

# BEAST tree

## Step 1: MSA

Beforehand, run the sequence in MEGA and ready a .nexus file with the MSA.

## Step 2 : BEAUtl input

1. Go to File -> “Import Alignment” and open the .nexus file. Or, to edit an already existing .xml file, File -> “Load” and open the .xml file

Partitions	Tip Dates	Site Model	Clock Model	Priors	MCMC			
Link Site Models	Unlink Site Models	Link Clock Models	Unlink Clock Models	Link Trees	Unlink Trees			
Name	File	Taxa	Sites	Data Type	Site Model	Clock Model	Tree	
1994_up_form	1994_up_form	129	1614	aminoacid	1994_up_form	1994_up_form	1994_up_form	

2. In the “Site Model” tab, choose 4 “Gamma Category Counts”, and JTT as “Site Model”.

Partitions	Tip Dates	Site Model	Clock Model	Priors	MCMC
Gamma Site Model					
Substitution Rate	1.0				<input type="checkbox"/> estimate
Gamma Category Count	4				
Shape	1.0				<input checked="" type="checkbox"/> estimate
Proportion Invariant	0.0				<input type="checkbox"/> estimate
Subst Model		JTT			

3. Choose the “Clock Model”. I am showing “Relaxed Clock Exponential”.

Partitions	Tip Dates	Site Model	Clock Model	Priors	MCMC
Relaxed Clock Exponential					
Number Of Discrete Rates	-1				
Rates					
<input type="checkbox"/> Normalize					
Clock.rate	1.0				<input checked="" type="checkbox"/> estimate

## Priors

1. Add the Yule Model.

Partitions Tip Dates Site Model Clock Model Priors MCMC

Tree.t:1994\_up\_form Yule Model

Birth Rate 1.0 ☒ estimate

☐ Conditional On Root

## 2. Add the birthrate 0-100.

Partitions Tip Dates Site Model Clock Model Priors MCMC

Tree.t:1994\_up\_form Yule Model

birthRate.t:1994\_up\_form Uniform initial = [1.0] [-∞,∞] Yule speciation process birth rate of partition t:1994\_up\_form

Lower 0.0

Upper 100.0

Offset 0.0

2.5% Quantile 2.50  
5% Quantile 5.00  
Median 50.0  
95% Quantile 95.0

## 3. Add the clock rate 0.001 - 1 substitutions/site/Myr.

ucldMean.c:1994\_up\_form Uniform initial = [1.0] [-∞,∞] uncorrelated lognormal relaxed clock mean of partition c:1994\_up\_form

Lower 0.001

Upper 1.0

Offset 0.0

2.5% Quantile 0.0260  
5% Quantile 0.0510  
Median 0.501  
95% Quantile 0.950

## 4. Add the MCMC priors.

Partitions Tip Dates Site Model Clock Model Priors MCMC

Tree.t:1994\_up\_form Yule Model

birthRate.t:1994\_up\_form Uniform initial = [1.0] [-∞,∞] Yule speciation process birth rate of partition t:1994\_up\_form

gammaShape.s:1994\_up\_form Exponential initial = [1.0] [-∞,∞] Prior on gamma shape for partition s:1994\_up\_form

ucldMean.c:1994\_up\_form Uniform initial = [1.0] [-∞,∞] uncorrelated exponential relaxed clock mean of partition c:1994\_up\_form

1A\_all.prior Normal ☐ monophyletic

Mean 550.0 ☐ estimate

Sigma 50.0 ☐ estimate

Offset 0.0

2.5% Quantile 452  
5% Quantile 468  
Median 550  
95% Quantile 637

☐ Tiponly

☐ Use Originate

## MCMC

1. Add chain length desired.

Partitions	Tip Dates	Site Model	Clock Model	Priors	MCMC
Chain Length	<input type="text" value="50000000"/>				
Store Every	<input type="text" value="-1"/>				
Pre Burnin	<input type="text" value="0"/>				
Num Initialization Attempts	<input type="text" value="10"/>				
<input checked="" type="checkbox"/> tracelog					
<input checked="" type="checkbox"/> screenlog					
<input checked="" type="checkbox"/> treeiog.t:1994_up_form					

