

# Bayesian Phylogenetic Inference using RevBayes:

## Introduction to RevBayes

Sebastian Höhna

GeoBio-Center  
Ludwig-Maximilians Universität, München





Bastien Boussau



Tracy Heath



John Huelsenbeck



Fredrik Ronquist



## RevBayes Team



Michael Landis



Nicolas Lartillot



Brian Moore



Sebastian Höhna



Will Freyman



April Wright



Mike May



Will Pett



RevBayes Team<sup>2</sup>



Andy Magee



Jeremy Brown



Rachel Warnock



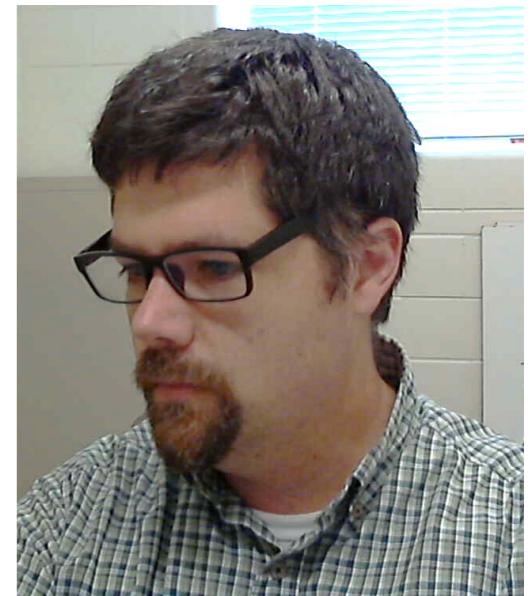
Wade Dismukes



Dominik Schrempf



Joëlle Barido-Sottani



Ben Redelings



Jiansi Gao



RevBayes Team<sup>3</sup>

# Which software to choose

---



# Which software to choose

---

Does the software run the model/analysis?

Am I able to understand the software and to use it?

Is the software fast enough to give me an answer?

# Why a single software

---

1. You only need to learn one software.
2. Lower chance of incompatibility.
3. Possibility to combine many models, methods & analysis.
4. Joint inferences are (theoretically) superior and easily feasible.
5. No redundant coding is necessary for recurring methods
  1. MCMC algorithm
  2. Likelihood computation
  3. Data structures
  4. etc.

# Aims for RevBayes

---

## 1) General and flexible model specification

- a) Availability of (common) models
- b) Extendability

## 2) Transparent

- a) No black-box!
- b) Well structured model specification
- c) Explicit models
- d) Documentation, examples and tutorials

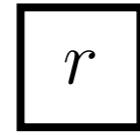
## 3) Fast & Efficient

- a) Fast likelihood calculators
- b) Efficient (MCMC) algorithms, e.g., tree proposals

# Graphical Model (DAG)

---

a)



```
# constant node  
r <- 10
```

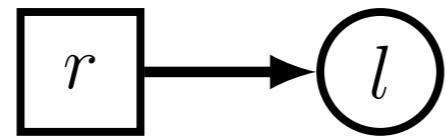
# Graphical Model (DAG)

---

a)



b)

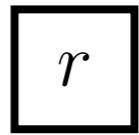


```
# constant node  
r <- 10
```

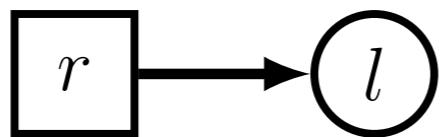
```
# stochastic node  
l ~ dnExp(r)
```

# Graphical Model (DAG)

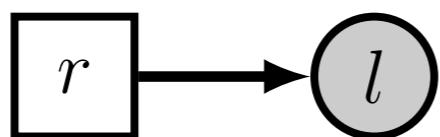
a)



b)



c)



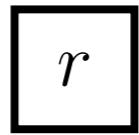
```
# constant node  
r <- 10
```

```
# stochastic node  
l ~ dnExp(r)
```

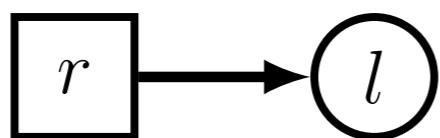
```
# stochastic node (observed)  
l.clamp(0.1)
```

# Graphical Model (DAG)

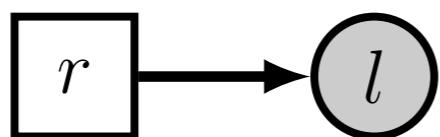
a)



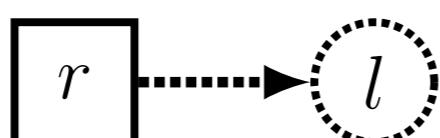
b)



c)



d)



```
# constant node  
r <- 10
```

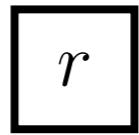
```
# stochastic node  
l ~ dnExp(r)
```

```
# stochastic node (observed)  
l.clamp(0.1)
```

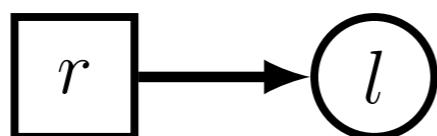
```
# deterministic node  
l := exp(r)
```

# Graphical Model (DAG)

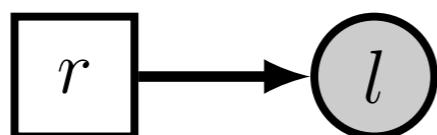
a)



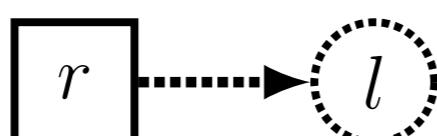
b)



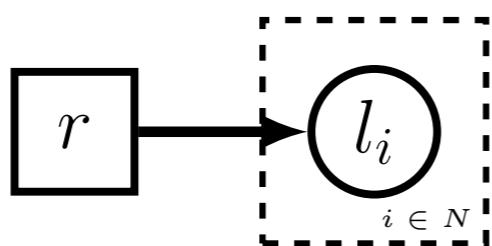
c)



d)



e)



```
# constant node  
r <- 10
```

```
# stochastic node  
l ~ dnExp(r)
```

