

## Phylogenomic Analyses Using Nectin4 example. Minimum Evolution, Maximum Parsimony, Maximum Likelihood and Bootstrap.

We will use PAUP\* as our teaching program as it has a rigorous heuristic search algorithm and substitution modeling for three major optimality criteria commonly used in phylogenetic reconstruction: Maximum Likelihood, Minimum Evolution (distance based) and maximum parsimony (character based).

In the future, as you deal with very large data sets, most researchers will use RAxML. RAxML (Randomized Accelerated Maximum Likelihood) is a program for Maximum Likelihood-based inference of large phylogenetic trees. The program is explicitly being developed to efficiently infer trees for extremely large datasets, either in terms of the number of taxa and/or the sequence length. For example, a 25,000-taxon alignment of protobacteria with an alignment length of 1,500 base pairs had a run time on a single CPU of the cluster of only 13.5 days, with a memory consumption of only 1.5GB. Much additional information can be found on <https://cme.h-its.org/exelixis/web/software/raxml/>

Another advancement with RAxML is CIPRES Science gateway [http://www.phylo.org/sub\\_sections/portal/](http://www.phylo.org/sub_sections/portal/)

PAUP remains an excellent program that can handle large data sets and offers a comprehensive suite of tools in both GUI and command line. We will focus on the GUI interface today, but note that PAUP will generate command line script for each command you use. An example of a PAUP command line file is provided in this exercise. You can enter these commands in the command line box on the bottom of the GUI display.

The tree files generated with PAUP can also be viewed in FIGTREE which is in GDWApps folder.

Website: <http://paup.phylosolutions.com>

**Part 1: Selection of empirical nucleotide substitution model using jModeltest.** This is a hierarchical testing of models which computes a likelihood score for each model. The model with the lowest score as selected by AIC (Akaike Information criterion) or BIC (Bayesian information criterion)

**Step 1:** Open jmodeltest folder in GDWApps

**Step 2:** Start program by selecting **jModelTest.jar**

**Step 3.** In the jmodeltest window open our Nectin4 alignment.

**File-Load DNA Alignment** /Desktop/PAUP\_Exercise /Infiles/Nectin32TaxaMuscle.fasta

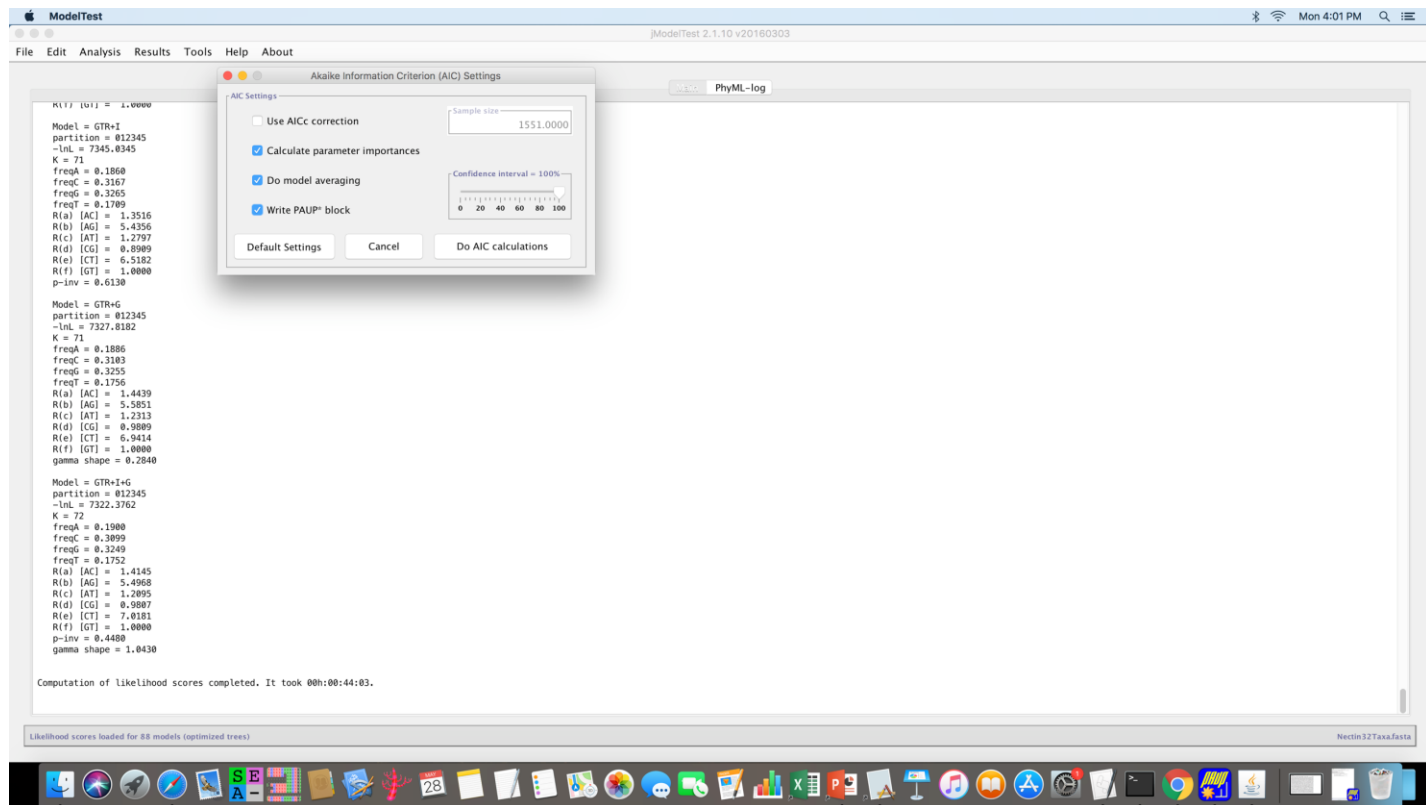
**Step 4:** Compute likelihood scores.