2. Change the name of the .log file as intended.

<input checked="" type="checkbox"/> tracelog	
File Name	<input type="text" value="50M_4_MRCA.log"/>
Log Every	<input type="text" value="1000"/>
Mode	<input type="text" value="autodetect"/>
Sort	<input type="text" value="smart"/>
<input checked="" type="checkbox"/> Sanitise Headers	

3. Change the name of the .trees file as intended.

<input checked="" type="checkbox"/> treeiog.t:1994_up_form	
File Name	<input type="text" value="50M_4_MRCA.trees"/>
Log Every	<input type="text" value="1000"/>
Mode	<input type="text" value="tree"/>
Sort	<input type="text" value="none"/>
<input type="checkbox"/> Sanitise Headers	
TreeWithMetaDataLogger.t:1994_up_form	

## Step 3: Beast run

### Move Files to Spartan HPC

1. Put all XML files in Desktop.
2. Make Desktop the cd.
3. Secure copy (scp) the XML files to HPC.

```
cd OneDrive\ -\ The\ University\ of\ Melbourne\Desktop
```

```
scp *.xml sabrinai@spartan.hpc.unimelb.edu.au:/data/cephfs/punim1012/ch_1/1994
```

### Run in HPC

1. In the same directory, there should be following files:
  - a. Beast (beast.jar)
  - b. Launcher (launcher.jar)
2. Create a .slurm file. Put the codes in the .slurm file.

```
#!/bin/sh
#SBATCH -J 1994Str          # The job name
#SBATCH --ntasks=8
#SBATCH --mem-per-cpu=16384  # Memory in MB ie. ? * 1024
#SBATCH --cpus-per-task=4    # 4 OpenMP Threads
#SBATCH --time=0-100:00:00

# Load required modules
module load beagle-lib/3.1.2
module load java/1.8.0_241
module load beast/2.6.2

# Launching BEAST
echo "Strict clock, 30M MCMC"
srun java -Xmx10g -Dbeast.load.jars=true -Djava.library.path=$BEAGLE_LIB_PATH -jar beast.jar
-threads 4 -beagle_SSE -seed 777 1994_str_30M_4_MRCA-90.xml
```

```
#!/bin/sh
#SBATCH -J 1994ERlx        # The job name
#SBATCH --ntasks=8
#SBATCH --mem-per-cpu=16384  # Memory in MB ie. ? * 1024
#SBATCH --cpus-per-task=4    # 4 OpenMP Threads
#SBATCH --time=0-200:00:00

# Load required modules
module load beagle-lib/3.1.2
module load java/1.8.0_241
module load beast/2.6.2

# Launching BEAST
echo "Exponential Relaxed clock, 50M MCMC"
srun java -Xmx10g -Dbeast.load.jars=true -Djava.library.path=$BEAGLE_LIB_PATH -jar beast.jar
-threads 4 -beagle_SSE -seed 777 1994_rlxE_50M_4_MRCA-90.xml
```

```
#!/bin/sh
#SBATCH -J 1994LRlx        # The job name
```

```
#SBATCH --ntasks=8
#SBATCH --mem-per-cpu=16384    # Memory in MB ie. ? * 1024
#SBATCH --cpus-per-task=4      # 4 OpenMP Threads
#SBATCH --time=0-200:00:00

# Load required modules
module load beagle-lib/3.1.2
module load java/1.8.0_241
module load beast/2.6.2

# Launching BEAST
echo "Lognormal Relaxed clock, 55M MCMC"
srun java -Xmx10g -Dbeast.load.jars=true -Djava.library.path=$BEAGLE_LIB_PATH -jar beast.jar
-threads 4 -beagle_SSE -seed 777 1994_rlxL_55M_5_MRCA-90.xml
```

