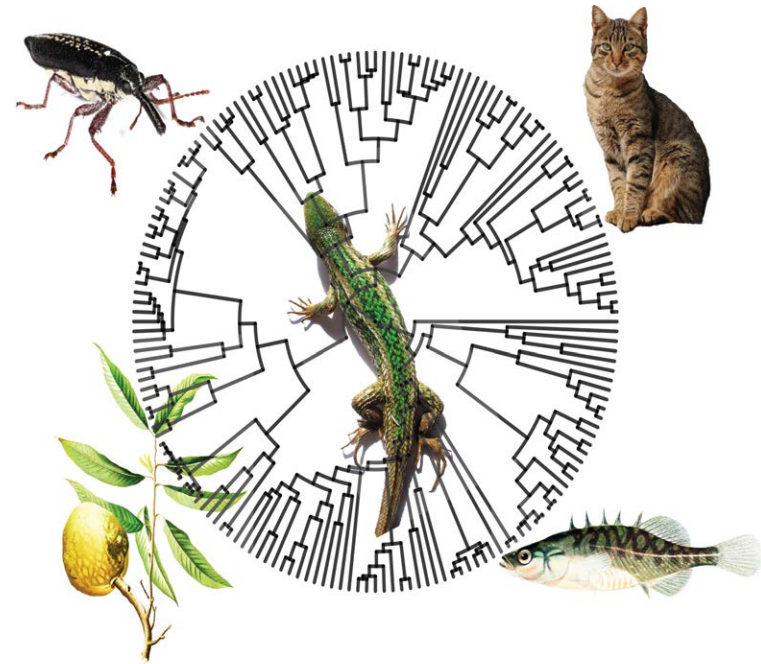


# Biology 559R: Introduction to Phylogenetic Comparative Methods

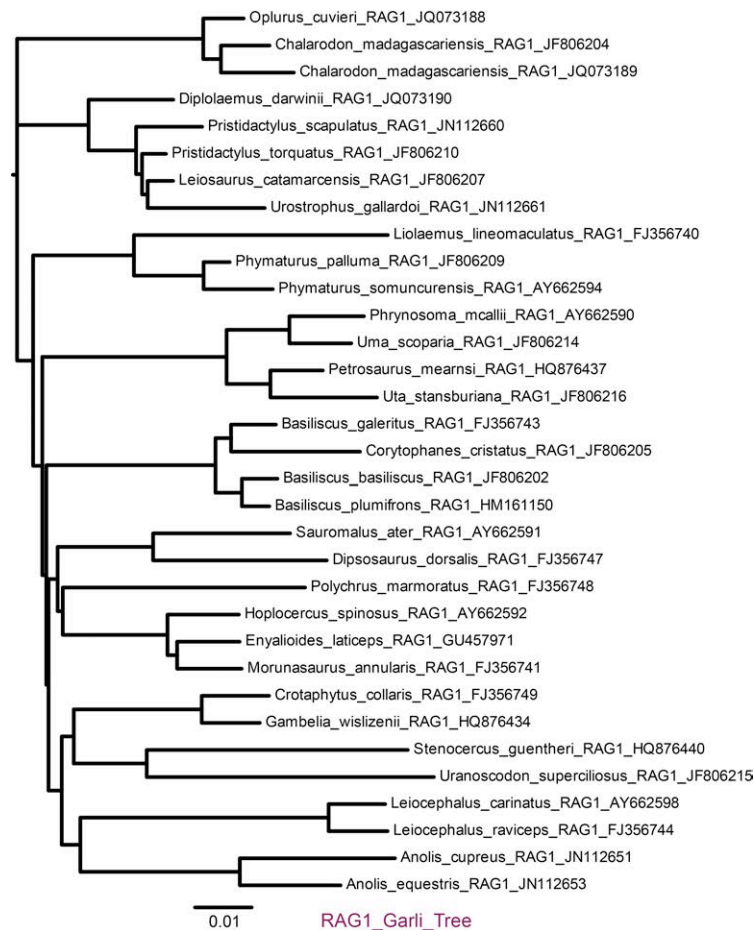
## Topics for this week (Feb 3 & 5):

- Chronogram estimation:  
Penalized Likelihood Approach  
BEAST
- Presentations of your projects



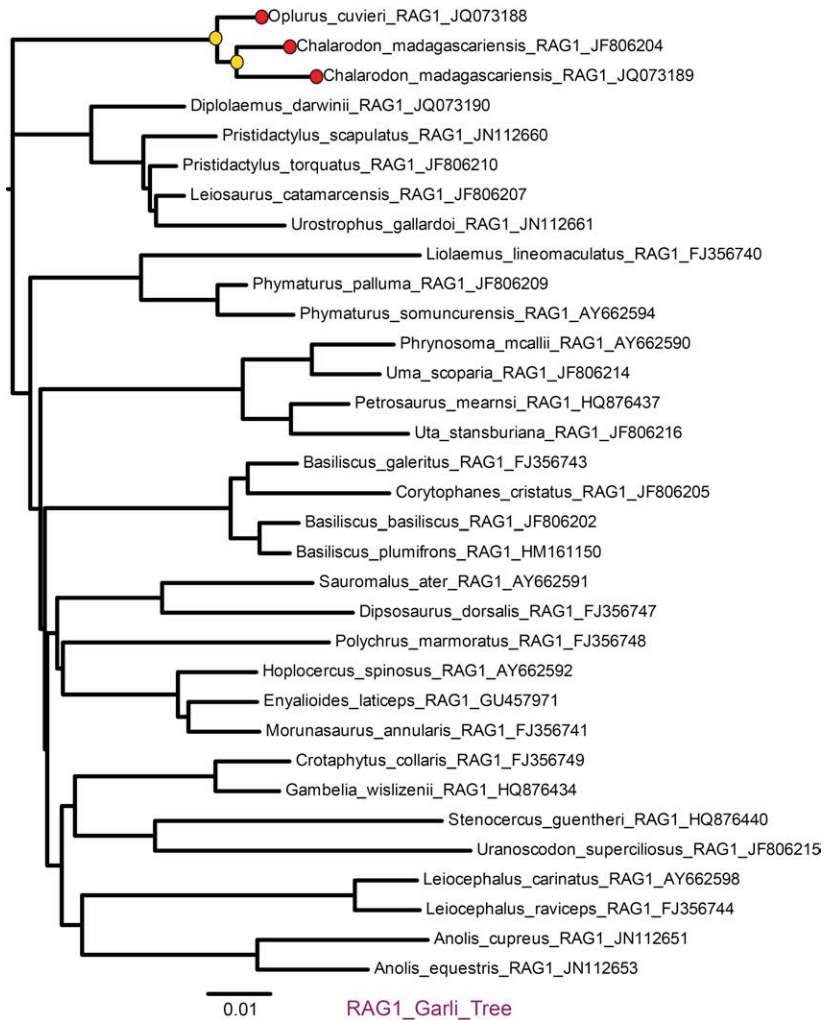
# The Anatomy of Trees

- The trees (phylogenies) represent the evolutionary relationships among a group of organisms of interest.
- A phylogeny is graph formed by a set of nodes and a set of edges that connect pairs of nodes.



# The Anatomy of Trees

- A node in a phylogeny represents an extant taxon (i.e., a tip or leaf) or a common ancestor for a set of taxa (internal nodes).



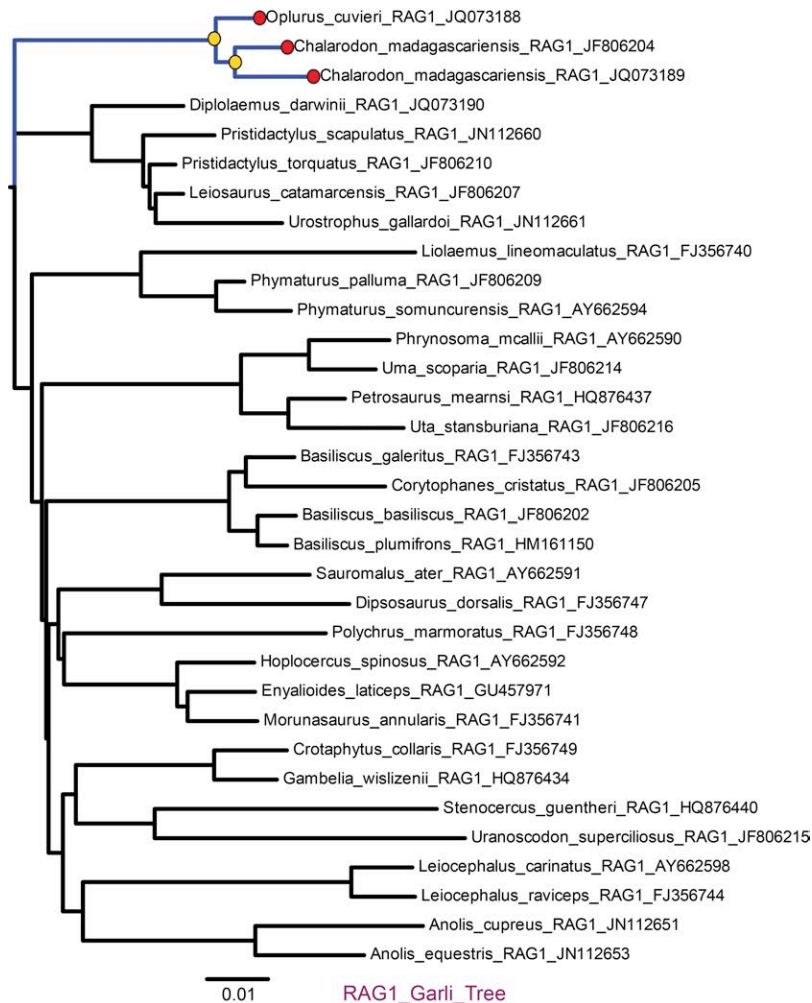
Nodes:

extant taxon (i.e., a tip or leaf)

common ancestor (internal nodes)

# The Anatomy of Trees

- An edge is the part of the phylogeny that connects two nodes (or tips), and is represented by a connecting line (i.e., a branch). We interpret the length of an edge as the **amount of evolutionary change** or time for transition from an ancestral taxon to a descendant taxon.

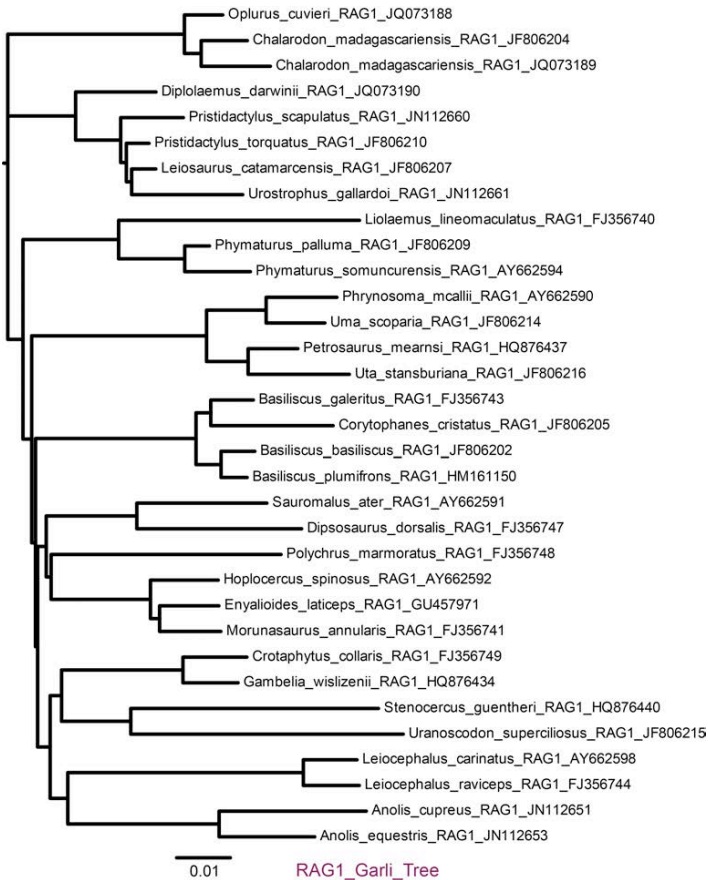


Edges or branches  
(see scale bar for reference)

# The Anatomy of Trees

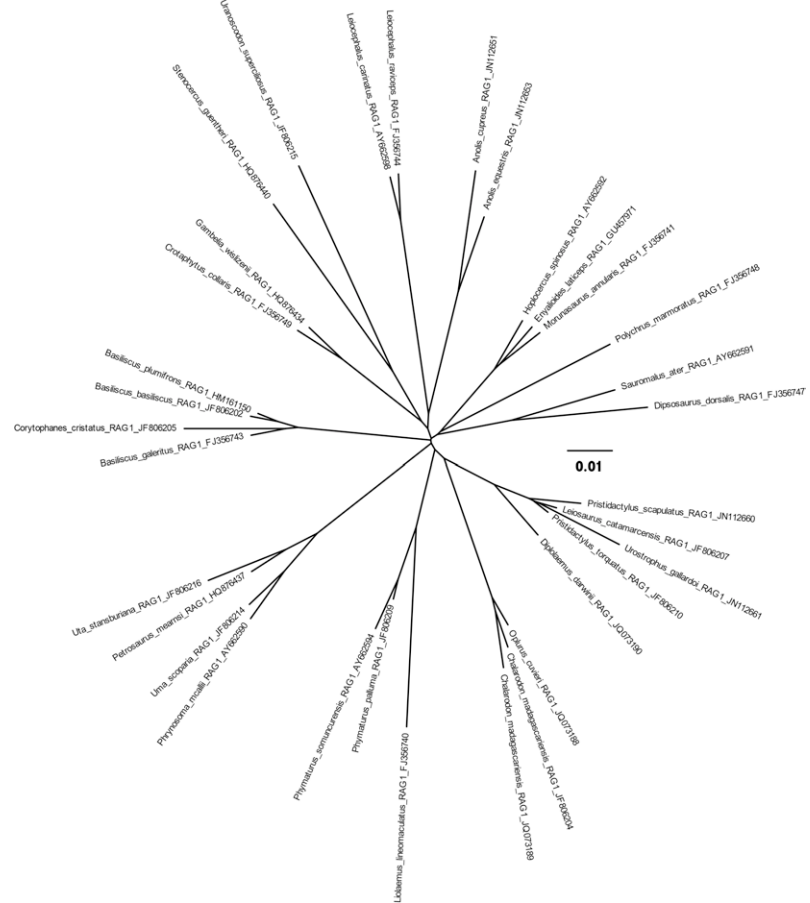
- The root is the node that represents the last common ancestor of all taxa in the tree.

Rooted tree:



A phylogeny where the time is represented by a single direction (bottom-up)

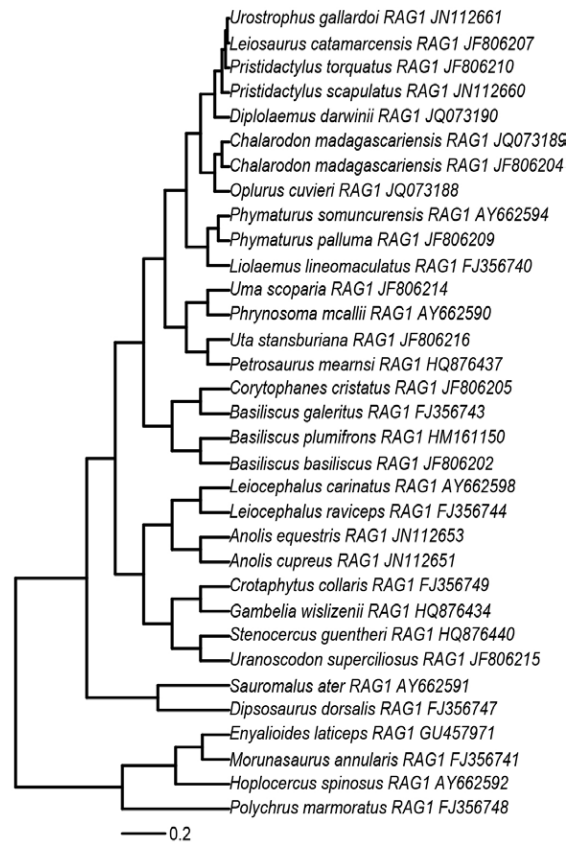
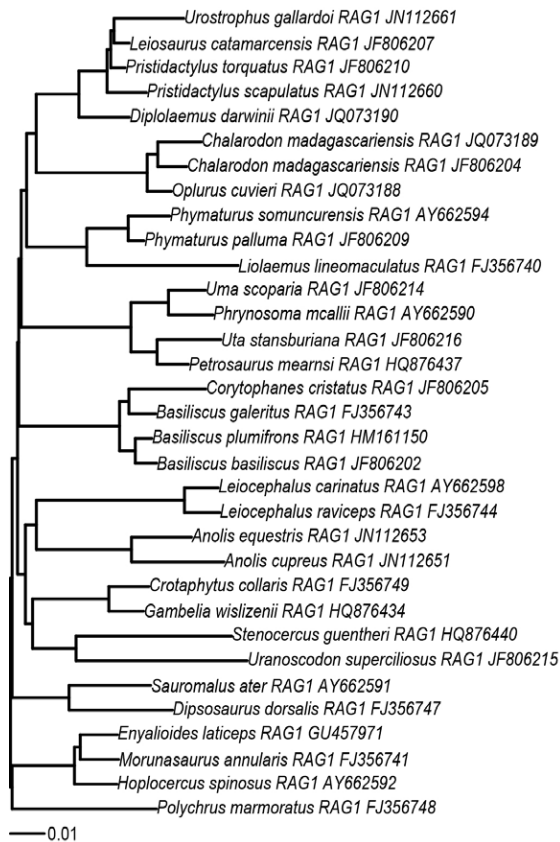
Unrooted tree