**Analysis-compute likelihood scores**

**Step 5:** Identify the model with the lowest likelihood score using AIC (or BIC).

**Analysis-Do AIC calculations**

**Select the following options shown here:**



**Step 6: Copy and paste the PAUP commands block as shown here: (you may have slightly different values)**

BEGIN PAUP;

Lset base=(0.1900 0.3107 0.3250 ) nst=6 rmat=(1.3621 5.5733 1.3621 1.0000 7.1275) rates=gamma  
shape=1.0570 ncat=4 pinvar=0.4500;

END;

**Step 7: You can save your results in an html file.**

**Results-Build HTML log**

Save into *Desktop/PAUP\_Exercise/Student\_files/Nectin32TaxaMuscle.jmodeltest.html*

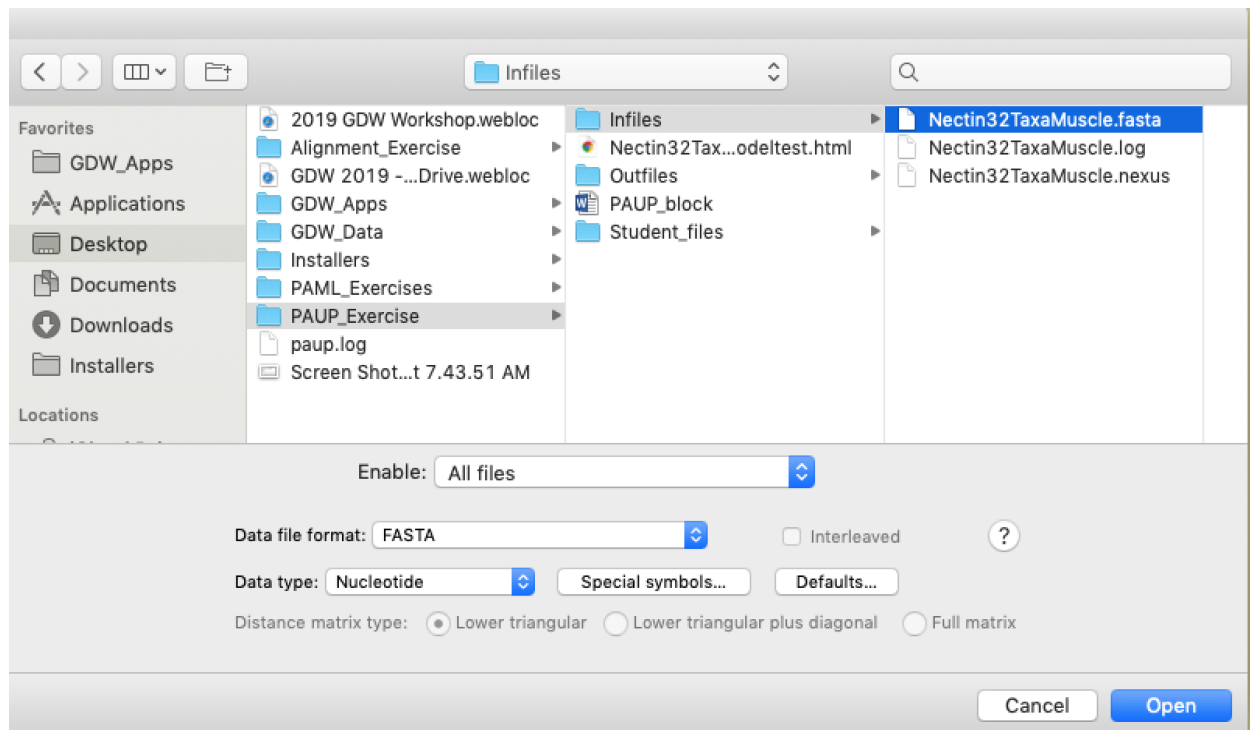
**Quit**

**Part 2: Construct phylogenetic tree using PAUP and nucleotide model selected by jModelTest and Maximum Likelihood optimality criterion**

Open PAUP in the GDWapps folder

