

## **Undergraduate course: Computational Evolutionary Biology**

### **Class Meeting**

Lectures:

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

Lab-session:

Tuesdays and Thursdays 12:30-2:00 PM Dirac Science Library Room 152

### **Instructor**

Peter Beerli

Office: 150-T DSL

Email: beerli@fsu.edu

Phone: (850) 559-9664

### **Teaching Assistant**

Michal Palczewski

Office: 150-J DSL

Email: michalp@gmail.com

Phone: (850) 345-6017

### **Office Hours**

Monday 1:00-3:00 PM or by appointment.

### **Objectives**

This course will introduce students to methods used in phylogenetics and population genetics and writing computer programs using 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, and learn about the details of often used methods in phylogenetic analysis and population genetics analysis.
2. to introduce students to the programming aspects of the field. Students will learn and use the programming language Python to develop scripts and to understand details of the methods.
3. to empower students to develop programming and analysis skills that involve development of scripts to change data format, execute applications, and analyze results.

## Rational for the objectives

Current biological studies do not emphasize the landslide changes in data analysis that are happening in biology. Within a few years, most biologists, graduate students or biology majors, will be at a disadvantage if no courses exist that improve the computational skills of students. This course prepares students for such a challenge.

## 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, the students will get hands-on experience in developing computational software implementing these methods. We expect that the students leave the course with the necessary skills to develop their own ideas and are able to develop projects that are based on simulated data sets and scripts.

## Textbook

Yang, Z. 2006. Computational Molecular Evolution. Oxford University Press, Oxford, England. (Book website: <http://abacus.gene.ucl.ac.uk/CME/>)

The textbook includes material for almost all of the major topics that we expect to cover. There is a small amount of material that will be presented in class that is not covered by the textbook. This textbook is a great starting point for further exploration of the subject and we certainly will not have time to explore all chapters.

## Grading

- Grades will be based on students' execution of the 5 programming assignments, each of which involves understanding the algorithms, code design, and program documentation [10 points each]
- Each student will work a project on their own during the last 4 weeks of the semester and also give a short presentation of their work in the last regular lab-meeting. [15 points for the report and 15 points for the presentation]
- We will have a theory test on November 8th (midterm). [20 points]
- There will be no final exam, the project substitutes for a final examination.

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".

### **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 per day for late assignments.

### **University Attendance Policy**

Excused absences include documented illness, deaths in the family and other documented crises, call to active military duty or jury duty, religious holy days, and official University activities. These absences will be accommodated in a way that does not arbitrarily penalize students who have a valid excuse. Consideration will also be given to students whose dependent children experience serious illness.

### **Academic Honor Policy**

The Florida State University Academic Honor Policy outlines the University's expectations for the integrity of students' academic work, the procedures for resolving alleged violations of those expectations, and the rights and responsibilities of students and faculty members throughout the process. Students are responsible for reading the Academic Honor Policy and for living up to their pledge to . . . be honest and truthful and . . . [to] strive for personal and institutional integrity at Florida State University. (Florida State University Academic Honor Policy, found at <http://dof.fsu.edu/honorpolicy.htm>.)

### **Americans With Disabilities Act**

Students with disabilities needing academic accommodation should: (1) register with and provide documentation to the Student Disability Resource Center; and (2) bring a letter to the instructor indicating the need for accommodation and what type. This should be done during the first week of class. This syllabus and other class materials are available in alternative format upon request. For more information about services available to FSU students with disabilities, contact the:

Student Disability Resource Center  
874 Traditions Way 108  
Student Services Building  
Florida State University  
Tallahassee, FL 32306-4167  
(850) 644-9566 (voice)  
(850) 644-8504 (TDD)  
[sdrc@admin.fsu.edu](mailto:sdrc@admin.fsu.edu)  
<http://www.disabilitycenter.fsu.edu/>

**Free Tutoring from FSU**

For tutoring and writing help in any course at Florida State University, visit the Academic Center for Excellence (ACE) Tutoring Services comprehensive list of tutoring options - see <http://ace.fsu.edu/tutoring> or contact [tutor@fsu.edu](mailto:tutor@fsu.edu) for more information. High-quality tutoring is available by appointment and on a walk-in basis. These services are offered by tutors trained to encourage the highest level of individual academic success while upholding personal academic integrity.