A phylogeny without a reference of time

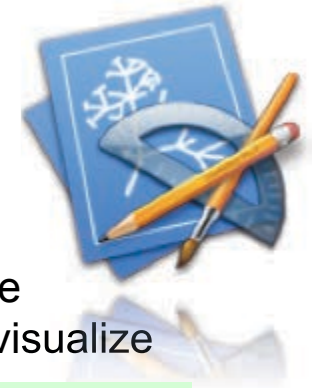
# Time-calibrated Phylogenies: Chronograms

- The trees that we have inferred have branch lengths that correspond to the amount of change that has happened on each tip and edge
- Most methods and analyses of character evolution are better suited if we use phylogenies whose branch lengths correspond with time (i.e., these trees are **ultrametric**)

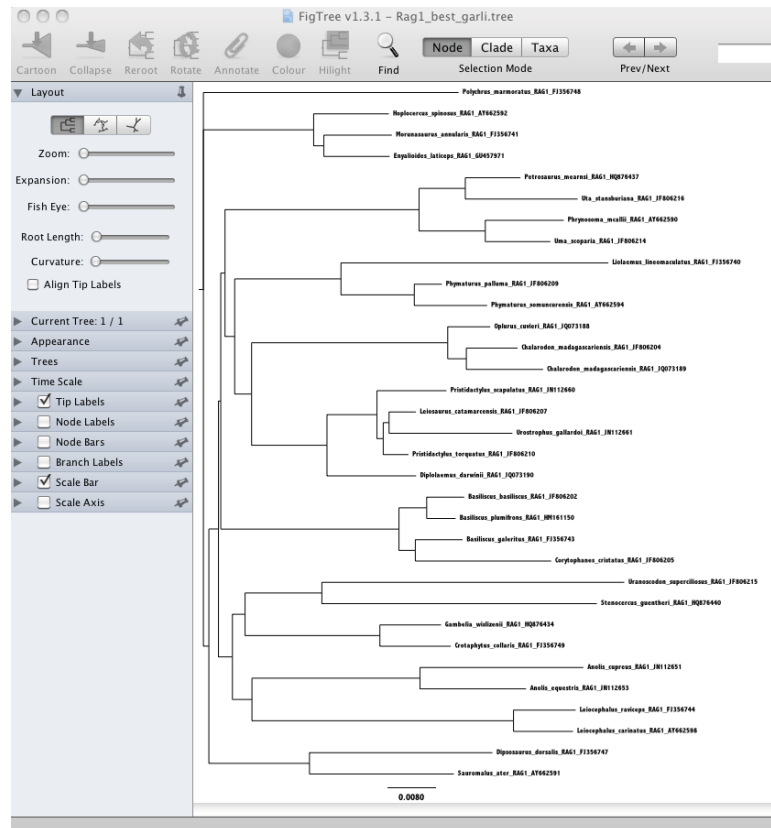


ultrametric tree is one in which the distances from the root to every branch tip are equal.

# Chronogram Estimation: A Penalized Likelihood Approach



- We are going to use R and FigTree to estimate an starting chronogram to be input in BEAST. You can use only R for this purpose, but I think it is useful to visualize these first steps.



← the manipulation menu is not available

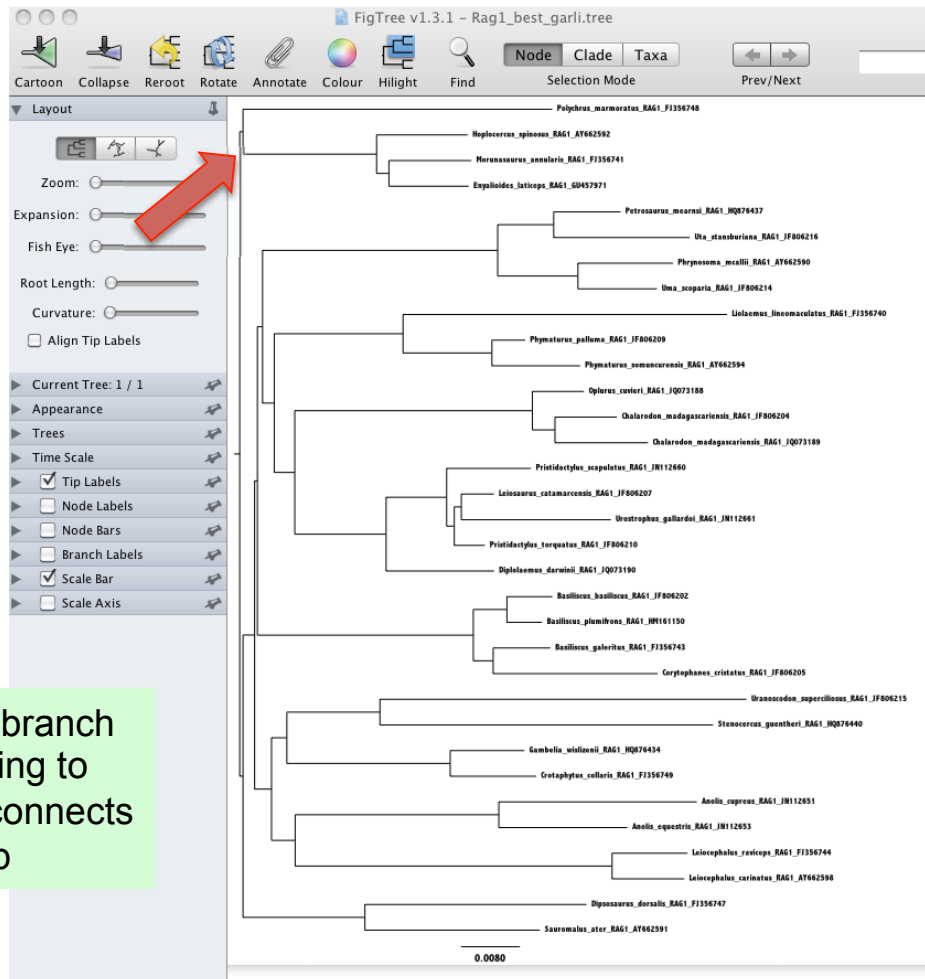
- Open the [Rag1\\_best\\_garli.tree](#) from the course website that we estimated using garli



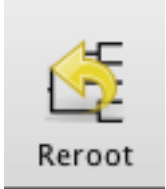
# Chronogram estimation: A Penalized Likelihood Approach



- We are going to 'root' (i.e., give a directionality to) the tree



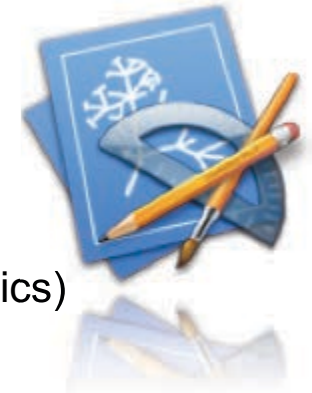
2) the manipulation menu will be activated  
3) Then click on the reroot tool



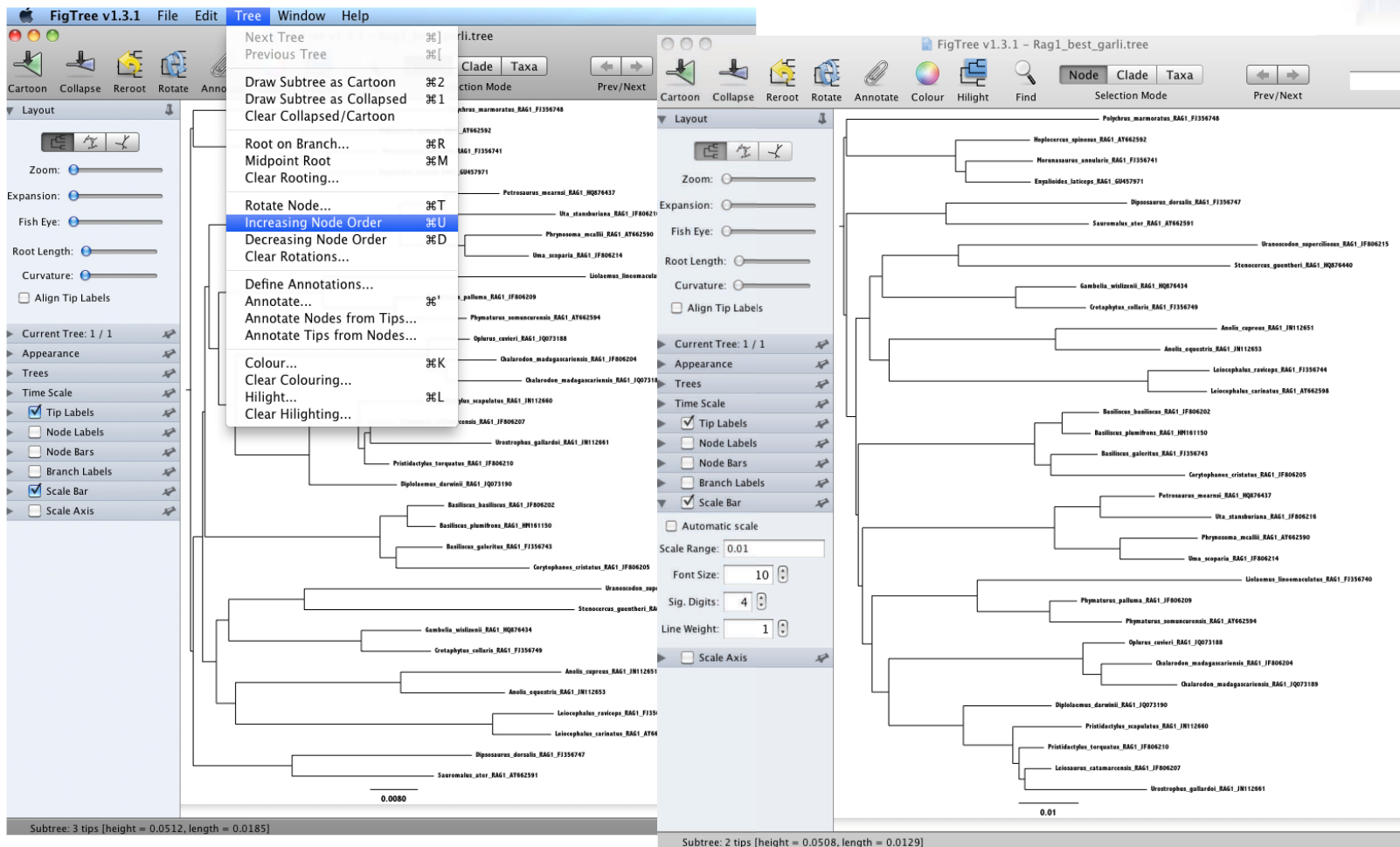
1) click on the branch that we are going to consider that connects to the outgroup



# Chronogram estimation: A Penalized Likelihood Approach



- Our tree will change in shape and then increase the node order (for aesthetics)



# Chronogram estimation: A Penalized Likelihood Approach



- Then save it in amenable format for R (i.e., newick)

The screenshot shows the FigTree v1.3.1 interface. The 'File' menu is open, and 'Export Trees...' is selected. The 'Export Trees' dialog box is open, showing 'Tree file format: Newick' and 'Save as currently displayed' checked. The background shows a phylogenetic tree with various species names and a scale bar.

Species names visible in the tree include: *Polychrus marmoratus*, *Hoplacercus spinosus*, *Morone muroneus*, *Eupholidius laticeps*, *Uma scoparia*, *Lioleum lineoculatum*, *Phymaturus pallens*, *Phymaturus somnarensis*, *Oplurus cavirostris*, *Chalarodon madagascariensis*, *Diploleum darwini*, *Pristidactylus scapellato*, *Pristidactylus torquatus*, *Leliosaurus catamarconensis*, and *Urostrophus gallardoi*.

# Chronogram estimation: A Penalized Likelihood Approach



- This is how a tree in newick looks like

```
((Polychrus_marmoratus_RAG1_FJ356748:0.04283221,(Hoplocercus_spinosus_RAG1_AY662592:0.01261769,
(Morunasaurus_annularis_RAG1_FJ356741:0.011411379999999999,Enyalioides_laticeps_RAG1_GU457971:0.011038409999999999
9):0.0017101999999999995):0.018509200000000003):4.5878E-4,
((Dipsosaurus_dorsalis_RAG1_FJ356747:0.03070635,Sauromalus_ater_RAG1_AY662591:0.02422809):0.016886570000000004,
(((Uranoscodon_superciliosus_RAG1_JF806215:0.050760970000000001,Stenocercus_guentheri_RAG1_HQ876440:0.04615523):
0.0128970299999999997,
(Gambelia_wislizenii_RAG1_HQ876434:0.0102686599999999999,Crotaphytus_collaris_RAG1_FJ356749:0.0119067):
0.0226298299999999997):0.00212305000000000013,
((Anolis_cupreus_RAG1_JN112651:0.0274269799999999997,Anolis_equestris_RAG1_JN112653:0.0178015000000000005):
0.0282085800000000004,
(Leiocephalus_raviceps_RAG1_FJ356744:0.0104387100000000004,Leiocephalus_carinatus_RAG1_AY662598:0.009998069999999999
98):0.04389986):0.003281209999999999927):0.002340889999999999982,
(((Basiliscus_basiliscus_RAG1_JF806202:0.0062940599999999997,Basiliscus_plumifrons_RAG1_HM161150:0.00486095999999999
975):0.00465753,
(Basiliscus_galeritus_RAG1_FJ356743:0.0079349999999999998,Corytophanes_cristatus_RAG1_JF806205:0.02278752):
0.00274980000000000036):0.0299112699999999997,
(((Petrosaurus_mearnsi_RAG1_HQ876437:0.0092565,Uta_stansburiana_RAG1_JF806216:0.0188545000000000003):
0.0077617599999999995,
(Phrynosoma_mcallii_RAG1_AY662590:0.01316908,Uma_scoparia_RAG1_JF806214:0.0109275100000000001):
0.0111173799999999996):0.0325743,((Liolaemus_lineomaculatus_RAG1_FJ356740:0.04486896,
(Phymaturus_palluma_RAG1_JF806209:0.00461683000000000002,Phymaturus_somuncurensis_RAG1_AY662594:0.012232880000000000
2):0.0123057799999999995):0.0178392200000000003,((Oplurus_cuvieri_RAG1_JQ073188:0.00713529000000000025,
(Chalarodon_madagascariensis_RAG1_JF806204:0.00859730000000000002,Chalarodon_madagascariensis_RAG1_JQ073189:0.01288
305):0.0031266100000000002):0.0329204800000000001,(Diplolaemus_darwinii_RAG1_JQ073190:0.0149654800000000003,
(Pristidactylus_scapulatus_RAG1_JN112660:0.0116696100000000004,
(Pristidactylus_torquatus_RAG1_JF806210:0.0042293799999999998,
(Leiosaurus_catamarcensis_RAG1_JF806207:0.0045133099999999995,Urostrophus_gallardoi_RAG1_JN112661:0.0207045499999
9995):9.772099999999992E-4):0.00105465000000000042):0.008388519999999997):0.0126527000000000003):
0.00283630999999999947):0.00162806000000000006):7.123299999999971E-4):3.857399999999953E-4):
0.001610090000000000085):4.587799999999978E-4);
```

# Chronogram estimation: A Penalized Likelihood Approach



- You can use r8s: <http://loco.biosci.arizona.edu/r8s/> See Sanderson (2002) and the r8s manual for further information. However, most of it has been implemented in R.
- We are going to use 'ape' package to ultrametrize our tree. Load the following package

```
library(ape)
```

- Make a working directory and select it as your working. Suggestion: you can also use `setwd()`.

