

Bayesian Phylogenetic Inference using RevBayes:

Introduction to RevBayes

Sebastian Höhna

Division of Evolutionary Biology
Ludwig-Maximilians Universität, München





Bastien Boussau



Tracy Heath



John Huelsenbeck



Fredrik Ronquist

RevBayes Team



Brian Moore



Michael Landis



Nicolas Lartillot



Sebastian Höhna



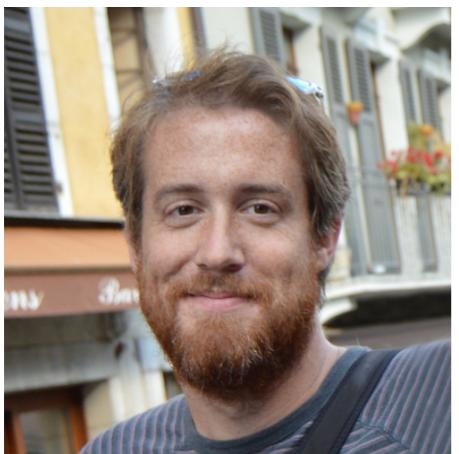
Will Freyman



April Wright



Mike May



Will Pett

RevBayes Team²

Lyndon Coghill



Jeremy Brown

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) Easy to learn

- a) Well structured model specification
- b) Explicit models
- c) Documentation, examples and tutorials

3) Fast & Efficient

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

Graphical Model (DAG)

a)



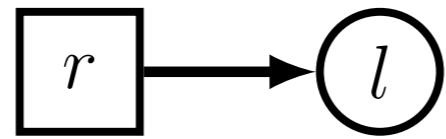
```
# constant node  
r <- 10
```

Graphical Model (DAG)

a)



b)



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

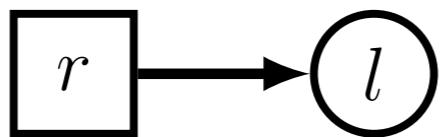
```
# stochastic node  
l ~ dnExp(r)
```

Graphical Model (DAG)

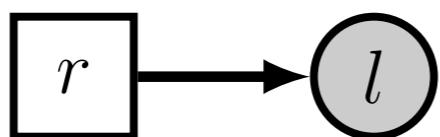
a)



b)



c)



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

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

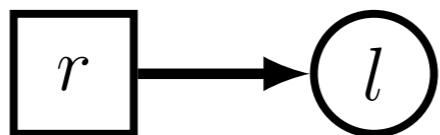
```
# stochastic node (observed)  
l.clamp(0.1)
```

Graphical Model (DAG)

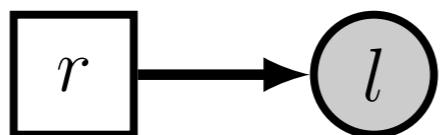
a)



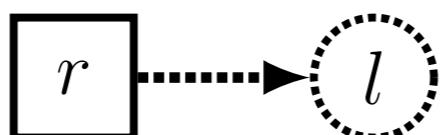
b)



c)



d)



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

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

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

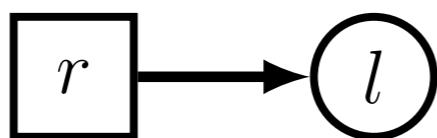
```
# deterministic node  
l := exp(r)
```

Graphical Model (DAG)

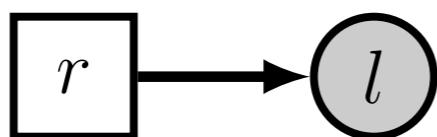
a)



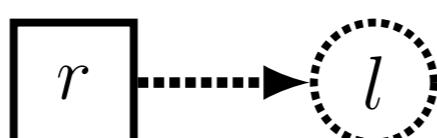
b)



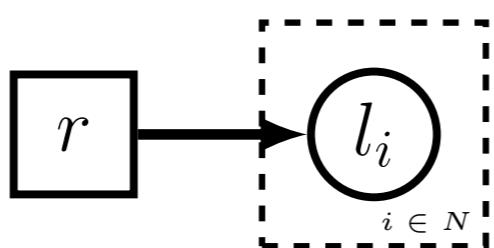
c)



d)



e)



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

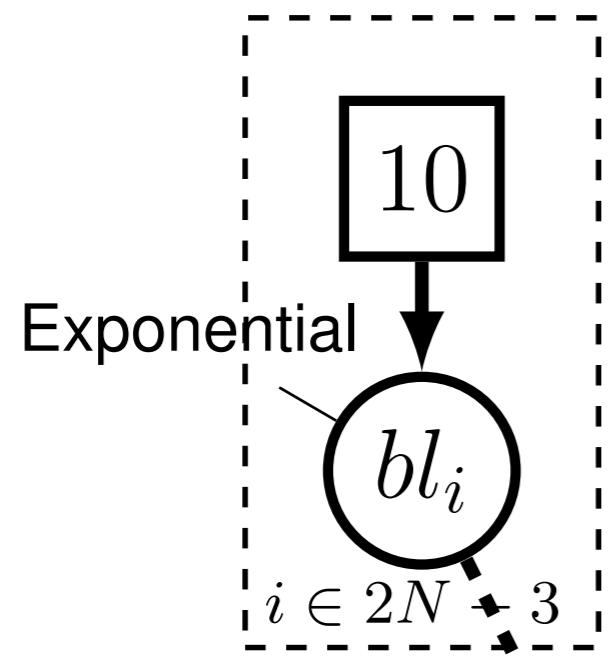
```
# stochastic node  
l ~ dnExp(r)
```

