# Phylogenetic Course: Exercises

*Basic phylogenetic tree making*

We are going to use this shows **basic concepts** and few exercises to learn tree making using some popular methods. Later we will use the workflows via ete toolkit version 3 or *ete3*

# 1. Distance Based Method

- We are going to use Neighbour Joining method using `FastTree` (http://www.microbesonline.org/fasttree (http://www.microbesonline.org/fasttree))
- We can also use MEGA 6 that is avaible on your windows machines please check your MEGA version will be slightyly different than `MEGA 6`
- Tree building usually requires several external programs to compute trees, sequence alignments and to perform other tasks.
- We have all the tools needed to build the trees and visual on your machine. One of the recommended tool is *etetoolkit* (http://etetoolkit.org/download/ (http://etetoolkit.org/download/))

**We check if the tools are working OK ?** Hint: select the following and press shift + Enter ** We are going to a file conversion from Fasta to Phylip format that is needed by Raxml and Fastree. It is very compact and good for viewing alignments in text format too. We use unaligned.fast(as input) and and the output file we suggest is unaligned2.fasta

In [12]:

```
%run fast2phyl.py unaligned.fasta unaligned2.phy
```

In [13]:

```
!cat unaligned2.phy | head -n 5
```

```
 16 152
FER1_SPIOL_Ferr MAATTTTMMGMATTFVPKPQAPPMMAALPSNTGRSLFGLKTGSRGGRMTM
AAYKVTLVTPTGNVEFQCPDDVYILDAAEEEGIDLPYSCRAGSCSSCAGKLKTGSLNQDDQSFLDD
DQIDEGWVLTCAAYPVSDVTIETHKEEELTA-----
Q93Z60_ARATH_At MASTALSSAIVSTSFLRRQQTPISLRSLPFANTQSLFGLKSSTARGGRVT
AMATYKVKFITPEGEQEVECEEDVYVLDAAEEAGLDLPYSCRAGSCSSCAGKVVSGSIDQSDQSFL
DD---------------------------------
FER_CAPAN_Ferre MASVSATMISTSFMPRKPAVTSLKPIPNVGEALFGLKSANGGKVTCMASY
KVKLITPDGPIEFDCPDNVYILDQAEEAGHDLPYSCRAGSCSSCAGKIAGGAVDQTDGNFLDDDQL
EEGWVLTCVAYPQSDVTIETHKEAELVG--------
FER1_MESCR_Ferr MAATTAALSGATMSTAFAPKTPPMTAALPTNVGRALFGLKSSASRGRVTA
MAAYKVTLVTPEGKQELECPDDVYILDAAEEAGIDLPYSCRAGSCSSCAGKVTSGSVNQDDGSFLD
DDQIKEGWVLTCVAYPTGDVTIETHKEEELTA----
```

In [ ]:

In [15]:

```
%%bash
ete3 build check
```

```
Current Toolchain path: /Users/cheemaj/anaconda/bin/ete3_apps
Current Toolchain version: 2.0.3

      clustalo: OK - 1.2.1
        codeml: OK - AAML in paml version 4.8a, July 2014
      fasttree: OK - FastTree Version 2.1.8 Double precision (No
SSE3), OpenMP (1 threads)
        kalign: OK - Kalign version 2.04, Copyright (C) 2004, 200
5, 2006 Timo Lassmann
         mafft: OK - MAFFT v6.861b (2011/09/24)
        muscle: OK - MUSCLE v3.8.31 by Robert C. Edgar
         phyml: OK - . This is PhyML version 20160115.
     pmodeltest: OK - pmodeltest.py v1.4
         prank: OK - prank v.100802. Minimal usage: 'prank sequen
ce_file'
       probcons: OK - PROBCONS version 1.12 - align multiple prote
in sequences and print to standard output
         raxml: OK - This is RAxML version 8.1.20 released by Ale
xandros Stamatakis on April 18 2015.
 raxml-pthreads: OK - This is RAxML version 8.1.20 released by Ale
xandros Stamatakis on April 18 2015.
        readal: OK - readAl v1.4.rev6 build[2012-02-02]
           slr: OK - # SLR "Sitewise Likelihood Ratio" selection
detection program. Version 1.4.3
        statal: OK - statAl v1.4.rev6 build[2012-02-02]
       tcoffee: OK - PROGRAM: T-COFFEE Version_11.00.8cbe486 (201
4-08-12 22:05:29 - Revision 8cbe486 - Build 477)
        trimal: OK - trimAl v1.4.rev6 build[2012-02-02]

Dialign-tx not supported in OS X
```

One of the popular format to use a simple sequence file(a dummy file) in a Fasta Format.

- It is a simple text based file format where a header line begins with as '>' and followed by specied or taxa name and the following line is the sequence.
- Make sure that your data is correctly encoded as a FASTA formatted file.
- It is recommended that the header of the FASTA file does not contain weird symbols and that sequence names are not duplicated.
- You can use amino acids or nucleotide sequences. For this example, we will use amino-acid sequences: *protein5.fa* We can check the content of the above fasta file by the following instructions(or command)

We can also use the MEGA to convert into a Phylip fomat [protein5.phy.

Which is a rather much simpler format: Number of taxa or sequence (5) and the length each sequence is (20) appear in the first line followed by name of taxa and it corresponding sequence as a one line each. The first line of the input file contains the number of taxa and the number of sequence characters separated by blanks. The information for each species follows, starting with a ten-character species# name (which can include punctuation marks and blanks), and continuing with the characters for that species. For example, protein5.phy:

```
5 20
Cow          MAYPMQLGFQ DATSPIMEEL
Carp         MAHPTQLGFK DAAMPVMEEL
Chicken      MANHSQLGFQ DASSPIMEEL
Human        MAHAAQVGLQ DATSPIMEEL
Loach        MAHPTQLGFQ DAASPVMEEL
```

You can also use various online tools to convert them into other formats too. For example: (http://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html (http://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html))

Please free to use to get a Neighbour Joining (NJ) distance based tree with MEGA 6 on your window machine.

Please goto file open and select protein sequences. Once the sequences are loaded select Phylogeny and then Test/ Construct Neighbour Joining Tree. Go ahead and compute the Tree And Save the displayed Tree as newick formatted tree this will have an extension nwk or newick formatted Tree.

This is a very simple format using parenthesis to separate taxa and their relationship. For example a phylogenetic tree:

```
                /-A
          /---|
          |     \-B
    ----|
          |     /-C
          \---|
                \-D
```

Can be simply represented using:

(A,B),(C,D));

If you really have to construct a large scale sequences which will be the case with the data these days. NJ tree can be Heuristic based Neighbor-Joining using FastTree. It is really simple to use. Visit the site (http://www.microbesonline.org/fasttree/ (http://www.microbesonline.org/fasttree/)).

**We are going to use fastree here to get the NJ (distance based) tree for protein5.phy as following:**

```
In [16]:
```

```
%%bash
cat protein5.phy
```

```
  5 20
Cow           MAYPMQLGFQ DATSPIMEEL
Carp          MAHPTQLGFK DAAMPVMEEL
Chicken       MANHSQLGFQ DASSPIMEEL
Human         MAHAAQVGLQ DATSPIMEEL
Loach         MAHPTQLGFQ DAASPVMEEL
```