**Syllabus Change Policy**

Except for changes that substantially affect implementation of the evaluation (grading) statement, this syllabus is a guide for the course and is subject to change with advance notice.

## Lectures: Topic overview

### 1. Processes and patterns

- Population genetics: Wright-Fisher population models, Moran model, coalescence theory, genetic drift, gene flow, selection;
- Phylogenetics: tree structures, speciation, birth/death models, Gene tree versus Species tree
- 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
- Model selection
- Bootstrap/Jackknife

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

## Laboratory

This list of assignments is an example, difficulty of assignments will depend on the overall class programming skills. Each Lab topic will be introduced in detail during the lab, parts of the assignment will be coded during lab time.

1. Transform a microsatellite spread sheet file into a datafile for another computer program (for example MIGRATE or Arlequin)
2. Write a tree structure
3. Simulate data on a branch
4. Simulate coalescence times of a sample of  $k$  individuals
5. TBA
6. Project: The project will discuss either a complex analysis of data or software development. The project consists of two parts, a report (of not more than 8 pages) and a presentation of 10 minutes. During class and lab we will develop ideas for the project.

## Lecture Schedule

1. Introduction. Trees and tree representation (Aug. 30)
2. Parsimony. Counting evolutionary change (Sep 1)
  - Fitch parsimony
  - Wagner parsimony
  - Other parsimony variants
  - Sankoff cost matrices
3. Searching for the best tree(s) (Sep 6)
  - How many trees are there
  - Explicit enumeration
  - Branch and bound
  - Heuristic methods
4. Stochastic models of evolution I (Sep 8)
  - Nucleotide models: JC, F81, HKY, GTR
  - Rate variation across sites: Gamma
5. Stochastic models of evolution II (Sep 13)
  - Protein models
  - Codon models
  - Other DNA type model
6. Stochastic models of evolution III (Sep 15)
  - Restriction sites
  - Microsats
  - AFLP
7. Stochastic models of evolution IV (Sep 20)
  - Quantitative characters
  - Morphology
8. Hidden Markov models (Sep 22)
  - General introduction to
  - Rate variation across sites: the autocorrelated gamma
  - Mixture models
9. Maximum likelihood inference I (Sep 27)

- General principles
  - Coin-tossing
  - Statistical consistency and efficiency
10. Maximum likelihood inference II (Sep 29)
- Conditional likelihoods
  - Calculating likelihoods of evolutionary models
11. Bayesian inference (Oct 4)
- General principles
  - Coin tossing
  - Statistical consistency and efficiency
  - Priors
12. Markov chain Monte Carlo I (Oct 6)
- Gibbs and Metropolis samplers
  - Calculating Hastings ratios
13. Markov chain Monte Carlo II (Oct 11)
- Convergence and mixing
  - Metropolis-coupling
14. Trees and tree models I (Oct 13)
- Molecular clocks
  - Relaxed clock models
15. Trees and tree models II (Oct 18)
- Birth-death process
  - Covarion models
16. Population models (Oct 20)
17. The coalescent I (Oct 25)
- The simple n-coalescent
  - Relation to other population genetic models
18. The coalescent II (Oct 27)
- The structured coalescent
  - Coalescent and selection and other complications

19. Implementing the coalescent (Nov 1)
  - ML implementations
  - Bayesian implementations
20. Review session (Nov 3)
21. Mid term (Nov 8)
22. Lab (Nov 10)
23. Phylogeography (Nov 15)
  - Haplotype networks
  - Nested clade analysis
  - Statistical developments
24. Gene trees and species trees 1 (Nov 17)
25. Gene trees and species trees 2 (Nov 22)
26. Model selection and model averaging I (Nov 29)
  - Hierarchical and nonhierarchical models
  - Hierarchical likelihood ratio test
  - Akaike information criterion
  - Bayes information criterion
27. Model selection and model averaging II (Dec 1 )
  - The Bayesian approach: Bayes factors
  - Model averaging
28. Bootstrapping and jackknifing (Dec 6)
  - Nonparametric and parametric approaches to confidence
  - Phylogenetic implementations
  - Bootstrap corrections
  - Bootstraps and posterior probabilities
29. Statistical multiple sequence alignment (Dec 8)
  - Heuristic dynamic programming approaches
  - Bayesian MCMC sampling of multiple alignments