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

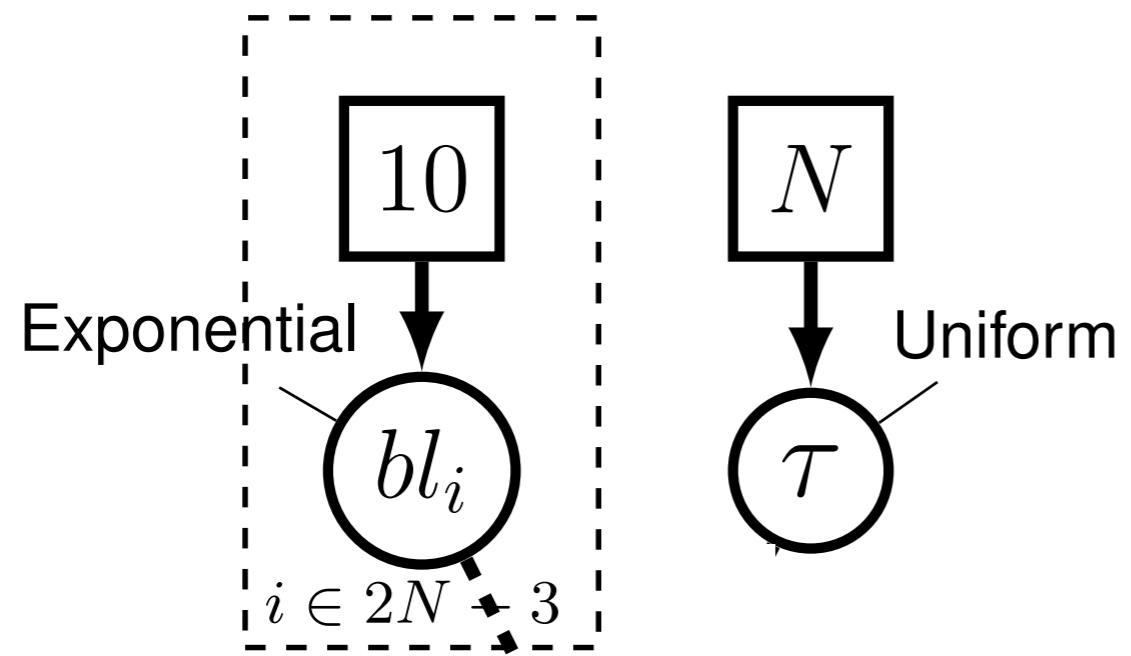
```
# deterministic node  
l := exp(r)
```