**Step 1. Import and construct datafile in nexus format**

**File-Import** *PAUP\_Exercise/Infiles/Nectin32TaxaMuscle.fasta*.



**Step 2:** Edit the PAUP datafile by copying and pasting the PAUP block from jModeltest log file to the end of the PAUP file so it looks like this:



```
paup> Set criterion=likelihood;
```

**Step 8** Review the settings for the analyses with the submenus. You will see where the information contained in the PAUP block from jmodeltest is now incorporated into the settings.

#### Analysis-Likelihood settings

```
paup> LSet;
```

**Step 9** Perform the heuristic search. Examine the submenu settings for various options you might try but keep defaults for now.

#### Analysis-Heuristic search

```
paup> HSearch;
```

**Step 10:** We want to see what our tree looks like.

#### Trees-Print/view trees

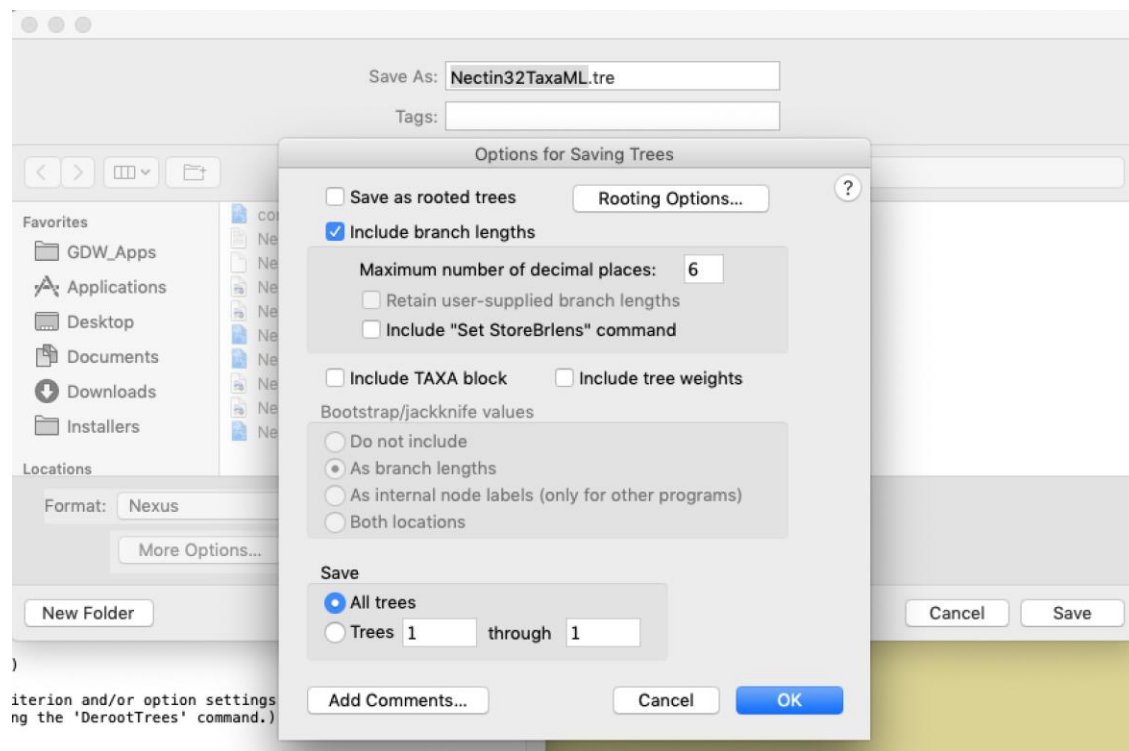
In this view we can change the taxon font to be smaller.

