

PHYG – TME5

2019-2020

YASSER MOHSENI
YASSER.MOHSENI_BEHBAHANI@UPMC.FR

5 December 2019

General rules

- Reports must be sent by e-mail, by the 13th of December, using the subject “[PHYG] TME5”, mentioning the names of the students who worked on it.
- Multiple files should be grouped in a compressed archive (`.tar.gz` or `.zip`)
- Your report *must be* in PDF format and named `student1_student2_TME5.pdf`. It should be simple, clear and well organized. Answers should be given in an exhaustive manner.
- All required materials can be found in the repository <https://github.com/yassermmb/PHYG2019.git>

How to run SplitsTree4

Unzip the file `splitstree4.zip` and run the command `SplitsTree` from the extracted directory.

Exercise 1 – Networks with a toy example

1. Select “*File → Enter data*” and copy-paste the content of `artificial1.txt` (an artificial alignment of DNA sequences). Click “*Execute*” to load the data. Then, select “*SplitDecomposition*” in the Networks menu and click “*Apply*”.
2. Do the same for `artificial2.txt`. You can select an edge and click “*Hide selected splits*” to remove the edges associated with a split. This allows you to select edges to get one of the possible trees. Try to do that to get the same tree as with `artificial1`. You can display again all edges by selecting “*Draw → Redraw all splits*”.

3. Try “*Draw → ClusterNetwork*” view (after having selected “*SplitDecomposition*”). What are the blue edges?
4. Can `artificial1` be explained by a tree? What about `artificial2`? Which hypothesis would you make for `artificial2`?

Exercise 2 – Networks with mammals

1. Select “*File → Enter data*” and copy-paste the contents of `CFTR_in_mammals.fasta` (the alignment of the CFTR protein in mammals from TME1). Click “*Execute*” to load the data, select “*SplitDecomposition*” in the *Networks* menu, and click “*Apply*”.
2. Export the network image and include it in your report (*File → Export image*).
3. What do you observe for rat and mouse with this network? Does this agree with UPGMA and NJ trees? (*Note: You can recompute them from the “Tree” menu.*)
4. Now look at pig (*Sus scrofa*), cow (*Bos taurus*), and horse (*Equus caballus*). The truth is that pig is nearer to cow than to horse (considering mouse for example as the outgroup). Previously, we found out that UPGMA got it wrong (pig nearer to horse) while NJ got it right. What does the network computed by `SplitsTree` find?
5. How do you explain these results? Why do you think we didn’t find a simple tree?

Exercise 3 – Bootstrapping with mammals

1. Load the CFTR alignment as in the previous exercise and create a UPGMA tree. Look at pig, cow and horse. As you can see pig and horse are together, while cow is further away (*i.e.*, nearer to an outgroup like mouse). This is not correct because pig and cow should be together (as seen in TME1). We will now use some bootstrapping: select “*Analysis → Bootstrap*” and choose 1000 runs.
2. Is the tree correct for pig, cow and horse? Include the image in your report (zoom in order to see these three species properly).
3. What do the numbers on the branches represent?
4. Let X and Y be the 2 species among pig, cow, and horse that are grouped together in your tree. What is the number on the branch that leads to the subtree with X and Y ? Include an image where it is possible to see this value in your report.
5. What is your interpretation of this value? (no need to make a very long answer)
6. Now, run the analysis using 10 runs of bootstrap. What is the value on the branch with X and Y in the consensus tree?

7. Build the NJ tree and run the bootstrap analysis (1000 runs). What are the 2 grouped species (as in point 4.)? What is the value now? Include a screenshot where it is possible to see this value and compare it with the value obtained with the UPGMA approach.

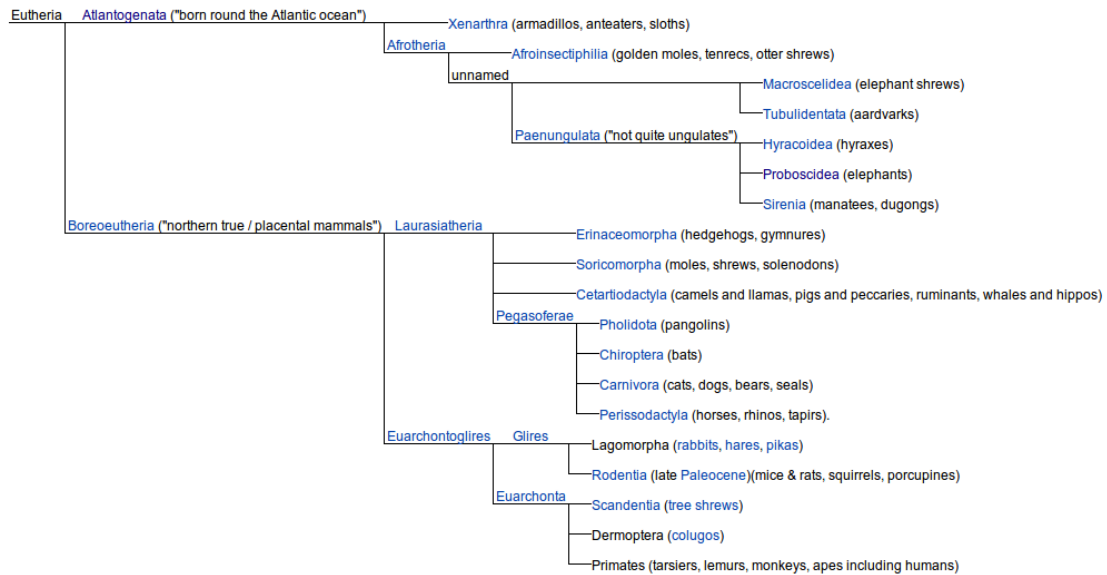


Figure 1: Correct phylogenetic tree of placental mammals