Introduction to Bayesian Phylogenetic Software

Tracy Heath

Department of Integrative Biology, University of California, Berkeley

2013 Workshop on Molecular Evolution
Český Krumlov, Czech Republic
Outline

Overview
Introduction to Bayesian software programs for phylogenetics
  • MrBayes: history and current version
  • Other programs: PhyloBayes, BEAST/*BEAST
  • RevBayes: graphical models and Bayesian phylogenetics

break

Tutorial – Conor Meehan and Instructors
Phylogenetic reconstruction in MrBayes
  • Basic phylogenetic inference and MCMC

dinner
Overview
Convergence and MCMC diagnosis
• tools and metrics for evaluating MCMC samples

Tutorial — Conor Meehan and Instructors
Phylogenetic reconstruction in MrBayes
• MCMC diagnostics in Tracer
• Partitioned phylogenetic inference in MrBayes

beer
A Bayesian inference program for phylogenetic inference and model selection

John Huelsenbeck

Fredrik Ronquist

also Maxim Teslenko, Paul van der Mark, Daniel Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc Suchard

Availability: http://mrbayes.net/
Phylogenetic inference under a wide range of models

- **Unrooted** trees
  - joint estimation of topology, branch length, and model parameters

- **Rooted** — time-calibrated trees
  - joint estimation of topology, branch rates, branch times, and model parameters, and gene-tree/ species-tree inference in BEST

- **Data types**
  - discrete characters — binary (0, 1) or multi-state (0, 1, ..., 9)
  - DNA — 4-state nucleotide, doublet, or codons
  - amino acid

Availability: http://mrbayes.net/
MrBayes — Models

MrBayes v3.2 manual

Data Type  
Restriction 0-1  
Standard 0-9

State Frequencies (Substitution Rates)  
fixed/estimated (Dirichlet) prset statefreqapr  
equal/estimated (SymmDir) prset symdirhyperpr

Across-Site Rate Variation  
equal/gamma lset rates  
equal/gamma lset rates

Coding Bias  
all/variable/nopresencesites/noabsencesites lset coding  
all/variable/informative lset coding

Misc.  
unordered/ordered ctype

Data Type  
DNA A C G T

Model Type  
4by4 lset nucmodel  
doublet lset nucmodel  
codon lset nucmodel

State Frequencies  
fixed/est. (Dirichlet) prset statefreqapr  
fixed/est. (Dirichlet) over 16 states prset statefreqapr  
fixed/est. (Dirichlet) over 61 states prset statefreqapr

Substitution Rates  
F81/HKY/GTR/mixed lset nst=1/2/6/mixed  
F81/HKY/GTR/mixed lset nst=1/2/6/mixed  
F81/HKY/GTR/mixed lset nst=1/2/6/mixed

Across-Site Rate Variation  
equal/gamma/propinv/lnvgamma/adgamma lset rates  
equal/gamma/propinv/lnvgamma lset rates  
equal/gamma/propinv/lnvgamma lset rates

Across-Tree Rate Variation  
yes/no lset covarion  
Across-Site Omega Variation

From the MrBayes v3.2 manual
Bayesian phylogenetic reconstruction under non-parametric mixture models


Huelsenbeck, Suchard. 2007. **A nonparametric method for accommodating and testing across-site rate variation.** *Syst. Biol.* 56:975–987

The Dirichlet process mixture model partitions sites into different rate categories.

No a priori specification of data partitions necessary.

Information from the data leads to the estimation of rate category assignment and the number of rate categories.

Broad Phylogenomic Sampling and the Sister Lineage of Land Plants (Timme et al. *PLoS1* 2012)
Joint Bayesian inference of tree topology \textit{(rooted)} and divergence times

Bayesian Evolutionary Analysis Sampling Trees

- population size
- growth/decline in population
- bottlenecks/transition points
- gene trees/species trees
- virus transmission dynamics
- recombination
- migration
- founder effects
- epidemiological tracking
- phylogeography
- trait evolution
- dates of MRCAs
- lineage rates
- ancestral character state reconstruction
- times of bottlenecks/transitions

http://beast.bio.ed.ac.uk/ (Drummond, Suchard, Xie, & Rambaut, \textit{MBE}, 2012)
## Program Features

<table>
<thead>
<tr>
<th>Method/Model/Feature</th>
<th>MrBayes v3.2</th>
<th>PhyloBayes</th>
<th>BEAST/ *BEAST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrooted trees</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Joint est. topology &amp; times</td>
<td>✓</td>
<td>X*</td>
<td>✓</td>
</tr>
<tr>
<td>Gene-tree/species-tree</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Dataset partitioning</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bayes factors</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Morphological data</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Demography/phylogeography</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>DPP mixture on site-rates/models</td>
<td>X</td>
<td>✓</td>
<td>X**</td>
</tr>
<tr>
<td>Continuous traits</td>
<td>X</td>
<td>✓***</td>
<td>✓</td>
</tr>
<tr>
<td>Graphical-user-interface (GUI)</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
</tbody>
</table>

* divergence time estimation on fixed topology
** in BEAST2
*** in companion program: Coevol
RevBayes

A flexible programming environment for model-based (primarily Bayesian) phylogenetic inference

Sebastian Höhna  John Huelsenbeck  Fredrik Ronquist

also Bastien Boussau, Tracy Heath, Michael Landis, Brian Moore, Ben Redelings, Chi Zhang (and others)
Graphical Models in RevBayes

Models are represented by directed acyclic graphs (DAGs)

A generic graphical model
Graphical Models in RevBayes

Models are represented by *directed acyclic graphs* (DAGs)

A generic graphical model
Graphical Models in RevBayes

Models are represented by directed acyclic graphs (DAGs)

A generic graphical model
A generic graphical model
Graphical Models in RevBayes

A tree is a graphical model (DAG)
Graphical Models in RevBayes

The GTR-Γ model
The Rev Language

An R-like language for specifying models and MCMC analysis

# An example (partial)

# read the data
D <- readCharacterData("data/primates.nex")[1]

# substitution model priors
bf <- v(1,1,1,1)
e <- v(1,1,1,1,1,1)
pi ~ dirichlet(bf)
er ~ dirichlet(e)

# Moves on substitution model parameters
moves[1] <- mSimplex(pi, 10.0, 4, true, 2.0)
moves[2] <- mSimplex(er, 10.0, 6, true, 2.0)

...
The RevBayes GUI
Specifying the analysis work-flow
The RevBayes GUI

Specifying the graphical model
Overview
Introduction to Bayesian software programs for phylogenetics
• MrBayes: history and current version
• Other programs: PhyloBayes, BEAST/*BEAST
• RevBayes: graphical models and Bayesian phylogenetics

break

Tutorial — Conor Meehan and Instructors
Phylogenetic reconstruction in MrBayes
• Basic phylogenetic inference and MCMC

dinner