In rooting, change to midpoint root and then select rectangular plot type.

**Save as pdf** in *PAUP\_Exercise /Student\_files/Nectin32taxaML.pdf*  
**Close**

**Step 11.** We also want to save this treefile in a treefile format including branch lengths.

**Trees-Save trees to file, format = Nexus Open More Options and select Include branch lengths**  
*PAUP\_Exercise folder/Student\_files/Nectin32taxaML.tre (this is the nexus tree)*

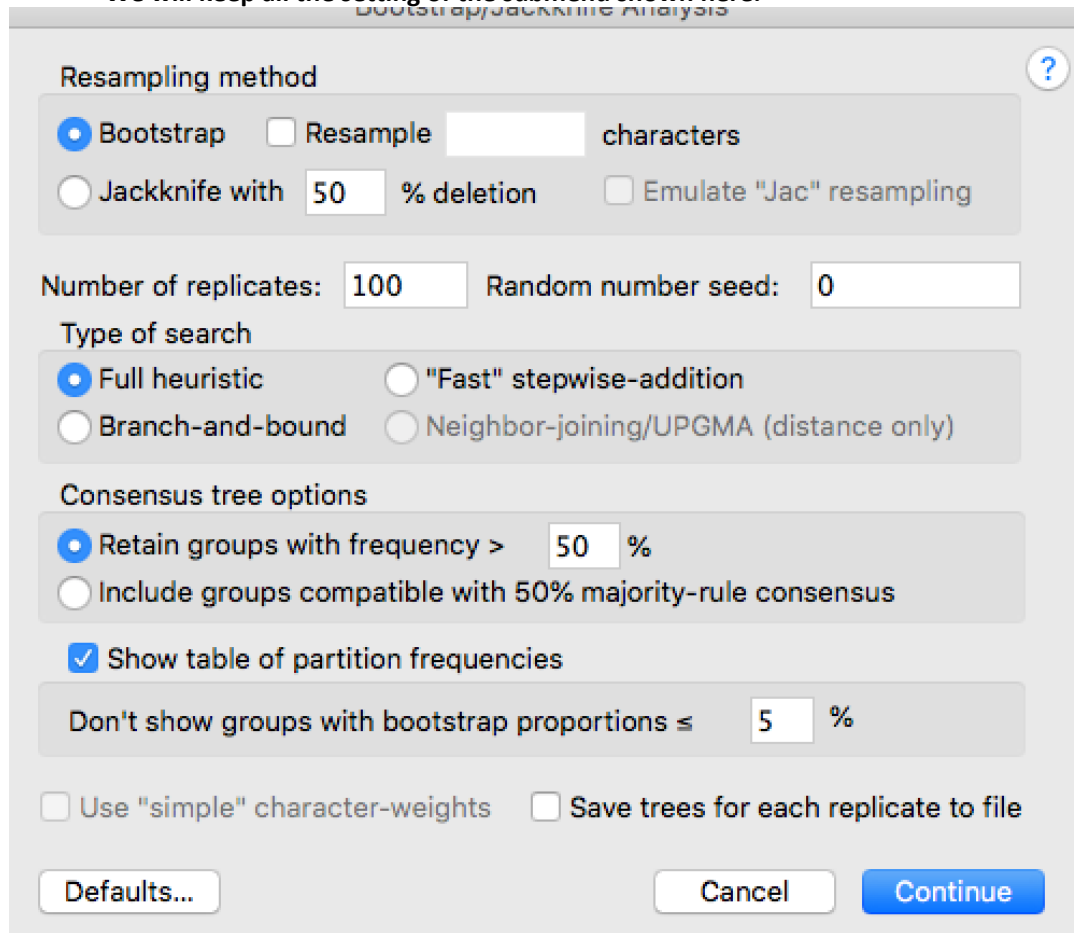


```
paup> SaveTrees
file='/Users/gdw/Desktop/PAUP_exercise/Student_files/Nectin32TaxaML.tre'
brlens=yes;
```

**Bootstrap Analysis:** We have to test the support of the nodes within the topology through bootstrapping. We keep all the parameters the same for maximum likelihood.