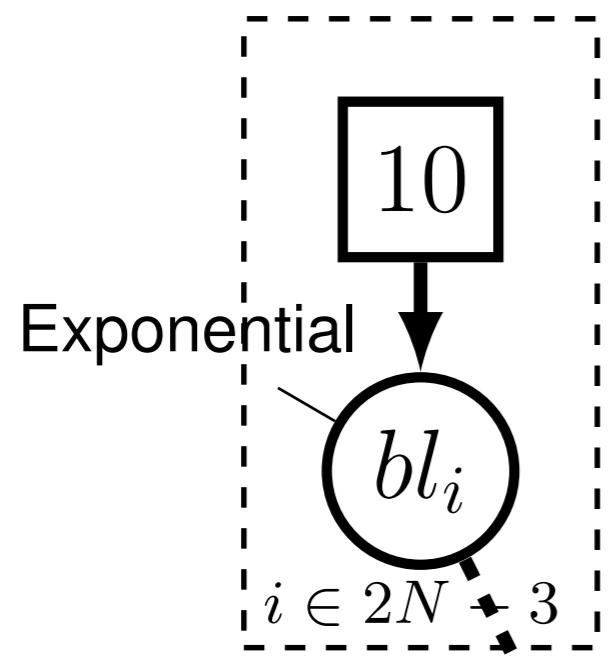
```
# stochastic node (observed)  
l.clamp(0.1)
```

```
# deterministic node  
l := exp(r)
```

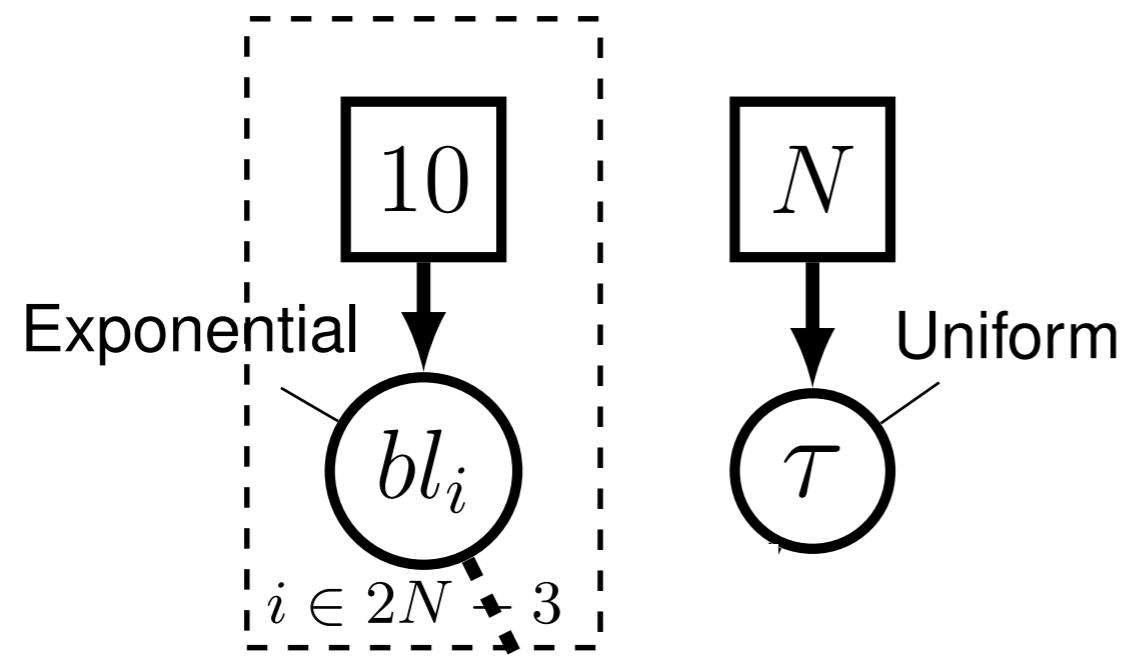
```
# stochastic nodes (iid)  
for (i in 1:N) {  
    l[i] ~ dnExp(r)  
}
```