R top menus: Misc>Change Working Directory (select the directory that has our file)

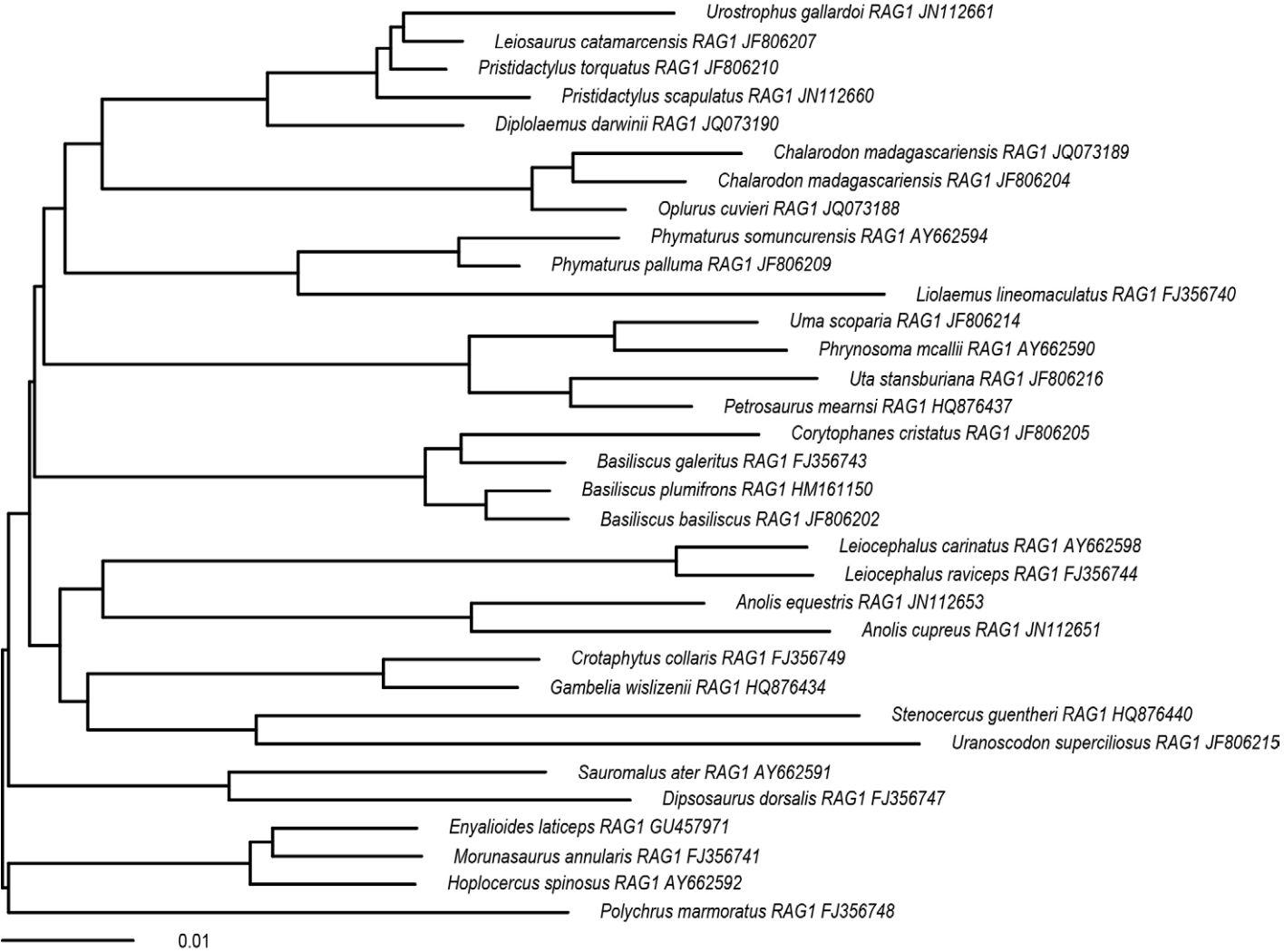
- Read our RAG1-phylogeny

```
rag1_rooted_garli <-read.tree(file = "Rag1_best_garli_rooted.newick")
```

- Let's take a look to our phylogeny

```
plot(rag1_rooted_garli)  
add.scale.bar(x=0, y=0)
```

# Chronogram estimation: A Penalized Likelihood Approach



# Chronogram estimation: A Penalized Likelihood Approach



- Let's get names of variables in phylogeny

```
names(rag1_rooted_garli)
```

- Let's get tip labels

```
rag1_rooted_garli$tip.label
```

- **Suggestion:** for comparative analyses is better to drop outgroups, so we limit the effect of long branches on the penalized likelihood function that renders our tree ultrametric

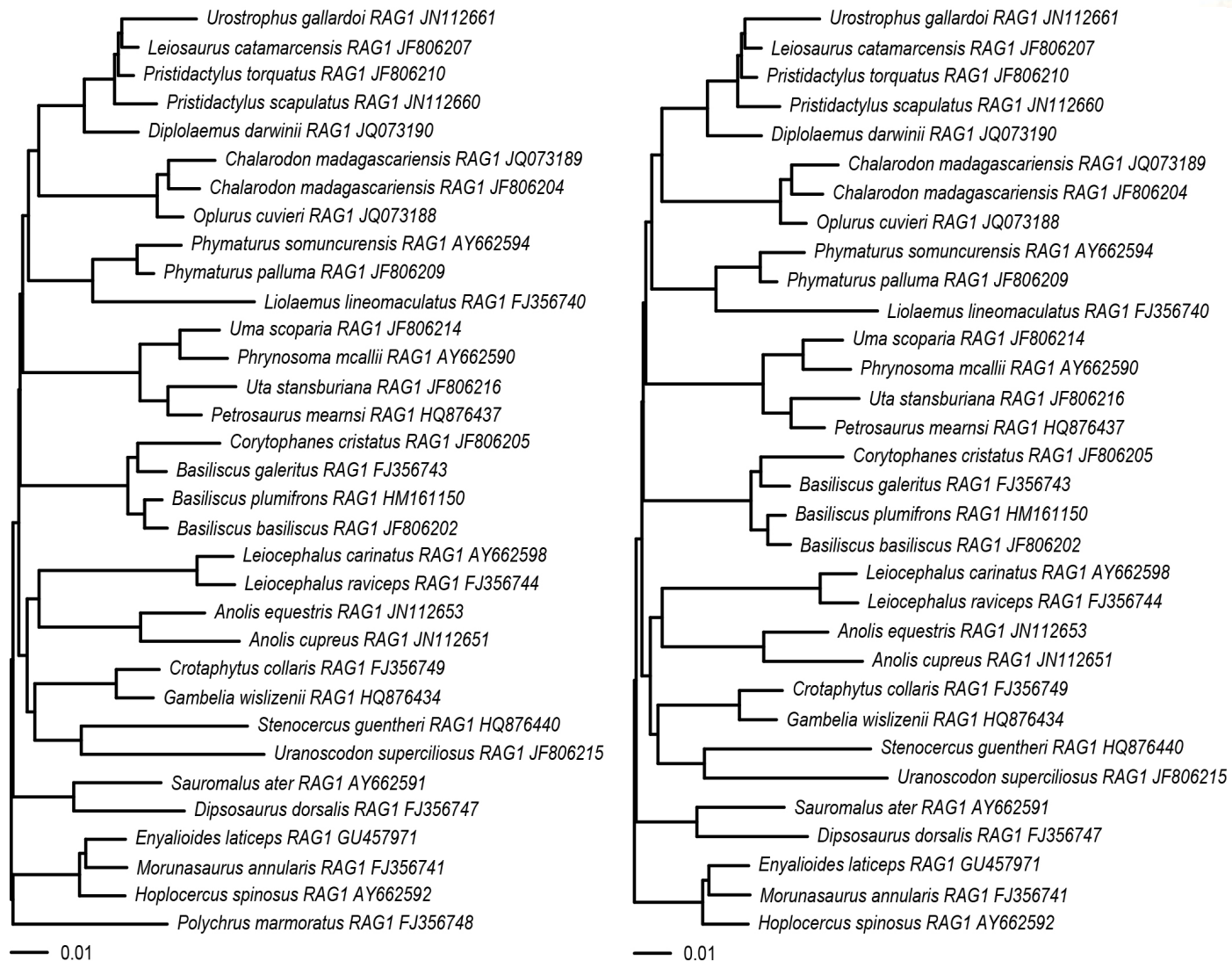
I will drop one tip, but for the starting BEAST tree we will use the **full tree**

```
rag1_rooted_garli_no_out <- drop.tip(rag1_rooted_garli, tip  
="Polychrus_marmoratus_RAG1_FJ356748")
```

- Let's take a look to our tree with and without that tip

```
par(mfrow=c(1,2))  
plot(rag1_rooted_garli)  
add.scale.bar(x=0, y=0)  
plot(rag1_rooted_garli_no_out)  
add.scale.bar(x=0, y=0)
```

# Chronogram estimation: A Penalized Likelihood Approach





## Chronogram estimation: A Penalized Likelihood Approach



- We can force the tree to be ultrametric using the `chronos()` function in 'ape'.
- If we don't have a calibration point, we can set the age of the whole tree to "1". First, we need to define the internal function 'makeChronosCalib(rag1\_rooted\_garli)' of the function 'chronos()'.

For relative age chronogram let's fix root at 1

```
cal_relative <- makeChronosCalib(rag1_rooted_garli, node = "root", age.min = 1, age.max = 1, interactive = FALSE, soft.bounds = FALSE)
```

```
cal_relative
```

For absolute age chronogram, we need to provide a range of possible ages of the reference nodes

**Notice** the change in the parameter to `interactive = TRUE` and `soft.bounds = TRUE`

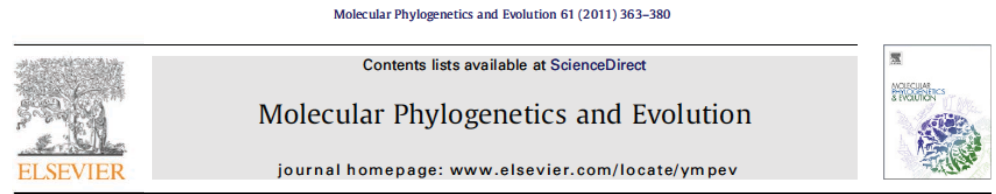
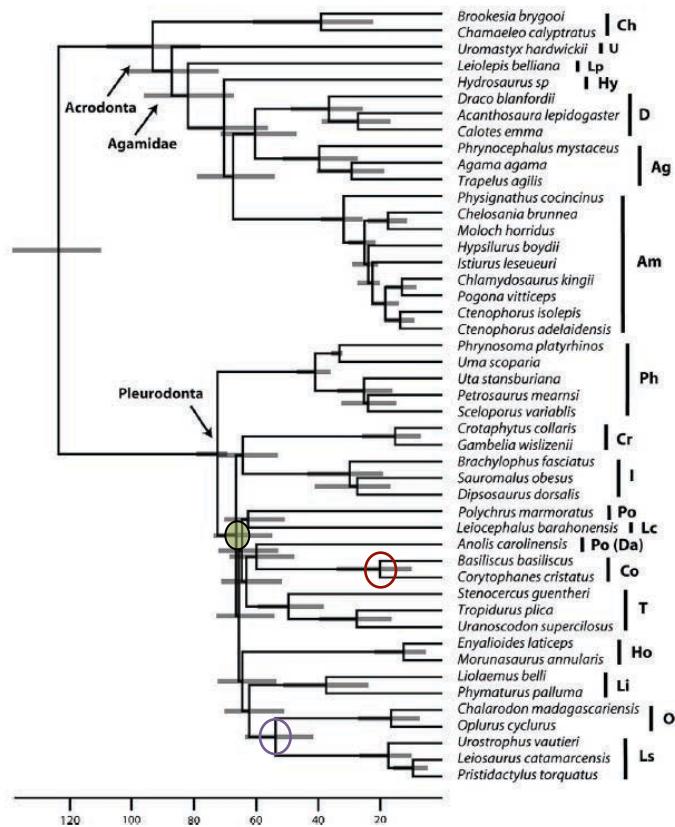
```
cal_absolute <- makeChronosCalib(rag1_rooted_garli, node = "root", age.min = 1, age.max = 1, interactive = TRUE, soft.bounds = TRUE)# in macs use righth-click (control + click) on the plot to exit
```

# Chronogram estimation: A Penalized Likelihood Approach



- For absolute age chronogram, we need to provide a range of possible ages of the reference nodes

```
cal_absolute <- makeChronosCalib(ragl_rooted_garli, node = "root", age.min = 1, age.max = 1, interactive = TRUE, soft.bounds = TRUE)# in macs use right-click (control + click) on the plot to exit
```



Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation

Ted M. Townsend<sup>a,\*</sup>, Daniel G. Mulcahy<sup>b</sup>, Brice P. Noonan<sup>b</sup>, Jack W. Sites Jr.<sup>b</sup>, Caitlin A. Kuczynski<sup>c</sup>, John J. Wiens<sup>c</sup>, Tod W. Reeder<sup>a</sup>

<sup>a</sup> Department of Biology, San Diego State University, San Diego, CA 92182-4614, United States  
<sup>b</sup> Department of Biology, Brigham Young University, Provo, UT 84602, United States  
<sup>c</sup> Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, United States

- (node number: 49): younger age is 15 and older age 35
- (node number: 59): younger age is 40 and older age 62
- root (node number: 34): younger age is 55 and older age 76

Fig. 4. Bayesian chronogram produced by BEAST analyses with 29 genes (based on mean age estimates). Non-iguanian taxa have been removed for clarity. Shaded bars represent 95% highest probability densities. Numbers on scale bar are millions of years before present. Clade name abbreviations as in Fig. 3.

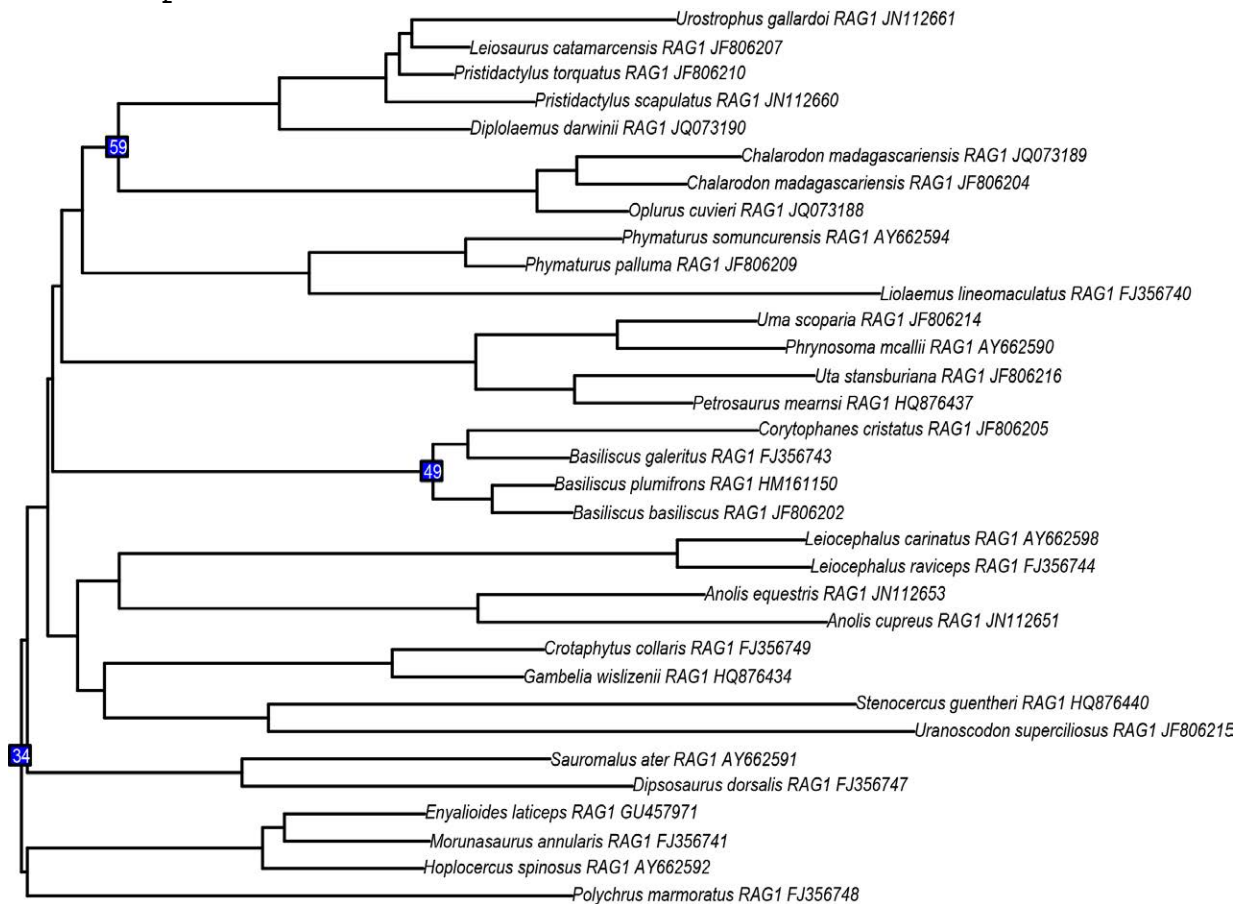
# Chronogram estimation: A Penalized Likelihood Approach



- For absolute age chronogram, we need to provide a range of possible ages of the reference nodes

```
cal_absolute <- makeChronosCalib(rag1_rooted_garli, node = "root", age.min = 1, age.max = 1, interactive = TRUE, soft.bounds = TRUE)# in macs use right-click (control + click) on the plot to exit
```

- (node number: 49): younger age is 15 and older age 35
- (node number: 59): younger age is 40 and older age 62
- root (node number: 34): younger age is 55 and older age 76



## Chronogram estimation: A Penalized Likelihood Approach



- Both `cal_relative` and `cal_absolute` are **data frames**, so you can **modify** these data without the interactive mode

```
cal_relative  
cal_absolute
```

- Let's estimate our chronogram using Penalized Likelihood and Maximum Likelihood. We need to select a method, I prefer a **relaxed clock model** to account for heterogeneity among branches.
- We also need to select a smoothing parameter `lambda` (see Sanderson, 2002)

If `lambda = 0`, then **the parametric component dominates and rates** vary as much as possible among branches, whereas for **increasing values of lambda**, the **variation in the rates are smoother** and tend to a **clock-like model (same rate for all branches)**

- However, I prefer **BEAST** for my chronogram estimation given that it is more parametrized and used the actual data to estimate the chronogram. So, you can pick an arbitrary `lambda` or test of different `lambdas` for this chronogram (This is **just a starting tree** for our BEAST analysis)

# Chronogram estimation: A Penalized Likelihood Approach



- Relative age chronogram with three smoothing parameters

```
rag1_chrono_rel_garli_lambda_0 <- chronos(rag1_rooted_garli, model = "relaxed", lambda =  
0, calibration = cal_relative)  
rag1_chrono_rel_garli_lambda_0_1 <- chronos(rag1_rooted_garli, model = "relaxed", lambda  
= 0.1, calibration = cal_relative)  
rag1_chrono_rel_garli_lambda_1_0 <- chronos(rag1_rooted_garli, model = "relaxed", lambda  
= 1.0, calibration = cal_relative)
```

```
rag1_chrono_abs_garli_lambda_0 <- chronos(rag1_rooted_garli, model = "relaxed", lambda =  
0, calibration = cal_absolute)  
rag1_chrono_abs_garli_lambda_0_1 <- chronos(rag1_rooted_garli, model = "relaxed", lambda  
= 0.1, calibration = cal_absolute)  
rag1_chrono_abs_garli_lambda_1_0 <- chronos(rag1_rooted_garli, model = "relaxed", lambda  
= 1.0, calibration = cal_absolute)
```

- Check the 'Penalised log-lik' values, the you want the **largest likelihood**; i.e., largest likelihood or **smallest negative log-likelihood**. Negative numbers are OK. A log-likelihood of -2 is better than -4.

