Computational Biology HW 1

Probst Jennifer, 16703423

12. October 2020

Question 1

Gaps in sequence alignments stand for indels/delection during sequence evolution and mismatches for a nucleotide mutation on one of the genes. Generally, gaps are more penalized, because the probability of a mismatch to happen is biologically higher than the occurrence of indels. In addition, sequencing errors can also lead to mismatches.

Question 2

In local alignments, the goal is to obtain the best match of partial sequences and can therefore omit indels (gaps) at the start and end.

Question 3

There could be multiple good local alignments for a really short sequence and it will be hard to choose the optimal one.

Question 4

The NW-algorithm for two sequences length m and n, takes $O(nm)$ steps, which would be 10^4 with $n = 100$ and $m = 100$. This would be roughly 10^{71} times lower than the brute-force method with roughly 10^{75} steps.

Question 5

For k sequences the NW matrix could be extended to k dimensions $m_1 \times m_2 \times ... \times m_k$, where m_i stands for the length of the ith sequence. I believe this is computationally only feasible for low k, as the runtime increases as the product of m_i, e.g. $O(m_1 * m_2 * ... * m_k)$ for k sequences.