

Homework #5 (due 11/14)

1. Download the small sample set of sequences from the course website.
2. Modify your previous global alignment code to compute the Hamming distance (number of mismatches) between all pairs of sequences in the small sample ignoring indels. Output the resulting simple matrix to the screen, and place the program, the sample set, and instructions to compile/run this code in your dropbox (5 points)
3. Generate by hand the distance-based tree for the small set of sequences using the UPGMA algorithm. You may either hand in your answer as a hardcopy in Prof. Emrich's mail box or as an ecopy in your dropbox (8 points).
4. Download and install the phylip package (<http://evolution.genetics.washington.edu/phylip/>) and the larger set of sequences from the course website. All programs below are part of phylip.
5. Generate a distance-based tree using phylip for the larger dataset using both KITSCH and NEIGHBOR. You will need to compute a distance matrix first using DNADIST. Place the raw tree files into your dropbox and briefly comment on differences if any. (5 points)
6. Run the Clustal Omega multiple sequence alignment tool and one additional tool of your choice on the large dataset (#4) using this link: <http://www.ebi.ac.uk/Tools/msa/>. Place a text version of the resulting "alignment files" in your drop box (6 points) Are the trees largely similar compared to the KITSCH and NEIGHBOR on a quick glance (3 points)? Are they roughly similar to each other? (3 points)