*How many species the above file has ? What is the sequence length for each ?* Now let us try to construct the distance base Neighbour Joining(NJ) tree for the above sequence. Input file is given to fasttree as symbol '<' and output is obtained via '>' to the FastTree. The output will be a newick formatted tree representation.

```
In [17]:
```

```
%%bash
fasttree -fastest < protein5.phy > protein5-fastNJ.nwk
```

```
FastTree Version 2.1.9 Double precision (No SSE3)
Alignment: standard input
Amino acid distances: BLOSUM45 Joins: balanced Support: SH-like 10
00
Search: Fastest+2nd +NNI +SPR (2 rounds range 10) +ML-NNI opt-each
=1
TopHits: 1.00*sqrtN close=default refresh=0.50
ML Model: Jones-Taylor-Thorton, CAT approximation with 20 rate cat
egories
Initial topology in 0.00 seconds
Refining topology: 9 rounds ME-NNIs, 2 rounds ME-SPRs, 5 rounds ML
-NNIs
Total branch-length 0.710 after 0.00 sec
ML-NNI round 1: LogLk = -126.185 NNIs 0 max delta 0.00 Time 0.00
Switched to using 20 rate categories (CAT approximation)
Rate categories were divided by 0.740 so that average rate = 1.0
CAT-based log-likelihoods may not be comparable across runs
Use -gamma for approximate but comparable Gamma(20) log-likelihood
s
ML-NNI round 2: LogLk = -121.100 NNIs 1 max delta 0.00 Time 0.01
Turning off heuristics for final round of ML NNIs (converged)
ML-NNI round 3: LogLk = -121.100 NNIs 1 max delta 0.00 Time 0.01 (
final)
Optimize all lengths: LogLk = -121.100 Time 0.01
Total time: 0.02 seconds Unique: 5/5 Bad splits: 0/2
```

We can now check the newick formatted output Tree obtained above.

```
%%bash
ete3 view --text -t  protein5-fastNJ.nwk
```

```
      /-Cow
     |
--|--Human
     |
     |    /-Chicken
     \-|
          |    /-Carp
          \-|
               \-Loach
```

# 2. Parsimony Tree

***Constructing a Maximum Parsimony Tree***

We are going to use RAxML developed by Alexandros Stamakakis for computing Maximum parsimony trees. There is a tool called Parsimonator: A fast open-source parsimony program

http://sco.h-its.org/exelixis/web/software/parsimonator/index.html (http://sco.h-its.org/exelixis/web/software/parsimonator/index.html)

Put simply, the method chooses (amongst many alternative trees) a tree that maximises the probability of observing the data, given the evolutionary model. The model choosesout of all possible trees, the maximum likelihood tree is the tree that makes the evolution of our sequence alignment the most probable and its expressed in the log probabilities scale.

Parsimony and maximum likelihood analysis calculate a score (say log- likelihood) for many different trees. There is no way to evaluate all possible tree s, unless you're looking at very few taxon. Therefore, we need to employ many clever heuristic methods- essentially starting with Neighbor-Joining trees, and rearranges taxon until the likelihood no longer improves. Maximum likelihood method (and the conceptually related, the Bayesian method) are the most computationally intensive of phylogenetic methods. Theoretically, if the evolutionary model used is realistic this method should give us the best chance of finding the "actual" tree.

We are going to use RAxML as we have precompiled binaries of RAxML for Windows, Linux x86. The linux executables are already on our Virtual machine.

We can invoke making parsimonator tree, which generally an approach used by RAxML to produce candidate Trees.

```
raxml  -y -s alignment_file> -m yourmodel -n  outputname  -p  Randomsee
d
```

and we run as follows:

```
In [20]:
```

```
%%bash
raxml -y -s protein5.phy  -m PROTGAMMAAUTO  -n  protein5-parsim.nwk    -p 123
4
```

This is RAxML version 8.2.9 released by Alexandros Stamatakis on J
uly 20 2016.

With greatly appreciated code contributions by:
Andre Aberer        (HITS)
Simon Berger        (HITS)
Alexey Kozlov       (HITS)
Kassian Kobert      (HITS)
David Dao           (KIT and HITS)
Sarah Lutteropp     (KIT and HITS)
Nick Pattengale     (Sandia)
Wayne Pfeiffer      (SDSC)
Akifumi S. Tanabe   (NRIFS)
Charlie Taylor      (UF)


Alignment has 17 distinct alignment patterns

Proportion of gaps and completely undetermined characters in this
alignment: 0.00%

RAxML rapid hill-climbing mode

Using 1 distinct models/data partitions with joint branch length o
ptimization


Executing 1 inferences on the original alignment using 1 distinct
randomized MP trees

All free model parameters will be estimated by RAxML
GAMMA model of rate heteorgeneity, ML estimate of alpha-parameter

GAMMA Model parameters will be estimated up to an accuracy of 0.10
00000000 Log Likelihood units

Partition: 0
Alignment Patterns: 17
Name: No Name Provided
DataType: AA
Substitution Matrix: AUTO
Using fixed base frequencies


RAxML was called as follows:

raxml -y -s protein5.phy -m PROTGAMMAAUTO -n protein5-parsim.nwk -
p 1234

```
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 17.
Please go and fix your data!

Partition: 0 with name: No Name Provided
Base frequencies: 0.087 0.044 0.039 0.057 0.019 0.037 0.058 0.083
0.024 0.048 0.086 0.062 0.020 0.038 0.046 0.070 0.061 0.014 0.035
0.071
```

In [21]:

```
%%bash
#rm RAxML_info.protein5-parsim.nwk
cp RAxML_parsimonyTree.protein5-parsim.nwk protein5-parsim.nwk
ete3 view --text -t  protein5-parsim.nwk
```

```
   /-Chicken
  |
  |        /-Loach
  |    /-|
--|--|    \-Carp
  |   |
  |    \-Human
  |
   \-Cow
```

# 3. Calculating Maximum Likelihood Trees with RAxML

The simplest method of running RAxML is to copy the binary appropriate directory and we can cd to working directory where you have saved it: Sequences are in PHYLIP format. [ protein5.phy]

To create the Maximum Likelihood tree, type the following:

```
    Windows:    raxml.exe -f d -m PROTGAMMAAUTO -s protein5.phy -n PRT5   -#
    1
    Linux :    raxml    -f  d -m PROTGAMMAAUTO -s protein5.phy -n PRT5   -#
    1
```

```
    The -f option refers to the type of algorithm you want RAxML  to execut
    e.
    In this case we are specifying the default (d) hill-climbing
    algorithm.

    The -m option denoted the substitution model to be used.
    The -s option refers the sequence data file which must be in PHYLIP for
    mat

    The -n option gives the output file prefix

    Finally, the -# option
    Is the number of alternative runs on distinct starting trees.

    The program should take about 2 minutes to create one replicate of the
    maximum likelihood phylogenetic tree over our reduced set of sequences.

    As it completes each tree, it assigns a likelihood score to each.

    At the end, it prints which tree is the most likely to the screen.

    This information is also found at the end of the file RAxML info.*

    Typically RAxML chooses the best tree out of thousands of
    possible replicates.

    The completed maximum likelihood phylogenetic tree is
    found in the  RAxML_bestTree.PRT5
```

We need to estimate program parameters, the program parameter PROTGAMMAAUTO, the optimal models of substitution can be identified.

We will make use of 10 different starting trees for a given alignment using the parsimonator algorithm of RAxML with 10 different random seeds. We will estimate 10 maximum likelihood analyses on different starting trees.

Next, we will conduct the standard bootstraps and performed ML analysis on 100 replicates. Finally, we select the tree with best likelihood score and draw bootstrap supports onto each selected tree.

The newick-formatted trees will be read and processed using the ETE3 library (Huerta-Cepas et al, 2010) and the scripts will be in Python.

# Calculating Bootstrapping Values with RAxML

Bootstrapping is a method to assign confidence scores in your phylogenetic tree.

Bootstrapping involves generating many phylogenetic trees using random resamplings of the input sequences (Baldauf, 2003).

Each phylogenetic tree is then compared with the other trees, and the structural similarities are determined.

If a particular subtree persists across many phylogenetic trees despite many sequence permuta- tions, it is assigned a high bootstrapping score.

For example, if 100 bootstrapping trees are computed, and a particular subtree is found in 80 of those trees, that subtree will be assigned a bootstrapping value of 80, indicating 80% of trees contained that subtree. Typically, a bootstrapping value above 70 is required for confidence that a particular subtree accurately represents the phylogeny of the underlying organisms.

The next step is to compute the bootstrapping values for the best likelihood tree. You need to provide RAxML with the original alignment in PHYLIP format and the best likelihood tree from the previous step.

Maximum parsimony using RAxML and then doing ML searches on them and assigning bootstrap values.