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

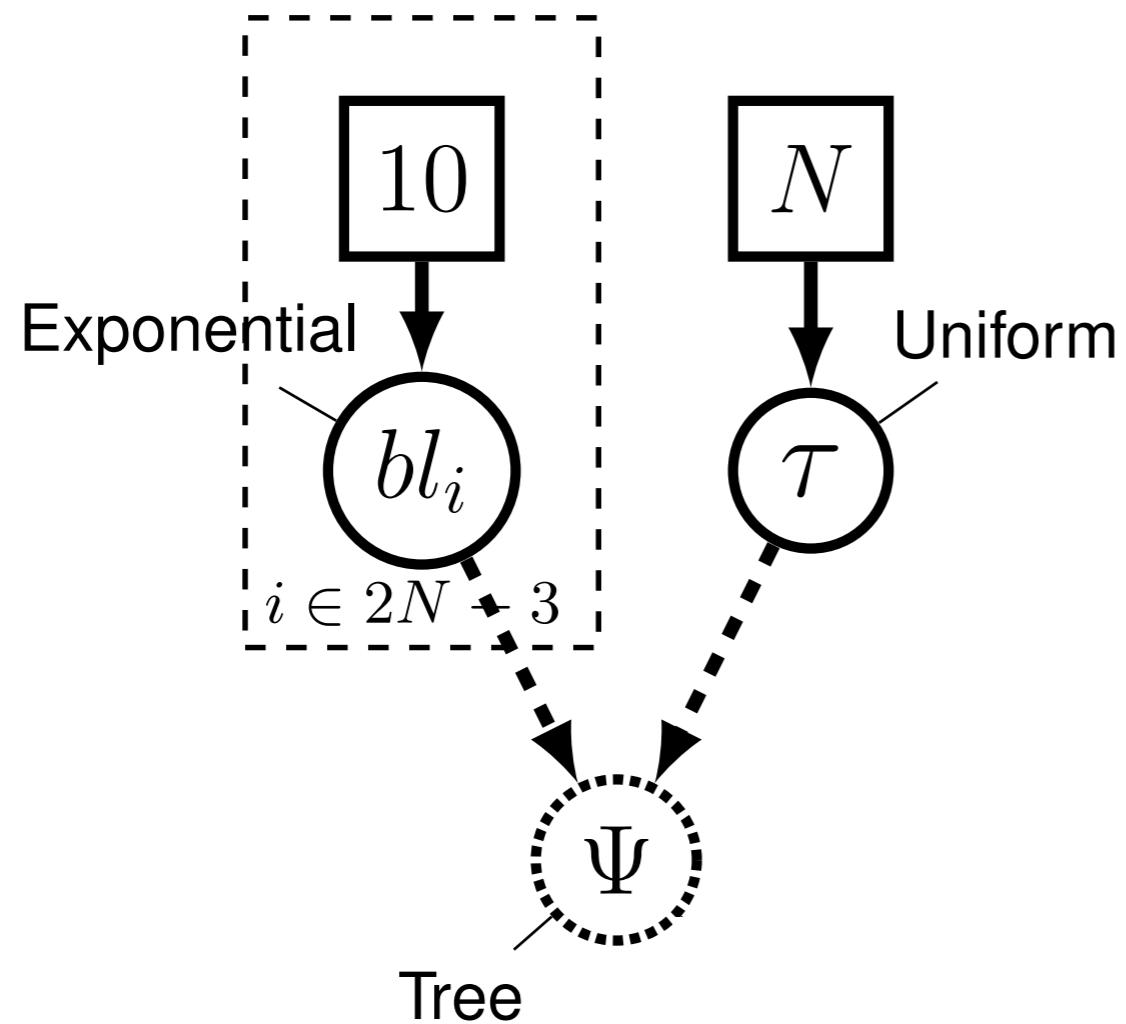
GTR+Gamma Model with Prior



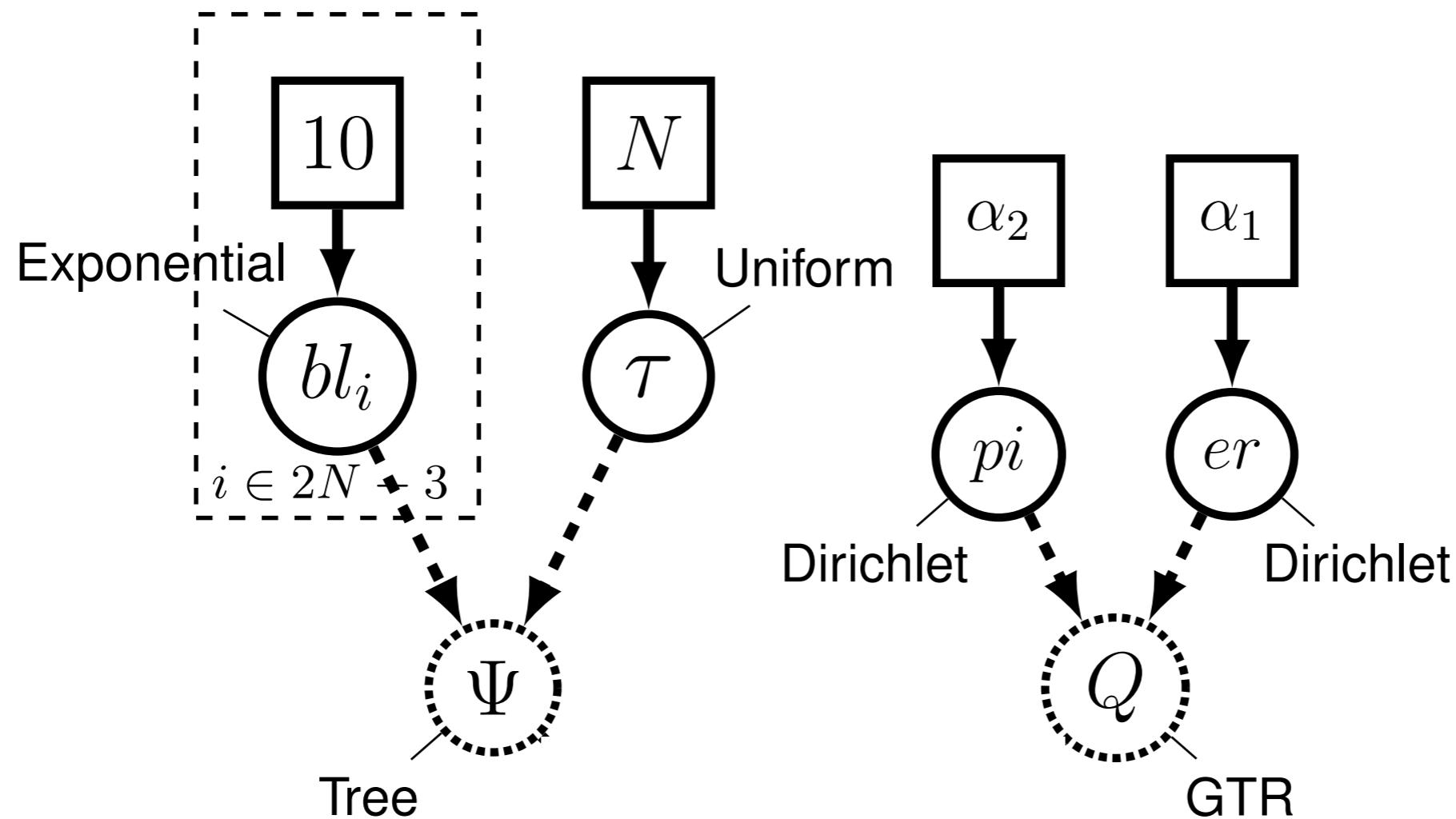
GTR+Gamma Model with Prior



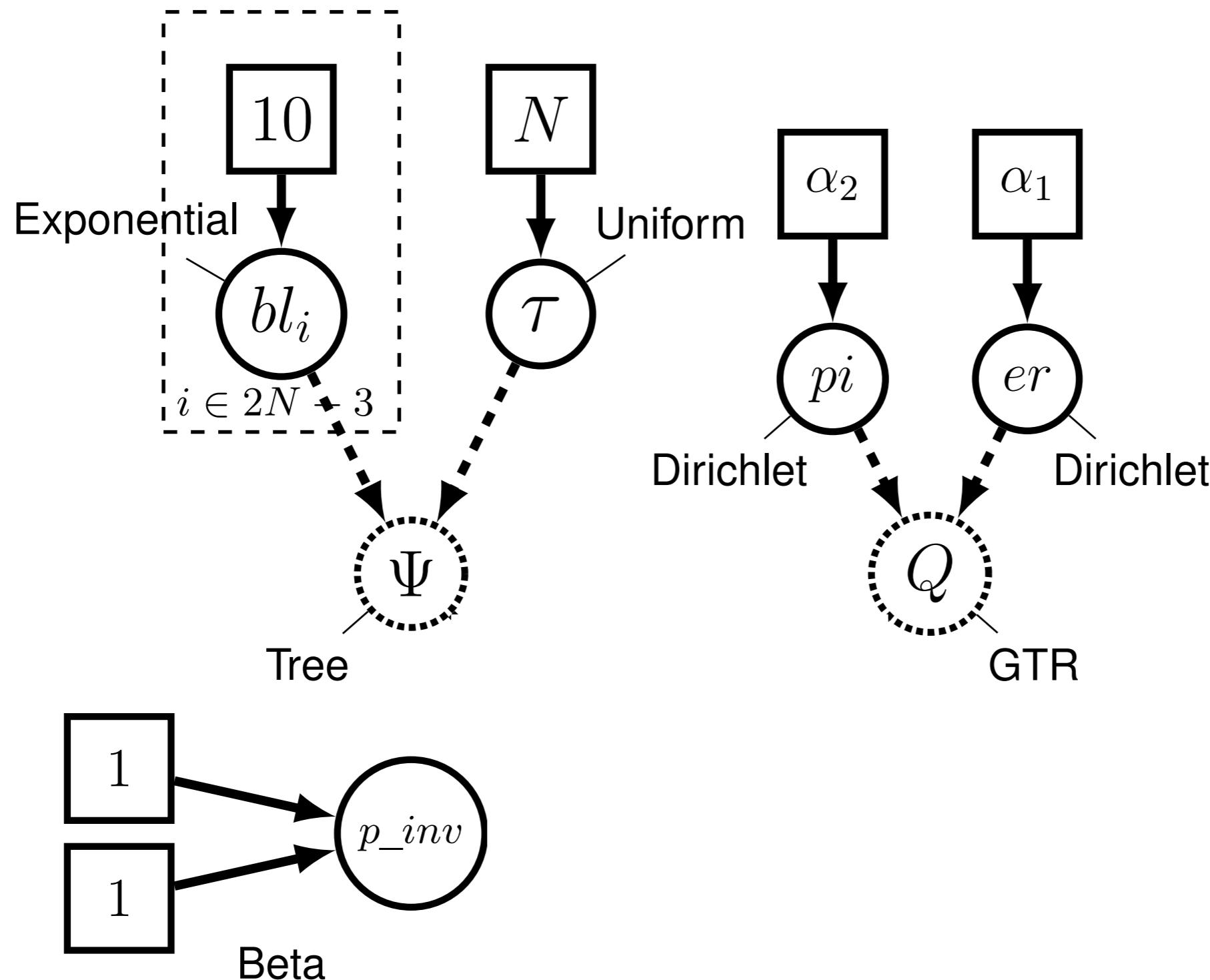
GTR+Gamma Model with Prior



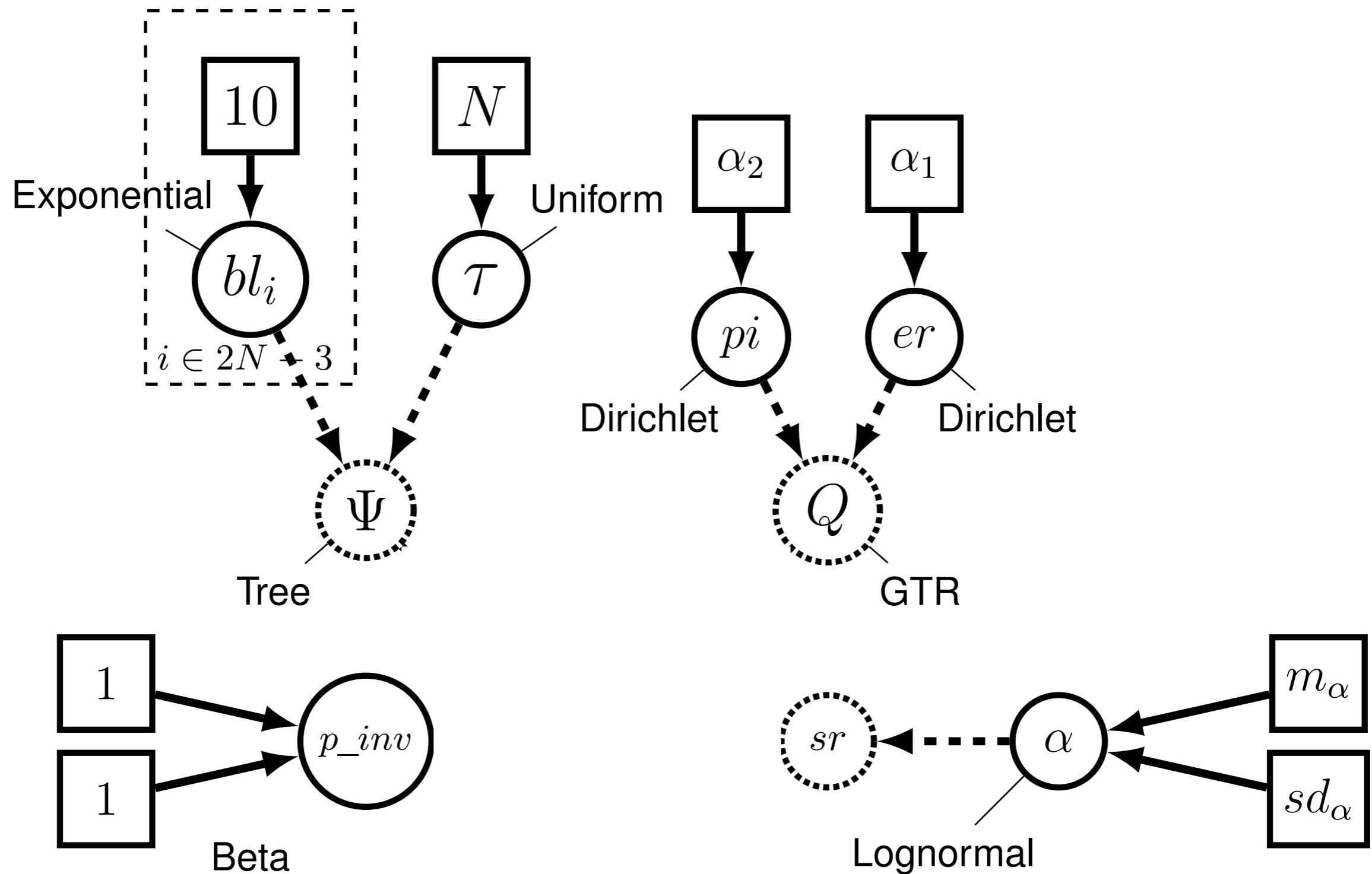
GTR+Gamma Model with Prior



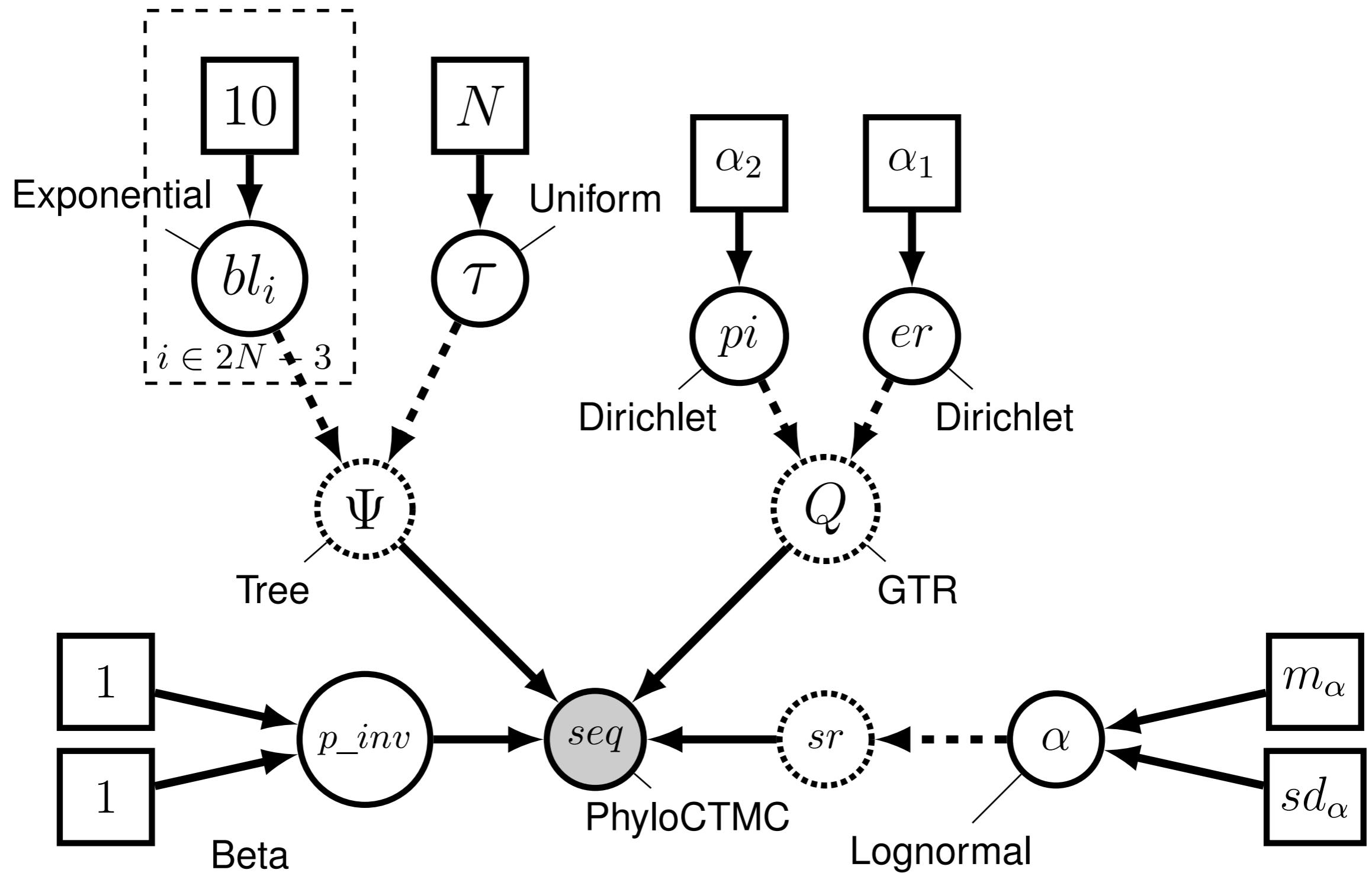
GTR+Gamma Model with Prior



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