**Step 12: Analysis-Bootstrap/jackknife**

We will keep all the setting of the submenu shown here:



The screenshot shows the 'Bootstrap/Jackknife Analysis' dialog box in PAUP. The 'Resampling method' section has 'Bootstrap' selected with a radio button, and 'Resample' is unchecked. The 'Number of replicates' is set to 100, and the 'Random number seed' is 0. The 'Type of search' section has 'Full heuristic' selected. The 'Consensus tree options' section has 'Retain groups with frequency > 50 %' selected, and 'Show table of partition frequencies' is checked. The 'Don't show groups with bootstrap proportions ≤ 5 %' option is also visible. At the bottom, there are buttons for 'Defaults...', 'Cancel', and 'Continue'.

Resampling method

☒ Bootstrap ☐ Resample  characters

☐ Jackknife with  % deletion ☐ Emulate "Jac" resampling

Number of replicates:  Random number seed:

Type of search

☒ Full heuristic ☐ "Fast" stepwise-addition

☐ Branch-and-bound ☐ Neighbor-joining/UPGMA (distance only)

Consensus tree options

☒ Retain groups with frequency >  %

☐ Include groups compatible with 50% majority-rule consensus

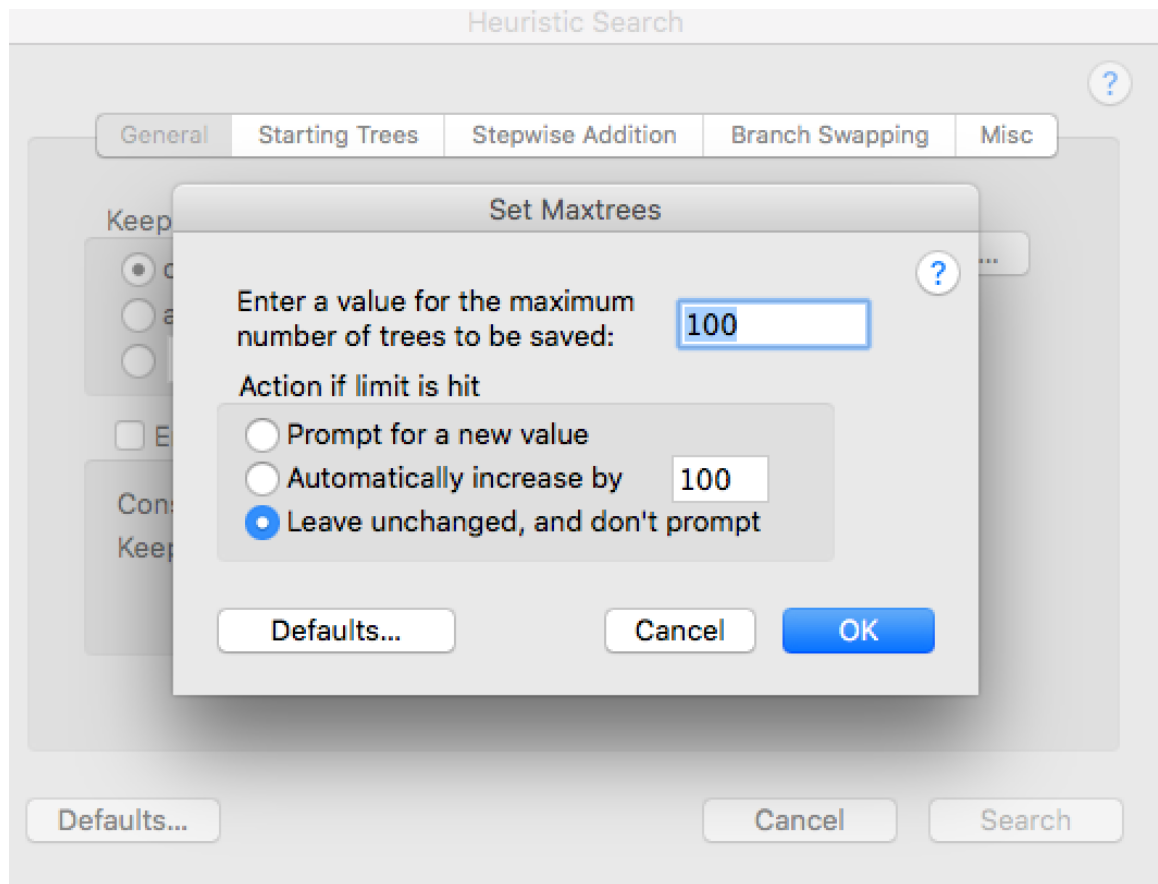
☒ Show table of partition frequencies

Don't show groups with bootstrap proportions ≤  %

☐ Use "simple" character-weights ☐ Save trees for each replicate to file

Defaults... Cancel Continue

We have to set max trees to 100 per replicate, and to not increase, as shown here:



```
paup> Bootstrap; Set increase=no;
```

