Homework 1 – Computational Biology

John Oehninger

17.10.2022

1) Gaps are generally more penalized than mismatches, as they result in an entire frameshift which results in an entirely different result than having a mismatch. But one must bear in mind that gaps are only introduced if the overall score is maximized by it!

2) In the local alignment we only have a negative score for the mismatch and as soon as a score becomes negative, it is set to zero. Also, the matrix is predefined, which wouldn't allow this anyway.

3) If we align a very short sequence to a much larger one, we could potentially receive many possible alignments with the same score, therefore our certainty in the alignment decreases.

4) This would result I roughly 10^{14} iterations, considering the iterations in my code. This would come down to roughly $2*10^{62}$ times less steps. And this is exactly why we have these algorithms.

5) The Needleman-Wunsch algorithm could be extended to k sequences (for k>2) by calculating the pairwise similarity between all sequences. But as we would climb up the dimensional later, this would become impractical, as the computational work that needs to be invested would grow exponentially. It is much easier to take a probabilistic approach to the problem, combined with model of sequence evolution. And since online reference databases are quite common nowadays, this is the best approach to the problem.