

Module 2: Phylogenetic inference training session – Dec. 9, 2014

Datasets

- **Dataset 1** : amino-acid alignment of the **phosphoglycerate kinase (pgk) protein** in 21 bacterial species of the firmicutes group (formats : clustalw, fasta, phylip, nexus).
- **Dataset 2** : nucleic alignment of the **beta glucosidase gene (bglA)** in 9 bacterial species of the *Staphylococcus* and *Listeria* genus (bglA_listeriaStpah.*, formats : clustalw, fasta, phylip and nexus).

Softwares

- **Modelgenerator** :

command: modelgenerator.jar

doc: http://genoweb.toulouse.inra.fr/~formation/11_Phylogeny/doc/KeaneBMCbioinfo2006.pdf

- **Seaview** :

command : seaview &

- **FigTree**

command : figtree &

- **Phylip**

command : dnadist, protdist...etc

doc : <http://evolution.genetics.washington.edu/phylip.html>

- **BioNJ**

command : BIONJ

doc : http://genoweb.toulouse.inra.fr/~formation/11_Phylogeny/doc/Gascuel_bionj_1997.pdf

- **PhyML**

command : phym1

doc : phym1 -help or https://phym1.googlecode.com/files/phym1_manual.pdf

- **Mr Bayes**

module load bioinfo/MrBayes_3.2.2

command : mb

doc : http://mrbayes.sourceforge.net/mb3.2_manual.pdf

Exercise 1: Model selection

Use *modelgenerator* to choose the best-fit evolutionary model for the two datasets.

Exercise 2: Practice a distance method and root a tree

- **Question 1.** Use *seaview* or *phylip/bioNJ* to build a Neighbor-Joining tree for the two datasets.
- **Question 2** Use *figtree* to root the trees using two different methods : midpoint rooting and arbitrary outgroup (for instance Staph_epider for the nucleic dataset and Myco_geni for the proteic dataset)

Exercise 3: Practice a parsimony method

Use *seaview* or *phylip* to build a parsimony tree for the two datasets.
Use *figtree* to compare the parsimony and the bioNJ trees. What kind of differences can you observe ?

(afternoon)

Exercise 4: Practice a Maximum likelihood method

Use *phyml* or *seaview* to build a Maximum Likelihood tree for the two datasets. Be careful to choose the best-fitted model (results of exercise 1).
Use *seaview* and *figtree* to identify differences with the bioNJ and parsimony trees. What can you observe ?

Exercise 5: Practice a Bayesian method

Use *MrBayes* to build a tree for the the nucleic dataset *bgIA_listeriaStpah.nex*. Be careful, input file has to be in nexus format. Comment the obtained trees.

Exercise 6: Evaluate a tree using bootstrap

Use *seaview* or *phylip* to compute bootstrap values for Neighbor Joining trees and the Maximum likelihood trees. Are the obtained trees robust ?