

TME5 - PHYG

Witold PODLEJSKI & Jérémie PERRIN

December 5, 2019

1 Exercise 1: Networks with toy example

1.1 Question 1

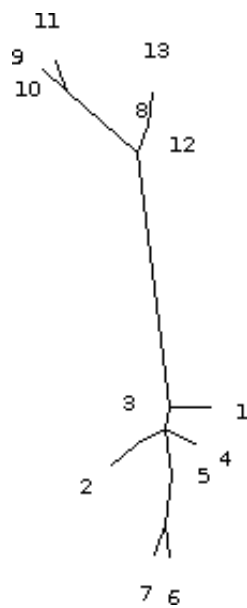


Figure 1: The tree inferred with the Split-Decomposition algorithm on generated data

1.2 Question 2

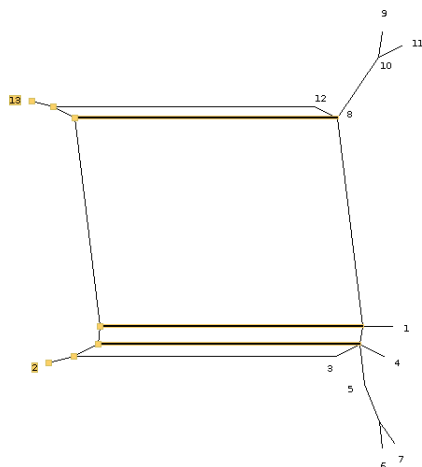


Figure 2: The network inferred with the Split-Decomposition algorithm on generated data

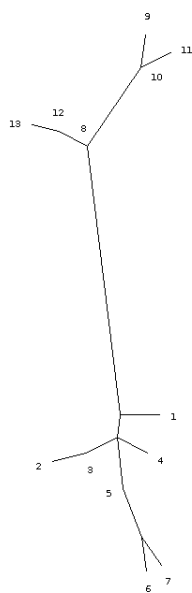


Figure 3: The tree obtained by cutting the last network

We deleted the selected edges in Figure 2 to build a tree (Figure 3) similar to the one in Figure 1.

1.3 Question 3

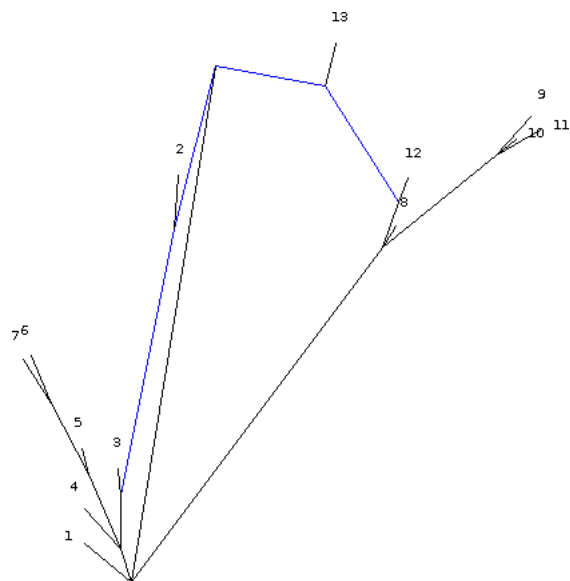


Figure 4: The network obtained with *Cluster Network*

Blue edges are the edges preventing our phylogenetic network to be a tree. They

1.4 Question 4

`artificial1.txt` can be explained by a tree but `artificial2.txt` cannot as the reconstructed trees using each gene are not compatible to a tree structure. We thus make the hypothesis that there have been two reticulation events.