# Make ML tree from 10 runs starting with a different parsimony tree

***making 5 ML trees each starting from a different parsimonious tree will make max 10 runs on each***

```
raxmlHPC -f d  -m PROTGAMMAJTT -s protein5.phy -N 10  -n ML1 -p 1234
raxmlHPC -f d  -m PROTGAMMAJTT -s protein5.phy -N 10  -n ML2 -p 2565
raxmlHPC -f d  -m PROTGAMMAJTT -s protein5.phy -N 10  -n ML3 -p 3445
raxmlHPC -f d  -m PROTGAMMAJTT -s protein5.phy -N 10  -n ML4 -p 4565
raxmlHPC -f d  -m PROTGAMMAJTT -s protein5.phy -N 10  -n ML5 -p 5676
```

We run RAxML:

In [22]:

```
%%bash
raxml -f d  -m PROTGAMMAJTT -s protein5.phy -N 10  -n ML1 -p 1234
```

This is RAxML version 8.2.9 released by Alexandros Stamatakis on July 20 2016.

With greatly appreciated code contributions by:
Andre Aberer        (HITS)
Simon Berger        (HITS)
Alexey Kozlov       (HITS)

Kassian Kobert      (HITS)
David Dao          (KIT and HITS)
Sarah Lutteropp    (KIT and HITS)
Nick Pattengale    (Sandia)
Wayne Pfeiffer     (SDSC)
Akifumi S. Tanabe  (NRIFS)
Charlie Taylor     (UF)

Alignment has 17 distinct alignment patterns

Proportion of gaps and completely undetermined characters in this alignment: 0.00%

RAxML rapid hill-climbing mode

Using 1 distinct models/data partitions with joint branch length optimization


Executing 10 inferences on the original alignment using 10 distinct randomized MP trees

All free model parameters will be estimated by RAxML
GAMMA model of rate heteorgeneity, ML estimate of alpha-parameter

GAMMA Model parameters will be estimated up to an accuracy of 0.1000000000 Log Likelihood units

Partition: 0
Alignment Patterns: 17
Name: No Name Provided
DataType: AA
Substitution Matrix: JTT
Using fixed base frequencies



RAxML was called as follows:

raxml -f d -m PROTGAMMAJTT -s protein5.phy -N 10 -n ML1 -p 1234


Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 17. Please go and fix your data!

Partition: 0 with name: No Name Provided
Base frequencies: 0.077 0.052 0.043 0.052 0.020 0.041 0.062 0.073 0.023 0.054 0.092 0.059 0.024 0.040 0.051 0.069 0.059 0.014 0.032 0.066

Inference[0]: Time 0.147421 GAMMA-based likelihood -125.805980, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 17.

```
Please go and fix your data!

Inference[1]: Time 0.127145 GAMMA-based likelihood -125.805980, be
st rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 17.
Please go and fix your data!

Inference[2]: Time 0.119402 GAMMA-based likelihood -125.805972, be
st rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 17.
Please go and fix your data!

Inference[3]: Time 0.119506 GAMMA-based likelihood -125.805980, be
st rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 17.
Please go and fix your data!

Inference[4]: Time 0.133627 GAMMA-based likelihood -125.805980, be
st rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 17.
Please go and fix your data!

Inference[5]: Time 0.139968 GAMMA-based likelihood -125.805980, be
st rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 17.
Please go and fix your data!

Inference[6]: Time 0.124596 GAMMA-based likelihood -125.805972, be
st rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 17.
Please go and fix your data!

Inference[7]: Time 0.119041 GAMMA-based likelihood -125.805980, be
st rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 17.
Please go and fix your data!

Inference[8]: Time 0.131916 GAMMA-based likelihood -125.805980, be
st rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 17.
Please go and fix your data!

Inference[9]: Time 0.135800 GAMMA-based likelihood -125.805980, be
st rearrangement setting 2


Conducting final model optimizations on all 10 trees under GAMMA-b
ased models ....
```

```
Inference[0] final GAMMA-based Likelihood: -125.805980 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.0
Inference[1] final GAMMA-based Likelihood: -125.805980 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.1
Inference[2] final GAMMA-based Likelihood: -125.805972 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.2
Inference[3] final GAMMA-based Likelihood: -125.805980 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.3
Inference[4] final GAMMA-based Likelihood: -125.805980 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.4
Inference[5] final GAMMA-based Likelihood: -125.805980 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.5
Inference[6] final GAMMA-based Likelihood: -125.805972 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.6
Inference[7] final GAMMA-based Likelihood: -125.805980 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.7
Inference[8] final GAMMA-based Likelihood: -125.805980 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.8
Inference[9] final GAMMA-based Likelihood: -125.805980 tree writte
n to file /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/R
AxML_result.ML1.RUN.9


Starting final GAMMA-based thorough Optimization on tree 2 likelih
ood -125.805972 ....

Final GAMMA-based Score of best tree -125.805972

Program execution info written to /Users/cheemaj/JITU/TRAINING/PHY
LO-COURSE/trash-master/RAxML_info.ML1
Best-scoring ML tree written to: /Users/cheemaj/JITU/TRAINING/PHYL
O-COURSE/trash-master/RAxML_bestTree.ML1

Overall execution time: 1.362516 secs or 0.000378 hours or 0.00001
6 days
```

```
In [23]:
```

```bash
%%bash
#rm RAxML_info.protein5-parsim.nwk
#cp RAxML_bestTree.ML1 protein5-parsim.nwk
ete3 view --text -t RAxML_bestTree.ML1
```

```
   /-Human
  |
  |      /-Loach
  |   /-|
--|--|   \-Carp
  |  |
  |   \-Chicken
  |
   \-Cow
```

# Drawing Boot strap trees

- Make 100 boot trees

```
raxmlHPC  -f d -m PROTGAMMAJTT -s protein5.phy -b 12463  -p 54439 -n BO
OT100  -N 100
```

```
</pre>
```
The 100 bootstrapped trees will written to as a file called: RAxML_boot
strap.BOOT100

```
In [25]:
```

```bash
%%bash
raxml  -f d  -m PROTGAMMAJTT -s protein5.phy -b 12463  -p 54439 -n BOOT100   -N
100
```

```
This is RAxML version 8.2.9 released by Alexandros Stamatakis on J
uly 20 2016.

With greatly appreciated code contributions by:
Andre Aberer      (HITS)
Simon Berger      (HITS)
Alexey Kozlov     (HITS)
Kassian Kobert    (HITS)
David Dao         (KIT and HITS)
Sarah Lutteropp   (KIT and HITS)
Nick Pattengale   (Sandia)
Wayne Pfeiffer    (SDSC)
Akifumi S. Tanabe (NRIFS)
Charlie Taylor    (UF)
```

Alignment has 17 distinct alignment patterns

Proportion of gaps and completely undetermined characters in this alignment: 0.00%

RAxML rapid hill-climbing mode

Using 1 distinct models/data partitions with joint branch length optimization

Executing 100 non-parametric bootstrap inferences

All free model parameters will be estimated by RAxML
GAMMA model of rate heteorgeneity, ML estimate of alpha-parameter

GAMMA Model parameters will be estimated up to an accuracy of 0.1000000000 Log Likelihood units

Partition: 0
Alignment Patterns: 17
Name: No Name Provided
DataType: AA
Substitution Matrix: JTT
Using fixed base frequencies

RAxML was called as follows:

raxml -f d -m PROTGAMMAJTT -s protein5.phy -b 12463 -p 54439 -n BOOT100 -N 100

Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 15. Please go and fix your data!

Bootstrap[0]: Time 0.114746 seconds, bootstrap likelihood -122.427442, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 12. Please go and fix your data!

Bootstrap[1]: Time 0.102668 seconds, bootstrap likelihood -88.825481, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 14. Please go and fix your data!

Bootstrap[2]: Time 0.112220 seconds, bootstrap likelihood -139.016810, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 15. Please go and fix your data!

Bootstrap[3]: Time 0.136589 seconds, bootstrap likelihood -100.406
280, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[4]: Time 0.118063 seconds, bootstrap likelihood -133.174
744, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[5]: Time 0.099887 seconds, bootstrap likelihood -158.994
132, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[6]: Time 0.116077 seconds, bootstrap likelihood -115.772
185, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 11.
Please go and fix your data!

