

# Bayesian Phylogenetic Analysis Using *BEAST*

**Monday 12 July to Wednesday 14 July, 2010**

**Simon Ho  
School of Biological Sciences  
University of Sydney**

## Overview

This workshop provides an introduction to Bayesian phylogenetics, with a focus on the methods and models implemented in the software *BEAST*. The workshop will be fairly informal and attendance is free. Graduate students are especially welcome to attend.

## Registration

Registration for this workshop is essential. To register, please contact Simon Ho (simon.ho@sydney.edu.au).

## Lectures

The workshop will include 6 lectures, all of which will be held in the DT Anderson Lecture Theatre in the Heydon-Laurence Building (A08). The first two lectures will provide an introduction to Bayesian phylogenetics. The third lecture will describe some of the sophisticated models available in *BEAST*. The final three lectures will deal with advanced topics. To maximise the intake of information, lectures will be kept fairly short (<45 minutes), with a minimum of a 15-minute break between successive lectures. For this reason, there will be no official question time. Questions may be asked between or after lectures, or during the practical session or discussion group.

## Practical

A practical session will be held on the first day of the workshop in the Ashby Audio-Visual Laboratory in the Macleay Building (A12). This session will have a limit of 20 participants; students will be given priority. The practical session will involve a basic analysis of a prepared data set using Macintosh computers, but participants are welcome to bring their own data sets and laptops if preferred. There will be a repeat session of the practical on Tuesday 13 July for on-campus participants.

## Discussion Group

A discussion group will be held on the second day of the workshop in the Ashby Audio-Visual Laboratory in the Macleay Building (A12). This will involve detailed discussion of 2-3 papers illustrating important concepts in Bayesian phylogenetic analysis and the usage of sophisticated models in *BEAST*. The aim of this discussion group is to consolidate the material covered in the advanced lectures and will allow participants to ask detailed questions about the methods and to discuss them in detail. This session will have a limit of about 15 participants.

## Monday 12 July: Introductory topics

### **10.00 – 10.45    Lecture: Introduction to Bayesian phylogenetics (Part 1)**

- Introduction to phylogenetic analysis
- Maximum likelihood
- Review of non-Bayesian methods
- The Bayesian paradigm
- The prior

### **11.00 – 11.45    Lecture: Introduction to Bayesian phylogenetics (Part 2)**

- The posterior
- Processing the results
- Example: Cave lions
- Advantages and disadvantages

### **12.00 – 12.45    Lecture: Phylogenetic analysis using *BEAST***

- Estimating rates and time-scales
- Relaxed molecular clocks
- Implementing calibrations
- Handling intraspecific and interspecific data
- Using *BEAST*

### **14.00 – 17.00    Practical: A mysterious hominin from Siberia (off-campus participants)**

## Tuesday 13 July: Introductory topics

### **14.00 – 17.00    Practical: A mysterious hominin from Siberia (on-campus participants)**

## Wednesday 14 July: Advanced topics

### 10.00 – 10.30 Lecture: Reconstructing demographic history

Reconstructing demographic history  
Demographic models in *BEAST*  
Skyline-plot methods  
Interpreting skyline plots

### 10.45 – 11.15 Lecture: Bayesian phylogeography

Phylogeographic analysis  
The discrete model  
The continuous model  
Using the models

### 11.30 – 11.50 Lecture: Inferring species trees from gene trees

Gene trees  
Bayesian inference of species trees

### 14.00 – 16.30 Discussion Group: Advanced topics

1. Support for vicariant origins of the New Zealand Onychophora
2. Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics
3. Gene tree discordance, phylogenetic inference and the multispecies coalescent

# Abstracts

## Introduction to Bayesian phylogenetic analysis (Parts 1 and 2)

A brief outline of the various phylogenetic methods will be given, with a focus on maximum likelihood. The basic theory behind Bayesian phylogenetic methods will be introduced.

Bayesian phylogenetic methods have two key components: the likelihood function and the prior. These combine to form the posterior distribution of parameters and trees, which is what we wish to estimate. Despite sharing a basis in the likelihood function, there are important differences between maximum-likelihood and Bayesian methods.

The posterior is usually estimated using Markov chain Monte Carlo sampling. The output from MCMC sampling is then processed and the results are interpreted.

Some of the major criticisms of Bayesian phylogenetic methods will be described. One of the major concerns has been the influence of the priors, the choice of which can be somewhat subjective.

**Assumed:** Basic knowledge of phylogenetic analysis.

## Phylogenetic analysis using *BEAST*

This talk will present some of the models available in *BEAST*, and will describe the processes of creating an input file and interpreting results.

**Relaxed molecular clocks and calibration.** Estimating substitution rates and divergence times is an important component of many molecular evolutionary studies. Relaxed molecular-clock models in *BEAST* can allow for rate variation among lineages. In addition, a number of calibration techniques have been implemented in *BEAST*.

**Handling intraspecific and interspecific data.** Population-level and species-level data need to be analysed using different approaches. For example, a coalescent-based approach needs to be taken for intraspecific data, whereas we would expect species-level data to be described better by a speciation process.

**Using BEAUti and BEAST.** *BEAUti* provides a user-friendly, menu-based interface for creating input files for *BEAST*. The options available in the latest version of *BEAUti* can cater for the majority of standard *BEAST* analyses. *BEAST* produces a number of output files that need to be processed independently using further software.