# GTR+Gamma Model with Prior

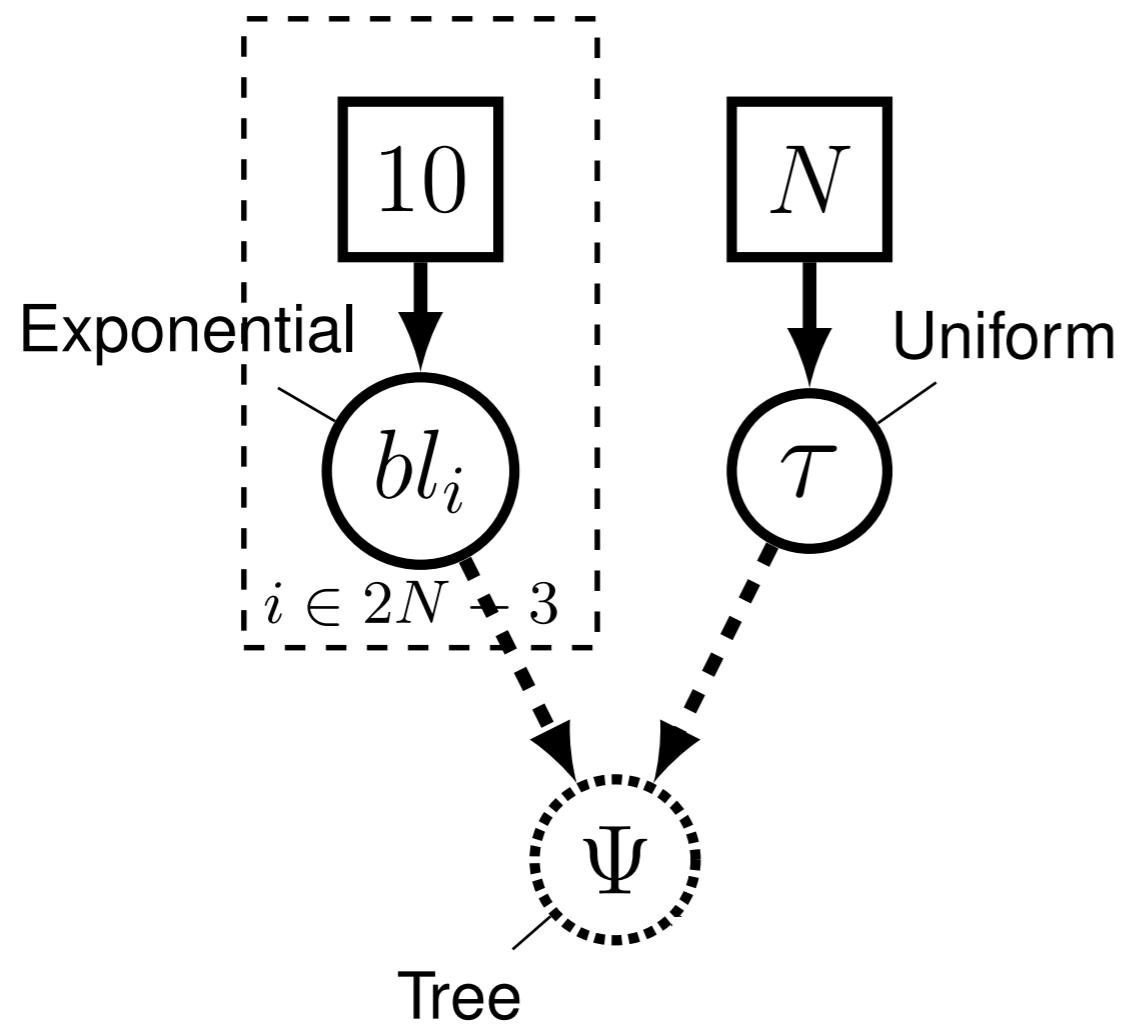
---



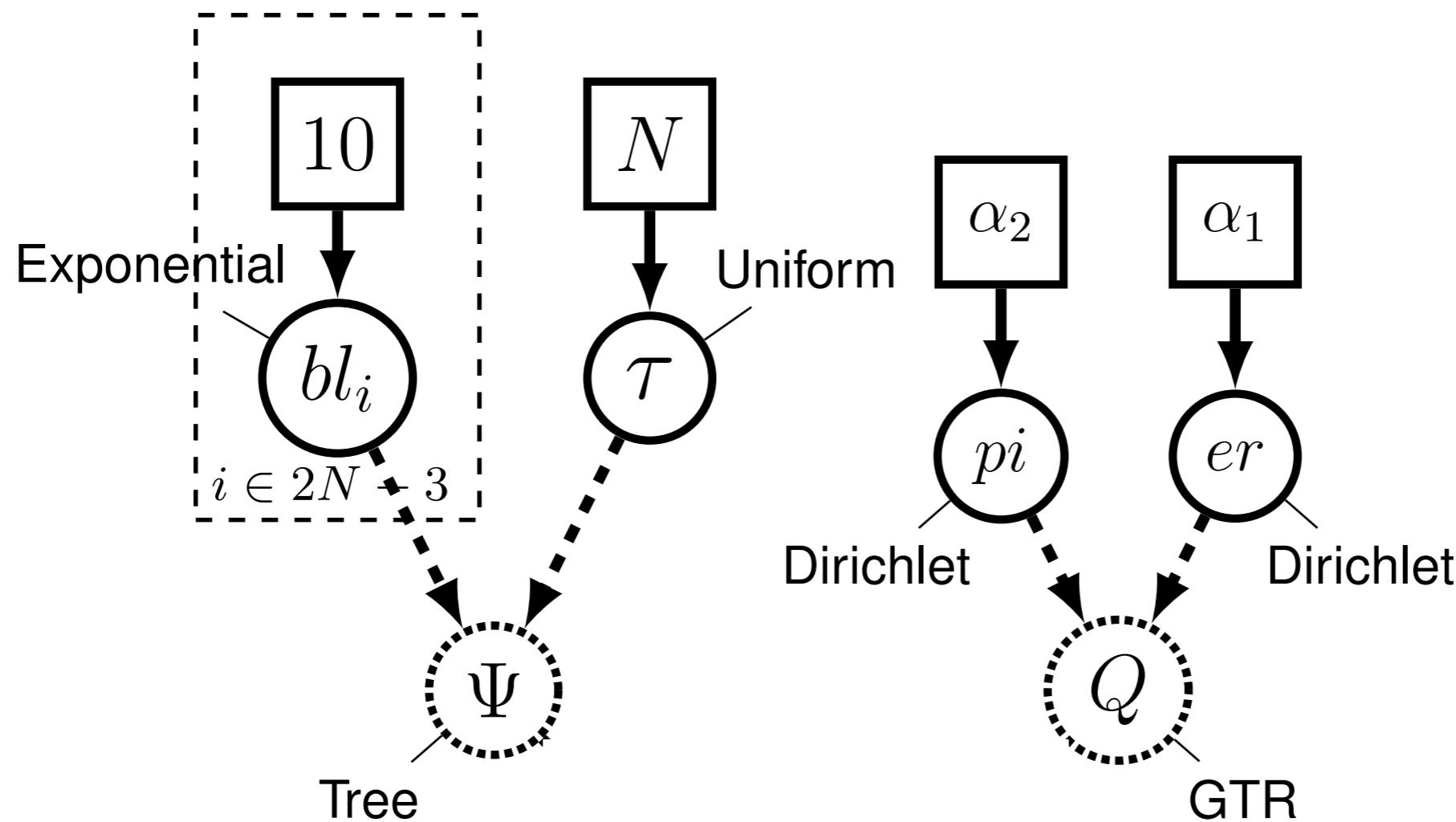
# GTR+Gamma Model with Prior



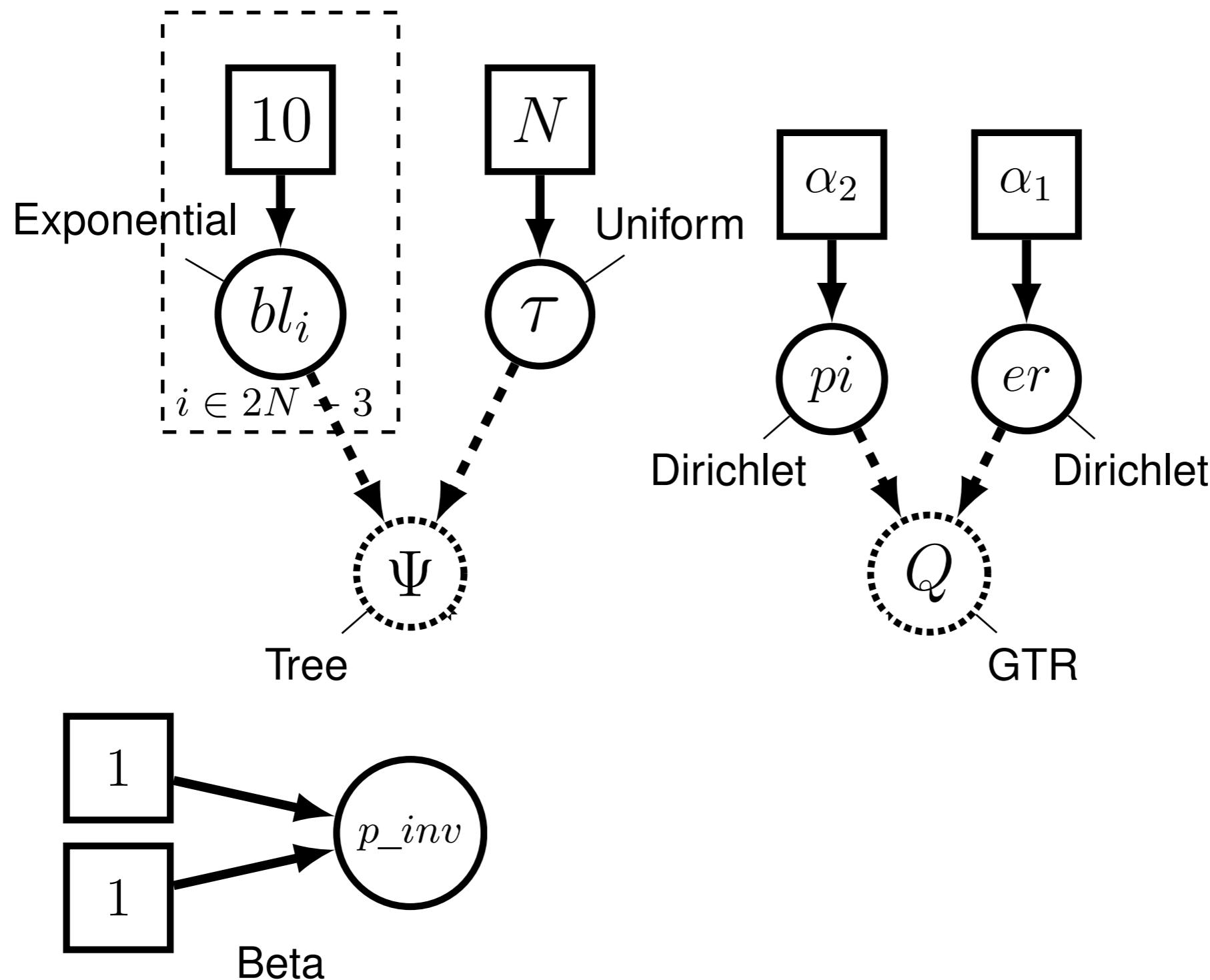
# GTR+Gamma Model with Prior



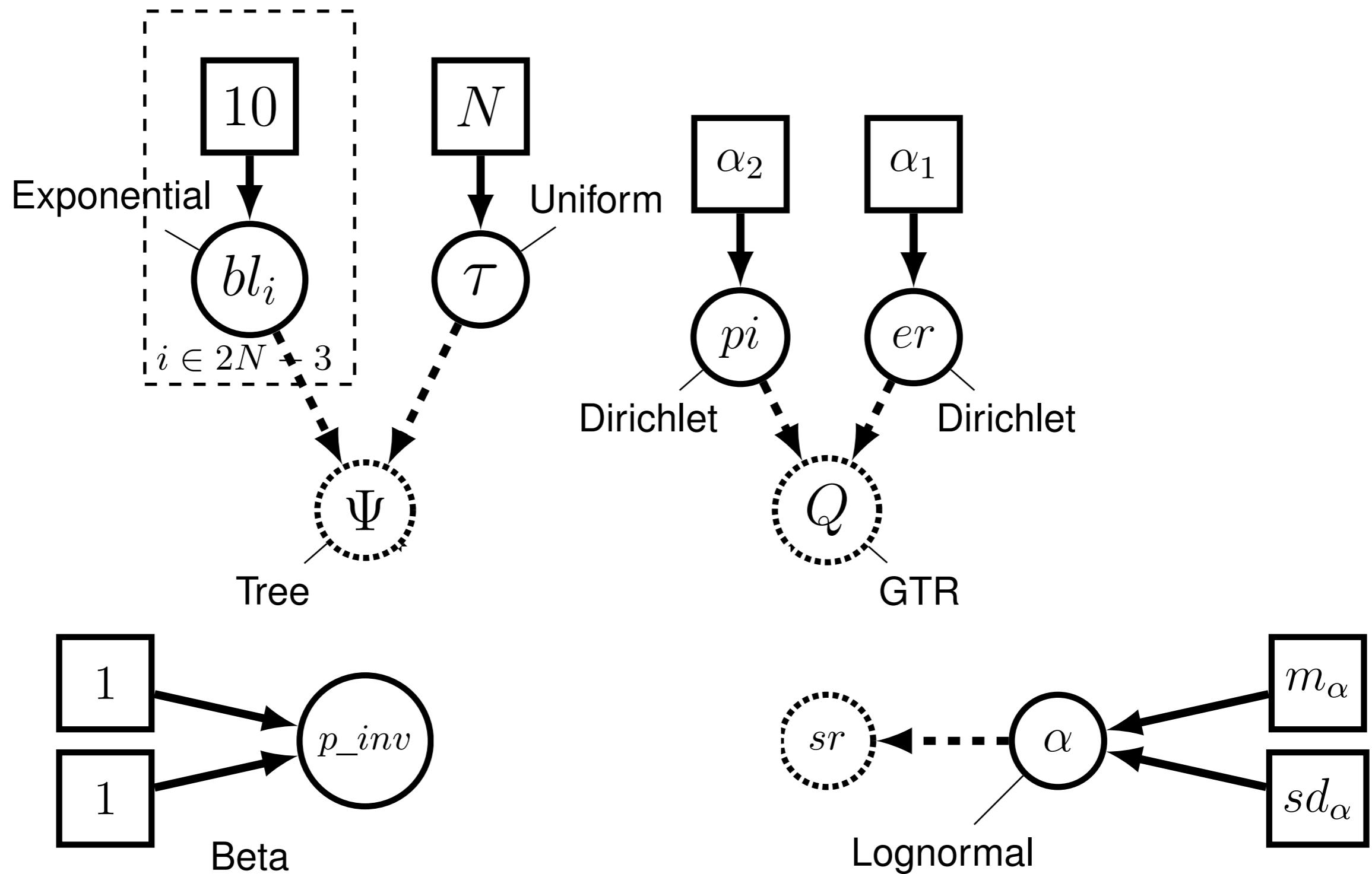
# GTR+Gamma Model with Prior



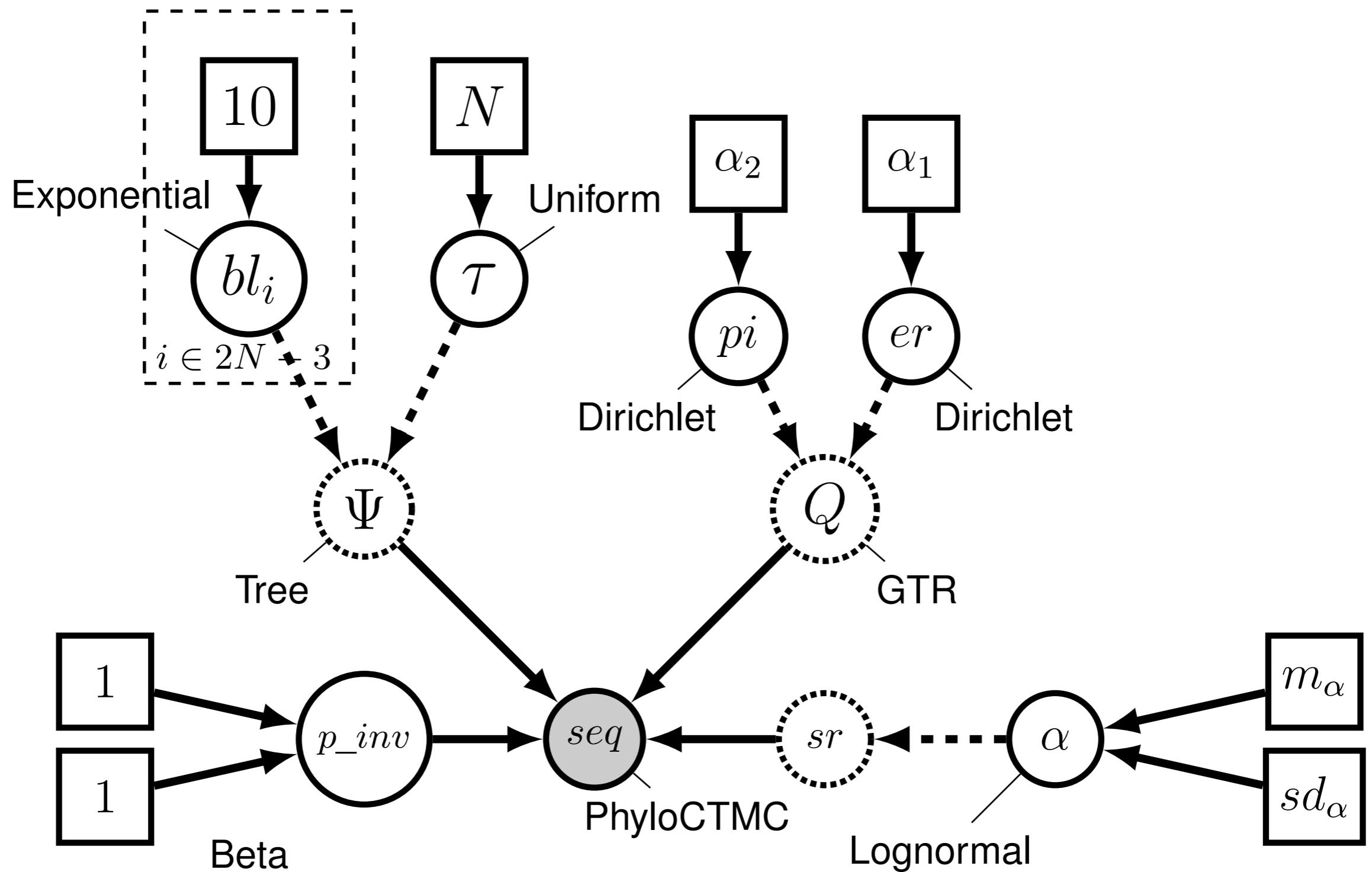
# GTR+Gamma Model with Prior



# GTR+Gamma Model with Prior



# GTR+Gamma Model with Prior



# Aims for RevBayes

---

## 1) General and flexible model specification

- a) Availability of (common) models
- b) Extendability



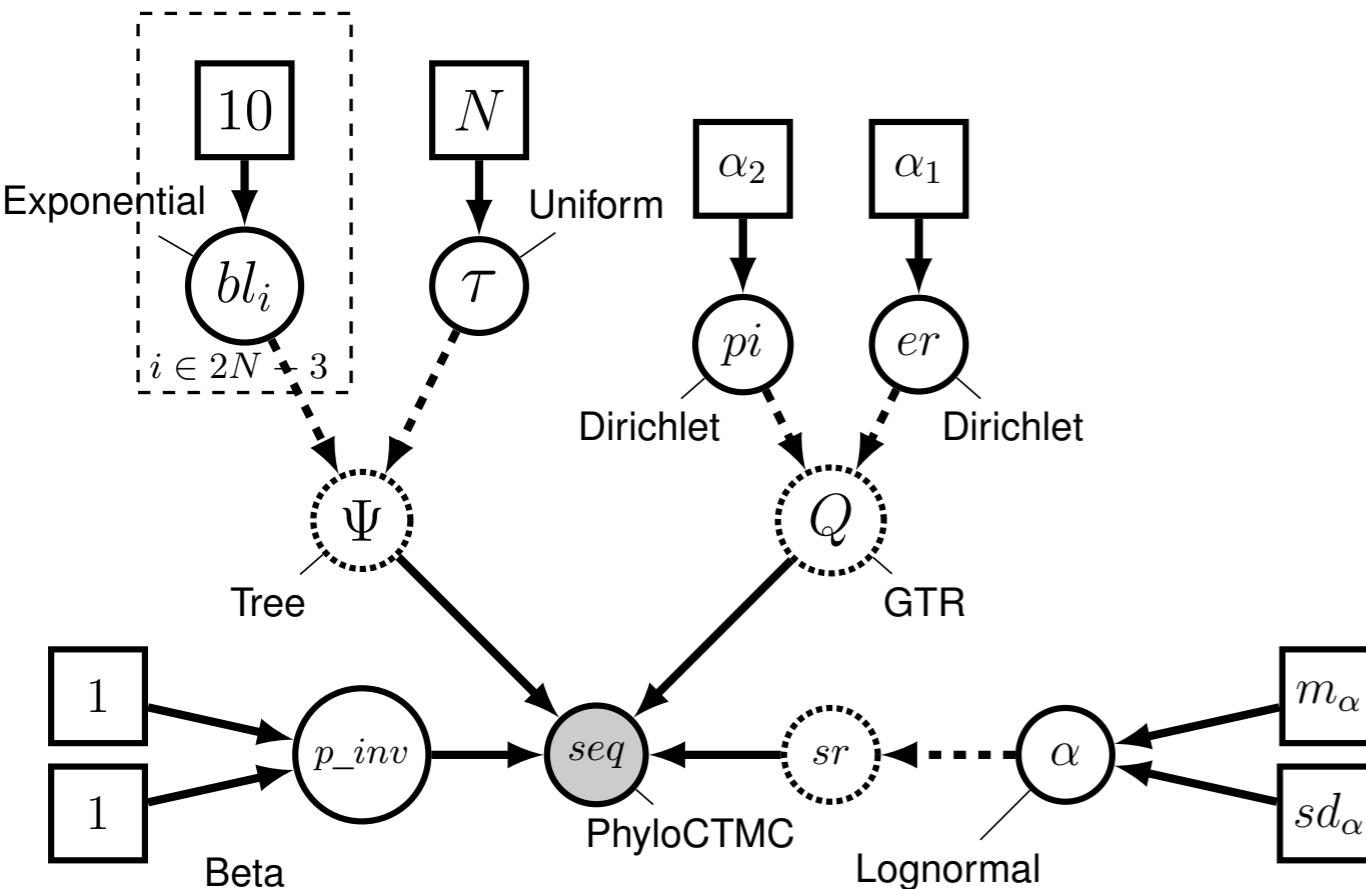
## 2) Transparent

- a) No black-box!
- b) Well structured model specification