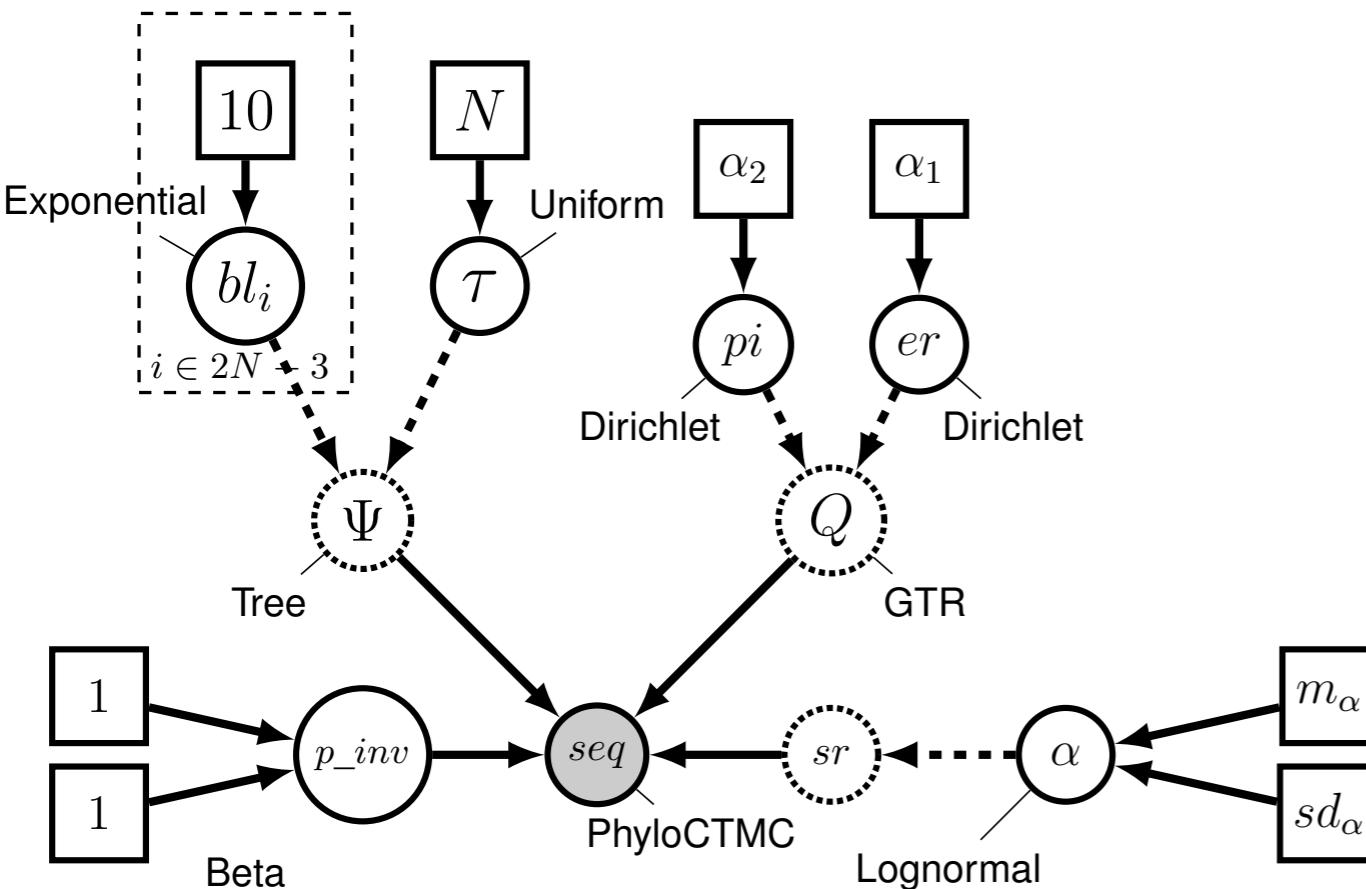
2) Easy to learn

- a) Well structured model specification
- b) Explicit models
- c) 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_inv,
                    siteRates=sr, type="DNA")
seq.clamp(data)
  
```

Manual (Compilation of 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

**Over 300
pages !!!**

Aims for RevBayes

1) General and flexible model specification