#### **STOP SEARCH**

To save time as ML Bootstrap takes a longer amount of time, the consensus bootstrap treefile has been saved in

*PAUP\_Exercise/Outfiles/Nectin32taxaMLBS.tre*

*PAUP\_Exercise/Outfiles/Nectin32taxaMLBS.pdf*

**But, when you are doing a real time analyses and you want to see the majority rule tree:**

#### **Print/view Consensus Tree**

```
paup> ConTree / majRule
treeFile='/Users/gdw/Desktop/PAUP_exercise/Student_files/Nectin32TaxaMLBS.
tre';
```

#### **Repeat Analyses using the Minimum Evolution Optimality Criterion (Steps 7-12)**

**Step 7** Select Distance optimality criterion:

##### **Analysis-Distance**

```
paup> Set criterion=distance;
```

**Step 8** Review the settings for the analyses with the submenus.

##### **Analysis-Distance settings**



## Distance Options

### Choose Distance



- ☐ Standard distances: Total character difference
- ☒ DNA/RNA distances: General time-reversible
- ☐ RFLP/AFIP distances: Upholt Length = 6
- ☐ User-input distance

#### Objective Function

Neg. Branch Lengths

A.S.R.V.

Missing/Ambig

Other

- ☒ Unweighted least-squares (power = 0)
- ☐ Weighted least squares with standard ("polynomial") weighting
- ☐ Inverse weighting (power = 1)
  - ☐ Fitch-Margoliash criterion (power = 2)
  - ☒ Weight using power = 0
- ☐ Exponentially weighted least squares with  $P' = 0$
- ☐ Weighted least squares with user-supplied weights matrix
- ☐ Balanced minimum evolution (BME)
- ☐ Minimum evolution (ME/OLS)

Defaults...

Cancel

OK



Distance Options

Choose Distance ?

☐ Standard distances: Total character difference

☒ DNA/RNA distances: General time-reversible

☐ RFLP/AFLP distances: Upholt Length = 6

☐ User-input distance

Objective Function   **Neg. Branch Lengths**   A.S.R.V.   Missing/Ambig   Other

☐ Constrain branch lengths to be nonnegative

☒ Allow negative branch lengths. When calculating tree scores:
 

- ☐ Leave alone
- ☒ Set to zero
- ☐ Set to absolute value

☐ Collapse branches of effectively zero length when searching

Defaults...   Cancel   OK

We can enter values from our jModletest for each of the submenus shown here

```
paup> DSet distance=GTR;
```

Use the options shown here:

Distance Options

Choose Distance ?

