15. October 2013

Exercise 1:   UPGMA can reconstruct each tree that fulfills the molecular-clock assumption from the distances of its leaves. Is this also true for hierarchical cluster methods that define the distance of two taxa groups as the minimal or the maximal pairwise distance instead of the mean distance that UPGMA is using? Give a proof or counterexamples.

Exercise 2:   Install Phylip and R with the ape package on your computer and explore with both programs how well UPGMA and Neighbor Joining can estimate trees from imprecise distance data. Start with several distance matrices that are compatible to trees. How much can you perturb the distances until UPGMA and/or Neighbor Joining give you a different tree? Explore this with some trees that fulfill the molecular clock assumption and with some that do not.

Exercise 3:   FOR BIOINFORMATICIANS: Implement UPGMA and Neighbor Joining (e.g. in R, Java, C++ or Python) and test your program with example data.