EE550

Computational Biology

Homework 1 – due 22.03.2022

Instructions: Please return your homework 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 of your work including the intermediary results if applicable.

Question (100 points)

For this question, generate a random transition rate matrix Q to model the rates of nucleic acid substitutions in such a way that the off-diagonal entries vary between 0 and 1 and the diagonal entries are set so that each row sums to 0.

a) (50 points) Calculate the transition probability matrix $P(t)$ corresponding to the transition rate matrix Q above and plot the corresponding transition probabilities between each nucleotide pair in a 4×4 table for $t \in [0,5]$.
(Hint: You may need to figure out how to calculate matrix exponentials. Feel free to extend the time span of the plots to observe convergence of the probability curves to a constant level.)

b) (50 points) Given the transition probability matrix $P(t)$ above, calculate the functional relationship between the sequence distance D and the evolutionary distance d between two nucleotide sequences. Show the calculated relationship as a graph of d versus D.
(Hint: Assume equal rates for all four nucleotides and carry out the associated summations for varying t.)