### 3. Run the .slurm files

```
sbatch 1994Str.slurm

sbatch 1994ERlx.slurm

sbatch 1994LRlx.slurm
```

## Move the Output Files to Device

### 1. There should be 3 types of file: .log, .trees, .xml.state

```
cd OneDrive\ -\ The\ University\ of\ Melbourne\Desktop

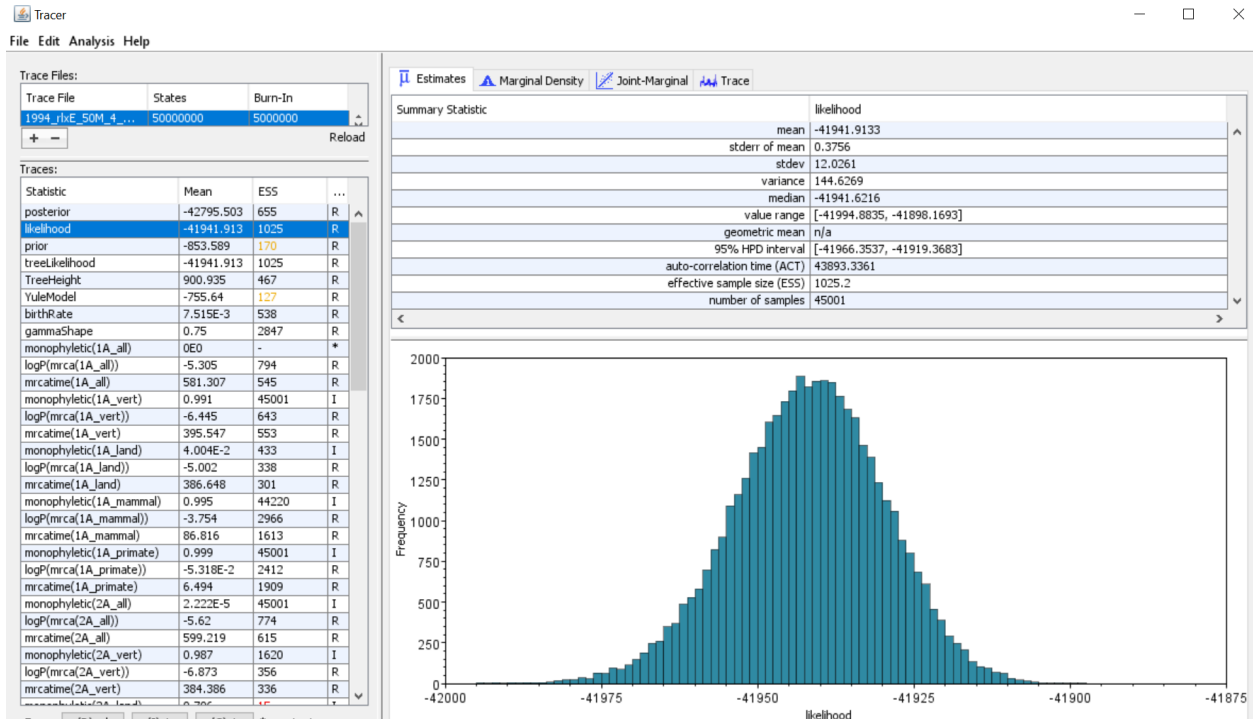
scp
sabrinal@spartan.hpc.unimelb.edu.au:/data/gpfs/projects/punim1012/ch_1/1994/1994_str_30M_4_
MRCA* .

scp
sabrinal@spartan.hpc.unimelb.edu.au:/data/gpfs/projects/punim1012/ch_1/1994/1994_rlxL_50M_4
_MRCA* .

scp
sabrinal@spartan.hpc.unimelb.edu.au:/data/gpfs/projects/punim1012/ch_1/1994/1994_rlxL_55M_5
_MRCA* .
```

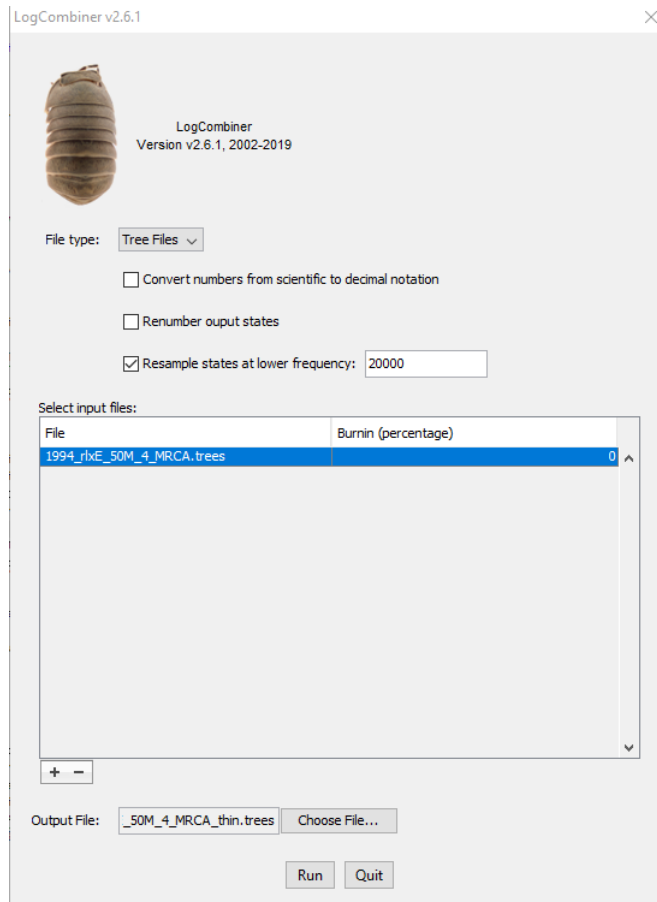
## Step 4: Tracer

1. Check all ESS. Majority of ESS should be >200. Especially Likelihood and Probability should be high.



## Step 5: LogCombiner

1. Choose the input file (.trees) and name the output file (thin.trees). Resample at the target length.



## Step 6: Tree Annotator

1. Chose 10 as “Burnin percentage”
2. Choose the this.trees as input file.
3. Name .tre as output file. **This .tre file is our actual output tree.**

TreeAnnotator v2.6.0

Burnin percentage:

Posterior probability limit:

Target tree type:

Node heights:

Target Tree File:

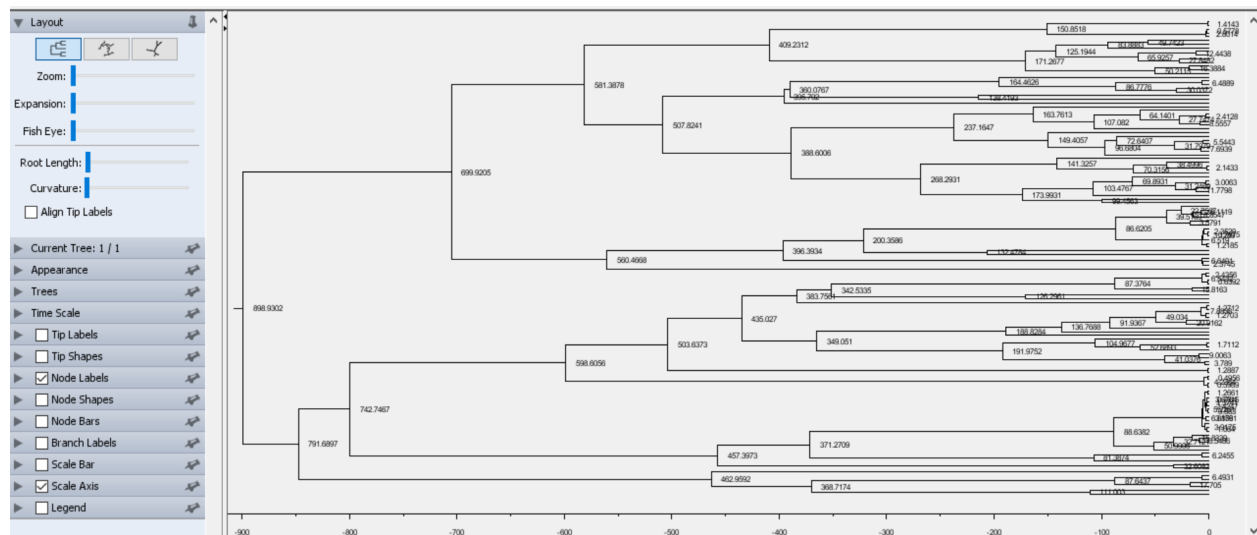
Input Tree File:

Output File:

Low memory: ☐

## Step 7: Figtree

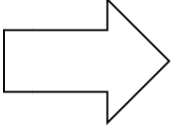
1. Using Figtree, look at the properties of the .tre file, such as the posterior values, and the heights of the nodes. These are our results.
2. To label the nodes we are interested in has a nicer tree image though, Figtree is not enough.



## Step 8: Edit Tre file

1. We would need ggtree for the tree figure. It will not color tips by species name if all tips have unique names (we needed this to produce the output). Now, we change each taxa name to only the name of the species.



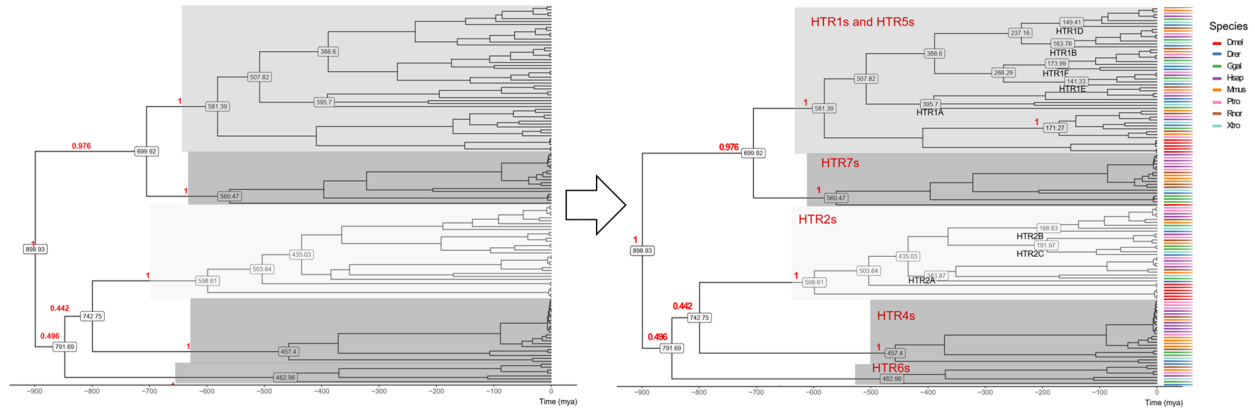
<pre> 3 Begin taxa; 4   Dimensions ntax=129; 5   Taxlabels 6     5-HTR_1A_B_Drosophila_melanogaster 7     5-HTR_1A_C_Drosophila_melanogaster 8     5-HTR_1A_Gallus_gallus 9     5-HTR_1A_Homo_sapiens 10    5-HTR_1A_Mus_musculus 11    5-HTR_1A_Pan_troglodytes 12    5-HTR_1A_Rattus_norvegicus 13    5-HTR_1A_Xenopus_tropicalis 14    5-HTR_1A_a_Danio_rerio 15    5-HTR_1A_b_Danio_rerio 16    5-HTR_1B_A_Drosophila_melanogaster 17    5-HTR_1B_D_Drosophila_melanogaster 18    5-HTR_1B_Danio_rerio 19    5-HTR_1B_E_Drosophila_melanogaster 20    5-HTR_1B_Gallus_gallus 21    5-HTR_1B_Homo_sapiens 22    5-HTR_1B_Mus_musculus 23    5-HTR_1B_Pan_troglodytes 24    5-HTR_1B_Rattus_norvegicus 25    5-HTR_1B_Xenopus_tropicalis 26    5-HTR_1D_Danio_rerio 27    5-HTR_1D_Gallus_gallus 28    5-HTR_1D_Homo_sapiens 29    5-HTR_1D_Mus_musculus 30    5-HTR_1D_Pan_troglodytes 31    5-HTR_1D_Rattus_norvegicus 32    5-HTR_1D_Xenopus_tropicalis 33    5-HTR_1E_Danio_rerio 34    5-HTR_1E_Gallus_gallus </pre>		<pre> 3 Begin taxa; 4   Dimensions ntax=129; 5   Taxlabels 6     Drosophila_melanogaster 7     Drosophila_melanogaster 8     Gallus_gallus 9     Homo_sapiens 10    Mus_musculus 11    Pan_troglodytes 12    Rattus_norvegicus 13    Xenopus_tropicalis 14    Danio_rerio 15    Danio_rerio 16    Drosophila_melanogaster 17    Drosophila_melanogaster 18    Danio_rerio 19    Drosophila_melanogaster 20    Gallus_gallus 21    Homo_sapiens 22    Mus_musculus 23    Pan_troglodytes 24    Rattus_norvegicus 25    Xenopus_tropicalis 26    Danio_rerio 27    Gallus_gallus 28    Homo_sapiens 29    Mus_musculus 30    Pan_troglodytes 31    Rattus_norvegicus 32    Xenopus_tropicalis 33    Danio_rerio 34    Gallus_gallus </pre>
---	---	--

## Step 9: Ggtree

1. Run the codes, inputting the .tre file

## Step 10: Inkscape

1. Edit clades squares to fit more nicely.
2. Edit node label positions and size.
3. Add taxa tip coloring.
4. Add HTR clade labels



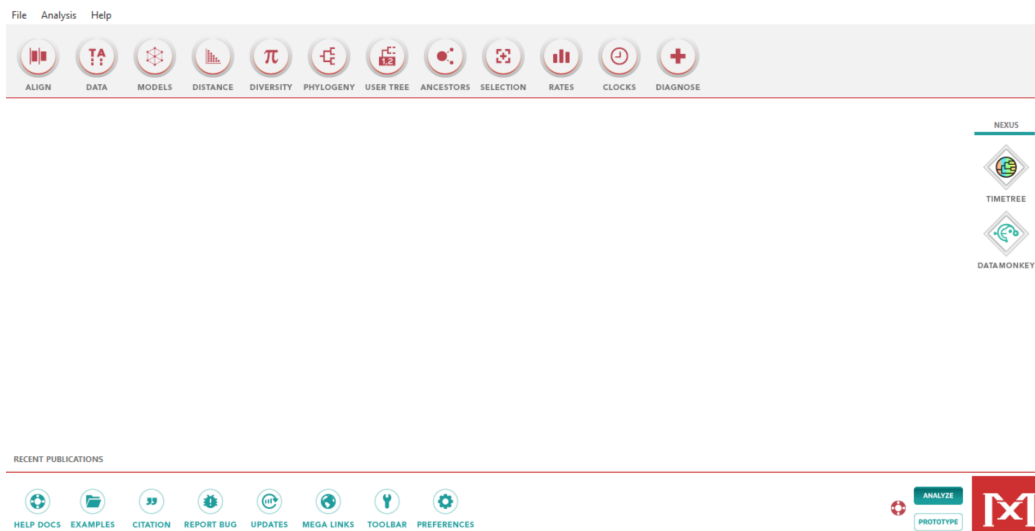
## ML tree with MEGA

### Step 1: MSA

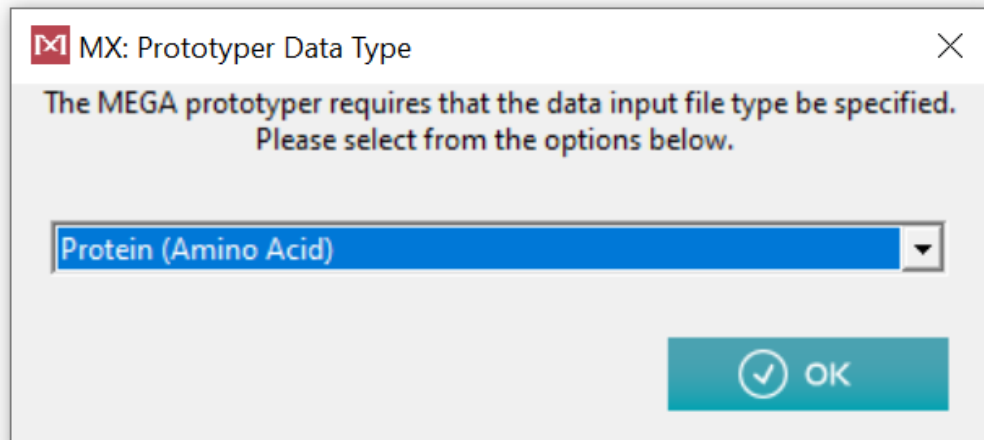
Perform MSA with the sequences and save them in .meg format.

### Step 2: MEGA Input (Generating the MEGA Analysis Options File)

1. On the main MEGA window, click the *Prototype* button, which is near the bottom right corner of the window, to switch MEGA to the prototyping model.



2. A dialog will be shown that prompts for the data type to be used. In this dialog, select Protein and click OK.



3. Select **Phylogeny | Construct/Test Maximum Likelihood Tree** from the main MEGA windows launch bar. Click the *Save Settings* button.
4. When prompted to save the .mao file, navigate to the target directory as infer\_ML\_amino\_acid\_1000.mao.
5. Close MEGA X.

