MATH 660: Theory of Phylogenetics
(Advanced Mathematical Modeling)

MWF 9:15 – 10:15
Duckering 352, F01 and video FM1
https://eallman.github.io/classes/660/660_2017.html

Instructor: Elizabeth S. Allman, Chapman 308B, e.allman@alaska.edu and 474-2479.

Office Hours: Mon 10:30 - 11:30, Wed 8:00 - 9:00, and by appointment. These office hours may change at a later date, depending on student demand and scheduling concerns. Please note that the best way to reach me is by e-mail.

Prerequisities: 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, there may be additional (out-of-class) tutorials on some of the mathematical tools we need, if students are unfamiliar with these topics.

Textbook: Course notes provided by the instructor. An excellent (optional) resource is <u>Inferring Phylogenies</u>, by J. Felsenstein, Sinauer Associates. This optional book is a comprehensive overview of phylogenetics by a leader in its development.

Grading: There will be one midterm, one project, and a final exam in MATH 660. In addition, there will be weekly homework assignments and several labs. Grades will be assigned using the following weights:

Homework and Labs	25 %
Midterm	25 %
Project	25 %
Final Exam	25 %

The midterm is tentatively scheduled for Friday, March 9. The final exam will take place as published in the schedule of courses, on Friday, May 4 from 8:00 - 10:00. Makeup exams will not be given.

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 charicature 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. With the availability of genomic data (full genomes or hundreds or thousands of genes from an individual), inference of a species trees from such large datasets is a hot topic of current research. Our class will study the gene tree/species tree problem as its final topic.

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.

Course Outline: MATH 660 is a three hour course, and class meetings will typically be in lecture format, but I encourage you to ask questions and participate actively in class. This makes the hour more interesting and lively for all, and ensures that students are actively engaged. Furthermore, any questions you might have on background you are lacking are likely to be shared with other (perhaps more shy) students. Asking for clarifications helps everyone. Questions on homework and labs should primarily be asked during office hours. If a particular topic is stumping the class, then we will of course make time clarifying it during lecture.

Homework will be assigned weekly on the class webpage and collected on Wednesdays at 5 pm. Homework papers may be turned into my box or slipped under my office door. Remote students are expected to post their homework assignment in their shared drive by 5:00 pm on Wednesdays. Remote students (only) can send me a copy of their homework electronically by the Wednesday deadline, and I will print it out. You should get into the habit of checking the class webpage regularly. Class announcements and other useful information will be posted there in addition to assignments. Late homeworks, both problems and labs, will not be accepted.

For a tentative schedule of topics and midterm date, check the last page of this syllabus. Mark your datebooks now with the time for our final exam: Friday, May 4, 8:00 - 10:00. Students must take the final exam at the scheduled time.

Project: Each student will complete a project in this course, which can be tailored to particular interests. The project will have both a written and possibly an oral component.

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. Other ideas are okay as well – the only thing ruled out is a 'straightforward' phylogenetic analysis of the sort a biologists 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.

Other Policies:

Course accommodations: If you need course adaptations or accommodations because of a disability, please inform your instructor during the first week of the semester, after consulting with the Office of Disability Services, 203 Whitaker (474-7403).

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, some of which are listed below. A complete listing can be found at
https://www.uaf.edu/dms/policies/.

Prerequisites: The prerequisite for MATH 660 is the permission of the instructor. Details were given above.

Late Withdrawal: This semester the last day for withdrawing with a 'W' appearing on your transcript is Friday, March 30. If, in my opinion, a student is not participating adequately in the class, I may elect to drop this student.

Graded Coursework: Please keep all graded work for MATH 660 until final grades have been assigned.

Academic Honesty: Academic dishonesty, including cheating and plagiarism, will not be tolerated. It is a violation of the Student Code of Conduct and will be punished according to UAF procedures.

Grade Bands: A, A- (90 - 100%), B+,B, B- (80 - 89%), C+, C, C- (70 - 79%), D+, D, D- (60 - 69%), F (0 - 59%). On rare occasion, I may lower the thresholds. Also, in an effort to reward the student who makes significant improvement over the course of the term, a stellar grade on the final may overcome a deficiency on the midterm and improve a student's final grade.

Courtesies: As a courtesy to your instructor and fellow students, please arrive to class on time, turn your cell phones and electronic gadgets off during class, and pay attention in class.

Tentative Schedule

	Dates	Content	Comments
Week 1	Jan 17, 19	Molecular evolution, combinatorics of trees	
Week 2	Jan 22, 24, 26	Combinatorics (cont.)	
		Parsimony, weighted and unweighted	
Week 3	Jan 29, 31, Feb 2	Splits, clades, consensus trees and supertrees	
Week 4	Feb 5, 7, 9	Distance algorithms	
Week 5	Feb 12, 14, 16	Probabilistic models of DNA mutation	
Week 6	Feb 19, 21, 23	Probabilistic models (cont.)	
Week 7	Feb 26, 28, Mar 2	Phylogenetic distances	
Week 8	Mar 5, 7, 9	Maximum likelihood	Midterm F Mar 9
	Mar 12 - 16	Spring break	
Week 9	Mar 19, 21, 23	Tree space	
Week 10	Mar 26, 28, 30	Rate-variation models, codon and protein models,	Friday, March 30 last day
		Covarion models, Lewis model, mixture models	to withdraw without a W
Week 11	Apr 2, 4, 6	Model testing, bootstrapping, consistency,	
		long branch attraction	
Week 12	Apr 9, 11, 13	Bayesian methods	
Week 13	Apr 16, 18, 20	Species tree/gene tree concordance	
		and the coalescent model	
Week 14	Apr 23, 25, 27	Species tree/gene tree inference methods	
Week +	Apr 30		
Final Exam	Friday, May 4		Summer!
	8:00 - 10:00		

Additional information for remote students.

To make things run smoothly, there are a number of things you will have to attend to from your remote location in Anchorage.

1. Finding a room at UAA for watching and participating in class lectures.

2. Homeworks: Due by Wednesdays at 5 pm. You can either send an electronic copy by e-mail, or place a copy in a shared drive.

3. Office Hours: Please send questions by e-mail and if we need more 'in person' time, we can try skyping or Facetime or some other collaborative method.

4. Exams: You must find a way to have your exams proctored, or somehow determine a way that you can take the exam in a secure environment. Please do not leave this to the last moment.

5. Please try to be online by 9:11 so that we can be sure that the connection is up and running before the start of class at 9:15.