The best in this case corresponds to the lambda = 0 (i.e., Penalised log-lik = -3.885571)

## Chronogram estimation: A Penalized Likelihood Approach



- After the calculations are done, we can write the trees in newick format for future use.

```
write.tree (ragl_chrono_rel_garli_lambda_0, file
="ragl_chrono_rel_garli_lambda_0.newick")
write.tree (ragl_chrono_rel_garli_lambda_0_1, file
="ragl_chrono_rel_garli_lambda_0_1.newick")
write.tree (ragl_chrono_rel_garli_lambda_1_0, file
="ragl_chrono_rel_garli_lambda_1_0.newick")
```

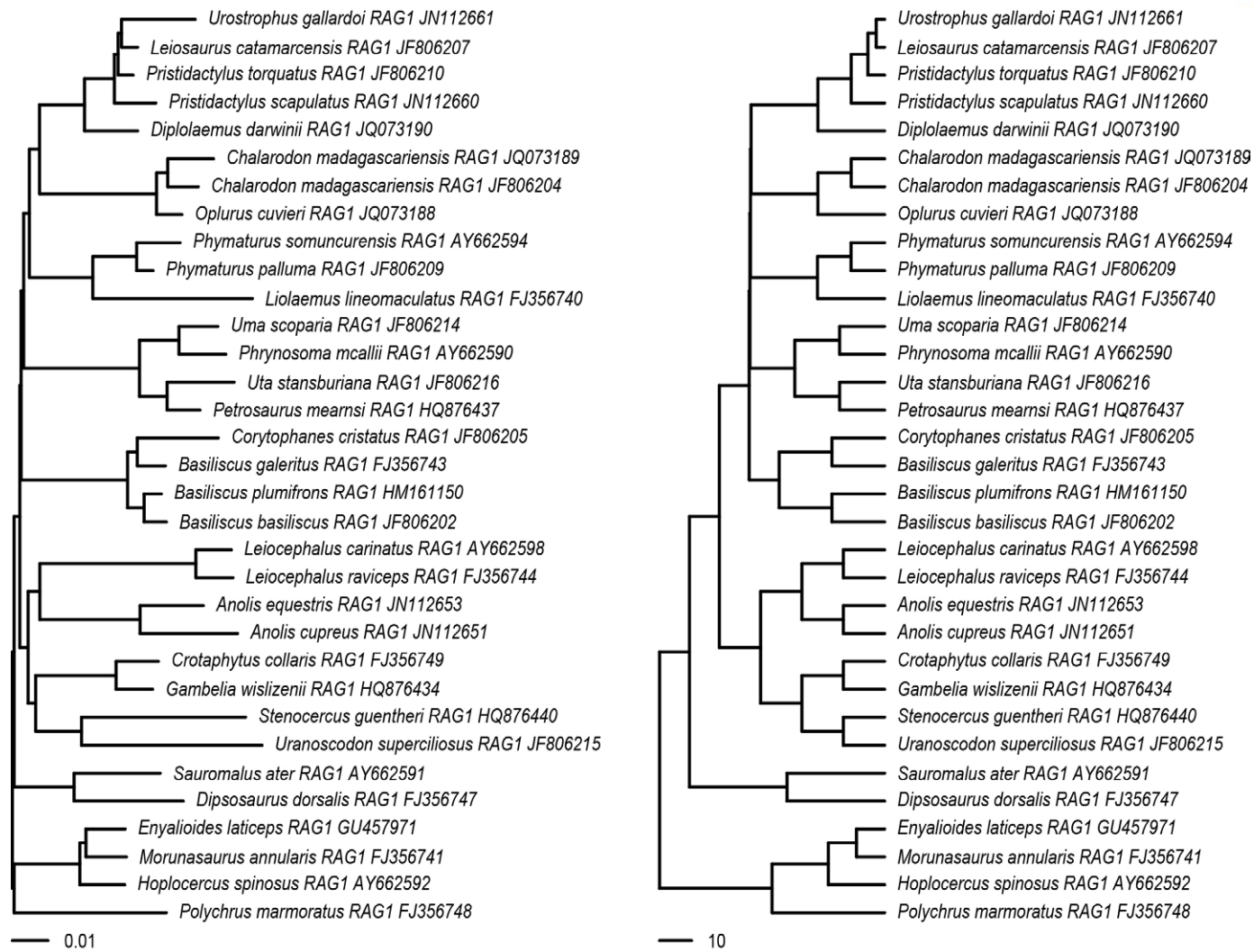
```
write.tree (ragl_chrono_abs_garli_lambda_0, file
="ragl_chrono_abs_garli_lambda_0.newick")
write.tree (ragl_chrono_abs_garli_lambda_0_1, file
="ragl_chrono_abs_garli_lambda_0_1.newick")
write.tree (ragl_chrono_abs_garli_lambda_1_0, file
="ragl_chrono_abs_garli_lambda_1_0.newick")
```

- Let's check that our trees are ultrametric

```
is.ultrametric(ragl_rooted_garli)
is.ultrametric(ragl_chrono_rel_garli_lambda_0)
is.ultrametric(ragl_chrono_rel_garli_lambda_0_1)
is.ultrametric(ragl_chrono_rel_garli_lambda_1_0)
is.ultrametric(ragl_chrono_abs_garli_lambda_0)
is.ultrametric(ragl_chrono_abs_garli_lambda_0_1)
is.ultrametric(ragl_chrono_abs_garli_lambda_1_0)
```



# Chronogram estimation: A Penalized Likelihood Approach





## Chronogram estimation: BEAST



- One of the most widely used software to estimate chronograms and most recently species trees

**Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 Molecular Biology And Evolution 29: 1969-1973**

<http://beast.bio.ed.ac.uk/>

From the authors:

BEAST is a cross-platform program for **Bayesian analysis of molecular sequences using MCMC**. It is entirely orientated **towards rooted, time-measured phylogenies** inferred using strict or relaxed molecular clock models. It can be used as a method of reconstructing phylogenies but is also a framework for testing evolutionary hypotheses without conditioning on a single tree topology. BEAST uses MCMC to average over tree space, so that each tree is weighted proportional to its posterior probability. We include a simple to use user-interface program for setting up (**BEAUti**) standard analyses and a suit of programs for analysing the results.

## Chronogram estimation: BEAST



- This software has detailed manuals and tutorials:

<http://beast.bio.ed.ac.uk/tutorials>

- After installation, you will also require the extensive set of auxiliary programs. Most of them are included in the BEAST package, but others require independent installation.

**Bayesian Evolutionary Analysis Utility (BEAUti):** GUI application for the preparation of the BEAST XML files.

**LogCombiner:** Program that allows you to combine log and tree files from multiple independent runs of BEAST.

**TreeAnnotator:** Program that assists you in summarizing the information from a sample of trees produced by BEAST.

**Tracer:** is graphical tool for visualization and diagnostics of MCMC output. It can read output files from MrBayes and BEAST (Not in BEAST package):

<http://tree.bio.ed.ac.uk/software/tracer/>

**FigTree:** You know this one.

## Chronogram estimation: BEAST



- This is very important, you need to install BEAGLE

BEAGLE is a **high-performance library** that can perform the core calculations at the heart of most Bayesian and Maximum Likelihood phylogenetics package. It can make use of highly-parallel processors such as those in 3D graphics boards found in many PCs.

<http://beast.bio.ed.ac.uk/BEAGLE>

MACs:

<https://code.google.com/p/beagle-lib/wiki/MacInstallInstructions>

Windows:

<https://code.google.com/p/beagle-lib/wiki/WindowsInstallInstructions>

## Preparing run for BEAST using a starting tree



- You need to create a folder with a nexus file with the sequences and the chronogram that we estimated using 'ape'.

Folder: [RAG1\\_BEST](#)

# Preparing run for BEAST using a starting tree



- Edit and add the chronogram at the end of the nexus file (I have added the edited file to the courser website): [RAG1\\_nucleotide\\_chrono.nex](#)

```
#NEXUS-
-
BEGIN DATA;
-
Δ DIMENSIONS : NTAX=33 NCHAR=2925;
Δ FORMAT DATATYPE = DNA GAP = - MISSING = ?;
Δ MATRIX-
Δ Phymaturus_palluma_RAG1_JF806209 .....
Δ Chalarodon_madagascariensis_RAG1_JF806204 .....
Δ Gambelia_wislizenii_RAG1_HQ876434 .....
Δ Basiliscus_plumifrons_RAG1_HM161150 .....
Δ Phymaturus_somuncurensis_RAG1_AY662594 .....
Δ Corytophanes_cristatus_RAG1_JF806205 .....
Δ Uta_stansburiana_RAG1_JF806216 .....
Δ Hoplocercus_spinosus_RAG1_AY662592 .....
Δ Dipsosaurus_dorsalis_RAG1_FJ356747 .....
Δ Anolis_cupreus_RAG1_JN112651 .....
Δ Stenocercus_quentheri_RAG1_HQ876440 .....
Δ Pristidactylus_torquatus_RAG1_JF806210 .....
Δ Enyalioides_laticeps_RAG1_GU457971 .....
Δ Urostrophus_gallardoi_RAG1_JN112661 .....
Δ Liolaemus_lineomaculatus_RAG1_FJ356740 .....
Δ Polychrus_marmoratus_RAG1_FJ356748 .....
Δ Crotaphytus_collaris_RAG1_FJ356749 .....
Δ Sauromalus_ater_RAG1_AY662591 .....
Δ Oplurus_cuvieri_RAG1_JQ073188 .....
Δ Anolis_equestris_RAG1_JN112653 .....
Δ Phrynosoma_mcallii_RAG1_AY662590 .....
Δ Chalarodon_madagascariensis_RAG1_JQ073189 .....
Δ Uranoscodon_supercilius_RAG1_JF806215 .....
Δ Leiosaurus_catamarcensis_RAG1_JF806207 .....
Δ Uma_scoparia_RAG1_JF806214 .....
Δ Basiliscus_basiliscus_RAG1_JF806202 .....
Δ Leiocephalus_carinatus_RAG1_AY662598 .....
Δ Leiocephalus_raviceps_RAG1_FJ356744 .....
Δ Basiliscus_galeritus_RAG1_FJ356743 .....
Δ Pristidactylus_scapulatus_RAG1_JN112660 .....
Δ Diplolaemus_darwini_RAG1_JQ073190 .....
Δ Morunasaurus_annularis_RAG1_FJ356741 .....
Δ Petrosaurus_mearnsi_RAG1_HQ876437 .....
```

```
-
;
-
END;
-
BEGIN SETS;
-
Δ CHARSET position_1 : = 2-2924\3;
Δ CHARSET position_2 : = 3-2925\3-2925;
Δ CHARSET position_3 : = 1-2923\3;
END;
-
begin trees;
-
Δ tree rag1_chrono_abs_garli_lambda_1_0 = ((Polychrus_marmoratus_RAG1_FJ356748:30.343504,(Hoplocercus_spinosus_RAG1_AY662592:15.38144795,(Morunasaurus_
end;
```

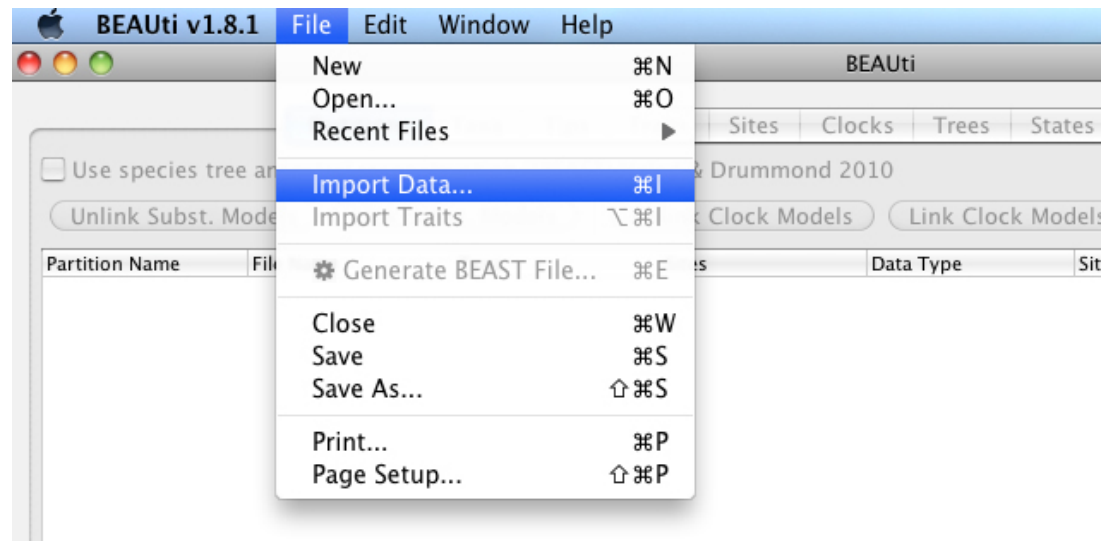
← Partitions by codon (if you have multiple genes)

← Our chronogram

## Preparing run for BEAST using a starting tree



- Open BEAUti v1.8.1 and import the nexus file with chronogram



# Preparing run for BEAST using a starting tree

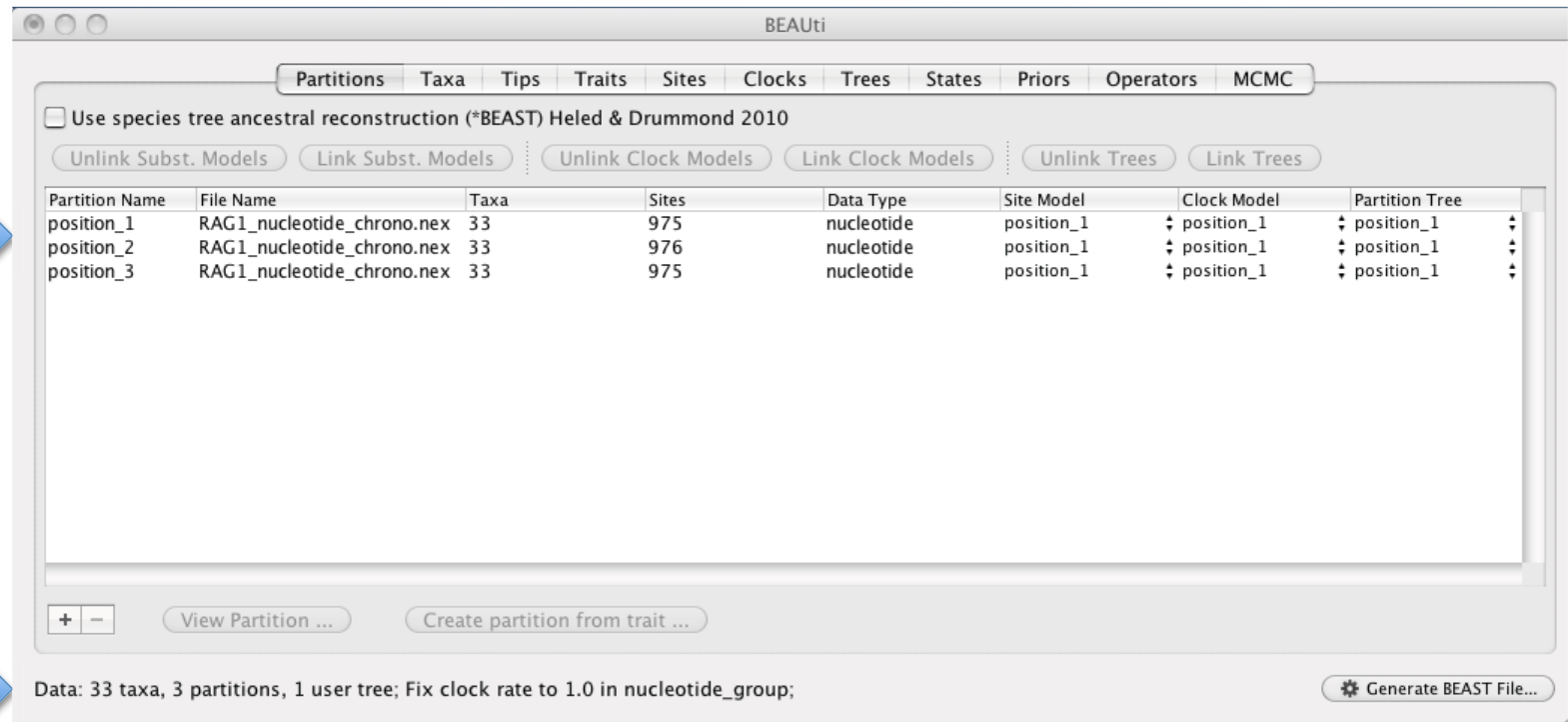


- Open BEAUti v1.8.1 and import the nexus file with chronogram

Our data file



Our starting chronogram and partitions (e.g., genes, codon positions)

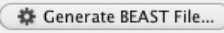


The screenshot shows the BEAUti v1.8.1 interface with the following components:

- Menu bar: Partitions, Taxa, Tips, Traits, Sites, Clocks, Trees, States, Priors, Operators, MCMC
- Checkbox:  Use species tree ancestral reconstruction (\*BEAST) Heled & Drummond 2010
- Buttons: Unlink Subst. Models, Link Subst. Models, Unlink Clock Models, Link Clock Models, Unlink Trees, Link Trees
- Table of partitions:

| Partition Name | File Name                  | Taxa | Sites | Data Type  | Site Model | Clock Model  | Partition Tree |
|----------------|----------------------------|------|-------|------------|------------|--------------|----------------|
| position_1     | RAG1_nucleotide_chrono.nex | 33   | 975   | nucleotide | position_1 | ↕ position_1 | ↕ position_1   |
| position_2     | RAG1_nucleotide_chrono.nex | 33   | 976   | nucleotide | position_1 | ↕ position_1 | ↕ position_1   |
| position_3     | RAG1_nucleotide_chrono.nex | 33   | 975   | nucleotide | position_1 | ↕ position_1 | ↕ position_1   |

Buttons at the bottom: +, -, View Partition ..., Create partition from trait ...