☐ Standard distances: Total character difference

☒ DNA/RNA distances: General time-reversible

☐ RFLP/AFLP distances: Upholt Length = 6

☐ User-input distance

Objective Function Neg. Branch Lengths **A.S.R.V.** Missing/Ambig Other

Proportion of sites assumed to be invariable = 0.45

☐ Remove in equal proportions for each base

☒ Remove in proportion to state frequencies estimated from

☒ all sites ☐ constant sites only

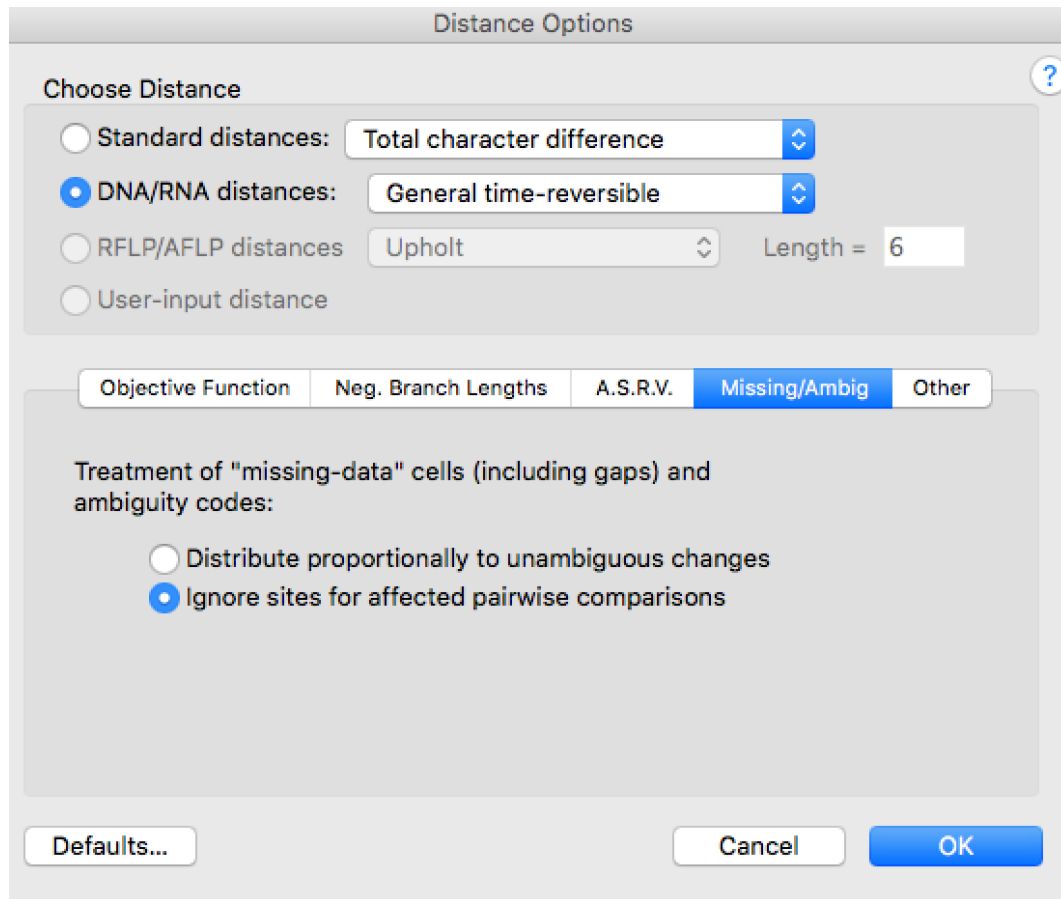
Rate for variable sites: ☐ Equal

☒ Gamma, shape = 1.0570

Defaults... Cancel OK

### How to treat indels (gaps)

Use the options shown here:



```
paup> DSet;
```

```
paup> DSet missDist=ignore;
```

**Step 9** Perform the heuristic search. Examine the submenu settings for various options you might try but keep defaults for now.

#### **Analysis-Heuristic search**

```
paup> HSearch;
```

**Step 10:** We want to see what our tree looks like.

#### **Trees-Print/view trees**

**Save as pdf** in *PAUP\_Exercise/Student\_files/Nectin32taxaME.pdf*

**Close**

**Step 11.** We also want to save this tree in a treefile format and make sure the branch lengths are being included (see ML above).

**Trees-Save trees to file, format = Nexus Select 'Include Branch Lengths'**

*PAUP\_Exercise/Student\_files/Nectin32taxaME.tre (this is the nexus tree)*

```
paup> SaveTrees
file='/Users/gdw/Desktop/PAUP_exercise/Student_files/Nectin32TaxaME.tre'
format=Nexus brlens=yes;
```

**Bootstrap Analysis:** We have to test the support of the nodes within the topology through bootstrapping. We keep all the parameters the same for maximum likelihood.

#### Step 12: Analysis-Bootstrap/jackknife

We will keep all the setting of the submenu shown here change to 1000 replicates

Use the options shown here:

**Bootstrap/Jackknife Analysis**

Resampling method ?

☒ Bootstrap ☐ Resample  characters

☐ Jackknife with  % deletion ☐ Emulate "Jac" resampling

Number of replicates:  Random number seed:

Type of search

☒ Full heuristic ☐ "Fast" stepwise-addition

☐ Branch-and-bound ☐ Neighbor-joining/UPGMA

Consensus tree options

☒ Retain groups with frequency >  %

☐ Include groups compatible with 50% majority-rule consensus

☒ Show table of partition frequencies

Don't show groups with bootstrap proportions ≤  %

☐ Use "simple" character-weights ☐ Save trees for each replicate to file

Defaults... Cancel Continue

Keep set max trees to 100 per replicate, and not increase.

