PHYLOGENY – TME2

2017-2018

Riccardo Vicedomini
riccardo.vicedomini@upmc.fr
https://github.com/vice87/PHYG

12 October 2017

General rules

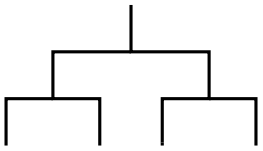
- Reports must be sent by e-mail, *mandatorily* using the subject `[PHYG] TME2` and stating in the body the names of the persons who worked on it (maximum two students per group).
- Multiple files should be grouped in a compressed archive (`.tar.gz` or `.zip`)
- Your report **must be** in PDF format and named `student1_student2_TME2.pdf`. It should be simple, clear and well organized. Answers should be given in an exhaustive manner. Consider adding at the beginning a summary indicating the page of each answer.
- Source code (if any) must be well explained, commented and, most importantly, it should run on my computer. Provide all needed information (*e.g.,* compiler/interpreter version) in a `README` file.

Exercise 1: Parsimony

1. What is the main idea of parsimony methods? Give an example.
2. What are the small and large parsimony problems? Which one is harder? Why?

3. Given the following sequences, topology and cost matrices, apply the Fitch and Sankoff's algorithms to calculate the scores.

$A =$ `ATCCTG` $B =$ `ATCCGG` $C =$ `ACGGCC` $D =$ `AGGGCA`



Fitch algorithm

	A	T	G	C
A	0	1	2	3
T	1	0	2	4
G	2	2	0	1
C	3	4	1	0

Sankoff algorithm

	A	T	G	C
A	0	3	4	9
T	3	0	2	4
G	4	2	0	4
C	9	4	4	0

4. What is the main idea of the nearest neighbor interchange algorithm? Why it is considered a heuristic method?

Exercise 2: Reconstruction using reversal distances

1. Go to the web page `http://cinteny.cchmc.org`, choose *human* and *mouse* and click start. Then, select *whole genome analysis* (using human genome as reference). For human, genes are colored by chromosome, while for mouse by chromosome of human's homologous genes. Include both figures in your report.

2. Start again with human and mouse but select *chromosome versus chromosome* for chromosome 1 in human and 4 in mouse. What is the reversal distance? Why a big part of each chromosome was left in white? Include the figure in your report.

3. Now start again with human, mouse, cow, and chimpanzee. Choose a whole genome analysis, write the matrix of reversal distances. Include this matrix in your report.

4. Use PHYLIP's command `neighbor` to compute NJ and UPGMA trees from this matrix. Are these trees correct? See Figure 1

5. Now, we want to do the same with all mammals. The distance matrix is already available as file `mammalsMatrix.txt` (`https://github.com/vice87/PHYG/`). Compute both NJ and UPGMA trees and include them in your report.

6. Are these trees correct? What is the limitation of the approach used?

Exercise 3: Reconstruction using characters

1. Use `pars` command (PHYLIP) with the matrix below in order to compute a tree based on characters. Attach the tree to your report. Is it correct?

2. Run the analysis again without considering the last column (*i.e.,* enlarged malleolus). What happens? Which character is responsible for this incorrect tree?

	Has placenta	Lives in water	Lays eggs	Single pair of incisors	Opposable thumb	Enlarged malleolus
Zebrafish	No	Yes	Yes	No	No	No
Opossum	No	No	No	No	Yes	No
Whale	Yes	Yes	No	No	No	Yes
Mouse	Yes	No	No	Yes	No	Yes
Rat	Yes	No	No	Yes	No	Yes
Chimp	Yes	No	No	No	Yes	Yes
Human	Yes	No	No	No	Yes	Yes

3. The presence of a character in two species can be explained either by a common ancestor or by convergent evolution. Find all cases of convergent evolution in the table.

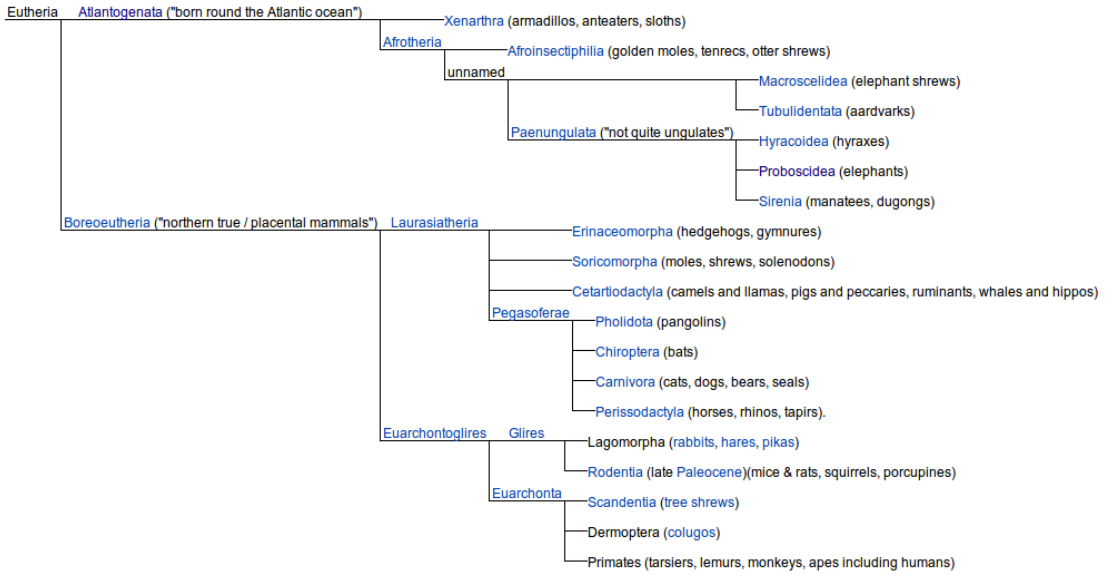


Figure 1: The correct phylogenetic tree of placental mammals (opossum does not appear because it is a not a placental mammal).