "Patterns of population structure and introgression among recently differentiated *Drosophila melanogaster* populations." Molecular Biology and Evolution 39.11 (2022): msac223.

### Module 4: Recombination & Selection

- **9MAR23:** The coalescent with recombination.
  - Assigned readings
    - Cahpter 5 from Hein, Jotun, Mikkel Schierup, and Carsten Wiuf. Gene genealogies, variation and evolution: a primer in coalescent theory. Oxford University Press, USA, 2004.
  - Module 3 quiz.
- **14MAR23:** Direct selection.
  - Assigned readings
    - Eyre-Walker, Adam. "The genomic rate of adaptive evolution." Trends in ecology & evolution 21.10 (2006): 569-575.
- **16MAR23:** Linked selection.
  - Assigned readings
    - Charlesworth, Brian, and Jeffrey D. Jensen. "Effects of selection at linked sites on patterns of genetic variability." Annu Rev Ecol Evol Syst 52.1 (2021): 177-197.
  - Problem set 3 due by 11:59 pm.
  - Problem set 4 assigned (due by 11:59 pm on 4APR23).
- **21MAR23:** Student-led paper presentation and discussion.
  - Assigned reading
    - Huerta-Sánchez, Emilia, et al. "Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA." Nature 512.7513 (2014): 194-197.
- **23MAR23:** Student-led paper presentation and discussion.
  - Assigned reading
    - Calfee, Erin, et al. "Selective sorting of ancestral introgression in maize and teosinte along an elevational cline." PLoS genetics 17.10 (2021): e1009810.

### Module 5: Advanced Topics in Population Genomics

- **28MAR23: NO CLASS!**
- **30MAR23: NO CLASS!**
- **4APR23:** Speciation genomics.
  - Assigned readings
    * Seehausen, Ole, et al. "Genomics and the origin of species." Nature Reviews Genetics 15.3 (2014): 176-192.
  - Problem set 4 due by 11:59 pm.
  - Problem set 5 assigned (due by 11:59 pm on 18APR23).
- **6APR23:** Student-led paper presentation and discussion.
  - Assigned reading
    * Brandvain, Yaniv, et al. "Speciation and introgression between *Mimulus nasutus* and *Mimulus guttatus*." PLoS genetics 10.6 (2014): e1004410.
  - Module 4 quiz.
- **11APR23:** Hidden Markov Models in population genomics.
  - Assigned readings
    * Kern, Andrew D., and David Haussler. "A population genetic hidden Markov model for detecting genomic regions under selection." Molecular biology and evolution 27.7 (2010): 1673-1685.
- **13APR23:** Student-led paper presentation and discussion.
  - Assigned reading
    * Corbett-Detig, Russell, and Rasmus Nielsen. "A hidden Markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy." PLoS genetics 13.1 (2017): e1006529.
- **18APR23:** Machine Learning in population genomics.
  - Assigned readings
    * Schrider, D. R., & Kern, A. D. (2018). Supervised machine learning for population genetics: a new paradigm. Trends in Genetics, 34(4), 301-312.
  - Problem set 5 due by 11:59 pm.
- **20APR23:** TBA
  - Assigned readings
    * TBA
- **25APR23:** TBA
  - Assigned readings
    * TBA

### Final Projects

- **28APR23:** Final project presentations.
- **2MAY23:** Final project presentations.
- **4MAY23:** Final project presentations.
- **9MAY23:** Final project presentations.
- **10MAY23:** Final project reports due.