Universität Greifswald
Institute für Mathematik and Informatik
Lecturer: Marc Hellmuth Tutor: Nikolai Nøjgaard

5. EXERCISE "BIOINFORMATICS", SS 17

Aufgabe 1: (5 Credits)
Show that the Edit-Distance of two strings u and v is the same as the Edit-Distance of the inverted strings u^{-1} and v^{-1}, i.e.,

$$\text{Edit-Distance}(u,v) = \text{Edit-Distance}(v^{-1},u^{-1}),$$

where $w^{-1} = w_k w_{k-1} \dots w_2 w_1$ whenever $w = w_1 w_2 \dots w_{k-1} w_k$.

Aufgabe 2: (5+1+2+3 =11 Credits)
Given the strings $u =$`GTTTAAG` $v =$`GAAGA` and cost function

$$\delta(a,b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } a \neq b \text{ and } a,b \neq - \\ 2 & \text{else} \end{cases}$$

(a) Compute D for the strings u and v.

(b) What is the optimal alignment score?

(c) Give one possible optimal alignment for u and v.

(d) How many optimal alignments are there for u and v - Explain shortly your results.

Aufgabe 3: (4 Credits)
Get familiar with BLAST (`http://blast.ncbi.nlm.nih.gov/Blast.cgi`), see e.g. `https://digitalworldbiology.com/blast` for a short tutorial.

Assume you have sequenced the strings ">Sequence_X" with $X \in \{1,2,3,4\}$ in the file seq.txt (see extra material on teaching homepage). Use the first best hit of BLAST program *nucleotide blast* with database option *Others / nucleotide collection* and optimization option *discontiguous megablast* to determine the organism the respective sequence is taken from.

Deadline: Tuesday - May 9, 2017