Status bar: Data: 33 taxa, 3 partitions, 1 user tree; Fix clock rate to 1.0 in nucleotide\_group; 



## Preparing run for BEAST using a starting tree



- We are going to change the partitions defaults for codon positions in a single gene, but you can do the same procedure for multiple genes. You need to do the following to allow to run a partitioned dataset:

1) On the **partitions tab** select all partitions or genes

2) We like to analyze each partition/gene under separate substitutions models, while assuming the clock and tree are linked

Highlight all three partitions/genes

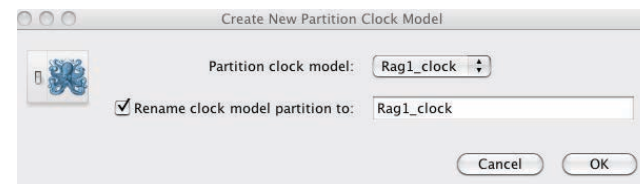
Unlink Subst. Models

3) Rename Partition Clock Model. Changes are done by highlighting all partitions

Highlight all three partitions/genes

Link Clock Models

Rename this to Rag1\_clock

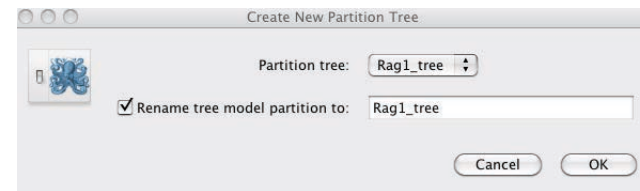


4) Rename Partition Tree. Changes are done by highlighting all partitions

Highlight all three genes

Link Trees

Rename this to Rag1\_tree



# Preparing run for BEAST using a starting tree



- After all of these changes you should have something like this:

BEAUti

Partitions Taxa Tips Traits Sites Clocks Trees States Priors Operators MCMC

Use species tree ancestral reconstruction (\*BEAST) Heled & Drummond 2010

Unlink Subst. Models Link Subst. Models Unlink Clock Models Link Clock Models Unlink Trees Link Trees

| Partition Name | File Name                  | Taxa | Sites | Data Type  | Site Model | Clock Model  | Partition Tree |
|----------------|----------------------------|------|-------|------------|------------|--------------|----------------|
| position_1     | RAG1_nucleotide_chrono.nex | 33   | 975   | nucleotide | position_1 | ⚡ Rag1_clock | ⚡ Rag1_tree    |
| position_2     | RAG1_nucleotide_chrono.nex | 33   | 976   | nucleotide | position_2 | ⚡ Rag1_clock | ⚡ Rag1_tree    |
| position_3     | RAG1_nucleotide_chrono.nex | 33   | 975   | nucleotide | position_3 | ⚡ Rag1_clock | ⚡ Rag1_tree    |

+ - View Partition ... Create partition from trait ...

Data: 33 taxa, 3 partitions, 1 user tree; Fix clock rate to 1.0 in nucleotide\_group; ⚙ Generate BEAST File...

## Preparing run for BEAST using a starting tree



- We will create the calibration nodes and identifying internal nodes in the **Taxa** tab

**Notice:** If not enforced monophyletic, then it is possible to define a node using a pair of descendant taxa and use the name of the node corresponding to their MRCA (most recent common ancestor) to define calibration times. This procedure will reduced the time editing the xml BEAST file

The screenshot shows the BEAUti software interface with the 'Taxa' tab selected. The interface includes a menu bar with options like Partitions, Taxa, Tips, Traits, Sites, Clocks, Trees, States, Priors, Operators, and MCMC. Below the menu bar, there are several panels. On the left, a table lists 'Taxon Set' with columns for 'Mono?', 'Stem?', 'Tree', and 'Age'. The main area is divided into 'Excluded Taxa' and 'Included Taxa' sections. The 'Excluded Taxa' list contains 33 entries, including various species names like Anolis\_cupreus, Basiliscus\_basiliscus, and Dinosaurius\_dorsalis. The 'Included Taxa' section is currently empty. At the bottom left, a green callout box with a blue arrow pointing to a '+' button contains the text: 'Click of + to start defining groups that will used to do calibrations'. The status bar at the bottom indicates 'Data: 33 taxa, 3 partitions, 1 user tree; Fix clock rate to 1.0 in nucleotide\_group;' and a 'Generate BEAST File...' button is visible on the right.

# Preparing run for BEAST using a starting tree

- We are going to use the same calibration nodes that we did in the Penalized Likelihood Approach

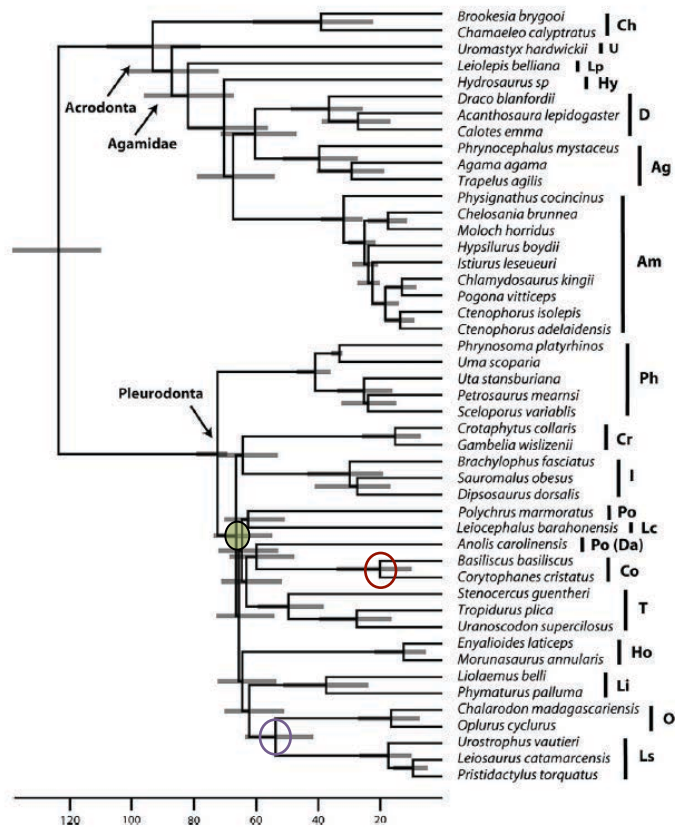
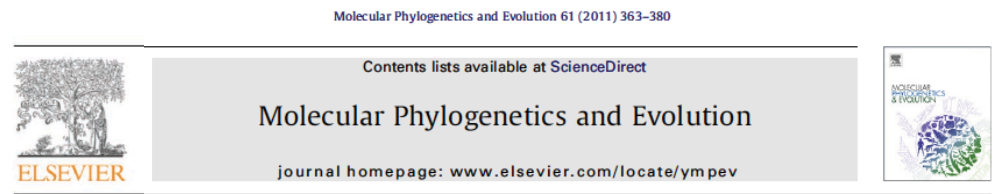


Fig. 4. Bayesian chronogram produced by BEAST analyses with 29 genes (based on mean age estimates). Non-iguanian taxa have been removed for clarity. Shaded bars represent 95% highest probability densities. Numbers on scale bar are millions of years before present. Clade name abbreviations as in Fig. 3.



## Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation

Ted M. Townsend<sup>a,\*</sup>, Daniel G. Mulcahy<sup>b</sup>, Brice P. Noonan<sup>b</sup>, Jack W. Sites Jr.<sup>b</sup>, Caitlin A. Kuczynski<sup>c</sup>, John J. Wiens<sup>c</sup>, Tod W. Reeder<sup>a</sup>

<sup>a</sup> Department of Biology, San Diego State University, San Diego, CA 92182-4614, United States

<sup>b</sup> Department of Biology, Brigham Young University, Provo, UT 84602, United States

<sup>c</sup> Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, United States

- (node number: 49): younger age is 15 and older age 35
- (node number: 59): younger age is 40 and older age 62
- root (node number: 34): younger age is 55 and older age 76

## Preparing run for BEAST using a starting tree



- We are going to use the same calibration nodes that we did in the Penalized Likelihood Approach

The screenshot shows the BEAUti software interface. The 'Taxa' tab is selected, and the 'Taxon set: Corytophanidae' is chosen. The interface is divided into two main sections: 'Excluded Taxa' and 'Included Taxa'. The 'Excluded Taxa' list contains 29 taxa, including *Anolis\_cupreus*, *Anolis\_equestris*, *Chalarodon\_madagascariensis*, *Crotaphytus\_collaris*, *Diplolaemus\_darwini*, *Dipsosaurus\_dorsalis*, *Enyalioides\_laticeps*, *Gambelia\_wislizenii*, *Hoplocercus\_spinosus*, and *Leiocephalus\_carinatus*. The 'Included Taxa' list contains 4 taxa: *Basiliscus\_basiliscus*, *Basiliscus\_galeritus*, *Basiliscus\_plumifrons*, and *Corytophanes\_cristatus*. At the bottom, a status bar indicates 'Data: 33 taxa, 3 partitions, 1 user tree; Fix clock rate to 1.0 in nucleotide\_group;' and a 'Generate BEAST File...' button is visible.

○ (node number: 49): younger age is 15 and older age 35



## Preparing run for BEAST using a starting tree



- We are going to use the same calibration nodes that we did in the Penalized Likelihood Approach

The screenshot shows the BEAUti software interface. The 'Taxa' tab is selected, displaying a list of taxa sets on the left and two lists of taxa for the 'Chalo\_Uros' taxon set in the center. The 'Excluded Taxa' list contains 31 taxa, and the 'Included Taxa' list contains 2 taxa. The status bar at the bottom indicates 'Data: 33 taxa, 3 partitions, 1 user tree; Fix clock rate to 1.0 in nucleotide\_group;'. A 'Generate BEAST File...' button is visible in the bottom right corner.

| Taxon Set      | Mon...                   | Stem?                    | Tree   | Age |
|----------------|--------------------------|--------------------------|--------|-----|
| Corytophanidae | <input type="checkbox"/> | <input type="checkbox"/> | Rag... | ↓   |
| Chalo_Uros     | <input type="checkbox"/> | <input type="checkbox"/> | Rag... | ↓   |

Taxon set: Chalo\_Uros

Excluded Taxa

- Prinyosoma\_mcaimi\_RAG1\_AY662590
- Phymaturus\_palluma\_RAG1\_JF806209
- Phymaturus\_somuncurensis\_RAG1\_AY662594
- Polychrus\_marmoratus\_RAG1\_FJ356748
- Pristidactylus\_scapulatus\_RAG1\_JN112660
- Pristidactylus\_torquatus\_RAG1\_JF806210
- Sauromalus\_ater\_RAG1\_AY662591
- Stenocercus\_guentheri\_RAG1\_HQ876440
- Uma\_scoparia\_RAG1\_JF806214
- Uranoscodon\_superciliosus\_RAG1\_JF806215
- Uta\_stansburiana\_RAG1\_JF806216

31 taxa excluded

Select:

Included Taxa

- Chalarodon\_madagascariensis\_RAG1\_JF806...
- Urostrophus\_gallardoi\_RAG1\_JN112661

2 taxa included

Select:

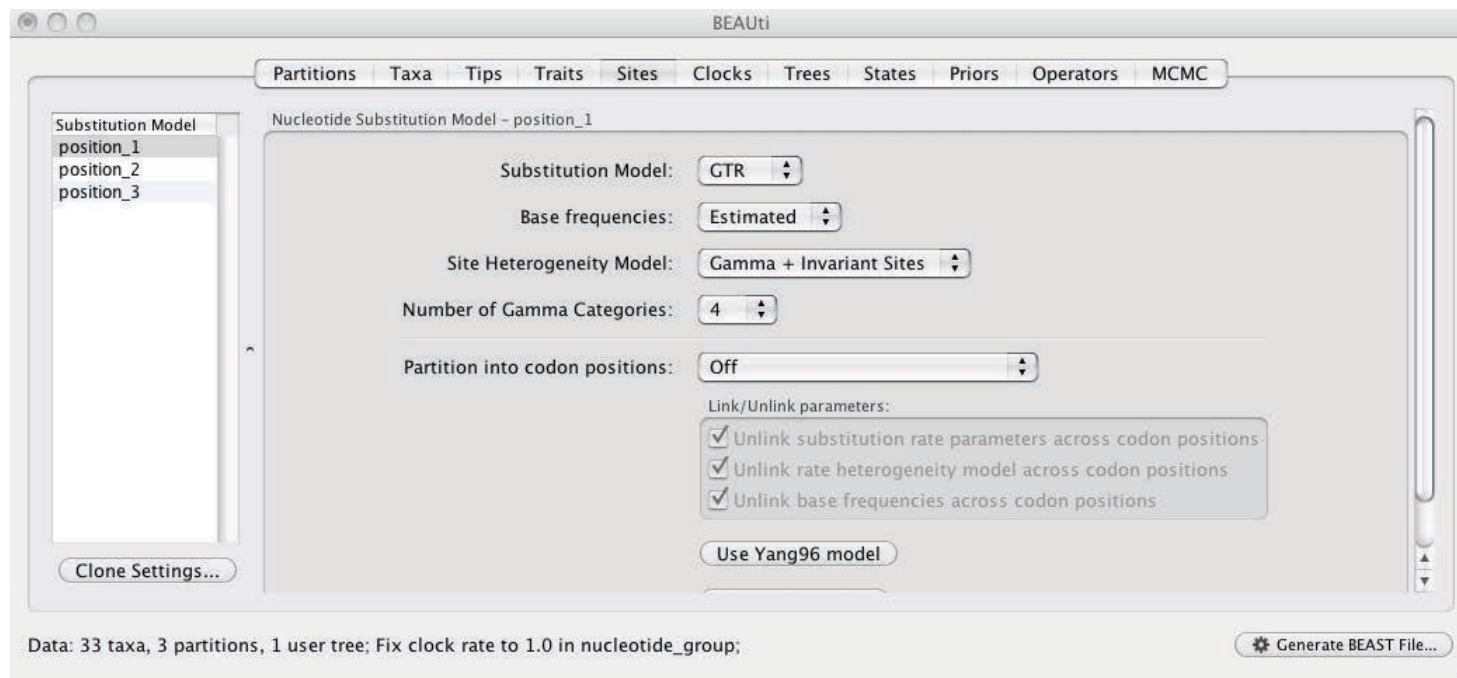
Data: 33 taxa, 3 partitions, 1 user tree; Fix clock rate to 1.0 in nucleotide\_group;

○ (node number: 59): younger age is 40 and older age 62

## Preparing run for BEAST using a starting tree



- Not necessary to worry about Tips (do not click)
- Not necessary to worry about Traits tab (do not click)
- On the **Sites tab** will select the substitution models for each partition/gene. We did this part for the entire RAG1 gene in jModelTest, but not for each of its codon positions. For this exercise, we will assume that each position has a GTR+G+I model. **Notice** the limited number of models





## Preparing run for BEAST using a starting tree



- On the **Clocks** tab will estimate the clock (click on **Estimate**)  
The model will be 'Lognormal relaxed clock (Uncorrelated)' which is fit for more distantly related taxa and low taxon sampling.