- c) Explicit models
- d) Documentation, examples and tutorials

## 3) Fast & Efficient

- a) Fast likelihood calculators
- b) Efficient (MCMC) algorithms, e.g., tree proposals

# Rev (the RevBayes language)



```

for (i in 1:n_branches) {
    bl[i] ~ dnExponential(10.0)
}
topology ~ dnUniformTopology(taxa)
psi := treeAssembly(topology, bl)

alpha1 <- v(1,1,1,1,1,1)
alpha2 <- v(1,1,1,1)
er ~ dnDirichlet(alpha1)
pi ~ dnDirichlet(alpha2)
Q := fnGTR(er, pi)

m_alpha <- ln(5.0)
sd_alpha <- 0.587405
alpha ~ dnLognormal(m_alpha, sd_alpha)
sr := fnDiscretizeGamma(alpha, alpha, 4, false)
p_inv ~ dnBeta(1,1)
seq ~ dnPhyloCTMC(tree=psi, Q=Q, pInv=p_invar,
                    siteRates=sr, type="DNA")
seq.clamp(data)
  
```

# Tutorials

---

## Currently covered topics:

1. Introduction
2. Models of molecular evolution.
3. Inference methods.
4. Divergence time estimation
5. Diversification rate estimation
6. Gene tree - species tree estimation
7. Historical biogeography
8. Comparative phylogenetic method

# Aims for RevBayes

---

## 1) General and flexible model specification

- a) Availability of (common) models
- b) Extendability



## 2) Transparent

- a) No black-box!
- b) Well structured model specification
- c) Explicit models
- d) Documentation, examples and tutorials