Search

```
paup> Bootstrap nreps=1000; Set increase=no;
```

Let's look at the 50% majority rule tree:

Trees-Print/view Bootstrap Consensus

Save Tree as pdf

*PAUP\_exercise/Student\_files/Nectin32TaxaMEBS.pdf'*

```
paup> ConTree / majRule
treeFile='/Users/gdw/Desktop/PAUP_exercise/Student_files/Nectin32TaxaMEBS.
tre';
```

#### **Trees-Save trees to file**

**If prompted to save tree to outgroup rooting say 'yes'**

*PAUP\_Exercise/Student\_files/Nectin32taxaMEBS.tre*

```
paup>set rootMethod=outgroup;
```

```
paup> SaveTrees
file='/Users/gdw/Desktop/PAUP_exercise/Student_files/Nectin32TaxaMPBS.tre'
;
```

#### **Repeat Analyses using the Maximum Parsimony Optimality Criterion (Steps 7-12)**

**Step 7** Select Maximum Parsimony optimality criterion:

##### **Analysis-Parsimony**

**Step 8** Review the settings for the analyses with the submenus.

##### **Analysis-Parsimony settings**

We will use the default settings. NOTE: The submenu has the option under the 'Gaps' tab to score indels as a 'fifth' nucleotide. This can be an important option for the most accurate phylogenies if you are **absolutely** confident in the indels within your alignment.

```
paup> Set criterion=parsimony;
paup> PSet;
```

**Step 9** Perform the heuristic search. Examine the submenu settings for various options you might try but keep defaults for now.

##### **Analysis-Heuristic search**

```
paup> HSearch;
```

**Step 10:** We want to see what our 2 trees looks like.

##### **Trees-Print/view trees**

In rooting, change to midpoint root and then select rectangular plot type.

**Save as pdf** in *PAUP\_Exercise/Student\_files/Nectin32taxaMP.pdf*

**Close**

**Step 11.** We also want to save this treefile in a treefile format with branch lengths.

#### **Trees-Save trees to file, format = Nexus**

*PAUP\_Exercise/Student\_files/Nectin32taxaMP.tre*

```
paup> SaveTrees
file='/Users/gdw/Desktop/PAUP_exercise/Student_files/Nectin32TaxaMP.tre';
```

**Bootstrap Analysis:** We have to test the support of the nodes within the topology through bootstrapping. We keep all the parameters the same for maximum likelihood.

**Step 12: Analysis-Bootstrap/jackknife**

Keep number of replicates at 1000

Keep set max trees to 100 per replicate, and to not increase.

Search

paup> Bootstrap;

**Let's Look at the 50% majority rule tree:**

**Trees-Print/view Bootstrap Consensus Tree**

```
paup> ConTree / majRule
treeFile='/Users/gdw/Desktop/PAUP_exercise/Student_files/Nectin32TaxaMPBS.
tre';
```

**Trees-Save trees to file**

*PAUP\_Exercise/Student\_files/Nectin32taxaMPBS.tre*

*If prompted to change to outgroup rooting answer yes*

```
paup>set rootMethod=outgroup;
```

```
paup> SaveTrees
file='/Users/gdw/Desktop/PAUP_exercise/Student_files/Nectin32TaxaMPBS.tre'
;
```

Merge BS values with tree found in heuristic search (\*.tre), and do not use BS tree. This is because BS tree is a consensus tree, and only saves the BS values, not branch lengths. Our pdfs of BS trees will be helpful to combine with treefile for our final figure.

If you click on each treefile found with heuristic search, it will open in the Program FIGTREE.

In FIGTREE

Interpretation of Results from MP, ME, ML

Which of the 3 different optimality criterion methods give the most similar phylogenies?

Are the phylogenies accurate with respect to systematic relationships of the taxa included in the study?

What nodes are not fully resolved, and why?

Would adding more taxa resolve these nodes?

How would you interpret Nectin4 evolution within Carnivores compared with primates or ruminants?

Can we make any interpretation about Nectin4 in dogs, cats (both known to have neurologic forms of CDV) relative to other species?

How to make final figure:

- 1) Open ML tree pdf in powerpoint. You can try other tree editing programs like TreeGraph
- 2) Label nodes with bootstrap values from ML/ME/MP bootstrap Trees
- 3) Include scale of branch length
- 4) Add Mr Bayes values when we get them next. Also add those values to nodes ML/ME/MP/Bayes