The screenshot shows the BEAUti software interface with the 'Clocks' tab selected. The 'Clock Model' section contains a table with the following data:

| Name       | Model        | Estimate                            | Rate | Group            |
|------------|--------------|-------------------------------------|------|------------------|
| Rag1_clock | Strict clock | <input checked="" type="checkbox"/> | 1.0  | nucleotide_group |

A dropdown menu is open for the 'Model' column, showing the following options:

- Lognormal relaxed clock (Uncorrelated)
- Exponential relaxed clock (Uncorrelated)
- Random local clock
- Fixed local clock

At the bottom of the window, the status bar reads: 'Data: 33 taxa, 3 partitions, 1 user tree; Estimate clock rate in nucleotide\_group;'. A 'Generate BEAST File...' button is visible in the bottom right corner.

## Preparing run for BEAST using a starting tree



- On the **Trees tab** we are going to select a tree prior and starting trees

**Tree Prior** -> Speciation: Birth-Death Incomplete Sampling (also Yule process is useful)  
Random starting tree if you do not have an starting tree

The screenshot shows the BEAUti software interface with the 'Trees' tab selected. The 'Link tree prior for all trees' checkbox is checked. In the 'Trees' list on the left, 'Rag1\_tree' is selected. The 'Tree prior shared by all tree models' section shows 'Tree Prior' set to 'Speciation: Birth-Death Incomplete Sampling' with a citation: 'Stadler, T; On incomplete sampling under birth-death models and connections to the sampling-based coalescent; JOURNAL OF THEORETICAL BIOLOGY (2009) 261:58-66'. The 'Tree Model - Rag1\_tree' section shows 'The estimated initial root height' set to 0.038. Under 'Starting tree' options, 'User-specified starting tree' is selected, and 'Select user-specified tree' is set to 'rag1\_chrono\_abs\_garli\_lambda\_1\_0'. At the bottom, the status bar indicates 'Data: 33 taxa, 3 partitions, 1 user tree; Estimate clock rate in nucleotide\_group;' and a 'Generate BEAST File...' button is visible.

# Preparing run for BEAST using a starting tree



- Not necessary to worry about States (do not click)
- On the **Priors tab** will input some a priori information to allow us find our time calibrated chronogram

Node upper and lower ages



parameters for the clock model



BEAUti

Partitions Taxa Tips Traits Sites Clocks Trees States Priors Operators MCMC

Priors for model parameters and statistics:

| Parameter                    | Prior   | Bound    | Description   |
|------------------------------|---|----------|---|
| tmrca(Corytophanidae)        | * Using Tree Prior                            | n/a      | tmrca statistic for taxon set untitled1 on tree Rag1_tree                 |
| tmrca(Chalo_Uros)            | * Using Tree Prior                            | n/a      | tmrca statistic for taxon set untitled2 on tree Rag1_tree                 |
| position_1.ac                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR A-C substitution parameter  |
| position_1.ag                | * Gamma [0.05, 20], initial=1                 | [0, ∞]   | GTR A-G substitution parameter  |
| position_1.at                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR A-T substitution parameter  |
| position_1.cg                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR C-G substitution parameter  |
| position_1.gt                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR G-T substitution parameter  |
| position_1.frequencies       | * Uniform [0, 1], initial=0.25                | [0, 1]   | base frequencies  |
| position_1.alpha             | * Exponential [0.5], initial=0.5              | [0, ∞]   | gamma shape parameter   |
| position_1.plnv              | * Uniform [0, 1], initial=0.5                 | [0, 1]   | proportion of invariant sites parameter                                   |
| position_2.ac                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR A-C substitution parameter  |
| position_2.ag                | * Gamma [0.05, 20], initial=1                 | [0, ∞]   | GTR A-G substitution parameter  |
| position_2.at                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR A-T substitution parameter  |
| position_2.cg                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR C-G substitution parameter  |
| position_2.gt                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR G-T substitution parameter  |
| position_2.frequencies       | * Uniform [0, 1], initial=0.25                | [0, 1]   | base frequencies  |
| position_2.alpha             | * Exponential [0.5], initial=0.5              | [0, ∞]   | gamma shape parameter   |
| position_2.plnv              | * Uniform [0, 1], initial=0.5                 | [0, 1]   | proportion of invariant sites parameter                                   |
| position_3.ac                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR A-C substitution parameter  |
| position_3.ag                | * Gamma [0.05, 20], initial=1                 | [0, ∞]   | GTR A-G substitution parameter  |
| position_3.at                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR A-T substitution parameter  |
| position_3.cg                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR C-G substitution parameter  |
| position_3.gt                | * Gamma [0.05, 10], initial=1                 | [0, ∞]   | GTR G-T substitution parameter  |
| position_3.frequencies       | * Uniform [0, 1], initial=0.25                | [0, 1]   | base frequencies  |
| position_3.alpha             | * Exponential [0.5], initial=0.5              | [0, ∞]   | gamma shape parameter   |
| position_3.plnv              | * Uniform [0, 1], initial=0.5                 | [0, 1]   | proportion of invariant sites parameter                                   |
| uclid.stdev                  | * Exponential [0.333333], initial=0.333333    | [0, ∞]   | uncorrelated lognormal relaxed clock stdev                                |
| uclid.mean                   | ? Not yet specified                           | [0, ∞]   | uncorrelated lognormal relaxed clock mean                                 |
| treeModel.rootHeight         | * Using Tree Prior in [0, ∞]                  | [0, ∞]   | root height of the tree   |
| birthDeath.meanGrowthRate    | * Uniform [0, 1E5], initial=550               | [0, 1E5] | Birth-Death speciation process rate                                       |
| birthDeath.relativeDeathRate | * Uniform [0, 1], initial=0.5                 | [0, 1]   | Birth-Death speciation process relative death rate                        |
| birthDeath.sampleProbability | * Beta [1, 1], initial=0.01                   | [0, 1]   | Birth-Death the proportion of taxa sampled from birth-death tree          |
| meanRate                     | * Indirectly Specified Through Other Param... | n/a      | The mean rate of evolution over the whole tree                            |
| covariance                   | * Indirectly Specified Through Other Param... | n/a      | The covariance in rates of evolution on each lineage with their ancest... |
| coefficientOfVariation       | * Indirectly Specified Through Other Param... | n/a      | The variation in rate of evolution over the whole tree                    |

Link parameters together | Link parameters into a hierarchical model | Unlink parameters

\* Marked parameters currently have a default prior distribution. You should check that these are appropriate.

Data: 33 taxa, 3 partitions, 1 user tree; Estimate clock rate in nucleotide\_group; Generate BEAST File...

# Preparing run for BEAST using a starting tree

- Nodes with an upper and lower within a uniform distribution



BEAUi

Partitions Taxa Tips Traits Sites Clocks Trees States Priors Op

Priors for model parameters and statistics:

| Parameter                    | Prior   | Bound  | Description                   |
|------------------------------|---|--------|-------------------------------|
| tmrca(Corytophanidae)        | * Using Tree Prior                            | n/a    | tmrca statistic for taxon set |
| tmrca(Chalo_Uros)            | * Using Tree Prior                            | n/a    | tmrca statistic for taxon set |
| position_1.ac                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR A-C substitution par      |
| position_1.ag                | * Gamma [0.05, 20], initial=1                 | [0, ∞] | GTR A-G substitution par      |
| position_1.at                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR A-T substitution par      |
| position_1.cg                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR C-G substitution par      |
| position_1.gt                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR G-T substitution par      |
| position_1.frequencies       | * Uniform [0, 1], initial=0.25                | [0, 1] | base frequencies              |
| position_1.alpha             | * Exponential [0.5], initial=0.5              | [0, ∞] | gamma shape parameter         |
| position_1.plnv              | * Uniform [0, 1], initial=0.5                 | [0, 1] | proportion of invariant sites |
| position_2.ac                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR A-C substitution par      |
| position_2.ag                | * Gamma [0.05, 20], initial=1                 | [0, ∞] | GTR A-G substitution par      |
| position_2.at                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR A-T substitution par      |
| position_2.cg                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR C-G substitution par      |
| position_2.gt                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR G-T substitution par      |
| position_2.frequencies       | * Uniform [0, 1], initial=0.25                | [0, 1] | base frequencies              |
| position_2.alpha             | * Exponential [0.5], initial=0.5              | [0, ∞] | gamma shape parameter         |
| position_2.plnv              | * Uniform [0, 1], initial=0.5                 | [0, 1] | proportion of invariant sites |
| position_3.ac                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR A-C substitution par      |
| position_3.ag                | * Gamma [0.05, 20], initial=1                 | [0, ∞] | GTR A-G substitution par      |
| position_3.at                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR A-T substitution par      |
| position_3.cg                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR C-G substitution par      |
| position_3.gt                | * Gamma [0.05, 10], initial=1                 | [0, ∞] | GTR G-T substitution par      |
| position_3.frequencies       | * Uniform [0, 1], initial=0.25                | [0, 1] | base frequencies              |
| position_3.alpha             | * Exponential [0.5], initial=0.5              | [0, ∞] | gamma shape parameter         |
| position_3.plnv              | * Uniform [0, 1], initial=0.5                 | [0, 1] | proportion of invariant sites |
| ucl.d.stdev                  | * Exponential [0.5], initial=0.5              | [0, ∞] | gamma shape parameter         |
| ucl.d.mean                   | * Exponential [0.5], initial=0.5              | [0, ∞] | gamma shape parameter         |
| treeModel.rootHeight         | * Using Tree Prior                            | n/a    | treeModel.rootHeight          |
| birthDeath.meanGrowthRate    | * Uniform [0, 1], initial=0.5                 | [0, 1] | birthDeath.meanGrowthRate     |
| birthDeath.relativeDeathRate | * Uniform [0, 1], initial=0.5                 | [0, 1] | birthDeath.relativeDeathRate  |
| birthDeath.sampleProbability | * Uniform [0, 1], initial=0.5                 | [0, 1] | birthDeath.sampleProbability  |
| meanRate                     | * Beta [1, 1], initial=1                      | [0, ∞] | meanRate                      |
| evolutionaryRate             | * Indirectly Specified Through Other Param... | n/a    | The mean rate of evolution    |
| evolutionaryRate             | * Indirectly Specified Through Other Param... | n/a    | The covariance in rates of    |

Prior for Parameter tmrca(Corytophanidae)

Select prior distribution for tmrca(Corytophanidae)

Prior Distribution: **Uniform**

Initial value: 25

Upper: 35.0

Lower: 5.0

Cancel OK

Prior for Parameter tmrca(Corytophanidae)

Select prior distribution for tmrca(Corytophanidae)

Prior Distribution: **Uniform**

Initial value: 25

Upper: 35.0

Lower: 5.0

Cancel OK

Prior for Parameter tmrca(Chalo\_Uros)

Select prior distribution for tmrca(Chalo\_Uros)

Prior Distribution: **Uniform**

Initial value: 51

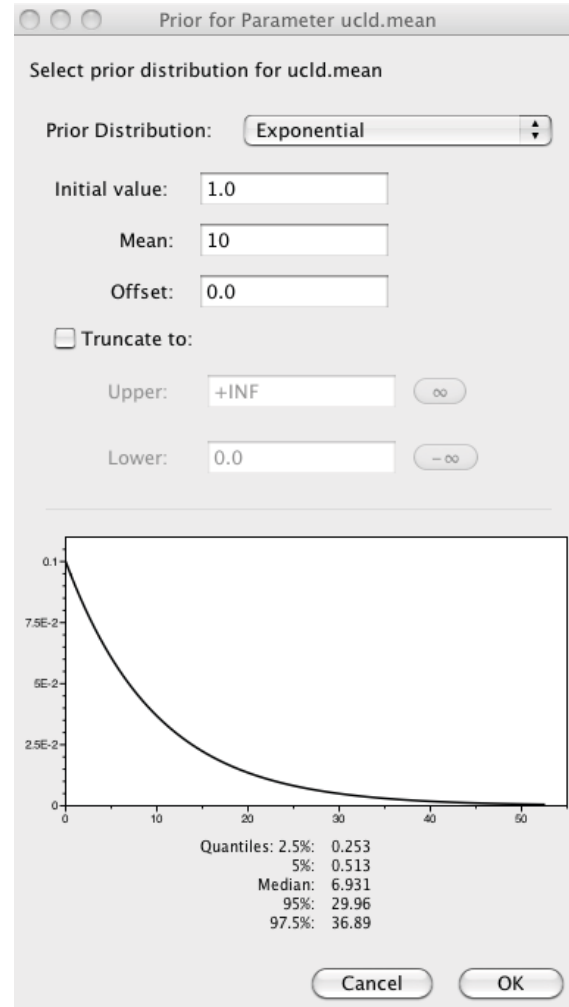
Upper: 62

Lower: 40

Cancel OK

## Preparing run for BEAST using a starting tree

- Parameter for the clock model



I trust the theoreticians and this values for this prior seems to work well

## Preparing run for BEAST using a starting tree



- On the tab Operators left them as they are, the software is going 'Auto optimize' while searching for the optimal chronogram.
- On the **MCMC tab** will like to have a really long chain > 90 million (depends on the computational disponibility and time).
- I usually run at least 4 independent starts (based on random numbers)

A screenshot of the BEAUti software interface, specifically the MCMC tab. The window title is "BEAUti". The tabs at the top are: Partitions, Taxa, Tips, Traits, Sites, Clocks, Trees, States, Priors, Operators, and MCMC. The MCMC tab is active. The settings are as follows:  
Length of chain: 1000000  
Echo state to screen every: 1000  
Log parameters every: 1000  
File name stem: RAG1\_nucleotide\_chrono  
 Add .txt suffix  
Log file name: RAG1\_nucleotide\_chrono.log  
Trees file name: RAG1\_nucleotide\_chrono.trees  
 Create tree log file with branch length in substitutions:  
Substitutions trees file name:  
 Create operator analysis file:  
Operator analysis file name: RAG1\_nucleotide\_chrono.ops  
 Sample from prior only - create empty alignment  
Select the option below to perform marginal likelihood estimation (MLE) using path sampling (PS) / stepping-stone sampling (SS) which performs an additional analysis after the standard MCMC chain has finished.  
 Perform marginal likelihood estimation (MLE) using path sampling/stepping-stone sampling  
A "Settings" button is located below the MLE options.  
At the bottom of the window, it says: "Data: 33 taxa, 3 partitions, 1 user tree; Internal node(s) calibrated in nucleotide\_group;" and a "Generate BEAST File..." button.



## Preparing run for BEAST using a starting tree



- Generate the BEAST file. But, **DO NOT CLOSE BEUTi**, this will allow you correct mistakes easily in case something needs a quick fix
- A tab about un changed defaults priors will appear, just click 'Continue'

The screenshot shows the BEAUti interface with the 'Unchanged Default Priors' dialog box open. The dialog lists various parameters and their default priors. A red arrow points to the 'Continue' button at the bottom of this dialog. Another red arrow points to the 'Save As' field in the 'Generate BEAST XML File...' dialog box, which is also open and shows the filename 'RAG1\_nucleotide\_chrono\_BEAST' and the location 'BEAST\_Rag1'.

| Parameter                    | Prior                                       | Bo... | Description                               |
|------------------------------|---|-------|---|
| position_1.ac                | * Gamma [0.05, 10], initial=1               | [0... | GTR A-C substitution parameter            |
| position_1.ag                | * Gamma [0.05, 20], initial=1               | [0... | GTR A-G substitution parameter            |
| position_1.at                | * Gamma [0.05, 10], initial=1               | [0... | GTR A-T substitution parameter            |
| position_1.cg                | * Gamma [0.05, 10], initial=1               | [0... | GTR C-G substitution parameter            |
| position_1.gt                | * Gamma [0.05, 10], initial=1               | [0... | GTR G-T substitution parameter            |
| position_1.frequencies       | * Uniform [0, 1], initial=0.25              | [0... | base frequencies                          |
| position_1.alpha             | * Exponential [0.5], initial=0.5            | [0... | gamma shape parameter                     |
| position_1.plnv              | * Uniform [0, 1], initial=0.5               | [0... | proportion of invariant sites parameter   |
| position_2.ac                | * Gamma [0.05, 10], initial=1               | [0... | GTR A-C substitution parameter            |
| position_2.ag                | * Gamma [0.05, 20], initial=1               | [0... | GTR A-G substitution parameter            |
| position_2.at                | * Gamma [0.05, 10], initial=1               | [0... | GTR A-T substitution parameter            |
| position_2.cg                | * Gamma [0.05, 10], initial=1               | [0... | GTR C-G substitution parameter            |
| position_2.gt                | * Gamma [0.05, 10], initial=1               | [0... | GTR G-T substitution parameter            |
| position_2.frequencies       | * Uniform [0, 1], initial=0.25              | [0... | base frequencies                          |
| position_2.alpha             | * Exponential [0.5], initial=0.5            | [0... | gamma shape parameter                     |
| position_2.plnv              | * Uniform [0, 1], initial=0.5               | [0... | proportion of invariant sites parameter   |
| position_3.ac                | * Gamma [0.05, 10], initial=1               | [0... | GTR A-C substitution parameter            |
| position_3.ag                | * Gamma [0.05, 20], initial=1               | [0... | GTR A-G substitution parameter            |
| position_3.at                | * Gamma [0.05, 10], initial=1               | [0... | GTR A-T substitution parameter            |
| position_3.cg                | * Gamma [0.05, 10], initial=1               | [0... | GTR C-G substitution parameter            |
| position_3.gt                | * Gamma [0.05, 10], initial=1               | [0... | GTR G-T substitution parameter            |
| position_3.frequencies       | * Uniform [0, 1], initial=0.25              | [0... | base frequencies                          |
| position_3.alpha             | * Exponential [0.5], initial=0.5            | [0... | gamma shape parameter                     |
| position_3.plnv              | * Uniform [0, 1], initial=0.5               | [0... | proportion of invariant sites parameter   |
| ucld.stdev                   | * Exponential [0.333333], initial=0.3333... | [0... | uncorrelated lognormal relaxed clock std  |
| treeModel.rootHeight         | * Using Tree Prior in [0, ∞]                | [0... | root height of the tree                   |
| birthDeath.meanGrowthRate    | * Uniform [0, 1ES], initial=0.2             | [0... | Birth-Death speciation process rate       |
| birthDeath.relativeDeathRate | * Uniform [0, 1], initial=0.5               | [0... | Birth-Death speciation process relative d |

• These priors listed above are still set to the default values and need to be reviewed.