Bootstrap[7]: Time 0.106212 seconds, bootstrap likelihood -152.590
500, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[8]: Time 0.098950 seconds, bootstrap likelihood -157.510
939, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 12.
Please go and fix your data!

Bootstrap[9]: Time 0.060278 seconds, bootstrap likelihood -116.058
068, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[10]: Time 0.098081 seconds, bootstrap likelihood -144.90
5457, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[11]: Time 0.121644 seconds, bootstrap likelihood -151.01
0180, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[12]: Time 0.134361 seconds, bootstrap likelihood -124.48
4909, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e

xpected states is 20 the number of states that are present is 16.
Please go and fix your data!

Bootstrap[13]: Time 0.082261 seconds, bootstrap likelihood -119.34
1614, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[14]: Time 0.070988 seconds, bootstrap likelihood -124.01
3108, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[15]: Time 0.177186 seconds, bootstrap likelihood -140.81
3869, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 11.
Please go and fix your data!

Bootstrap[16]: Time 0.181429 seconds, bootstrap likelihood -128.92
2183, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 9.
Please go and fix your data!

Bootstrap[17]: Time 0.121953 seconds, bootstrap likelihood -115.78
1552, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[18]: Time 0.073415 seconds, bootstrap likelihood -114.37
4592, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[19]: Time 0.113018 seconds, bootstrap likelihood -138.74
8460, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 8.
Please go and fix your data!

Bootstrap[20]: Time 0.132701 seconds, bootstrap likelihood -122.53
6286, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 11.
Please go and fix your data!

Bootstrap[21]: Time 0.061059 seconds, bootstrap likelihood -118.23
9196, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 12.
Please go and fix your data!

Bootstrap[22]: Time 0.078415 seconds, bootstrap likelihood -113.50
2903, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 12.
Please go and fix your data!

Bootstrap[23]: Time 0.151907 seconds, bootstrap likelihood -134.62
0387, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[24]: Time 0.108813 seconds, bootstrap likelihood -132.35
6029, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[25]: Time 0.087210 seconds, bootstrap likelihood -129.37
8426, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 16.
Please go and fix your data!

Bootstrap[26]: Time 0.092704 seconds, bootstrap likelihood -132.04
8046, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[27]: Time 0.128348 seconds, bootstrap likelihood -134.23
1787, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[28]: Time 0.144596 seconds, bootstrap likelihood -122.34
0146, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[29]: Time 0.151506 seconds, bootstrap likelihood -134.87
5697, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[30]: Time 0.100777 seconds, bootstrap likelihood -121.29
2033, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[31]: Time 0.087781 seconds, bootstrap likelihood -129.15
4042, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e

xpected states is 20 the number of states that are present is 11.
Please go and fix your data!

Bootstrap[32]: Time 0.133067 seconds, bootstrap likelihood -152.85
0944, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 11.
Please go and fix your data!

Bootstrap[33]: Time 0.134766 seconds, bootstrap likelihood -104.29
3444, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 9.
Please go and fix your data!

Bootstrap[34]: Time 0.061840 seconds, bootstrap likelihood -76.591
240, best rearrangement setting 5
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[35]: Time 0.075916 seconds, bootstrap likelihood -126.66
1474, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[36]: Time 0.107185 seconds, bootstrap likelihood -118.52
2428, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[37]: Time 0.083819 seconds, bootstrap likelihood -121.64
9447, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[38]: Time 0.094840 seconds, bootstrap likelihood -122.44
2610, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[39]: Time 0.087963 seconds, bootstrap likelihood -127.89
5735, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[40]: Time 0.148154 seconds, bootstrap likelihood -128.29
5156, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[41]: Time 0.106394 seconds, bootstrap likelihood -117.92
8230, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[42]: Time 0.133171 seconds, bootstrap likelihood -137.56
0911, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[43]: Time 0.136777 seconds, bootstrap likelihood -127.56
2913, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[44]: Time 0.113637 seconds, bootstrap likelihood -126.76
1501, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 16.
Please go and fix your data!

Bootstrap[45]: Time 0.095410 seconds, bootstrap likelihood -151.26
4811, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 16.
Please go and fix your data!

Bootstrap[46]: Time 0.113206 seconds, bootstrap likelihood -122.69
4888, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 12.
Please go and fix your data!

Bootstrap[47]: Time 0.128772 seconds, bootstrap likelihood -106.53
9963, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[48]: Time 0.100348 seconds, bootstrap likelihood -100.13
9292, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[49]: Time 0.114830 seconds, bootstrap likelihood -127.63
3564, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 16.
Please go and fix your data!

Bootstrap[50]: Time 0.078816 seconds, bootstrap likelihood -101.38
9912, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e

xpected states is 20 the number of states that are present is 15. Please go and fix your data!

Bootstrap[51]: Time 0.099056 seconds, bootstrap likelihood -140.447162, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 11. Please go and fix your data!

Bootstrap[52]: Time 0.080924 seconds, bootstrap likelihood -95.040452, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 15. Please go and fix your data!

Bootstrap[53]: Time 0.141984 seconds, bootstrap likelihood -130.615658, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 13. Please go and fix your data!

Bootstrap[54]: Time 0.083540 seconds, bootstrap likelihood -101.430130, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 10. Please go and fix your data!

Bootstrap[55]: Time 0.078212 seconds, bootstrap likelihood -144.086497, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 15. Please go and fix your data!

Bootstrap[56]: Time 0.088188 seconds, bootstrap likelihood -98.185769, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 12. Please go and fix your data!

Bootstrap[57]: Time 0.094437 seconds, bootstrap likelihood -135.700743, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 16. Please go and fix your data!

Bootstrap[58]: Time 0.074826 seconds, bootstrap likelihood -120.282833, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 14. Please go and fix your data!

Bootstrap[59]: Time 0.087562 seconds, bootstrap likelihood -118.052950, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of expected states is 20 the number of states that are present is 13. Please go and fix your data!

Bootstrap[60]: Time 0.178525 seconds, bootstrap likelihood -106.79
1223, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[61]: Time 0.087584 seconds, bootstrap likelihood -103.36
6675, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 11.
Please go and fix your data!

Bootstrap[62]: Time 0.079264 seconds, bootstrap likelihood -133.56
8005, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[63]: Time 0.094917 seconds, bootstrap likelihood -118.24
1972, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[64]: Time 0.095332 seconds, bootstrap likelihood -132.31
6185, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 16.
Please go and fix your data!

Bootstrap[65]: Time 0.079450 seconds, bootstrap likelihood -116.89
5079, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[66]: Time 0.076733 seconds, bootstrap likelihood -131.56
8642, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[67]: Time 0.095902 seconds, bootstrap likelihood -121.54
3008, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[68]: Time 0.104829 seconds, bootstrap likelihood -114.13
0281, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 11.
Please go and fix your data!

Bootstrap[69]: Time 0.115519 seconds, bootstrap likelihood -122.59
0234, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e

xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[70]: Time 0.118293 seconds, bootstrap likelihood -120.31
5707, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[71]: Time 0.091283 seconds, bootstrap likelihood -116.51
8573, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[72]: Time 0.097533 seconds, bootstrap likelihood -119.88
3596, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[73]: Time 0.090947 seconds, bootstrap likelihood -121.49
0424, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[74]: Time 0.088067 seconds, bootstrap likelihood -113.81
2121, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[75]: Time 0.096868 seconds, bootstrap likelihood -99.179
777, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 12.
Please go and fix your data!

Bootstrap[76]: Time 0.094770 seconds, bootstrap likelihood -105.60
3277, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[77]: Time 0.108800 seconds, bootstrap likelihood -127.02
8113, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[78]: Time 0.111159 seconds, bootstrap likelihood -123.51
6486, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[79]: Time 0.111965 seconds, bootstrap likelihood -146.96
2709, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[80]: Time 0.126618 seconds, bootstrap likelihood -131.02
6713, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 12.
Please go and fix your data!

Bootstrap[81]: Time 0.144272 seconds, bootstrap likelihood -116.74
7166, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[82]: Time 0.092222 seconds, bootstrap likelihood -159.36
2390, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 11.
Please go and fix your data!

