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

OUTLINE

Overview

Convergence and MCMC diagnosis

- tools and metrics for evaluating MCMC samples

break

Tutorial – Conor Meehan and Instructors

Phylogenetic reconstruction in MrBayes

- MCMC diagnostics in Tracer
- Partitioned phylogenetic inference in MrBayes

beer

MrBayes

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/`

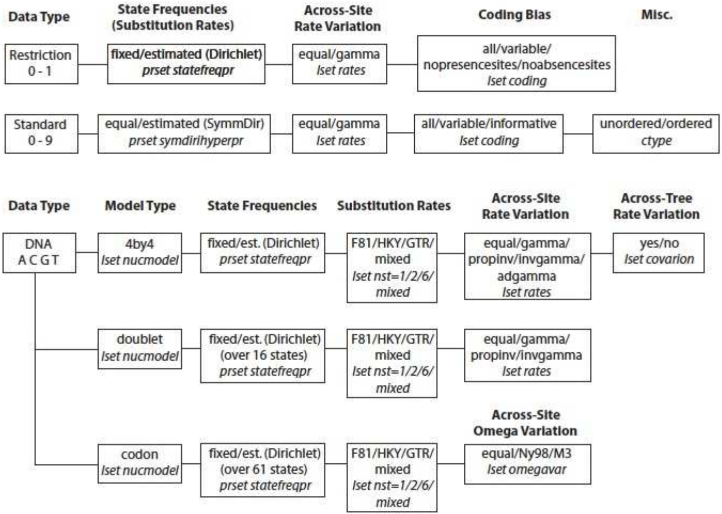
MrBayes

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 $(0, 1, \ldots, 9)$
  - DNA – 4-state nucleotide, doublet, or codons
  - amino acid

Availability: `http://mrbayes.net/`

MrBayes — Models

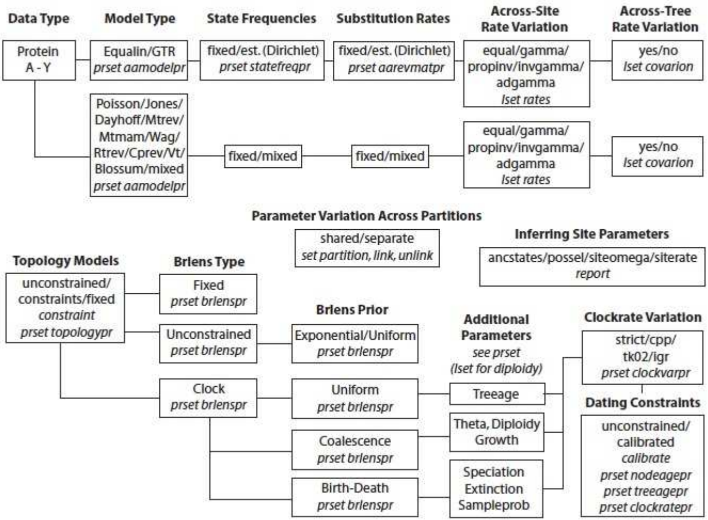


Data Type	State Frequencies (Substitution Rates)	Across-Site Rate Variation	Coding Bias	Misc.
Restriction 0 - 1	fixed/estimated (Dirichlet) *prset statefreqpr*	equal/gamma *lset rates*	all/variable/ nopresencesites/noabsencesites *lset coding*	
Standard 0 - 9	equal/estimated (SymmDir) *prset symdirihyperpr*	equal/gamma *lset rates*	all/variable/informative *lset coding*	unordered/ordered *ctype*

Data Type	Model Type	State Frequencies	Substitution Rates	Across-Site Rate Variation	Across-Tree Rate Variation
DNA A C G T	4by4 *lset nucmodel*	fixed/est. (Dirichlet) *prset statefreqpr*	F81/HKY/GTR/ mixed *lset nst=1/2/6/* *mixed*	equal/gamma/ propinv/invgamma/ adgamma *lset rates*	yes/no *lset covarion*
	doublet *lset nucmodel*	fixed/est. (Dirichlet) (over 16 states) *prset statefreqpr*	F81/HKY/GTR/ mixed *lset nst=1/2/6/* *mixed*	equal/gamma/ propinv/invgamma *lset rates*	
	codon *lset nucmodel*	fixed/est. (Dirichlet) (over 61 states) *prset statefreqpr*	F81/HKY/GTR/ mixed *lset nst=1/2/6/* *mixed*	**Across-Site Omega Variation** equal/Ny98/M3 *lset omegavar*	

from the MrBayes v3.2 manual

MRBAYES — MODELS



Data Type

Protein
A - Y

Model Type

Equalin/GTR
prset aamodelpr

Poisson/Jones/
Dayhoff/Mtrev/
Mtmam/Wag/
Rtrev/Cprev/Vt/
Blossum/mixed
prset aamodelpr

