

Mathematics 660: Advanced Mathematical Modeling
(Theory of Phylogenetics)
Spring 2009

Instructors: Elizabeth Allman, 303B Chapman, 474-2479, e.allman@uaf.edu

Office Hours: M 1–2, T 8:30-9:30, R 1–2

John Rhodes, 102 Chapman, 474-5445, j.rhodes@uaf.edu

Office Hours: T R 11:45–12:45, and by appointment

Web page: <http://www.dms.uaf.edu/~jrhodes/M660.html>

Prerequisites: Flexible. Useful background material includes linear algebra, probability and statistics, and some biology. However, a motivated student proficient in one area can pick up the basics as we proceed. In particular, we may offer additional (out-of-class) tutorials on some of the mathematical tools we need that biologists are not familiar with.

Credit Hours: 3.0

Textbook: Inferring Phylogenies, by J. Felsenstein, Sinauer Associates; and course notes provided by instructors

Class Meetings: T R 9:45 –11:15 in Chapman 107

Exams: Midterm: Thursday, March 19 (no class March 17),
Final: Saturday, May 9, 8-10am

Course overview and goals:

Phylogenetics is concerned with inferring evolutionary relationships between some collection of organisms (typically, representatives of different species). Most commonly these days, inference is based on DNA or protein sequence data, though historically morphological characteristics were used. Although phylogenetic problems arise from biology, they are rather quickly abstracted to mathematical ones. The techniques used to address them involve a mix of mathematics, statistics, and computer science, but informed by an understanding of what is biologically plausible.

The early mathematical approaches to analyzing phylogenetic data were primarily combinatorial in nature, with a strong algorithmic flavor. Although these methods were quickly adopted to provide new biological insights, and are still used by some, they lacked a firm statistical basis. Making evolutionary inferences in a statistical framework requires formulating an explicit mathematical model of the evolutionary process. Although such a model will at best be a caricature of the true process, by capturing the most important features of the process it can lead to better inference procedures. Understanding the limitations, both biological and mathematical, of phylogenetic models is necessary to use them wisely.

This course will survey the main approaches to phylogenetic inference, including parsimony and distanced-based methods, but emphasizing the model-based methods of maximum likelihood and Bayesian analysis. Throughout, the emphasis will be on understanding the mathematics behind these, so that strengths and weaknesses of each are highlighted. While we will touch on software implementing various approaches, the goal is not to make you a proficient user of any package, but rather to make the packages something other than ‘black boxes’ which analyze data for you. Although this course is definitely not one on how to perform phylogenetic inference in practice, it should be useful to those who expect to be primarily users of these methods.

Phylogenetics is interdisciplinary; biology, computer science, statistics, and mathematics all play important roles. The students in this class come from several of these fields, and can contribute by providing differing disciplinary perspectives. No one person is likely to have the ‘right’ background across the board. What is more important is a willingness to contribute and ask questions.

Mechanics of the course:

Lectures The instructors will each deliver about half the class lectures, according to a secret schedule. Both of us are happy to be interrupted with questions and comments at any point. It is particularly important to speak up if you do not understand some background material. You’ll be doing a favor to other students who are too shy to ask.

If you miss class, you should get notes from another student, and check the class web page for any handouts.

Textbook The Felsenstein book is a comprehensive overview of phylogenetics by a leader in its development. However, it is not a formal mathematics book, and does not always explain necessary background. It also lacks any exercises, though working through problems is essential to developing a solid understanding. We will therefore be supplementing it by course notes that are intended to provide a more methodical development. As these notes are being developed through the semester, we will hand them out as they become available.

Homework You are expected to read and assimilate assigned sections of the textbook/course notes. Readings may sometimes include material dealt with only cursorily, or not at all, in lectures.

Homework problems will usually be assigned daily, but only collected each Tuesday (due in class, but accepted until 5pm). Homework problems will be posted on the course web page as they are assigned.

Late homework that has not been approved ahead of time will not be accepted. *There will be no exceptions to this* other than for a genuine emergency (e.g., a death in the family, documented illness, etc.).

We encourage you to work with others on the homework, *but to write up the solutions independently*. In writing up your work, you should present your arguments in such a way that an intelligent, but ignorant, person can understand them. In particular complete sentences and a logical presentation are expected.

Examinations Both the midterm and final examinations will be in-class, closed-book exams. For missed examinations that are not approved in advance, no

make-up exams will be given, except in case of emergencies.

Project Each student will complete a project in this course, which can be tailored to particular interests. We encourage students to work in pairs on these.

Possibilities include delving more deeply into a topic introduced in class by reading and presenting appropriate research papers, developing your own ideas for a variation on some phylogenetic method, writing software to implement some calculation, exploring a real data set thoroughly through many different phylogenetic approaches, or using simulated data to investigate the limits of some of these approaches. We are open to other ideas as well – the only thing ruled out is a ‘straightforward’ phylogenetic analysis of the sort a biologist might do to obtain a single tree.

Midway through the course, you will submit for approval a brief description of the project you would like to undertake.

Grades Your performance will be evaluated based on 15% homework, 5% class participation, 25% midterm exam, 30% final exam, 25% project. Course grades will be determined according to the following cutoffs:

$$A \geq 90\%, \quad B \geq 80\%, \quad C \geq 70\%, \quad D \geq 60\%.$$

We reserve the right to move the cutoff points downward if particular exams turn out to be unexpectedly difficult. Note that you are not in competition with your peers – everyone in the class may get an *A*, or everyone may get an *F*.

University and Department Policies Your work in this course is governed by the UAF Honor Code. The Department of Mathematics and Statistics has specific policies on incompletes, late withdrawals, and early final exams which can be found at

<http://www.dms.uaf.edu/dms/Policies.html>.

If you have any disabilities that your instructors should know about, bring them to our attention soon so that we can work with the Office of Disability Services to set up any necessary accommodations.

Theory of Phylogenetics

MATH 660: Advanced Mathematical Modeling

TR 9:45-11:15

Tentative Schedule

Week	Dates	Topics
1	Jan 27, 29	Molecular evolution, combinatorics of trees
2	Feb 3, 5	Parsimony, tree space, tree metrics
3	Feb 10, 12	Distance algorithms
4	Feb 17, 19	Probabilistic models of DNA mutation
5	Feb 24, 26	Phylogenetic distances
6	Mar 3, 5	Maximum likelihood, moving through tree space,
	Mar 10, 12	Spring break
7	Mar 19	Midterm; no class Mar 17
8	Mar 24, 26	Rate-variation models, codon and protein models, Covarion models, Lewis model
9	Mar 31, Apr 2	Model testing, bootstrapping, consistency, long branch attraction
10	Apr 7, 9	Gene tree/species tree concordance, consensus methods, more combinatorics of trees
11	Apr 14, 16	Bayesian methods
12	Apr 21, 23	Networks
13	Apr 28, 30	Networks