Bootstrap[83]: Time 0.108294 seconds, bootstrap likelihood -138.73
6109, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 11.
Please go and fix your data!

Bootstrap[84]: Time 0.105900 seconds, bootstrap likelihood -102.43
0754, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[85]: Time 0.083673 seconds, bootstrap likelihood -100.98
8701, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[86]: Time 0.108145 seconds, bootstrap likelihood -127.49
0109, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 16.
Please go and fix your data!

Bootstrap[87]: Time 0.110267 seconds, bootstrap likelihood -136.91
4301, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 12.
Please go and fix your data!

Bootstrap[88]: Time 0.081574 seconds, bootstrap likelihood -83.408
196, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e

xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[89]: Time 0.136964 seconds, bootstrap likelihood -123.94
1687, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 13.
Please go and fix your data!

Bootstrap[90]: Time 0.103733 seconds, bootstrap likelihood -132.85
7260, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 10.
Please go and fix your data!

Bootstrap[91]: Time 0.127950 seconds, bootstrap likelihood -160.09
5056, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[92]: Time 0.085665 seconds, bootstrap likelihood -128.89
2027, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 15.
Please go and fix your data!

Bootstrap[93]: Time 0.079575 seconds, bootstrap likelihood -118.10
9008, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[94]: Time 0.099489 seconds, bootstrap likelihood -139.28
5561, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[95]: Time 0.093827 seconds, bootstrap likelihood -143.66
7259, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 14.
Please go and fix your data!

Bootstrap[96]: Time 0.102159 seconds, bootstrap likelihood -103.70
6500, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 12.
Please go and fix your data!

Bootstrap[97]: Time 0.083073 seconds, bootstrap likelihood -113.10
3594, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 16.
Please go and fix your data!

```
Bootstrap[98]: Time 0.094948 seconds, bootstrap likelihood -129.15
6631, best rearrangement setting 2
Partition No Name Provided number 0 has a problem, the number of e
xpected states is 20 the number of states that are present is 12.
Please go and fix your data!

Bootstrap[99]: Time 0.079589 seconds, bootstrap likelihood -111.94
4738, best rearrangement setting 2


Overall Time for 100 Bootstraps 10.515204


Average Time per Bootstrap 0.105152
All 100 bootstrapped trees written to: /Users/cheemaj/JITU/TRAININ
G/PHYLO-COURSE/trash-master/RAxML_bootstrap.BOOT100
```

**Drawing the boostrap values on the best of ML tree: reconcile or draw using option -z**

```
raxml  -m PROTGAMMAJTT -f b   -t RAxML_bestTree.ML1   -z  BOOT100.tre
-n  finalML1_boot100
```

In [26]:

```bash
%%bash
raxml  -m PROTGAMMAJTT -f b   -t RAxML_bestTree.ML1   -z  RAxML_bootstrap.BOOT
100 -nfinalML1_boot100
```

```
Found a total of 5 taxa in first tree of tree collection RAxML_boo
tstrap.BOOT100
Expecting all remaining trees in collection to have the same taxon
set
add [Chicken]
add [Human]
add [Loach]
add [Carp]
add [Cow]


WARNING: RAxML is not checking sequences for duplicate seqs and si
tes with missing data!



This is RAxML version 8.2.9 released by Alexandros Stamatakis on J
uly 20 2016.

With greatly appreciated code contributions by:
Andre Aberer        (HITS)
Simon Berger        (HITS)
Alexey Kozlov       (HITS)
Kassian Kobert      (HITS)
David Dao          (KIT and HITS)
Sarah Lutteropp     (KIT and HITS)
```

```
Nick Pattengale    (Sandia)
Wayne Pfeiffer     (SDSC)
Akifumi S. Tanabe  (NRIFS)
Charlie Taylor     (UF)
```

```
RAxML Bipartition Computation: Drawing support values from trees i
n file RAxML_bootstrap.BOOT100 onto tree in file RAxML_bestTree.ML
1
```

```
RAxML was called as follows:

raxml -m PROTGAMMAJTT -f b -t RAxML_bestTree.ML1 -z RAxML_bootstra
p.BOOT100 -nfinalML1_boot100
```

```
Found 1 tree in File RAxML_bestTree.ML1
```

```
Found 100 trees in File RAxML_bootstrap.BOOT100
```

```
Found 1 tree in File RAxML_bestTree.ML1
```

```
Time for Computation of Bipartitions 0.011846
Tree with bipartitions written to file:  /Users/cheemaj/JITU/TRAIN
ING/PHYLO-COURSE/trash-master/RAxML_bipartitions.finalML1_boot100
Tree with bipartitions as branch labels written to file:  /Users/c
heemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/RAxML_bipartitionsB
ranchLabels.finalML1_boot100
Execution information file written to :  /Users/cheemaj/JITU/TRAIN
ING/PHYLO-COURSE/trash-master/RAxML_info.finalML1_boot100
```

Viewing the Boostrapped drawn *RAxML_bipartitionsBranchLabels.finalML1_boot100* tree on the best Maximum likelihood tree:

In [27]:

```bash
%%bash
ete3 view  --raxml --text -t   RAxML_bipartitionsBranchLabels.finalML1_boot100
ete3 view   --raxml --image booted.png --show_support  -t   RAxML_bipartitions
BranchLabels.finalML1_boot100
```
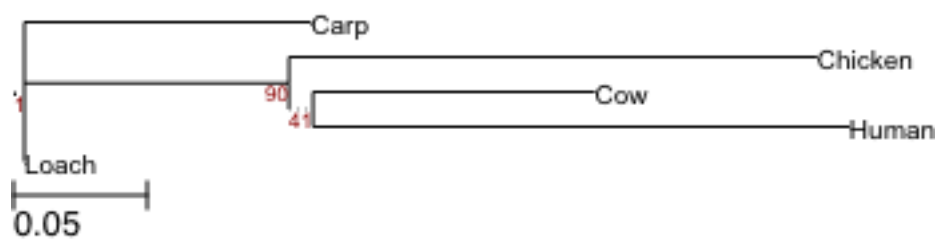
```
    /-Carp
   |
   |     /-Chicken
   |--|
--|  |     /-Cow
   |   \-|
   |      \-Human
   |
    \-Loach
```

In [28]:

```python
from IPython.display import Image
Image(filename='booted.png')
```

Out[28]:

# Tree manipulations

We will use a Python framework for the analysis and visualization of trees called ete toolkit. This is capable of loading, creating, traversing, search, prune, or modify hierarchical tree structures with ease using the ETE Python API. One of the rich feature it has is to make custom node(taxa) attributes that can be rendered as graphical elements. Choose among external images, charts, symbols, text labels, and to finally make publication quality trees.

- (http://etetoolkit.org/ (http://etetoolkit.org/))
- (http://etetoolkit.org/documentation/tools/ (http://etetoolkit.org/documentation/tools/))

# Reference and Acknowledgements

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 26882690 doi:10.1093/bioinformatics/btl446. Stamatakis A.( 2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312-3. http://dx.doi.org/10.1093/bioinformatics/btu033 (http://dx.doi.org/10.1093/bioinformatics/btu033) Huerta-Cepas J et al(2010) ETE: a python Environment for Tree Exploration. BMC Bioinformatics 11: 24. http://rogerlab.biochemistryandmolecularbiology.dal.ca/lab3_2011.html (http://rogerlab.biochemistryandmolecularbiology.dal.ca/lab3_2011.html) [Some of the material above is from this valuable resource please see details]

</pre>

# Further Exercises

We now try an additional exercise but this time using the Clustal Omega, Fasttree and ete3. This is from the ete3. We start with NUP62 homologues sequences. Let see first few lines:

In [29]:

```bash
%%bash
cat NUP62.fa | head -n 4
```

```
>Phy003I7ZJ_CHICK
TMSQFNFSSAPAGGGFSFSTPKTAASTTAATGFSFTPAPSSGFTFGGAAPTPASSQPVTP
FSFSTPASSALPTAFSFGTPATATTAAPAASVFPLGGNAPKLNFGGTSTTQATGITGGFG
FGTSAPTSVPSSQAAAPSGFMFGTAATTTTTTTAAQPGTTGGFTFSSGTTTQAGTTGFNI
```

