

EE550
Computational Biology
Homework 2 – due 29.03.2022

Instructions: Please return your homework as a single PDF document by the due date with “EE550” and your student number in the file name. Sharing of ideas and discussions is encouraged but sharing of results and/or text is not. Show the details your work including the intermediary results.

Question (100 points)

Consider the amino acid sequence data contained in the file AAseq.txt.

- a) (40 points) Compute the PAM substitution matrix for a unit evolution time in which 1% of the amino acids are expected to incur a substitution, as well as the corresponding log-odds scoring matrix. (Hint: Try to pick a suitable order and coefficients for the log-odds calculations.)
- b) (40 points) Compute the Blosum log-odds scoring matrix of the same sequence data. (Hint: Again, try to pick suitable coefficients for the log-odds calculations. You may have to figure out a way to deal with zeros.)
- c) (20 points) Compare and contrast the two scoring matrices in terms of the most and the least conserved amino acids along with the amino acid pairs most and least likely to be substituted for each other.