**Assumed:** Moderate knowledge of Bayesian phylogenetic analysis.

## Reconstructing demographic history

The demographic history of a population leaves a signature in the DNA of its modern representatives. There are several Bayesian methods for estimating demographic history from a sequence alignment. The 'skyline plot' family of methods will be described in detail.

**Demographic models in BEAST.** Various demographic models, such as constant size and exponential growth, are available in *BEAST* and can be tested using Bayes factors.

**Skyline-plot methods.** The development of skyline-plot methods will be traced, beginning with the classic skyline, and continuing with the generalised skyline, Bayesian spline, Bayesian skyline, Bayesian skyride, and extended Bayesian skyline.

**Assumed:** Basic knowledge of Bayesian phylogenetic analysis.

## Bayesian phylogeography

Ancestral geographic reconstruction can be performed in a Bayesian framework, allowing the simultaneous estimation of phylogenetic relationships, node times, ancestral states, and demographic history. This has recently been implemented in *BEAST*, with the introduction of geospatial models.

**Bayesian phylogeographic analysis.** The theory and motivation behind this method will be described, including the discrete and continuous models.

**Applications of the models.** The geospatial model has been used for several data sets. Three examples, involving African rabies virus, muskox, and raccoon rabies virus, will be shown in order to demonstrate the capabilities of the Bayesian phylogeographic approach in *BEAST*.

**Assumed:** Moderate knowledge of Bayesian phylogenetic methods and *BEAST*.

## Inferring species trees from gene trees

Trees estimated from individual genes can disagree with actual species trees, especially for closely related species. Even conflicting gene trees can provide useful information about the underlying species tree. Data from multiple loci can be analysed using an extension to *BEAST*, known as *\*BEAST*.

**Gene trees.** Trees estimated from unlinked loci represent independent realisations of a stochastic process (the coalescent). This can lead to discordance between gene trees which, in turn, can disagree with the species tree.

**Bayesian inference of species trees.** Several methods are available for estimation of species trees from gene trees. *\*BEAST* allows the co-estimation of gene trees, the species tree, ancestral population sizes, and divergence times.

**Assumed:** Moderate knowledge of Bayesian phylogenetic methods and *BEAST*.

## Useful references

### Introductory book chapters and journal articles

- **Inferring Phylogenies**  
Book by Joseph Felsenstein. Chapter 18 “Bayesian Inference of Phylogenies” contains an introduction to Bayesian phylogenetic methods. Most of the chapter is devoted to criticising Bayesian methods, particularly with respect to the role of priors.
- **Bayesian inference of phylogeny: a non-technical primer**  
Archibald JK, Mort ME, & Crawford DJ (2003) *New Trends Plant Syst*, 2003: 187-191.
- **Bayesian inference of phylogeny and its impact on evolutionary biology**  
Huelsenbeck JP, Ronquist F, Nielsen R, & Bollback JP (2001) *Science*, 294: 2310-2314.
- **Potential applications and pitfalls of Bayesian inference of phylogeny**  
Huelsenbeck JP, Larget B, Miller RE, & Ronquist F (2002) *Syst Biol*, 51: 673-688.

### Technical journal articles

Describes *BEAST* 1.4.6:

- **BEAST: Bayesian evolutionary analysis by sampling trees**  
Drummond AJ, & Rambaut A (2007) *BMC Evol Biol*, 7: 214.

Introduces the relaxed molecular-clock models in *BEAST*:

- **Relaxed phylogenetics and dating with confidence**  
Drummond AJ, Ho SYW, Phillips MJ, & Rambaut A (2006) *PLoS Biol*, 4: e88.

Describes the different calibration priors available in *BEAST*:

- **Calibrating molecular estimates of substitution rates and divergence times in birds**  
Ho SYW (2007) *J Avian Biol*, 38: 409-414.
- **Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times.**  
Ho SYW, & Phillips MJ (2009) *Syst Biol*, 58: 367-380.

Describes the skyline-plot methods in *BEAST*:

- **Skyline-plot methods for estimating demographic history from DNA sequences**  
Ho SYW, & Shapiro B (in prep.) – available soon
- **Bayesian coalescent inference of past population dynamics from molecular sequences**  
Drummond AJ, Rambaut A, Shapiro B, & Pybus OG (2005) *Mol Biol Evol*, 22: 1185-1192.
- **Bayesian inference of population size history from multiple loci**  
Heled J, & Drummond AJ (2008) *BMC Evol Biol*, 8: 289.

Describes the geospatial phylogeographic method in *BEAST*:

- **Bayesian phylogeography finds its roots**  
Lemey P, Rambaut A, Drummond AJ, & Suchard MA (2009) *PLoS Comp Biol*, 5: e1000520.
- **Phylogeography takes a relaxed random walk in continuous space and time**  
Lemey P, Rambaut A, Welch JJ, & Suchard MA (2010) *Mol Biol Evol*, in press.

Describes the the species-tree method in *\*BEAST*:

- **Bayesian inference of species trees from multilocus data**  
Heled J, & Drummond AJ (2009) *Mol Biol Evol*, 27: 570-580.

### BEAST resources

- **BEAST Google Group (<http://groups.google.com/group/beast-users>)**  
This site provides a forum allowing users to post questions about using BEAST.