Bioinformatic practical 7

# Molecular phylogenetics in practice

* ***Aim of the practical*: Inferring phylogenetic tree from the mRNA sequences (400-800 nt long subalignment) of six mammal species: human, chimp, mouse, dog, cow and opossum, using two different methods (neighbor-joining and maximum parsimony)**
* Work with your previously created alignment files ([filename].phy), if you do not have it, download it from here: <https://genetics.elte.hu/index.php/en/teaching/152-bioinformatics>
* In the practice you are going to use online tools to calculate the trees, but in case of internet failure, use the UGENE software (and the description for it)

**Task 0: properly format your sequences for online input (also good for UGENE):** [**http://phylogeny.lirmm.fr/phylo\_cgi/data\_converter.cgi**](http://phylogeny.lirmm.fr/phylo_cgi/data_converter.cgi)

 a) Copy or upload your phylip ([filename].phy) file into the webpage

 b) Set the Output format to **PHYLIP (sequential)**

 c) Click on **Convert** button, copy-paste and save the output to new file

**Task 1: Build a distance based – neighbor-joining (NJ) – tree using online version of PHYLIP dnadist and neighbor:**

[**http://www.trex.uqam.ca/index.php?action=trex&menuD=1&method=2**](http://www.trex.uqam.ca/index.php?action=trex&menuD=1&method=2)

(Here, the two subprogram is merged together accompanied with bootstrap as well. In the original command line (CLI) version, these have to be treated separately)

1. Click on **sequences** for data type, copy or upload the new phylip file
2. Set the following options below:

*Construction method*: Neighbour-Joining – Saitou and Nei (1987)

*Substitution model*: Kimura 2-parameters

*Missing bases option*: Ignore missing bases

*Validation*: Bootstrap, 100 replicates

1. Click on **Compute**
2. You are going to have three output files: distance matrix, fitting statistics, newick format tree and an option for viewing the tree. Download the fitting statistics and the newick file (copy paste to separate files)
3. Open Figtree\* software, then open the newick file in it. Root the phylogenetic tree.

Which species should be used for rooting the tree?

1. Did you get the same topology as Murphy et al. (Science, 2001)? See below.
If not, what are the differences?



1. Make bootstrap values visible in nodes: Node labels on, open → Display: Branch times (These are not branch times just this software treat them like this.)

Save the figure as PNG.

What are the lowest and the highest bootstrap values?

Which nodes are not reliable based on the bootstrap values (below 70%)? - If there is any.

Task1 if there is no internet:

**Build a distance based – neighbor-joining (NJ) – tree using the UGENE interpretation of PHYLIP dnadist and neighbor softwares.**

1. Calculate a simple distance based tree:

Push the Build tree button on the upper menu row.

*Tree building method*: PHYLIP Neighbor Joining

*Distance matrix model*: Kimura

*Display options* tab: Display tree in new window

1. Open the automatically saved Newick formatted tree (.nwk) in a text editor. Than read it to the *Figtree\** software.

Which species should be used for rooting the tree?

Root the tree with that species (outgroup) and save a figure of the tree: *File → Export PNG*

1. Did you get the same topology as Murphy et al. (Science, 2001)? See below.
If not, what are the differences?



1. Calculate a bootstrap neighbor-joining tree as well.

Use the same parameters as before, except on tab *Bootstrapping and Consensus tree* make Enable bootsrtapping on.

Choose the simple (not the *extended*) majority rule consensus tree method.

1. Open the result tree with *Figtree*.

Root it with the proper outgroup.

Did you get the same topology as without bootstrapping?

Make bootstrap values visible in nodes: Node labels on, open → Display: Branch times (These are not branch times just this software treat them like this.)

Save the figure as PNG.

What are the lowest and the highest bootstrap values?

Which nodes are not reliable based on the bootstrap values (below 65%)? - If there is any.

**Task 2: Calculate a Maximum Parsimony (MP) tree using the online version of PHYLIP dnapars:** [**http://www.trex.uqam.ca/index.php?action=phylip&app=dnapars**](http://www.trex.uqam.ca/index.php?action=phylip&app=dnapars)

1. Upload the subalignment sequences saved as *PHYLIP interleave* in last practical.

Open the subalignment file in a text editor to know what is the number (out of all species) of the outgroup. Give this number at *Outgroup root* part of the sofware.

Push Compute button.

How many MP tree did you get?

Download the *Outree file*.

1. Open the MP tree(s) in *Figtree*.

To see the root more clean (without multifurcation) reroot the tree with the outgroup.

If you got more MP tree: What are the differences between their topology?

Does any MP tree match with the neighbor-joining tree?

1. If you have more than one tree, create a consensus here:

<http://bar.utoronto.ca/webphylip/>

In the menu, go to the Consensus tree → Run

1. Upload the newick file, and click on submit

How much does it differ from NJ tree?

Did you get any multifurcations?

\*If Figtree is too slow, use this webpage: http://itol.embl.de/upload.cgi
Upload the tree (newick). Clicking on the outgroup species: *Tree structure* → *Reroot tree here*. Show bootstrap values: tab *Advanced*: *Branch lengths* → *Display*. Saving: tab *Export*.