## Fixing the BEAUTi file



- BEAST has tends to split long trees into pieces that need to put in a single line:

```
Δ Δ <!--The user-specified starting tree in a newick tree format.....-->~
Δ Δ <newick.usingDates="false">~
Δ Δ Δ ((Polychrus_marmoratus_RAG1_FJ356748:30.343504000000003,(Hoplocercus_spinosus_RAG1_AY662592:15.381447950000002,(Moruna
Δ Δ Δ G1_JN112660:10.4003656099999994,(Pristidactylus_torquatus_RAG1_JF806210:5.0764752749999993,(Leiosaurus_catamarcensis_RAG
Δ Δ </newick>~
Δ </rescaledTree>~
```

It should be like this:

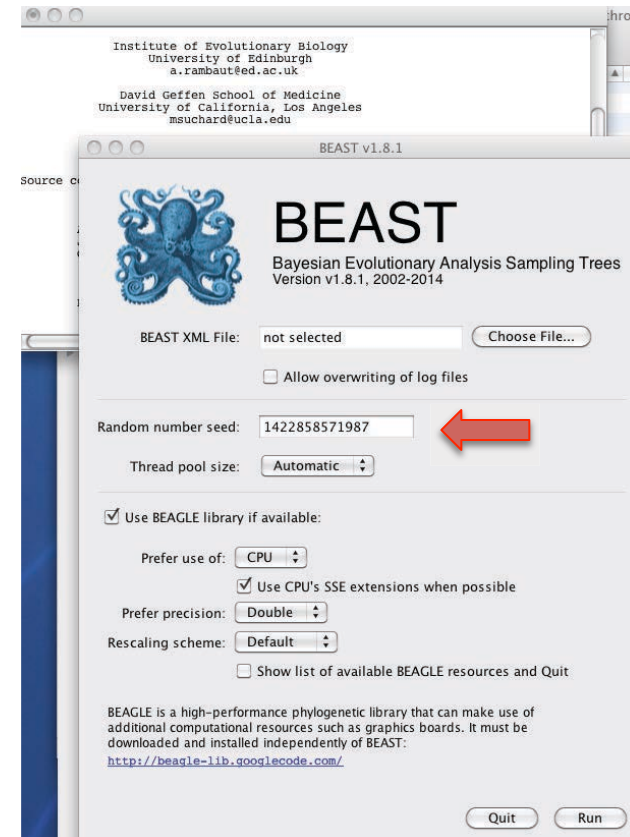
```
Δ Δ <!--The user-specified starting tree in a newick tree format.....-->~
Δ Δ <newick.usingDates="false">~
Δ Δ Δ ((Polychrus_marmoratus_RAG1_FJ356748:30.343504000000003,(Hoplocercus_spinosus_RAG1_AY662592:15.381447950000002,(Morunasc
Δ Δ </newick>~
Δ </rescaledTree>~
```

- During a test run I found that one of the lower bound of the uniform distribution of the Corytophanidae node was to high and I have to fix it to a lower value. It is very likely that you will have to review the BEAUTi file for such changes.

## Running BEAST



- If you made a good job, then BEAST will run for a long time and will produce several intermediate files. I recommend to create a folder for each run and have at least 4 independent starts.
- The same xml file will run in the cluster, so check that it runs in your computer before you send it to the cluster
- Click on BEAST v1.8.1, select you xml file and run.
- For independent starts, change the **Random Number Seed**



# Running BEAST



- Check the performance of the run and wait until it finishes

```
# BEAST v1.8.1, r6542
# Generated Sun Feb 01 23:33:41 MST 2015 [seed=1422858571987]
state      Posterior      Prior      Likelihood      rootHeight      uclid.mean      -
0          -28772.1693      -243.6374      -28528.5319      66.0000      1.00000      -
5000      -14593.1593      -213.7703      -14379.3890      71.7504      3.56228E-2      -
10000     -14407.9203      -215.8216      -14192.0986      88.9196      2.73428E-2      -
15000     -14310.7501      -216.0205      -14094.7296      86.6680      2.65093E-2      0.08 hours/million states
20000     -14194.7552      -209.9604      -13984.7948      82.4061      3.03578E-2      0.08 hours/million states
25000     -14161.6366      -205.8780      -13955.7586      74.9694      3.34211E-2      0.08 hours/million states
30000     -14099.9320      -215.6315      -13884.3006      110.228      1.89855E-2      0.08 hours/million states
35000     -14057.8879      -204.6979      -13853.1901      77.1003      1.87304E-2      0.08 hours/million states
40000     -14050.6278      -204.2057      -13846.4221      77.3917      1.72719E-2      0.08 hours/million states
45000     -14021.7850      -195.6160      -13826.1690      58.4651      1.69514E-2      0.08 hours/million states
50000     -14029.7041      -197.9477      -13831.7564      64.9551      1.80641E-2      0.08 hours/million states
55000     -14009.3050      -198.6718      -13810.6332      65.4767      1.43401E-2      0.08 hours/million states
60000     -14005.0601      -204.1453      -13800.9148      85.8936      1.13346E-2      0.08 hours/million states
65000     -14005.8682      -200.2025      -13805.6656      69.3222      9.35976E-3      0.08 hours/million states
70000     -13991.6573      -195.0670      -13796.5903      70.5298      1.19357E-2      0.08 hours/million states
75000     -13988.4780      -197.6260      -13790.8520      70.0007      6.46186E-3      0.08 hours/million states
80000     -13987.1500      -199.3932      -13787.7568      70.4576      5.48735E-3      0.08 hours/million states
85000     -13972.5891      -190.4962      -13782.0929      55.0260      6.45007E-3      0.08 hours/million states
90000     -13962.3126      -189.3563      -13772.9563      56.7294      6.85487E-3      0.08 hours/million states
95000     -13977.4897      -196.0226      -13781.4671      63.3262      4.49752E-3      0.08 hours/million states
100000    -13965.4051      -187.8012      -13777.6039      50.9620      4.16763E-3      0.08 hours/million states
105000    -13960.3971      -184.1117      -13776.2854      52.0825      3.69256E-3      0.08 hours/million states
110000    -13977.2950      -191.2086      -13786.0864      76.1688      2.36382E-3      0.08 hours/million states
115000    -13958.3660      -192.3255      -13766.0405      69.1255      2.38537E-3      0.08 hours/million states
120000    -13949.8271      -186.6823      -13763.1448      53.1830      2.97116E-3      0.08 hours/million states
125000    -13953.1798      -184.5700      -13768.6098      52.1756      2.18359E-3      0.08 hours/million states
130000    -13947.2597      -183.4735      -13763.7862      51.1387      1.78493E-3      0.08 hours/million states
135000    -13966.1815      -196.0768      -13770.1047      75.3659      1.03537E-3      0.08 hours/million states
140000    -13969.9735      -192.9568      -13777.0168      67.3515      1.15771E-3      0.08 hours/million states
```

# Running BEAST



- The run will finish with a report about the operators

```

-----
49970000      -13949.6312      -188.4724      -13761.1588      64.7572      1.11034E-3      0.09 hours/million st
49975000      -13958.1611      -196.4494      -13761.7117      90.5909      8.82895E-4      0.09 hours/million st
49980000      -13958.6557      -195.0457      -13763.6100      96.0201      8.47822E-4      0.09 hours/million st
49985000      -13939.4367      -182.4271      -13757.0096      57.6833      1.28911E-3      0.09 hours/million st
49990000      -13951.5338      -195.4775      -13756.0563      71.7646      1.02093E-3      0.09 hours/million st
49995000      -13956.5522      -185.5091      -13771.0432      59.1001      1.22199E-3      0.09 hours/million st
50000000      -13946.8169      -181.2748      -13765.5421      59.9515      1.33586E-3      0.09 hours/million st
-----

Operator analysis
Operator      Tuning      Count      Time      Time/Op      Pr(accept)
scale(position_1.ac)      0.405      45374      20757      0.46      0.2751
scale(position_1.ag)      0.466      45235      20661      0.46      0.2805
scale(position_1.at)      0.282      45269      20551      0.45      0.2896
scale(position_1.cg)      0.276      45146      20466      0.45      0.2786
scale(position_1.gt)      0.33      45140      20622      0.46      0.2531
position_1.frequencies      0.058      44996      20762      0.46      0.2716
scale(position_1.alpha)      0.482      45011      19448      0.43      0.2611
scale(position_1.pInv)      0.555      45173      19548      0.43      0.279
scale(position_2.ac)      0.318      45486      17998      0.4      0.2817
scale(position_2.ag)      0.322      45265      17843      0.39      0.2763
scale(position_2.at)      0.308      45259      17869      0.39      0.3081
scale(position_2.cg)      0.318      45595      18133      0.4      0.277
scale(position_2.gt)      0.245      45057      17774      0.39      0.2941
position_2.frequencies      0.06      45261      18128      0.4      0.2741
scale(position_2.alpha)      0.441      45247      16626      0.37      0.2579
scale(position_2.pInv)      0.648      45429      16790      0.37      0.2265
scale(position_3.ac)      0.483      45568      36966      0.81      0.2557
scale(position_3.ag)      0.662      45238      36589      0.81      0.2606
scale(position_3.at)      0.504      44789      36162      0.81      0.2615
scale(position_3.cg)      0.435      45320      36625      0.81      0.2722
scale(position_3.gt)      0.487      44959      36389      0.81      0.2731
position_3.frequencies      0.049      45339      36563      0.81      0.2568
scale(position_3.alpha)      0.464      45185      35342      0.78      0.2702
scale(position_3.pInv)      0.175      45194      35334      0.78      0.3163
scale(uclid.mean)      0.782      1356415      1254398      0.92      0.2212
scale(uclid.stdev)      0.54      1356488      1255277      0.93      0.2477
subtreeSlide(treeModel)      9.79      6786741      1842723      0.27      0.2334
Narrow Exchange(treeModel)      6787151      1269337      0.19      0.1021
Wide Exchange(treeModel)      1355796      186515      0.14      0.0037
wilsonBalding(treeModel)      1355947      368565      0.27      0.0066
scale(treeModel.rootHeight)      0.833      1359805      146776      0.11      0.2415
uniform(nodeHeights(treeModel))      13578475      3899661      0.29      0.2952
scale(birthDeath.meanGrowthRate)      0.502      1356241      50330      0.04      0.2622
scale(birthDeath.relativeDeathRate)      0.174      1355747      32609      0.02      0.2645
scale(birthDeath.sampleProbability)      0.261      1358500      47544      0.03      0.2578
up:uclid.mean down:nodeHeights(treeModel)      0.603      1356231      518721      0.38      0.2366
swapOperator(branchRates.categories)      4525097      1662619      0.37      0.4712
uniformInteger(branchRates.categories)      4525831      1325832      0.29      0.6072

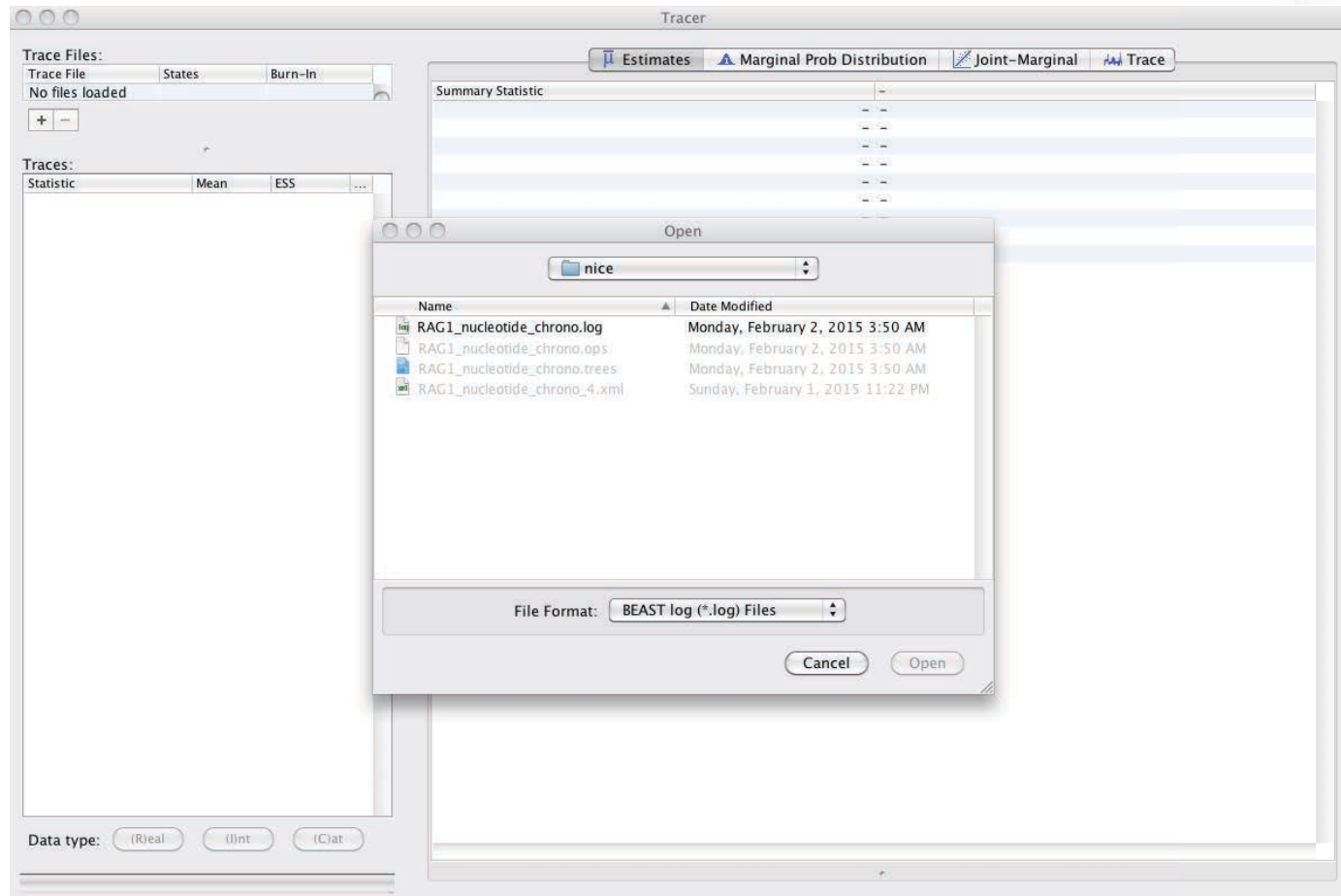
4.2769636111111111 hours

```

# BEAST: Visualizing the results in Tracer



- Open Tracer and explore the results of the course website:  
[RAG1\\_nucleotide\\_chrono.log.zip](#)





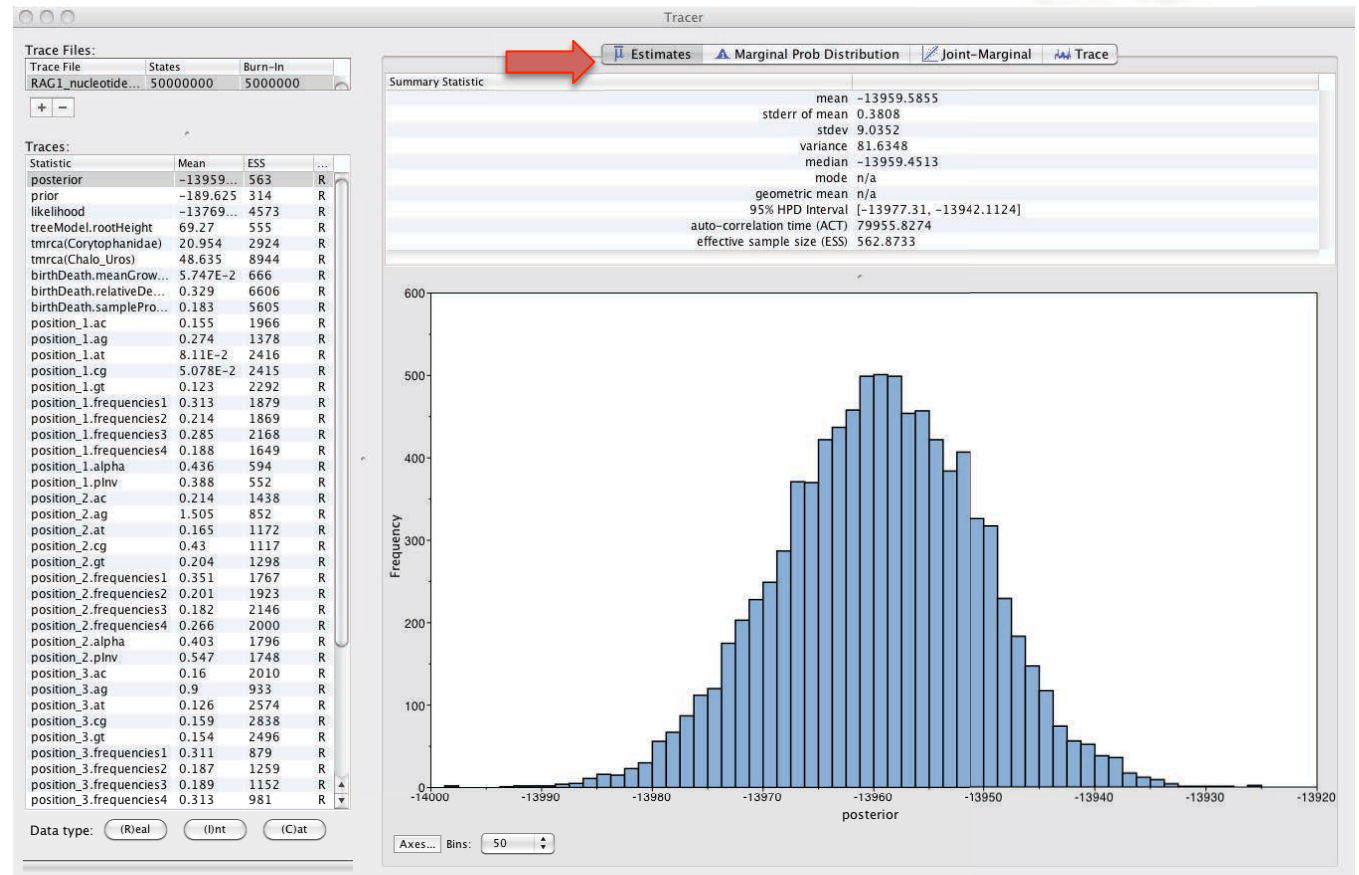
# BEAST: Visualizing the results in Tracer



- Open Tracer and explore the results of:  
`RAG1_nucleotide_chrono.log`

From the Manual: On the left hand side is a list of the different quantities that BEAST has logged. There are traces for the posterior (this is the log of the product of the tree likelihood and the prior probabilities), and the continuous parameters.