2 Exercise 2: Networks with mammals

2.1 Question 1-2

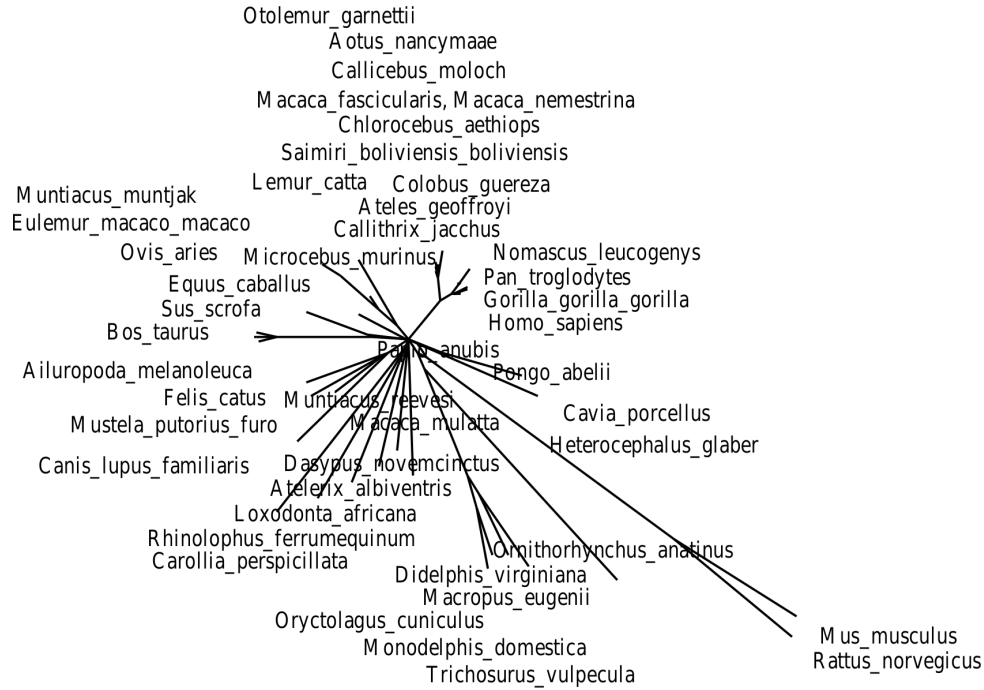


Figure 5: The tree obtained with *Split Decomposition* for all mammals

2.2 Question 3

In this network, the mouse (*Mus musculus*) and the rat (*Rattus norvegicus*) are each other's closest neighbour. This is compatible with the tree built *via* NJ and UPGMA.

2.3 Question 4

The network computed by **SplitsTree** puts the pig (*Sus scrofa*) closer to the horse (*Equus caballus*) than to the cow (*Bos taurus*). This is wrong according to the right tree built by Neighbor-joining.

2.4 Question 5

As it is very unlikely that a reticulation event has occurred. We can explain this result by some approximations due to the method or due to the data.

3 Exercise 2: Bootstrapping with mammals

3.1 Question 1-2

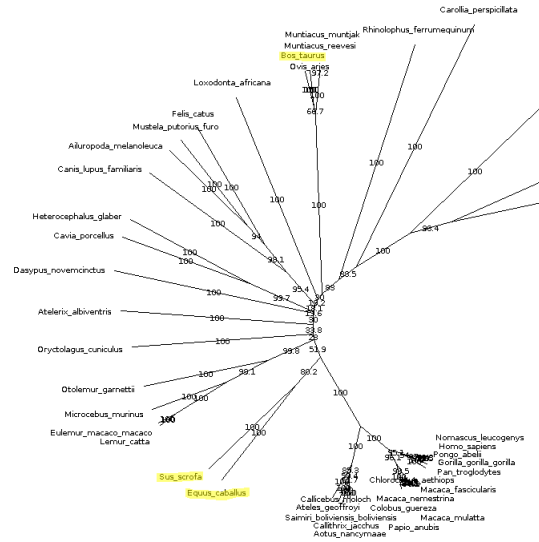


Figure 6: The tree obtained with *Bootstrapped UPGMA*

The bootstrapped tree is not right neither about pig, cow and horse.

3.2 Question 3

The numbers on the branches represent the percentage of simulations where this branch was present, amongst the 1000 simulations on random subsets of our dataset. It represents the stability of the branch according to our dataset.

3.3 Question 4

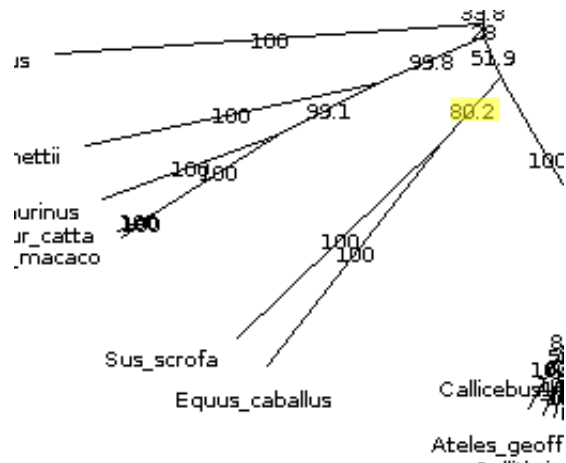


Figure 7: Zoom on *Sus scrofa*/*Equus caballus* branch

The branch leading to the node *Sus scrofa*/*Equus caballus* has a likelihood of 80.2%.

3.4 Question 5

This number indicates that UPGMA is not very assertive about the existence of this branch.

3.5 Question 6

If we run the bootstrap with only 10 simulations, the confidence of the branch from which the divergence *Sus scrofa*/*Equus caballus* emerges is 80%.

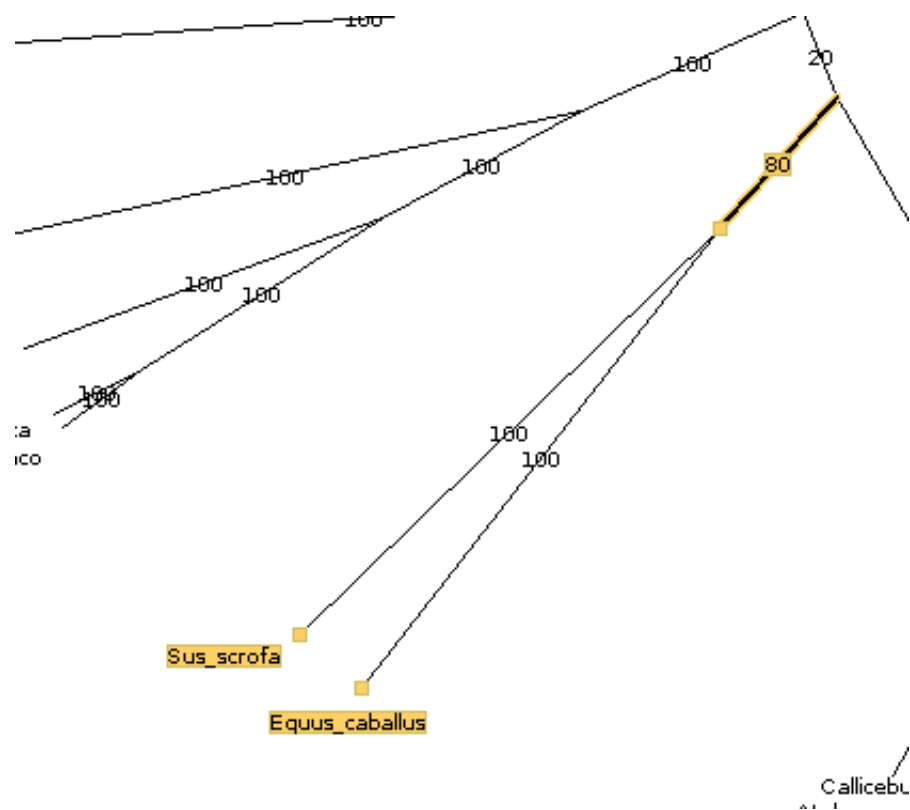


Figure 8: The same zoom as Figure 7 but with 10 run of bootstrap

3.6 Question 7

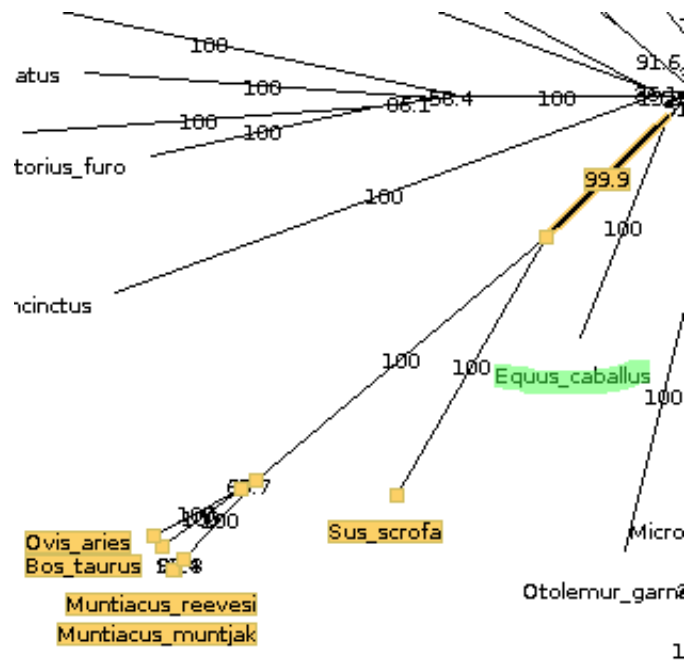


Figure 9: The tree with NJ bootstrapped 1000 times

With NJ, the two grouped species are *Sus scrofa* and *Bos taurus*, with a confidence of 99.9%, which is way more significant than the branch inferred by UPGMA.