***We pick a workflow name from the list of many from ete3 website and run it:***

In [30]:

```bash
%%bash
ete3 build workflows genetree
```

----------------------------------------------------------------------

```
---------------------------------------------------
                                      GeneTree Workf
lows
---------------------------------------------------
---------------------------------------------------
Worflow name                         | Description
| threads
===================================== | ================
===================================== | =========
phylomedb4                           | Worflow described
in Huerta-Cepas, et al (Nucleic Acid Resea | 1
                                     | rch, 2013) and use
d in http://PhylomeDB.org            |
------------------------------------- | ------------------
------------------------------------- | ----------
eggnog41                             | Workflow used in t
he EggNOG orthology database and based on | 1
                                     | the phylomedb4 wor
kflow                                |
------------------------------------- | ------------------
------------------------------------- | ----------
ensembl_small                        | Experimental (do n
ot use yet)                          | 1
------------------------------------- | ------------------
------------------------------------- | ----------
ensembl_large                        | Experimental (do n
ot use yet)                          | 1
------------------------------------- | ------------------
------------------------------------- | ----------
full_modeltest                       | Complete (and slow
) pipeline testing all evolutionary models | 1
                                     |  and inferring tre
e with Phyml                         |
------------------------------------- | ------------------
------------------------------------- | ----------
full_modeltest_bootstrap             | Complete (and slow
) pipeline testing all evolutionary models | 1
                                     |  and inferring tre
e with Phyml and 100 bootstraps      |
------------------------------------- | ------------------
------------------------------------- | ----------
full_fast_modeltest                  | Test all evolution
ary models using NJ inference and computes | 1
                                     |  final tree with P
hyml                                 |
------------------------------------- | ------------------
------------------------------------- | ----------
full_fast_modeltest_bootstrap        | Test all evolution
ary models using NJ inference and computes | 1
                                     |  final tree with P
hyml and 100 bootstraps              |
------------------------------------- | ------------------
------------------------------------- | ----------
full_ultrafast_modeltest             | Test all evolution
ary models using NJ inference with no F, I | 1
                                     |  and G estimation
and computes final tree with Phyml   |
```

| | | |
|---|---|---|
| full_ultrafast_modeltest ary models using NJ inference with no F, I and computes final tree with Phyml and 100 | 1 | Test all evolution and G estimation bootstraps |
| soft_modeltest ) pipeline testing 5 protein evolutionary ng tree with Phyml | 1 | Complete (and slow models and inferri |
| soft_modeltest_bootstrap ) pipeline testing 5 protein evolutionary ng tree with Phyml and 100 bootstraps | 1 | Complete (and slow models and inferri |
| soft_fast_modeltest lutionary models using NJ inference and co with Phyml | 1 | Test 5 protein evo mputes final tree |
| soft_fast_modeltest_bootstrap lutionary models using NJ inference and co with Phyml and 100 bootstraps | 1 | Test 5 protein evo mputes final tree |
| soft_ultrafast_modeltest lutionary models using NJ inference with n ation and computes final tree with Phyml | 1 | Test 5 protein evo o F, I and G estim |
| soft_ultrafast_modeltest_bootstrap lutionary models using NJ inference with n ation and computes final tree with Phyml a | 1 | Test 5 protein evo o F, I and G estim nd 100 bootstraps |
| standard_raxml ing ClustalOmega with default parameters a L | 1 | Computes a tree us s aligner and RAxM |
| standard_raxml_bootstrap ing ClustalOmega with default parameters a | 1 | Computes a tree us s alignerand RAxML |

| Pipeline | Description | |
| --- | --- | --- |
| | with 100 bootstraps | |
| standard_phyml | Computes a tree using ClustalOmega with default parameters as aligner and Phyml | 1 |
| standard_phyml_bootstrap | Computes a tree using ClustalOmega with default parameters as aligner and Phyml and 100 bootstraps | 1 |
| standard_fasttree | Computes a tree using ClustalOmega with default parameters as aligner and FastTree | 1 |
| standard_trimmed_raxml | Computes a tree using default ClustalOmega as aligner, trimAl for alg. cleaning and RAxML | 1 |
| standard_trimmed_raxml_bootstrap | Computes a tree using ClustalOmega with default parameters, trimAl for alg. cleaning and RAxML with 100 bootstraps | 1 |
| standard_trimmed_phyml | Computes a tree using ClustalOmega with default parameters, trimAl for alg. cleaning as aligner and Phyml | 1 |
| standard_trimmed_phyml_bootstrap | Computes a tree using ClustalOmega with default parameters, trimAl for alg. cleaning as aligner and Phyml and 100 bootstraps | 1 |
| standard_trimmed_fasttree | Computes a tree using ClustalOmega with default parameters, trimAl for alg. cleaning as aligner and FastTree | 1 |

Only three parameters are required:

- Use `-a` to provide an amino acid sequence file (or `-n` for nucleotides)
- `-o` to define the output directory (should not exist, otherwise `--resume` or `--clearall` flags will be necessary)
- `-w` to set the name of the workflow to be executed. For this example we will use the `standard_fasttree` workflow.

