

Computational Evolutionary Biology

Class Meeting

Lectures:

Tuesdays and Thursdays 2:00-3:15, Dirac Science Library Room 152

Lab session:

Mondays 12:20-2:25, Dirac Science Library Room 152

Instructor

Peter Beerli

150-T DSL

beerli@scs.fsu.edu

(850) 645-1324 or (850) 559-9664

Assistant

Michal Palczewski

150-J DSL

michalp@gmail.com

(850) 345-6017

Office Hours

Tuesday and Thursday after class or by appointment.

Objectives

This course will introduce students to methods used in phylogenetics and population genetics and explore the computational complexity of such methods. Primary objectives of the course are:

1. to expose students to a large set of modern methods used in the field of theoretical evolutionary biology
2. introduce students to the programming aspects of the field
3. enable students to develop test methods to compare different inference programs in phylogenetics and population genetics.

Content

Advanced computational methods are becoming increasingly important in biology. A wide range of applications — including, for instance, identifying pathogens, tracing viral transmission pathways, and reconstructing the geographic expansion of humans out of Africa — rely on evolutionary inference. This course will cover the methods currently used for evolutionary inference, the stochastic models and inference principles they are based on, and how they are implemented in practice. In a separate lab-session we will discuss and explore simulation of genetic data under different population genetic methods and finally we will compare available computer programs using these simulated data. These comparison will be the first step to write a manuscript that discusses the estimation biases of the inference programs when their model assumptions are violated.

Textbook

No textbook required. If you think you need a textbook to understand the methods then I suggest to use

- Felsenstein, Joseph. 2004. Inferring Phylogenies. Sinauer, Sunderland MA.
- Semple, Charles, and Steel, Mike. 2003. Phylogenetics. Oxford Lecture Series in mathematics and its application 24.
- Wakeley, John. 2008. Coalescent theory. Roberts & Company Publishers, Greenwood Village Colorado.

These textbook are great starting points for further exploration of the subject.

Grading

- Grades will be based on students' execution of the 5 assignments [5 points each]
- Each student will do a project during the last 5 weeks of the semester and also give a short presentation of their work in the last week of class. [20 project+10 presentation]
- We will have a theory test on October 15th (midterm) and a final test in finals week. [10 points, each]

A student who accumulates 90% or more of the possible points will receive a grade of "A", a student who accumulates between 80% and 89% of the possible points will receive a grade of "B", a student who accumulates between 70% and 79% of the possible points will receive a grade of "C", a student who accumulates between 60% and 69% of the possible points will receive a grade of "D", and a student who accumulates less than 60% of the possible points will receive a grade of "F".

Attendance

Attendance in class is strongly recommended. While there is no penalty for failure to attend, the instructor may not be able to review with you any material that you miss by nonattendance unless you have a valid reason for absence (illness, death of a family member, professional conflict, subpoena to court, varsity athletics, religious holiday) or make a prior arrangement.

Missed/Late Assignments

Deadlines for assignments will be announced in class; late assignments will be accepted for full grade only in cases of illness or death in the family. 10% of the points are deducted for each day late assignments.

Florida State Honor Code

Students are expected to uphold the Academic Honor Code published in The Florida State University Bulletin and the Student Handbook, The first paragraph says: The Academic Honor System of Florida State University is based on the premise that each student has the responsibility to uphold the highest standards of academic integrity in the student's own work.

ADA Policy, Students with disabilities

Students with disabilities needing academic accommodations should: (1) Register with and provide documentation to the Student Disability Resource Center [SDRC]) in room 108 in the Student Services Building (644-9566). (2) Bring me a letter from the SDRC indicating you need academic accommodations. This should be done within the first week of classes. For more information on this see the Resource Center's web site (<http://www.disabilitycenter.fsu.edu/index.html>).

If you need alternative accommodations for any reason, please notify the instructor.

Lectures: Topic overview

1. Processes and patterns

- Population genetics: Wright-Fisher population models, Moran model, coalescence theory, genetic drift, gene flow;
- Phylogenetics: tree structures, speciation, birth/death models
- Mutation models: real data (microsatellite markers, SNPs, sequences), modeling data, mutation/substitution model
- Simulation of data

2. Inference:

- Parsimony and Distance methods
- Maximum likelihood, Bayesian inference, Markov chain Monte Carlo,
- Approximate Bayesian Computation
- Assignment methods
- Bootstrap/Jackknife

Each topic will include computational algorithm, problematic aspects such as convergence, biases, main focus will be on Bayesian and maximum likelihood inference

Laboratory

The final goal is to simulate genetic data using particular population models (and model parameter sets) and analyze these data sets with commonly used computer programs. This study will allow to establish biases, and give guidance on how many loci or individuals are needed for a proper analysis. To achieve this goal we will start out by learning the basics of the programming language Python and also learn some more about the UNIX environment. We will have maximally 5 assignments and a presentation where the students report on their progress to simulate particular population genetic scenario, and their findings: details on the projects will be forthcoming. The assignments will be

1. Calculate the time of the most recent common ancestor of a sample
2. Print out a coalescent tree in Newick format where the inputs are the number of samples and the effective population size (the tree and its depths are dependent on these two)
3. Generate a new DNA sequence using a template sequence and a mutation model, for example HKY, F84, or GTR.
4. Generate datasets using the previous assignments
5. Construct a very simple ABC sampler to estimate the effective population size using a small sample of genetic data.

The lab section will be used to learn Python/Unix/.... and discuss papers that are related to the project. Programming may not really be needed to do the project but I hope that you will use these tools to summarize results of the outputs etc.