## 3) Fast & Efficient

- a) Fast likelihood calculators
- b) Efficient (MCMC) algorithms, e.g., tree proposals

# Performance Study — Likelihood Computation

## Primates:

- 12 taxa
- 898 sites
- 412 patterns

## MCMC:

- burnin of  $10^5$
- chain length of  $10^6$
- only substitution model parameters are updated

	HKY	HKY+G	GTR	GTR+G
BEAST v1.8	65.3	188.4	75.8	213.4
BEAST v1.8 - BEAGLE	41.2	105.2	47.5	107.4
MrBayes 3.2	78.2	177.7	76.9	169.9
MrBayes 3.2 - BEAGLE	92.5	221.2	91.4	222.7
RevBayes	46.9	161.3	62.5	181.2

\* times are given in seconds

# Performance Study — MCMC shortcuts

## Primates:

- 12 taxa
- 898 sites
- 412 patterns

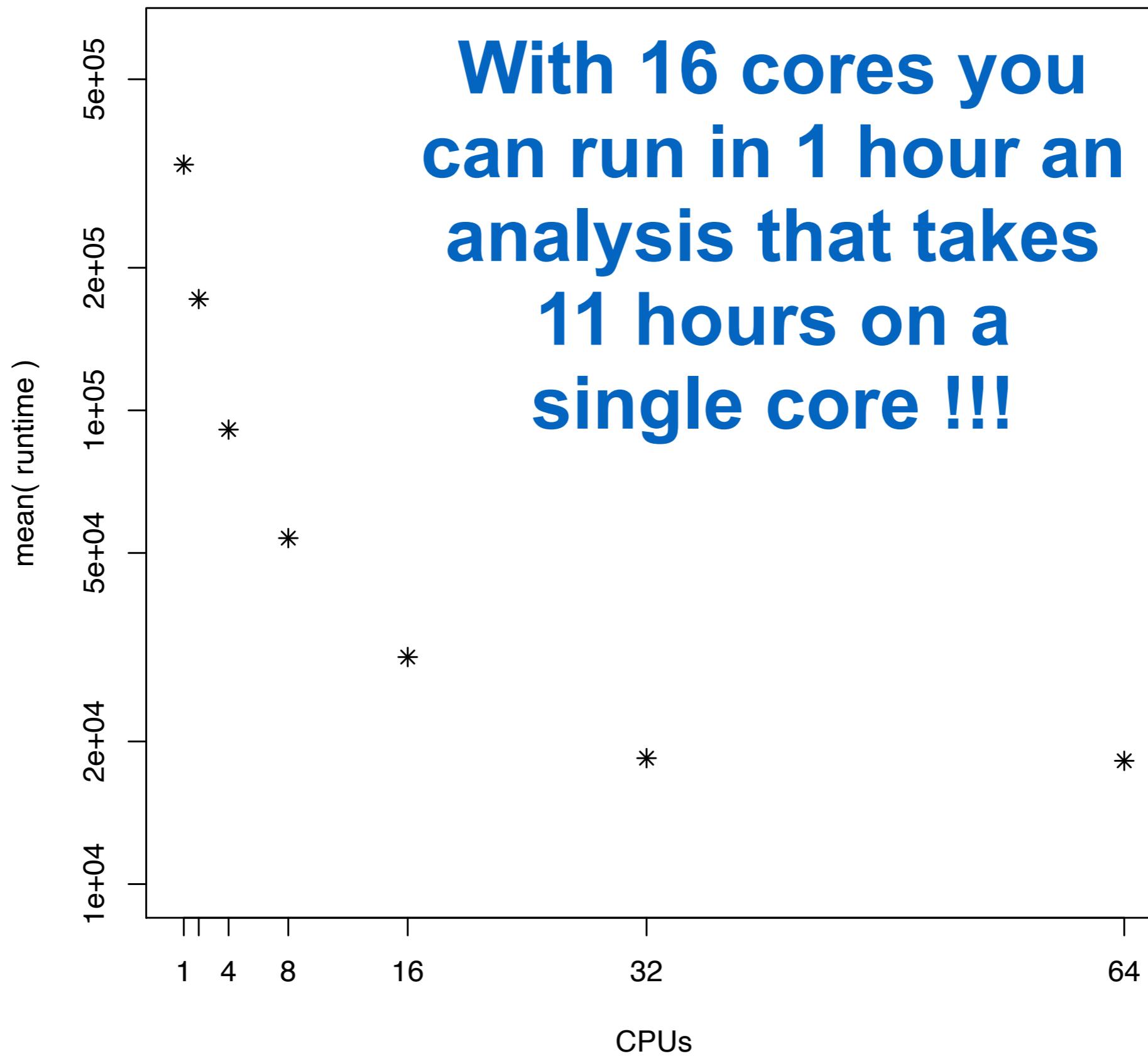
## MCMC:

- burnin of  $10^5$
- chain length of  $10^6$
- only topology or node ages are updated

	NNI	Node-Slide
BEAST v1.8	30.7	42.8
BEAST v1.8 - BEAGLE	21.0	28.3
MrBayes 3.2	37.2	38.1
MrBayes 3.2 - BEAGLE	42.6	31.9
RevBayes	17.8	23.5

\* times are given in seconds

# RevBayes on Computer Clusters



# Aims for RevBayes

---

## 1) General and flexible model specification

- a) Availability of (common) models
- b) Extendability



## 2) Transparent

- a) No black-box!
- b) Well structured model specification
- c) Explicit models
- d) Documentation, examples and tutorials



## 3) Fast & Efficient

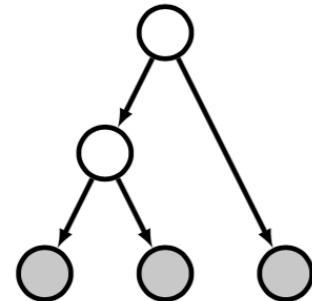
- a) Fast likelihood calculators
- b) Efficient (MCMC) algorithms, e.g., tree proposals



# Resources

---

- General website: [www.RevBayes.com](http://www.RevBayes.com)
- Code repository: [github.com/revbayes/revbayes](https://github.com/revbayes/revbayes)
- Tutorials: <http://revbayes.github.io/tutorials.html>
- Help: <https://revbayes.github.io/documentation/>
- Discussion-group/mailing-list:  
<https://groups.google.com/forum/#!forum/revbayes-users>

[Software](#)[Tutorials](#)[Workshops](#)[Jobs](#)[Developer](#)

## RevBayes

Bayesian phylogenetic inference using probabilistic graphical models and an interpreted language

### About

RevBayes provides an interactive environment for statistical computation in phylogenetics. It is primarily intended for modeling, simulation, and Bayesian inference in evolutionary biology, particularly phylogenetics. However, the environment is quite general and can be useful for many complex modeling tasks.

RevBayes uses its own language, Rev, which is a probabilistic programming language like [JAGS](#), [STAN](#), [Edward](#), [PyMC3](#), and related software. However, phylogenetic models require inference machinery and distributions that are unavailable in these other tools.

The Rev language is similar to the language used in R. Like the R language, Rev is designed to support interactive analysis. It supports both functional and procedural programming models, and makes a clear distinction between the two. Rev is also more strongly typed than R.

# Software



Software

Tutorials

Workshops

Jobs

Developer

## Download and Install RevBayes

Mac OS X

[Download Executable \(10.6+\)](#)

Windows

[Download Executable \(7+\)](#)

Source code

[GitHub Repository](#)

## Compiling from source

Linux

```
$ git clone https://github.com/revbayes/revbayes.git revbayes
$ cd revbayes/projects/cmake
$ ./build.sh
```

For the MPI version:

# Tutorials

[Software](#)[Tutorials](#)[Workshops](#)[Jobs](#)[Developer](#)

## RevBayes Tutorials

This list shows all of the RevBayes tutorials for learning various aspects of RevBayes and Bayesian phylogenetic analysis. Each one explicitly walks you through model specification and analysis set-up for different phylogenetic methods. These tutorials have been written for new users to learn RevBayes at home, at workshops, and in courses taught at the undergraduate and graduate levels. You may find that the styles are somewhat different between tutorials and that some have overlapping content.

Please see the [Tutorial Format](#) guide for details about how to read the tutorials.

Please see [Recommended Software](#) for links to various software programs you may need to download in order to follow the tutorials.

[Contribute!](#)

### Introduction to RevBayes and MCMC

# Tutorials

## Introduction to RevBayes

[Getting Started with RevBayes and Rev Language Syntax](#)

Overview of introductory topics

[Getting Started with RevBayes](#)

A very basic overview on how to use RevBayes

[Rev Language Syntax](#)

A very short introduction to the Rev language

[Introduction to Graphical Models](#)

A gentle introduction to graphical models, probabilistic programming, and MCMC using a simple linear regression example.

## Introduction to MCMC

[Introduction to Markov chain Monte Carlo \(MCMC\) Sampling](#)

Overview of topics on MCMC methods in RevBayes

[Introduction to MCMC](#)  
A simple Poisson regression example

[Introduction to MCMC using RevBayes](#)

Introduction to MCMC Simulation using a simple Binomial Model

[Introduction to MCMC using RevBayes](#)

A simple Archery example for building a hierarchical model and sampling under Markov Chain Monte Carlo

# Tutorials

---

## Model Selection and Testing

[General Introduction to Model selection](#)

Comparing relative model fit with Bayes factors

[Model selection of common substitution models for one locus](#)

Comparing relative model fit with Bayes factors

[Model selection of partition models](#)

Comparing relative model fit with Bayes factors

[Model averaging of substitution models](#)

Reversible-jump MCMC over substitution models

[Assessing Phylogenetic Reliability Using RevBayes and  \$P^3\$](#)

Model adequacy testing using posterior prediction (Data Version).

[Assessing Phylogenetic Reliability Using RevBayes and  \$P^3\$](#)

Model adequacy testing using posterior prediction (Inference Version).