State Frequencies

fixed/est. (Dirichlet)
prset statefreqpr

fixed/mixed

Substitution Rates

fixed/est. (Dirichlet)
prset aarevmatpr

fixed/mixed

Across-Site Rate Variation

equal/gamma/
propinv/invgamma/
adgamma
lset rates

equal/gamma/
propinv/invgamma/
adgamma
lset rates

Across-Tree Rate Variation

yes/no
lset covarion

yes/no
lset covarion

Parameter Variation Across Partitions

shared/separate
set partition, link, unlink

Inferring Site Parameters

ancstates/possel/siteomega/siterate
report

Topology Models

unconstrained/
constraints/fixed
constraint
prset topologypr

Brlens Type

Fixed
prset brlenspr

Unconstrained
prset brlenspr

Clock
prset brlenspr

Brlens Prior

Exponential/Uniform
prset brlenspr

Uniform
prset brlenspr

Coalescence
prset brlenspr

Birth-Death
prset brlenspr

Additional Parameters
see prset
(lset for diploidy)

Treeage

Theta, Diploidy
Growth

Speciation
Extinction
Sampleprob

Clockrate Variation

strict/cpp/
tk02/igr
prset clockvarpr

Dating Constraints

unconstrained/
calibrated
calibrate
prset nodeagepr
prset treeagepr
prset clockratepr

from the MrBayes v3.2 manual

PHYLOBAYES

Bayesian phylogenetic reconstruction under non-parametric mixture models

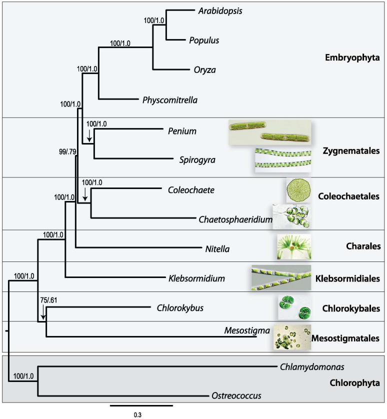
Lartillot, Philippe. 2004. **A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process.** *MBE* 21: :1095–1109.

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

Lartillot, Lepage, Blanquart. 2009. **PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating.** *Bioinformatics* 25: 2286–2288.

PhyloBayes

- 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)

BEAST/*BEAST

Joint Bayesian inference of tree topology (*rooted*) and divergence times

Bayesian **E**volutionary **A**nalysis **S**ampling **T**rees

- 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, *MBE*, 2012)

PROGRAM FEATURES

Method/Model/Feature	**MrBayes v3.2**	**PhyloBayes**	**BEAST/ *BEAST**
Unrooted trees	✓	✓	X
Joint est. topology & times	✓	X*	✓
Gene-tree/species-tree	✓	X	✓
Dataset paritioning	✓	✓	✓
Bayes factors	✓	✓	✓
Morphological data	✓	X	✓
Demography/phylogeography	X	X	✓
DPP mixture on site-rates/models	X	✓	X**
Continuous traits	X	✓***	✓
Graphical-user-interface (GUI)	X	X	✓

*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

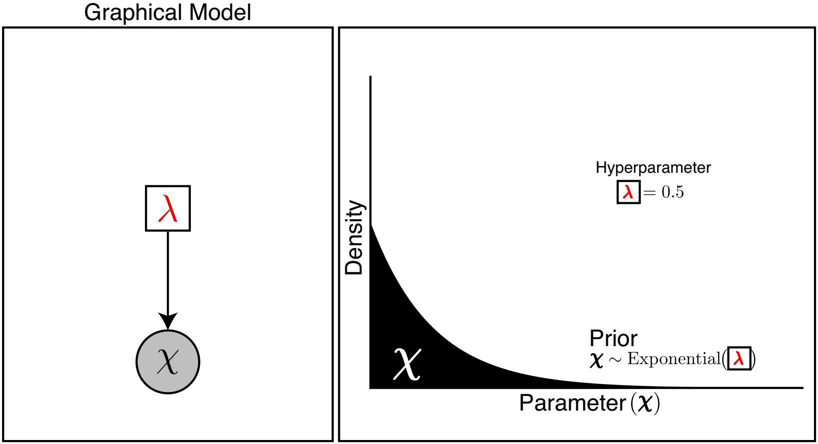
John Huelsenbeck

Fredrik Ronquist

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

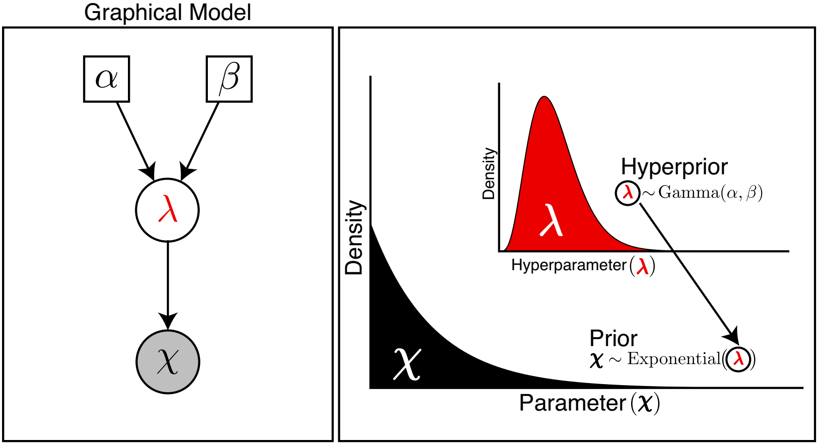
Graphical Models in RevBayes

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



GRAPHICAL MODELS IN REVBAYES

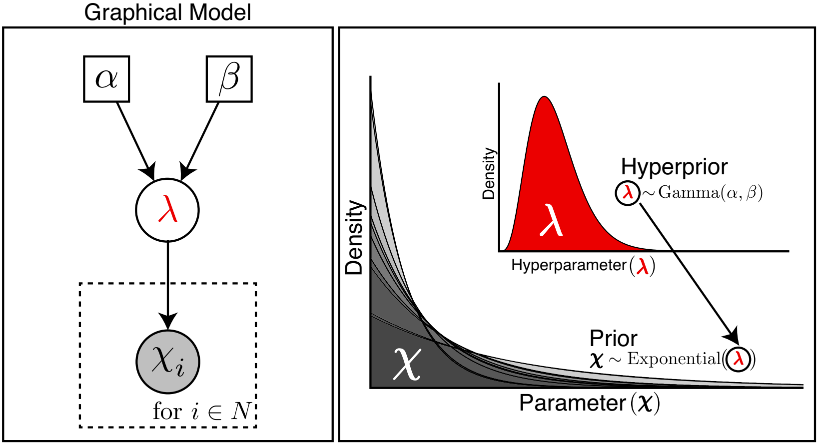
Models are represented by *directed acyclic graphs* (DAGs)



A generic graphical model

GRAPHICAL MODELS IN REVBAYES

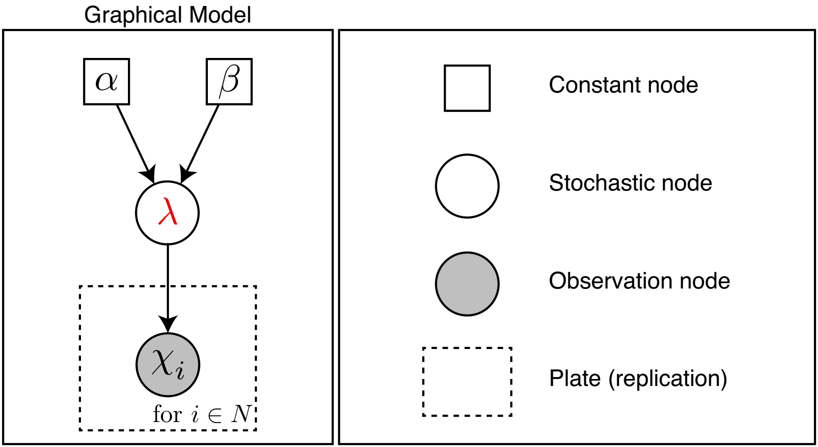
Models are represented by *directed acyclic graphs* (DAGs)



A generic graphical model

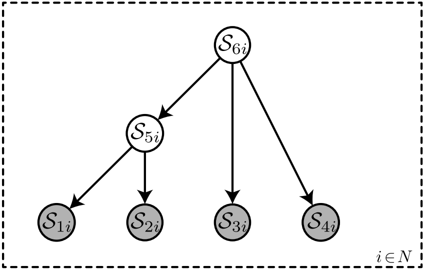
GRAPHICAL MODELS IN REVBAYES

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



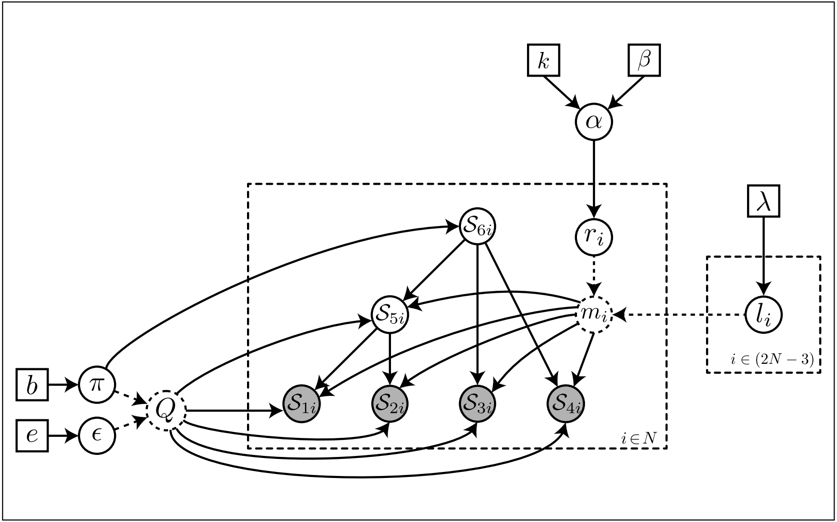
Graphical Models in RevBayes

A tree is a graphical model (DAG)



GRAPHICAL MODELS IN REVBAYES

The GTR-Γ model



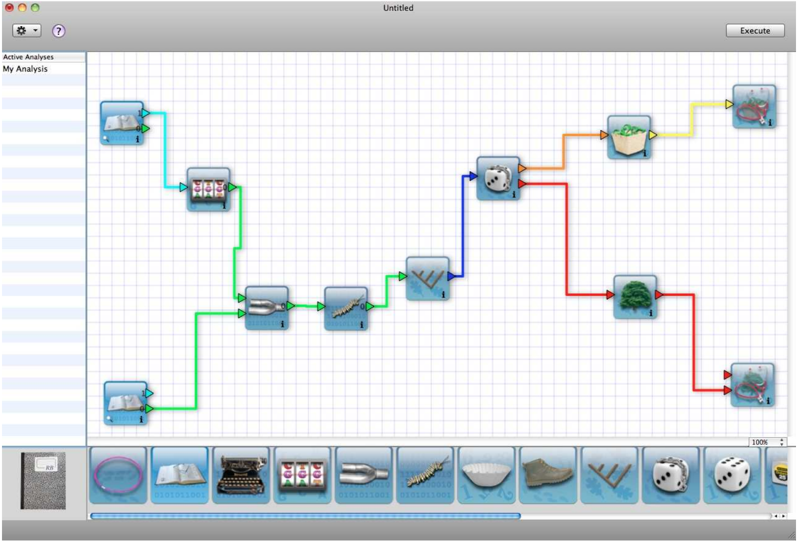
The Rev Language

An R-like language for specifying models and MCMC analysis

```
# An example (partial)

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

# substition 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)

...
```

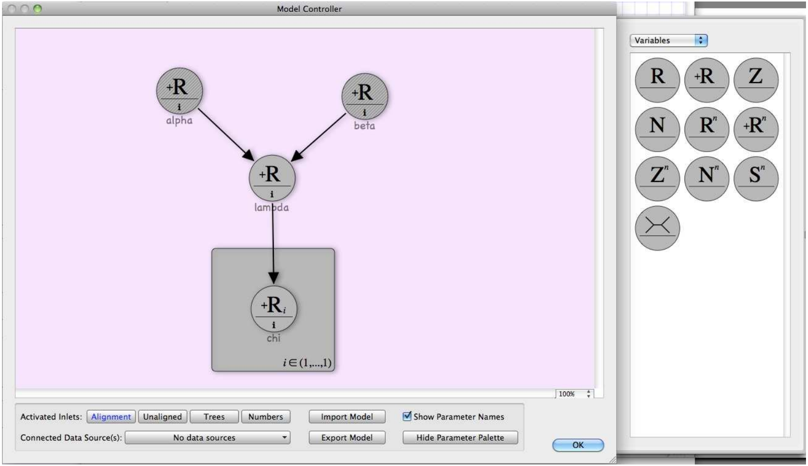
THE REVBAYES GUI

Specifying the analysis work-flow



THE REVBAYES GUI

Specifying the graphical model



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