In [31]:

```bash
%%bash
ete3 build -w standard_fasttree -a NUP62.fa -o NUP62_tree/ --clearall
```

```
Toolchain path: /Users/cheemaj/anaconda/bin/ete3_apps
Toolchain version: 2.0.3


      ------------------------------------------------------------
--------------------
                    ETE build (3.0.0b35) - reproducible phylogenetic
workflows

      Citation:

       Huerta-Cepas J, Serra F and Bork P. ETE 3: Reconstruction,
analysis and
        visualization of phylogenomic data. Mol Biol Evol (2016)
        doi:10.1093/molbev/msw046

      (Note that a list of the external programs used to complete
all necessary
        computations will be shown after workflow execution. Those p
rograms should
        also be cited.)


      ------------------------------------------------------------
--------------------

INFO -   Testing x86-64  portable applications...
        clustalo: OK - 1.2.1
Dialign-tx not supported in OS X
        fasttree: OK - FastTree Version 2.1.8 Double precision (No
SSE3), OpenMP (1 threads)
         kalign: OK - Kalign version 2.04, Copyright (C) 2004, 200
5, 2006 Timo Lassmann
          mafft: OK - MAFFT v6.861b (2011/09/24)
         muscle: OK - MUSCLE v3.8.31 by Robert C. Edgar
          phyml: OK - . This is PhyML version 20160115.
      pmodeltest: OK - pmodeltest.py v1.4
          prank: OK - prank v.100802. Minimal usage: 'prank sequen
ce_file'
        probcons: OK - PROBCONS version 1.12 - align multiple prote
in sequences and print to standard output
          raxml: OK - This is RAxML version 8.1.20 released by Ale
xandros Stamatakis on April 18 2015.
```

```
 raxml-pthreads: OK - This is RAxML version 8.1.20 released by Ale
xandros Stamatakis on April 18 2015.
        readal: OK - readAl v1.4.rev6 build[2012-02-02]
        statal: OK - statAl v1.4.rev6 build[2012-02-02]
       tcoffee: OK - PROGRAM: T-COFFEE Version_11.00.8cbe486 (201
4-08-12 22:05:29 - Revision 8cbe486 - Build 477)
        trimal: OK - trimAl v1.4.rev6 build[2012-02-02]
INFO -  Starting ETE-build execution at Mon Feb  6 23:28:38 2017
INFO -  Output directory /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE
/trash-master/NUP62_tree
INFO -  Erasing all existing npr data...
INFO -  Reading aa sequences from NUP62.fa...


WRNG -  25 target sequences
INFO -  ETE build starts now!
INFO -    Updating tasks status: (Mon Feb  6 23:28:38 2017)
INFO -  Thread clustalo_default-none-none-fasttree_full: pending t
asks: 1 of sizes: 25
INFO -    (W) MultiSeqTask (25 aa seqs, MSF, /clustalo_d...ttree_fu
ll)
INFO -    (D) MultiSeqTask (25 aa seqs, MSF, /clustalo_d...ttree_fu
ll)
INFO -  Waiting 2 seconds
INFO -    Updating tasks status: (Mon Feb  6 23:28:40 2017)
INFO -  Thread clustalo_default-none-none-fasttree_full: pending t
asks: 1 of sizes: 25
INFO -    (W) AlgTask (25 aa seqs, Clustal-Omega, /clustalo_d...ttr
ee_full)
INFO -  Waiting 2 seconds
INFO -    Updating tasks status: (Mon Feb  6 23:28:42 2017)
INFO -  Thread clustalo_default-none-none-fasttree_full: pending t
asks: 1 of sizes: 25
INFO -    (W) AlgTask (25 aa seqs, Clustal-Omega, /clustalo_d...ttr
ee_full)
INFO -    (D) AlgTask (25 aa seqs, Clustal-Omega, /clustalo_d...ttr
ee_full)
INFO -  Waiting 2 seconds
INFO -    Updating tasks status: (Mon Feb  6 23:28:44 2017)
INFO -  Thread clustalo_default-none-none-fasttree_full: pending t
asks: 1 of sizes: 25
INFO -    (W) TreeTask (25 aa seqs, FastTree, /clustalo_d...ttree_f
ull)
INFO -  Waiting 2 seconds
INFO -    Updating tasks status: (Mon Feb  6 23:28:46 2017)
INFO -  Thread clustalo_default-none-none-fasttree_full: pending t
asks: 1 of sizes: 25
INFO -    (W) TreeTask (25 aa seqs, FastTree, /clustalo_d...ttree_f
ull)
INFO -  Waiting 2 seconds
INFO -    Updating tasks status: (Mon Feb  6 23:28:48 2017)
INFO -  Thread clustalo_default-none-none-fasttree_full: pending t
asks: 1 of sizes: 25
INFO -    (R) TreeTask (25 aa seqs, FastTree, /clustalo_d...ttree_f
ull)
INFO -  Waiting 2 seconds
INFO -    Updating tasks status: (Mon Feb  6 23:28:50 2017)
```

```
INFO -   Thread clustalo_default-none-none-fasttree_full: pending t
asks: 1 of sizes: 25
INFO -     (R) TreeTask (25 aa seqs, FastTree, /clustalo_d...ttree_f
ull)
INFO -     (D) TreeTask (25 aa seqs, FastTree, /clustalo_d...ttree_f
ull)
INFO -   Waiting 2 seconds
INFO -    Updating tasks status: (Mon Feb  6 23:28:52 2017)
INFO -   Thread clustalo_default-none-none-fasttree_full: pending t
asks: 1 of sizes: 25
INFO -     (W) TreeMergeTask (25 aa seqs, TreeMerger, /clustalo_d...
ttree_full)
INFO -     (D) TreeMergeTask (25 aa seqs, TreeMerger, /clustalo_d...
ttree_full)
INFO -   Waiting 2 seconds
INFO -   Assembling final tree...
INFO -   Done thread clustalo_default-none-none-fasttree_full in 1
iteration(s)
INFO -   Writing final tree for clustalo_default-none-none-fasttree
_full
   /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/NUP62_tr
ee/clustalo_default-none-none-fasttree_full/NUP62.fa.final_tree.nw
   /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/NUP62_tr
ee/clustalo_default-none-none-fasttree_full/NUP62.fa.final_tree.nw
x (newick extended)
INFO -   Writing final tree alignment clustalo_default-none-none-fa
sttree_full
   /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/NUP62_tr
ee/clustalo_default-none-none-fasttree_full/NUP62.fa.final_tree.us
ed_alg.fa
INFO -   Writing root node alignment clustalo_default-none-none-fas
ttree_full
   /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/NUP62_tr
ee/clustalo_default-none-none-fasttree_full/NUP62.fa.final_tree.fa
INFO -   Generating tree image for clustalo_default-none-none-fastt
ree_full
   /Users/cheemaj/JITU/TRAINING/PHYLO-COURSE/trash-master/NUP62_tr
ee/clustalo_default-none-none-fasttree_full/NUP62.fa.final_tree.pn
g
INFO -   Done
INFO -   Deleting temporal data...


====================================================================
======
        The following published software and/or methods were used
.
            *** Please, do not forget to cite them! ***

====================================================================
======
   Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez
R,
      McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. F
ast,
      scalable generation of high-quality protein multiple sequenc
e
      alignments using Clustal Omega. Mol Syst Biol. 2011 Oct 11;7
```

:539.
     doi: 10.1038/msb.2011.75.
  Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, analysi
s and
     visualization of phylogenomic data. Mol Biol Evol (2016) doi
:
     10.1093/molbev/msw046
  Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximu
m-
     likelihood trees for large alignments. PLoS One. 2010 Mar
     10;5(3):e9490.

After a few minutes you should get the process done and get a list of generated files and references to the software used. All results will be stored in the provided output directory, whose structure is the following:

In [35]:

```
%%bash
ls NUP62_tree/
```

```
clustalo_default-none-none-fasttree_full
db
ete_build.cfg
tasks
```

- the `ete_build.cfg` is a copy of the configuration file used (including all workflow options, etc.)
- `db/` and `tasks/` are temporary directories used to run different processes. They are used in case you want to resume an analysis or if you need to debug any issue. Every job (i.e. FastTree, ClustalO, etc. will generate one or more directories in `tasks`)
- **your results** are stored in the folder matching the raw name of the workflow chosen. `standard_fastree` translates into a workflow using clustal Omega and Fasttree, and skipping model testing and alignment trimming, so the result dir is `clustalo_default-none-none-fasttree_full`

Note that a full path pointer to the final tree and alignment is also printed when ete3 finishes.
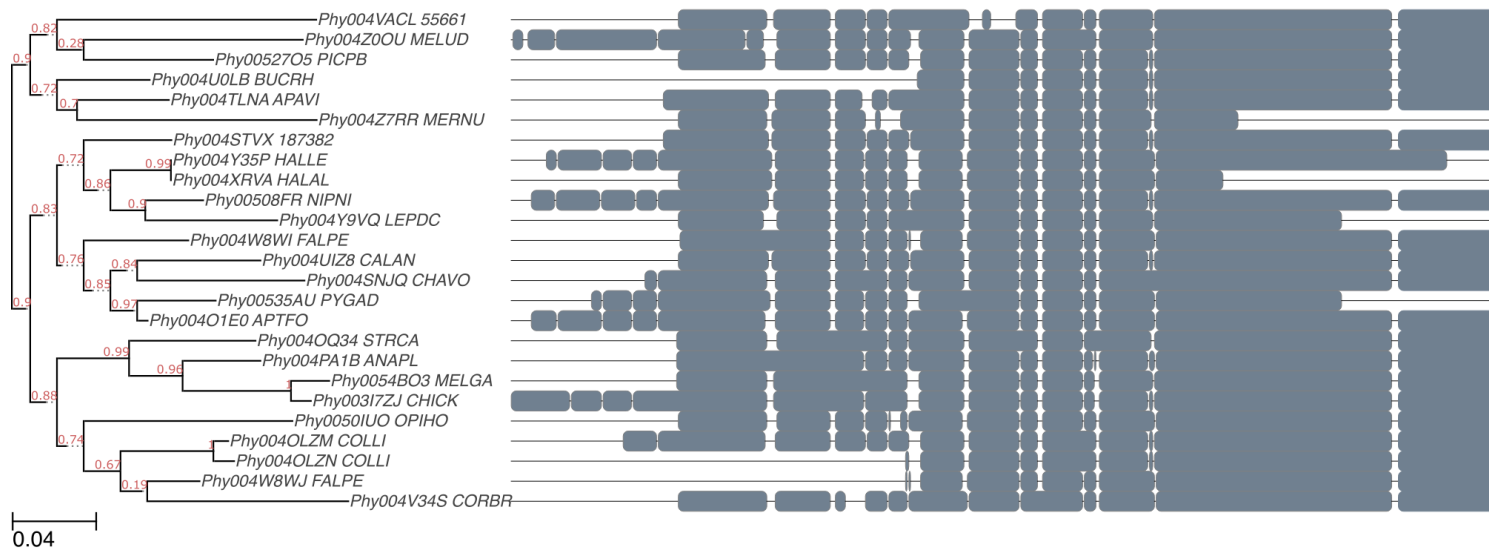