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



2) Easy to learn

- a) Well structured model specification
- b) Explicit models
- c) 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

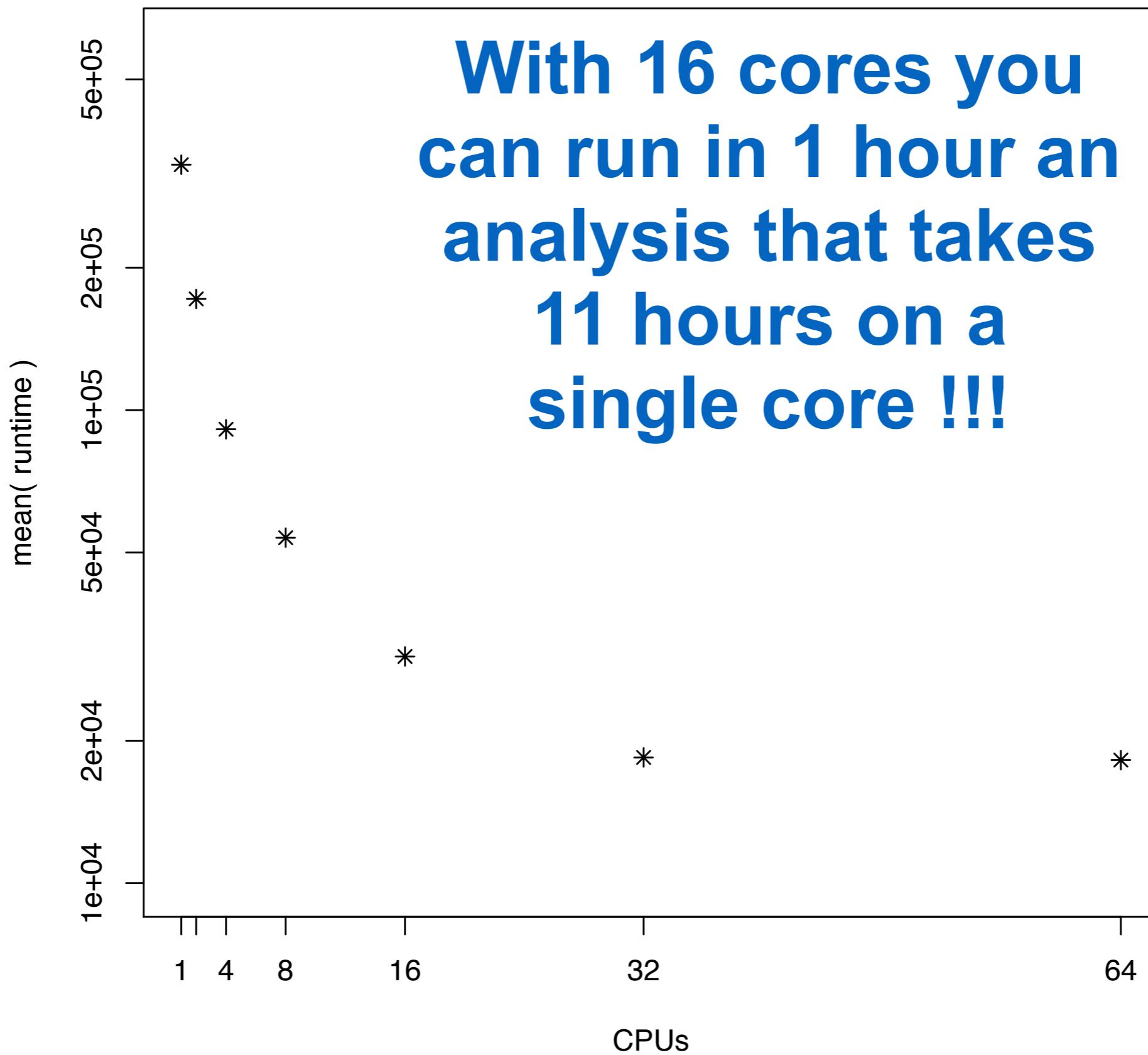
MCMC:

- 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) Easy to learn

- a) Well structured model specification
- b) Explicit models
- c) 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
- Code repository: github.com/revbayes/revbayes
- Tutorials: <http://revbayes.github.io/tutorials.html>
- Help: <http://revbayes.github.io/docs.html>
- Discussion-group/mailing-list:
<https://groups.google.com/forum/#!forum/revbayes-users>

RevBayes.com

RevBayes

Bayesian graphical models

About

Software

Introduction

Docs

Tutorials

Validation

Forum

Citation

 github

 #RevBayes

Software

Software

Windows

Supported versions: 7+

Mac OS X

Supported versions: 10.6+

Source

Public git repository

Installing RevBayes from source on Mac OS X

1) Make sure that you have a c++ compiler installed on your computer. GCC 4.2 (or higher) and Apple LLVM version 6.0 have been used successfully. If you don't have a c++ compiler, you can get it on a Mac when you download and install [XCode](#).

2) Make sure that you have [CMake](#) installed. One option to install CMake is using homebrew:

```
ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"  
brew install cmake
```

3) Download RevBayes from our [github repository](#). You can either clone the repository using git by running the following command in the terminal

```
git clone https://github.com/revbayes/revbayes.git revbayes-master
```

or you can download the zipped source code directly by clicking in the link somewhere on the right side. In this case you need to unzip the source files.

Remember where you downloaded the sources. You will need this directory later.

4) Open a terminal and go to the RevBayes source directory if you haven't already done so, e.g.,

```
cd revbayes-master
```

Next, go into the projects and then the cmake subdirectory:

```
cd projects/cmake
```

Now you can either build the standard version, e.g., for older Mac OS or Linux distributions using

```
./build.sh
```

This builds the standard RevBayes executable **rb**. Alternatively, you can build the MPI version using instead the

Tutorials

Tutorials

Manual

The complete manual containing all tutorials.

Introduction

Getting Started

Getting Started with RevBayes.

Rev basics

Basic language features and simple statistical (Poisson) regression model example.

Tutorial

Basic Introduction to Markov chain Monte Carlo.

Example script (Simple MCMC)

Example script (Moves)

Example script (RevBayes MCMC)

Basic phylogeny reconstruction and substitution models

Tutorial

Models of molecular character evolution (substitution models).

Primates data

Example script (Jukes-Cantor)

Example script (GTR+Gamma+Inv)

Tutorials

MCMC Strategies and Model Selection

Tutorial	Running and diagnosing an MCMC analysis		
Tutorial	Compute marginal likelihoods to selection between models	Primates data	Example script

More substitution models, data partitioning, and discrete morphology

Tutorial	Strategies for partitioning molecular data matrices, e.g., by genes and codon position.	Primates cytB data	Example script for MCMC (uniform partition)
		Primates cox2 data	Example script for MCMC (by-gene partition)
			Example script for MCMC (by-codon partition)
			Example script for marginal likelihood (uniform partition)

Tutorial	Models of morphological character evolution (substitution models).	Primates data	Example scripts.
--------------------------	--	-------------------------------	----------------------------------

Tutorials

Divergence Time Estimation

Tutorial	Divergence time estimation and node calibrations	Primates data	Example scripts.
Tutorial	Divergence time estimation and relaxed clocks	Primates data	Example scripts.
Tutorial	Divergence time estimation using tip-dating	Primates data	Example scripts.
Tutorial	Time trees and divergence time estimation using relaxed clocks	Bears data Bears tree	

Diversification Rate Estimation

Tutorial	Basic diversification using constant-rate pure birth and birth-death models.	Primates phylogeny	Example scripts.
Tutorial	Diversification rates through-time using episodic birth-death models.	Primates phylogeny	Example scripts.
Tutorial	Incomplete taxon sampling.	Primates phylogeny	Example scripts.
Tutorial	Branch-specific diversification rate estimation.	Primates phylogeny	Example scripts.
Tutorial	Character-dependent diversification rate estimation.	Primates phylogeny Primates morphology	Example MCMC script.

Tutorials

Biogeography

Tutorial	Biogeographic dating under the DEC model	Silversword data	Example scripts
			Example output

Species tree estimation

Tutorial	Concatenation of multiple genes	Primates data	Example script
Tutorial	Multi-species coalescent analysis	Primates data	Example script

Continuous trait evolution

Tutorial	Continuous trait evolution models: Brownian Motion	Primates data	Example script
Tutorial	Continuous trait evolution models: Omstein-Uhlenbeck models	Primates data	Example scripts

Model Fit

Tutorial	Assessing Phylogenetic Reliability	Example data	Example scripts
--------------------------	------------------------------------	------------------------------	---------------------------------

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

 ntanax 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

Recently viewed

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

Recent searches