# Tutorials

## Standard tree inference

[Understanding Continuous-Time Markov Models](#)

Simulating DNA sequence evolution with a die

[Nucleotide substitution models](#)

Phylogenetic inference of nucleotide data using RevBayes

[Partitioned data analysis](#)

Current Protocols in Bioinformatics - Phylogenetic Inference using RevBayes (Protocol #2)

[Discrete morphology - Tree Inference](#)

Phylogenetic inference with discrete morphological data

[Molecular dating](#)

Estimating species divergence times using RevBayes

## Relaxed Clocks & Time Trees

Comparing relaxed clock models & estimating rooted time trees

## Complex hierarchical models for phylogenetic inference

[Combined-Evidence Analysis and the Fossilized Birth-Death Process for Stratigraphic Range Data](#)

Joint inference of divergence times and phylogenetic relationships of fossil and extant taxa

[Combined-Evidence Analysis and the Fossilized Birth-Death Process for Analysis of Fossil and Extant Specimens](#)

Joint inference of divergence times and phylogenetic relationships of fossil and extant taxa

# Tutorials

## Diversification Rate Estimation

### Introduction to Diversification Rate Estimation

Overview of Analyses, Models and Theory

### Background on state-dependent diversification rate estimation

An introduction to inference using state-dependent speciation and extinction (SSE) branching processes

### Simple Diversification Rate Estimation

Comparing different constant-rate models of lineage diversification

### Episodic Diversification Rate Estimation

Estimating Rate-Variation Through Time in Diversification Rates

### Environmental-dependent Speciation & Extinction Rates

Estimating Correlation between Diversification Rates and Environmental Characters

### Branch-Specific Diversification Rate Estimation

How to estimate branch-specific shifts in diversification rates

### Macroevolutionary Analysis of Stratigraphic Range Data

Inference of diversification rates using the fossilized birth-death range process

### Diversification Rate Estimation with Missing Taxa

How to estimate diversification rates with incomplete taxon sampling

### State-dependent diversification with BiSSE and MuSSE

Inference using the binary/multiple state-dependent speciation and extinction (BiSSE/MuSSE) branching process

### State-dependent diversification with HiSSE

Inference using the hidden character binary/multiple state-dependent speciation and extinction with (HiSSE) branching process

### State-dependent diversification with HiSSE and ClaSSE

Hidden state and cladogenetic state-dependent speciation and extinction (HiSSE & ClaSSE) branching process

### Chromosome Evolution

Modeling chromosome evolution with ChromEvol, BiChroM, and ChromoSSE

# Tutorials

## Comparative methods

[Introduction to Models of Continuous-Character Evolution](#)

Overview of Standard Models of Continuous-Character Evolution

[Simple Ornstein-Uhlenbeck Models](#)

Estimating optima under Ornstein-Uhlenbeck evolution

[Simple Brownian Rate Estimation](#)

Estimating rates of Brownian-motion evolution

[Relaxed Brownian Rate Estimation](#)

Estimating branch-specific rates of Brownian-motion evolution

[State-Dependent Brownian Rate Estimation](#)

Estimating state-dependent rates of Brownian-motion evolution

[Multivariate Brownian Motion](#)

Accounting for correlations among continuous traits

[Relaxed Ornstein-Uhlenbeck Models](#)

Estimating lineage-specific optima under Ornstein-Uhlenbeck evolution

[Discrete morphology - Ancestral State Estimation](#)

Ancestral State Estimation and Testing for Irreversibility

# Tutorials

---

## Biogeography

### Introduction to Phylogenetic Models of Discrete Biogeography

Overview of the Dispersal-Extirpation-Cladogenesis (DEC) model

### Simple Phylogenetic Analysis of Historical Biogeography

Estimating ancestral ranges using the Dispersal-Extirpation-Cladogenesis (DEC) model

### Advanced Phylogenetic Analysis of Historical Biogeography

Ancestral range estimates using the Dispersal-Extirpation-Cladogenesis (DEC) model while incorporating time-dependent and region-dependent effects on dispersal rates

### Biogeographic Dating of Divergence Times

Estimating divergence times with molecular, biogeographic, and paleogeographic evidence under the Dispersal-Extirpation-Cladogenesis (DEC) model

# RevScripter

## RevScripter

Home

Data

Substitution Model

Tree Model

MCMC

Scripts

### RevScripter: An Introductory Tool for Creating RevBayes Analysis Scripts

RevScripter is a web-based tool that enables researchers to create scripts describing phylogenetic models and analyses in the Rev language. These script files then can be executed in the program [RevBayes](#). RevBayes is a statistical inference tool for Bayesian phylogenetics.

RevBayes is extremely flexible and the Rev language and graphical model framework enable researchers to apply a very rich set of complex evolutionary models. The development of RevScripter was motivated by a need to provide a familiar introductory interface for RevBayes and the Rev language. It will not be possible to maintain a menu-driven graphical user interface that encompasses the wide range of models and methods available in RevBayes. Thus, this tool will be useful for those new to RevBayes to set up standard analyses and easily see how elements of the model are reflected in the generated Rev language script.

Currently, the options available in RevScripter are limited and the tool is still very much in development. The tool can create scripts for running an unrooted analysis using nucleotide data under standard substitution models. For more details about the available models and analyses, please see the README file in the [RevScripter source repository](#).

### How to use RevScripter

RevScripter is set up to guide you through the options that must be specified in the script. Start with the *Data* tab or click the *Get Started* button below. Complete your model specification and MCMC options. When you have specified a complete analysis, go to the *Scripts* tab and generate the Rev script. You will then be able to download the script as a text file or copy it to a file that you have saved locally. With a completed analysis script for RevBayes, you can run the file on your own computer (or HPC resources).

To run the generated script in RevBayes, be sure that the file names and paths in the generated script are correct. Then open RevBayes and type the command: `source myscript.Rev`, where `myscript.Rev` is the name of the file you downloaded from RevScripter.

### Source

The source code (Javascript) for RevScripter is hosted on GitHub at <https://github.com/revbayes/revscripter>

Get Started

# Forum/Mailing group

Google Search for messages  Groups 3 

Groups NEW TOPIC C Mark all as read Actions Filters  

My groups Shared publicly revbayes-users 30 of 93 topics (89 unread)  Manage · Members · About

Welcome to revbayes-users, a forum and mailing list to discuss RevBayes-related topics, including: RevBayes installation and use, scripting and programming, phylogenetics, population genetics, models of evolution, graphical models, etc.

If you wish to report a bug in RevBayes, use our [GitHub Issues](#) page.

Edit welcome message Clear welcome message

 **ntaxa error & drawing new initial states** (3)  
By Eva Drukker - 3 posts - 20 views 8 Mar

 **Fixed ancestral states tutorial?** (2)  
By Simon Frost - 2 posts - 14 views 6 Mar

 **Updating installation page** (2)  
By Brian O'Meara - 2 posts - 19 views 6 Mar

 **Interest in RevBayes notebooks?** (4)  
By Simon Frost - 4 posts - 14 views 3 Mar

Click on a group's star icon to add it to your favourites

Favourites

Recently viewed

- Google C++ Testi...
- revbayes-develop...
- revbayes-users**
- Bay Area Populati...
- beast-users

Recent searches