## 5. Explore your tree

As we did *not* use the `--noimg` flag, a tree and alignment image was generated automatically.

```
In [36]:
```

```
from IPython.display import Image
Image(filename='NUP62_tree/clustalo_default-none-none-fasttree_full/NUP62.fa.f
inal_tree.png')
```

```
Out[36]:
```



We can also have a quick look at the resulting tree from terminal: `ete3 view --text`

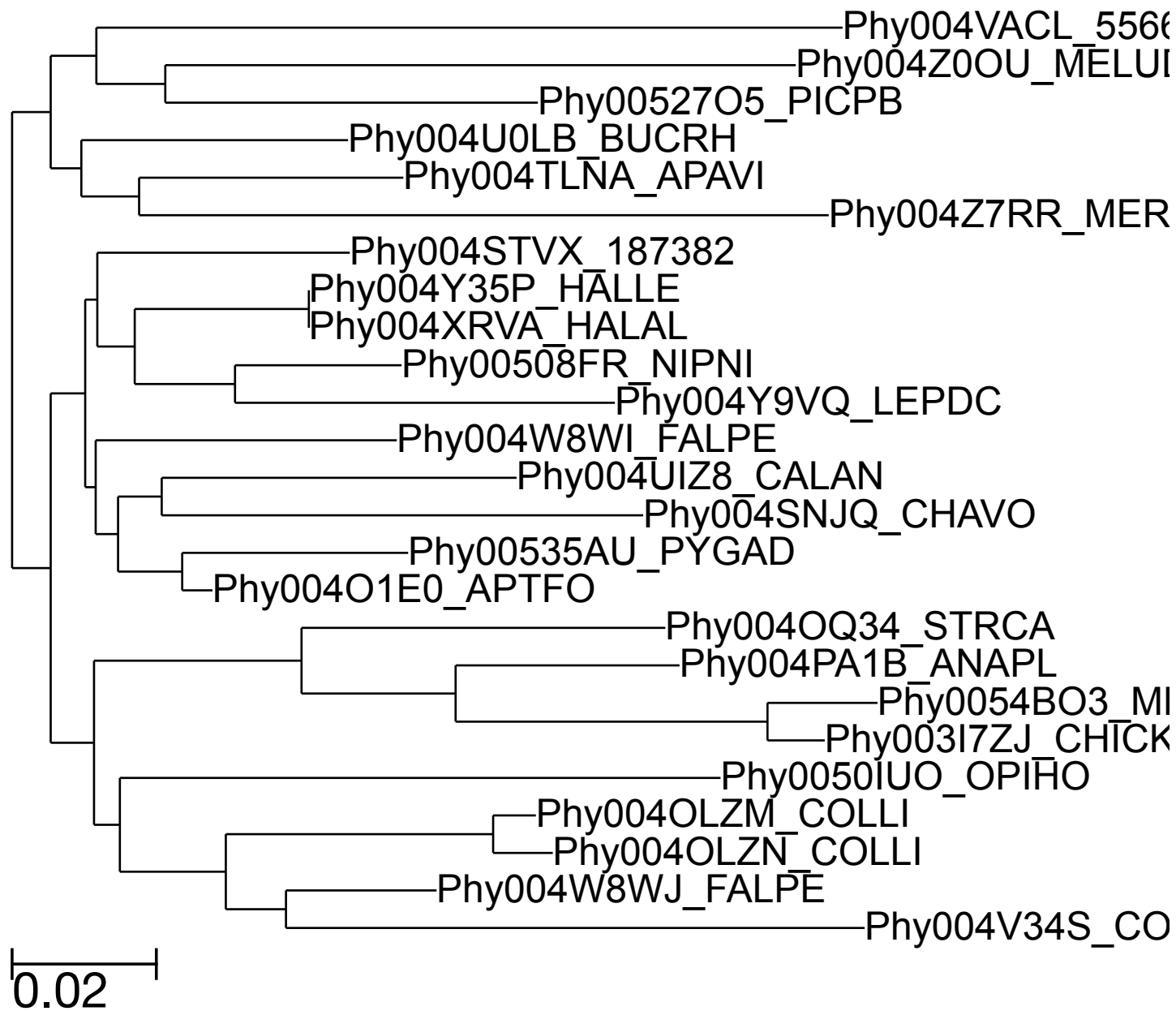or generate an SVG figure from the command line

```
In [37]:
```

```
%%bash
ete3 view --image tree.svg -t NUP62_tree/clustalo_default-none-none-fasttree_f
ull/NUP62.fa.final_tree.nw
```

```python
from IPython.display import SVG
SVG(filename='tree.svg')
```

```
                                          Phy004VACL_5566
                                        Phy004Z0OU_MELUI
                                  Phy00527O5_PICPB
                          Phy004U0LB_BUCRH
                          Phy004TLNA_APAVI
                                          Phy004Z7RR_MER
                        Phy004STVX_187382
                      Phy004Y35P_HALLE
                      Phy004XRVA_HALAL
                        Phy00508FR_NIPNI
                            Phy004Y9VQ_LEPDC
                      Phy004W8WI_FALPE
                          Phy004UIZ8_CALAN
                            Phy004SNJQ_CHAVO
                    Phy00535AU_PYGAD
                  Phy004O1E0_APTFO
                              Phy004OQ34_STRCA
                            Phy004PA1B_ANAPL
                                  Phy0054BO3_MI
                                  Phy003I7ZJ_CHICK
                            Phy0050IUO_OPIHO
                        Phy004OLZM_COLLI
                        Phy004OLZN_COLLI
                      Phy004W8WJ_FALPE
                                Phy004V34S_CO
```

```
0.02
```

## Loading the tree and alignment using the ete3 API for further analysis

The ete3 toolkit allows you to do much more, from rooting, traversing or node manipulation to custom visualization. Following is the exmaple verbatim from the ete3 toolkit, just ot show the alignment and the tree side by side. Impressive !!

In [39]:

```python
from ete3 import PhyloTree
tree = PhyloTree("NUP62_tree/clustalo_default-none-none-fasttree_full/NUP62.fa
.final_tree.nw")
tree.link_to_alignment("NUP62_tree/clustalo_default-none-none-fasttree_full/NU
P62.fa.final_tree.used_alg.fa")
tree.render("%%inline")
```
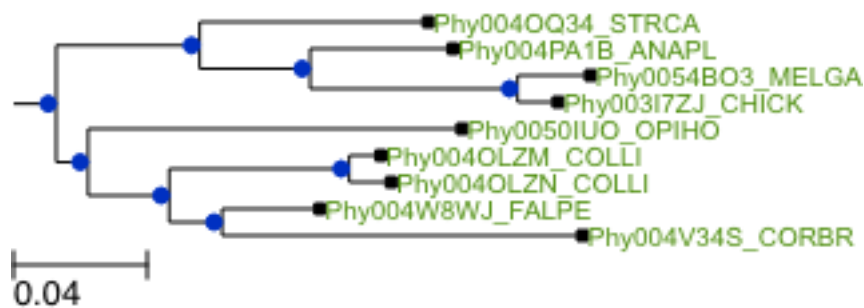
Out[39]:
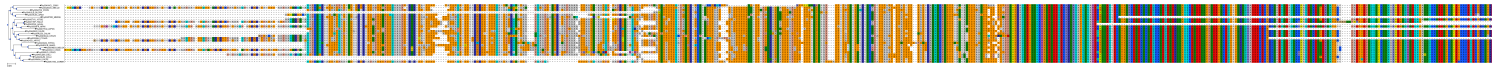


In [40]:

```python
tree.get_common_ancestor("Phy004W8WJ_FALPE", "Phy004OQ34_STRCA").render("%%inl
ine", layout="basic")
```

Out[40]:



In [41]:

```python
print 'DONE'
```

DONE

In [ ]: