Mathematical and Statistical Phylogenetics
Biomath/Human Genetics M211 – Spring 2013
MW 9-11AM, LS 5236

Instructor: Prof. Marc Suchard, Rm 6558, Gonda

Office Hours: By request (email `msuchard@ucla.edu` to make appointment).

Course Topics: Theoretical models in molecular evolution focusing on phylogenetic techniques. Topics include evolutionary tree reconstruction methods, studies of viral evolution, phylogeography and coalescent approaches. Examples provided from evolutionary biology and evolutionary medicine, with a unique focus on implications for human disease processes. I would appreciate suggestions for special topics that may interest the class.

Prerequisites: An undergraduate courses in statistics and probability. Some computer programming experience and familiarity with genetics terms is highly recommended. For those looking for a probability refresher, I suggest *Introduction to Probability Models* by Sheldon Ross.

Course Texts: The main text for this course is *Computational Molecular Evolution* by Yang. A supplementary book, *Analysis of Phylogenetics and Evolution in R* by Paradis, is more hands-on. Also highly recommended are *Inferring Phylogenies* by Felnsenstein and *Statistical Methods in Molecular Evolution* edited by Nielsen. *Biological Sequence Analysis: Probabilistic Models of Protein and Nucleic Acids* by Durbin, Eddy, Krogh and Mitchison provides a sound introduction to tree reconstruction. Two books with practical examples of how to use current phylogenetic software are *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny* edited by Lemey and Vandamme and *Phylogenetic Trees Made Easy* by Hall. For those more mathematically inclined, try *Mathematics of Evolution & Phylogeny* edited by Gascuel and *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory* by Hein, Schierup and Wuif.

Computing: Many homework problems and all (unless highly theoretical) final projects will require phylogenetic computation. I have reserved the Biomath computer lab on Wednesdays for sporadic in-class data analysis examples and time for students to familiarize themselves with several existing phylogenetic programs. Free packages include:

- Geneious – convenient framework for analyzing molecular sequence data, excellent user-interface and plugins to run most evolutionary biology analyses, `http://www.geneious.com` (Biomatters)

- PHYLIP – large collection of different parsimony-, distance- and likelihood-based programs, `http://evolution.genetics.washington.edu/phylip.html` (Felsenstein)

- PAML – likelihood- and empirical Bayes-based nucleotide and codon programs, `http://abacus.gene.ucl.ac.uk/software/paml.html` (Yang)

- MrBayes – Bayesian sampler for standard continuous-time Markov chain models, `http://mrbayes.csit.fsu.edu/index.php` (Ronquist, Huelsenbeck, Larget, Simon, Suchard)

- BEAST – Bayesian sampler for coalescent models, `http://evolve.zoo.ox.ac.uk/beast/` (Drummond, Rambaut, Suchard)

- DualBrothers – Bayesian sampler for multiple change-point models for recombination and horizontal gene transfer, `http://www.biomath.ucla.edu/msuchard/DualBrothers/` (Minin, Dorman, Kitchen, Suchard)

- BAli-Phy – Bayesian sampler for joint estimation of alignment and phylogeny, `http://www.biomath.ucla.edu/msuchard/bali-phy/` (Redelings, Suchard)

- FigTree – flexible tree drawing program, `http://tree.bio.ed.ac.uk/software/figtree/` (Rambaut)

Internet: There is a web page for this course at `http://www.biomath.ucla.edu/msuchard/bm211`. Please visit it often to check for class news and useful links. This page will also have an archive of class handouts and readings. The web page links to the above software. Also note that `http://evolution.gs.washington.edu/book/datasets.html` provides all datasets used in *Inferring Phylogenies*.

Grading: Grades will be based on the homework (50%) and final phylogenetic analysis project (50%).

Homework: I plan to give semi-weekly problem sets taken primarily from the readings. These assignments will be due one week after their distribution; due dates will be clearly marked on the assignment sheets and late homework will be accepted only under the most dire circumstances.

Data Analysis Project: You are expected to perform an original phylogenetic data analysis project during this course and present your model developments or findings in a mini-seminar series (approximately 15-20 mins per student) during the last week of classes. Additionally, a written report of your work (max. 10 typed pages) is due on 6/14 by 5PM via email (PDF). You are responsible for finding an appropriate data set and designing your study. One page project proposals are due

5/17 by 5PM via email (PDF). You should plan on attending several Biostatistics (Wednesdays at 3:30PM) and Human Genetics (Mondays at 11:00AM) department seminars to get a feeling for the appropriate format. Philippe Lemey and his group are visiting UCLA during Finals weeks and are happy to discuss projects with students.

Table 1: Course Outline

Week	Class Dates	Tentative Topic	Due
1	4/1, 4/3	Intro/Trees and Parsimony	
2	4/8, 4/10	Beyond Trees/Distance Methods	
3	4/15, 4/17	Continuous-Time Markov Models	
4	4/22, 4/24	Likelihood Methods	
5	4/29, 5/1	Bayesian Phylogenetics	
6	5/6, 5/8	Model Selection	
7	5/13, 5/15	Comparative Methods/Phenotype Models	Final Proposal (5/17)
8	5/20, 5/22	Recombination/Phylogeography	
9	5/27, 5/29	Holiday/Joint Alignment and Phylogeny	
10	6/3, 6/5	Mini-Seminars	Final Project (6/14)

Course learning objectives and competencies:

1. Understand the theory and application of phylogenetics to research problems in evolutionary biology and public health
   Ph.D. competencies:

   - A2 Formulate a public health scientific question in statistical terms

2. Perform a phylogenetic data analysis to answer scientific or public health objectives
   Ph.D. competencies:

   - A5 Conduct appropriate statistical analyses of study data and interpret results
   - B5 Develop ability to evaluate and incorporate new and evolving computational and digital technologies into biostatistical work

3. Communicate a phylogenetic analysis to research peers
   Ph.D. competencies:

   - C2 Effectively communicate statistical concepts and reasoning to public health collaborators
   - D6 Learn to write and publish substantive field publications and communicate the statistical portion of the methodology to a substantive field audience

Course learning objectives and competencies:

1. Understand the basic principles of statistical inference

MS competencies

- A2 Formulate a public health question in statistical terms
- A3 Identify the strengths and weaknesses of different study designs to address public health and scientific questions; communicate these issues to public health researchers
- A6 Conduct appropriate statistical analyses of study data and interpret the results

2. Understand the principles of statistical reasoning and their use in understanding, analyzing and developing new statistical models and formal arguments

MS competencies

- B3 Learn to write and disseminate substantive field publications and communicate the statistical portion of the methodology to a substantive field audience.
- C2 Develop ability to comprehend and be engaged in seminars and presentations on biostatistical research