Bayesian phylogenetics using *BEAST*

Workshop on Bayesian phylogenetic analysis, molecular dating, demographic recontruction, and phylogeography

Simon Ho*

December 1 – 3, 2010. Botan, University of Gothenberg

Wednesday Dec 1st

9.30   Workshop introduction

9.35   Seminar 1. Introduction to Bayesian phylogenetic analysis
(includes 10-minute break)

11.00   Coffee break

11.30   Seminar 2. Phylogenetic analysis using *BEAST*

13.00   Introductory computer workshop: "A mysterious hominin from Siberia"
Using *BEAUti*, *BEAST*, and *Tracer* – work through practical exercise or bring your own
data set

Thursday Dec 2nd

9.30   Seminar 3. Demographic reconstruction

10.30   Coffee break

11.00   Seminar 4. Bayesian phylogeography

13.00   Advanced computer workshop: "Phylogeography of the European brown bear"
Work through practical exercise or bring your own data set

Friday Dec 3rd

9.30   Discussion of article by Allwood *et al*. (2010)

10.30   Coffee break

11.00   Discussion of article by Campos *et al*. (2010)

12.00   Students' lunch with Simon

14.00   Workshop ends

* Dr Simon Ho is a Senior Lecturer (Molecular Evolution & Phylogenetics group) in the School of
Biological Sciences at the 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 include 4 seminars. The first seminar will provide an introduction to Bayesian phylogenetics. The second seminar will describe some of the sophisticated models available in *BEAST*. The final two seminars will deal with advanced topics.

Seminar 1. Introduction to Bayesian phylogenetic analysis

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.

Seminar 2. Introduction to Bayesian phylogenetic analysis

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.

Seminar 3. 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.

Seminar 4. 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*.

Discussion Paper 1

Support for vicariant origins of the New Zealand Onychophora

Allwood J, Gleeson D, Mayer G, Daniels S, Beggs JR, and Buckley TR (2010) *Journal of Biogeography*, 37: 669-681.

This paper involves the following topics of relevance:

- Model selection
- Estimation of divergence times
- Usage of age calibrations (parametric prior distributions)
- Tree priors
- Interpretation of estimated trees and dates

Discussion Paper 2

Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics

Campos PF, Willerslev E, Sher A, Orlando L, Axelsson E, Tikhonov A, Aaris-Sørensen K, Greenwood AD, Kahlke R-D, Kosintsev P, Krakhmalnaya T, Kuznetsova T, Lemey P, MacPhee R, Norris CA, Shepherd K, Suchard MA, Zazula GD, Shapiro B, and Gilbert MTP (2010) *Proceedings of the National Academy of Sciences of the USA*, 107: 5675-5680.

This paper involves the following topics of relevance:

- Ancient DNA and the impact of sequence damage
- Metropolis-coupled MCMC
- Reconstructing demographic history
- Bayesian skyline plot
- Bayesian skyride plot
- Discrete geospatial analysis (Bayesian phylogeography)
- Phylogeographic structure

Useful references

Introductory journal articles

- **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 (2011) *Mol Ecol Res*, accepted with minor revision.

- **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*, 27: 1877-1885.

- **Three roads diverged? Routes to phylogeographic inference**
  Bloomquist EW, Lemey P, & Suchard MA (2010) Trends Ecol Evol, 25: 626-632.