## Step 3: MEGA run

Move files to Spartan HPC

1. Put all files in the Desktop.
2. Make Desktop the cd.
3. Secure copy (scp) the XML files to HPC.

```
cd OneDrive\ -\ The\ University\ of\ Melbourne\Desktop  
  
scp *.mao sabrinai@spartan.hpc.unimelb.edu.au:/data/cephfs/punim1012/ch_1/mega  
  
scp *.meg sabrinai@spartan.hpc.unimelb.edu.au:/data/cephfs/punim1012/ch_1/mega
```

Run in HPC

1. In the same directory, there should be following files: Megacc
2. Create a .slurm file. Put the codes in the .slurm file.

```
#!/bin/sh  
#SBATCH -J MLTree          # The job name  
#SBATCH --ntasks=8  
#SBATCH --mem-per-cpu=16384 # Memory in MB ie. ? * 1024  
#SBATCH --cpus-per-task=4   # 4 OpenMP Threads  
#SBATCH --time=0-500:00:00
```

```
# Load required modules
module load mega/10.0.5
```

```
# Launching MEGA
```

```
echo "MEGA ML tree"
```

```
megacc -a infer_ML_amino_acid_1000.mao -d all_HTR_1000ML.meg -o MLTree
```

Move the Output files to Device

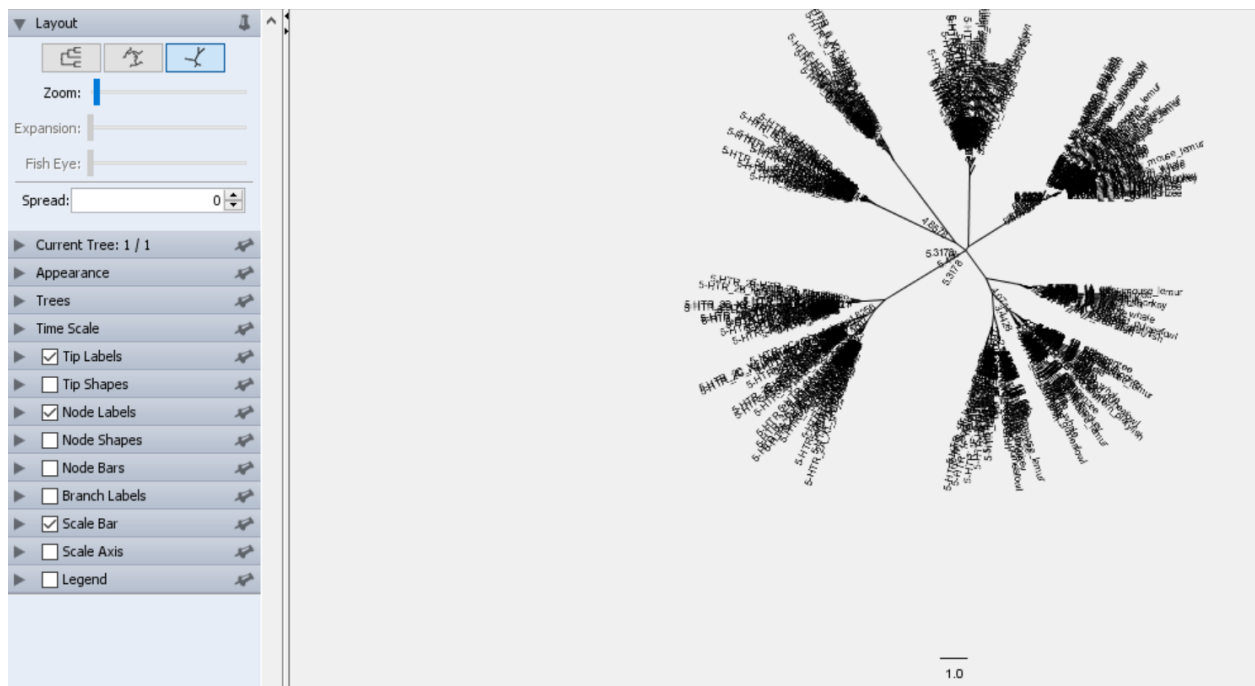
1. There should be 4 types of file: consensus.nwk, .nwk, partitions.txt, summary.txt

```
cd OneDrive\-\ The\ University\ of\ Melbourne\Desktop
```

```
scp sabrinai@spartan.hpc.unimelb.edu.au:/data/gpfs/projects/punim1012/ch_1/mega/MLTree*
```

## Step 4: Figtree

1. Visualize the tree in Figtree and save as PDF.
2. Note down bootstrap and node height.



## Step 5: Inkscape

1. Shorten tip labels if it looks better.

2. Highlight clades.
5. Edit node label color and size.
6. Add HTR clade labels