Selecting a trace on the left brings up analyses for this trace on the right hand side depending on tab that is selected.



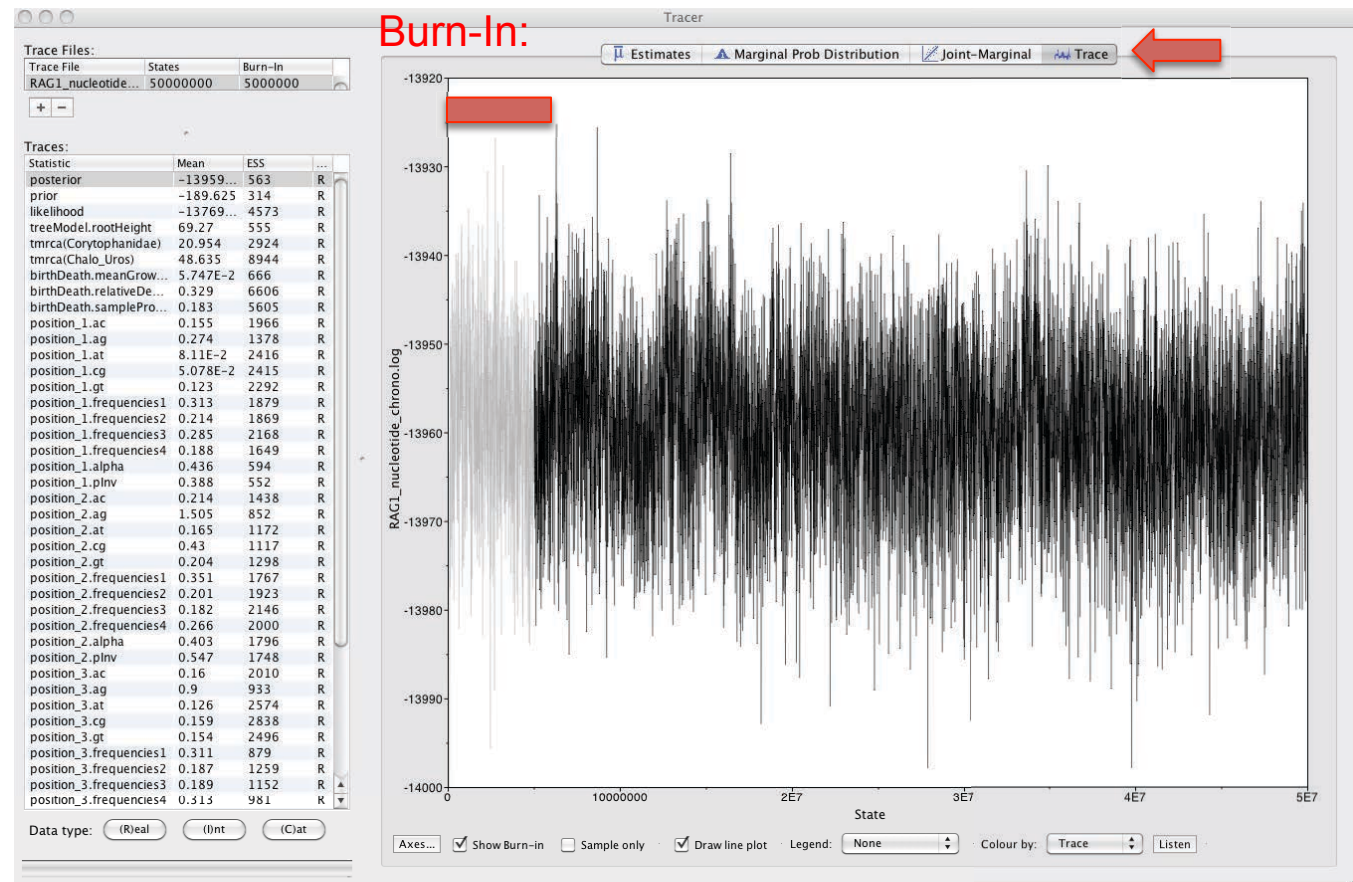
# BEAST: Visualizing the results in Tracer



- Open Tracer and explore the results of:  
`RAG1_nucleotide_chrono.log`

From the Manual: On the left hand side is a list of the different quantities that BEAST has logged. the posterior (this is the log of the product of the tree likelihood and the prior probabilities), and the continuous parameters.

**Burn-In:** We need to discard early very bad estimates that usually do not add information to our parameter estimation. These wrong estimates are usually visualized using the Trace Tab





## BEAST: Visualizing the results

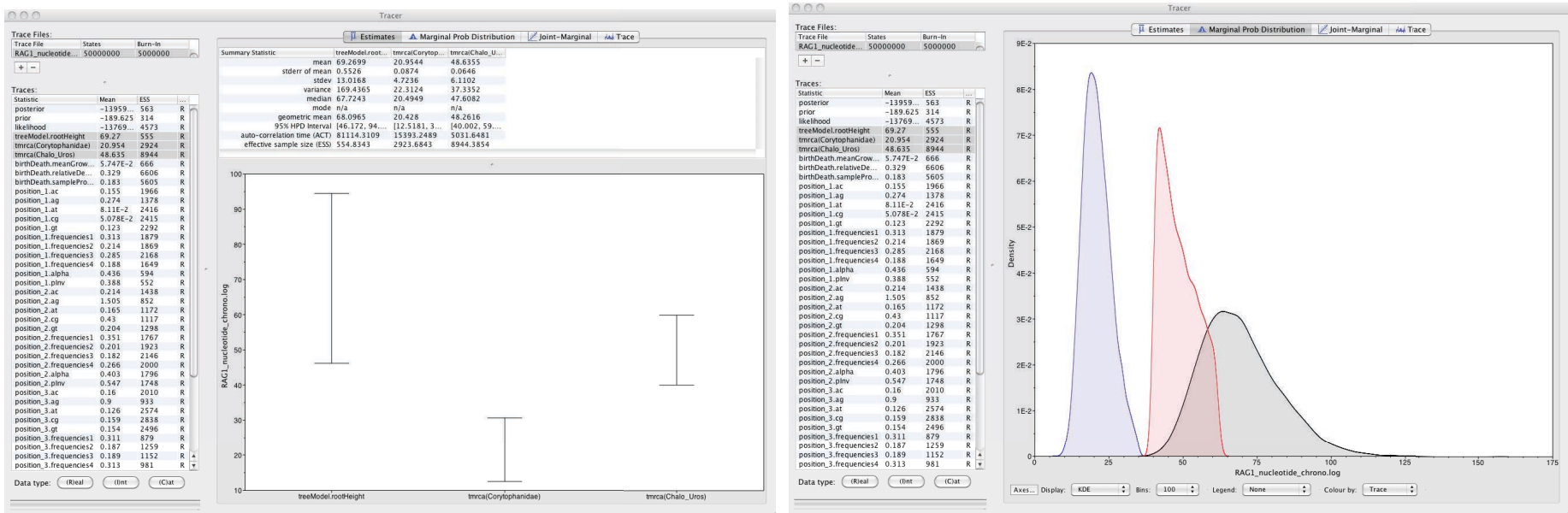


- Tracer will display summary statistics (e.g., means) and in the estimates in each parameter as the run progresses (graph).
- We are interested in **Effective Sample Size (ESS)**. This value indicates the size of a set of independent data points with the same information.
- If ESS is good if it is high enough (i.e.,  $> 100$ ) it will also be indicated in black.
- A low ESS (e.g.,  $< 100$ ) means that even though we may have sampled many data points, our search is rejecting **most of its proposals** (e.g., node ages) and thus **may not represent the posterior distribution well**. Other reason is that the proposals it accepts are all very close together so that it is not moving freely across the surface.
- Tracer will color **Effective Sample Size (ESS)** statistic under 100 as **RED** indicating a run which is **too short**. We need to run longer our chains (as many times the lowest ESS to get  $> 100$  ESS for that parameter: if ESS = 25 you should run  $> 4$  times the chain).
- It is suggested to have ESS  $> 200$  to be confident in our estimates. For more information read: <http://beast.bio.ed.ac.uk/analysing-beast-output>

# BEAST: Visualizing the results



- Let's explore the results of our nodes that we assigned calibration dates



- This BEAST search look in general good, so lets estimate our summary chronogram.

## BEAST: Summary Tree



- If you have more than 1 run, you need to be combined the tree and log files using **LogCombiner v1.8.1** (see Manual).
- Use Tracer to determine if your total runs are long enough to get a high ESS
- Then, open TreeAnnotator v1.8.1 and select the file from the course website:

[RAG1\\_nucleotide\\_chrono.trees.zip](#)

to create a summary tree for the graphs make sure:

- 1) to define the number of trees to burn-in with at least stationary posterior, in our case **5000000**
- 2) posterior probability limit at **0.0**
- 3) set to **maximum clade credibility tree**
- 4) set to **Mean or Median heights**

# BEAST: Summary Tree



```
TreeAnnotator v1.8.1, 2002-2014
MCMC Output analysis
by
Andrew Rambaut and Alexei J. Drummond
Institute of Evolutionary Biology
University of Edinburgh
a.rambaut@ed.ac.uk
Department of Computer Science
University of Auckland
alexei@cs.auckland.ac.nz
```

TreeAnnotator v1.8.1

Specify the burnin as the number of states

Burnin (as states):

Specify the burnin as the number of trees

Burnin (as trees):

Posterior probability limit:

Target tree type:

Node heights:

---

Target Tree File:

Input Tree File:

Output File:

# BEAST: Summary Tree



```
TreeAnnotator v1.8.1, 2002-2014
MCMC Output analysis
by
Andrew Rambaut and Alexei J. Drummond

Institute of Evolutionary Biology
University of Edinburgh
a.rambaut@ed.ac.uk

Department of Computer Science
University of Auckland
alexei@cs.auckland.ac.nz

Reading trees (bar assumes 10,000 trees)...
0          25          50          75          100
|-----|-----|-----|-----|
*****

Total trees read: 10001
Ignoring first 5000000 states (1000 trees).
Total unique clades: 303

Finding maximum credibility tree...
Analyzing 9001 trees...
0          25          50          75          100
|-----|-----|-----|-----|
*****

Best tree: STATE_11300000 (tree number 2261)
Highest Log Clade Credibility: -8.93969567825113
Collecting node information...
0          25          50          75          100
|-----|-----|-----|-----|
*****

Annotating target tree...
Writing annotated tree....
Finished - Quit program to exit.
```

We will get in our folder the annotated tree:

[RAG1\\_nucleotide\\_chrono\\_summary.tre](#)

We can now visualize our chronogram in FigTree and this can be used for comparative analyses

## BEAST: Summary Tree



- We can explore this tree or the one in the course website:

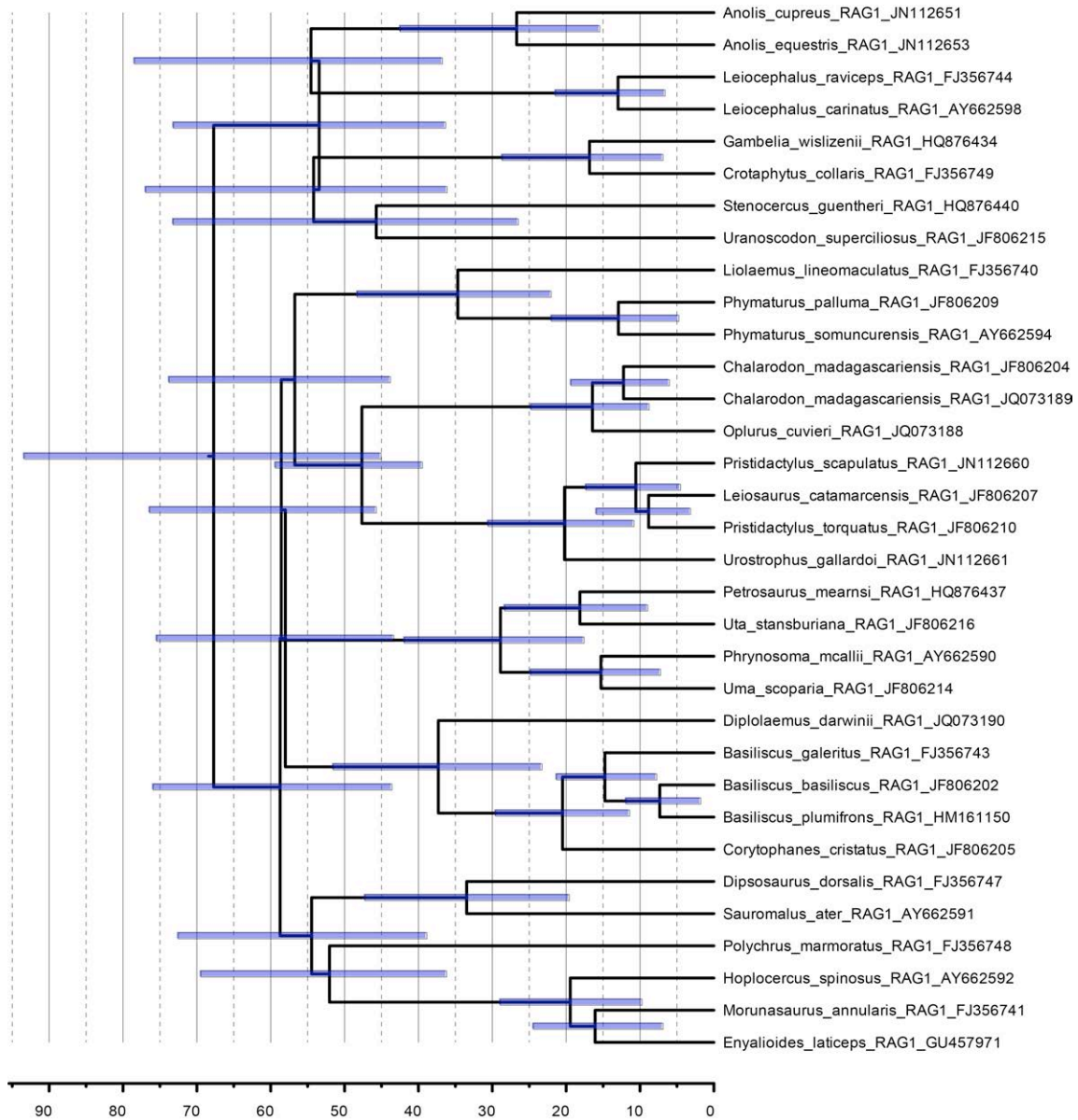
[RAG1\\_nucleotide\\_chrono\\_summary.tre](#)

and see the nodal 95% HPD (highest posterior density interval) which represents the most compact interval on the selected parameter that contains 95% of the posterior probability.

- These 95% HPD can be thought of as a Bayesian analog to a confidence interval.



# BEAST: Summary Tree



# Presentations

- This next Thursday (Feb 5<sup>th</sup>)
- You have 5 minutes (3 min for presentation and 2 min for questions)
- You will need to send me or bring a pdf of your presentation (so I put it in my computer)
- You will need to give me an abstract of your project:

Title: 150 characters (including spaces)

Text: 300 words maximum for your abstract.

This should include:

A brief background introduction

Your main question or hypothesis

Summary or names of comparative methods that you plan to use

Your expected results