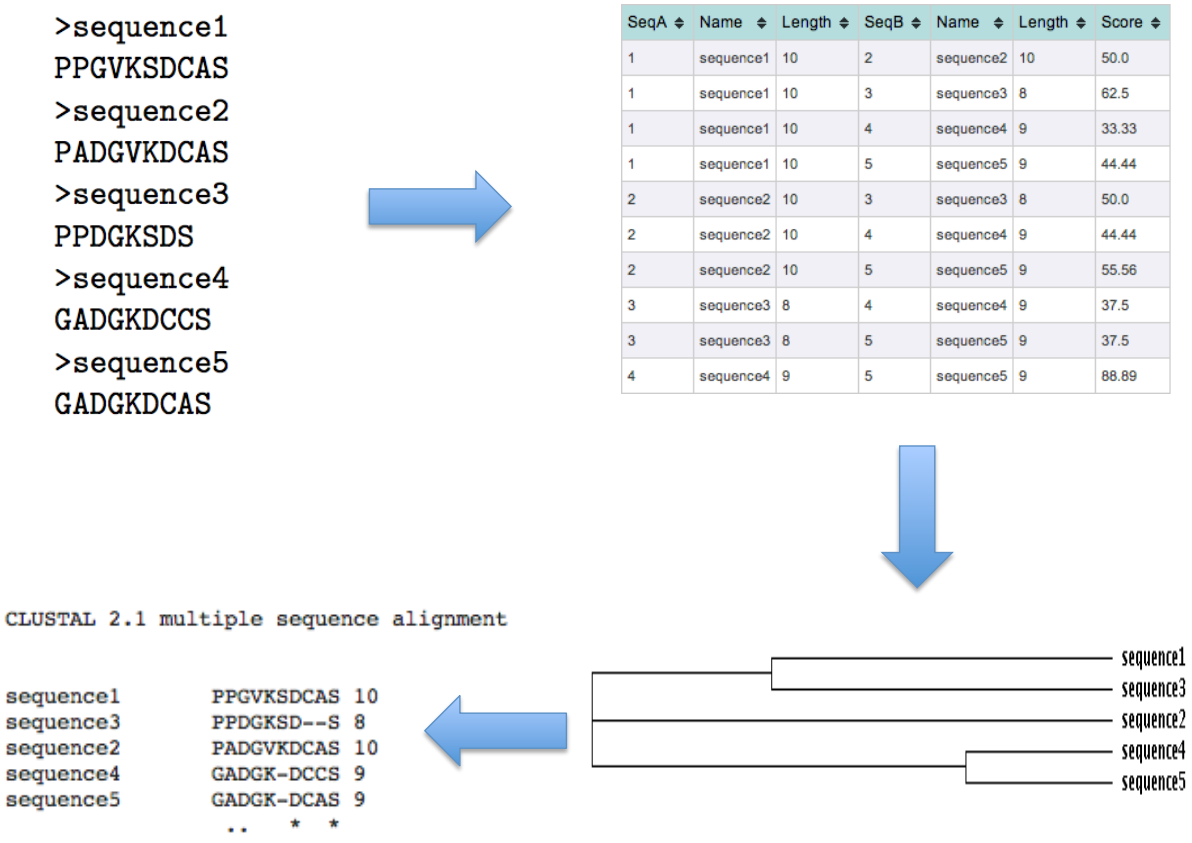
Exercises - Week 8

Genomics and bioinformatics

1 MSA with ClustalW

Here is a summary of the main steps followed by ClustalW to generate the MSA. ClustalW first computes the optimal global alignment for every pair of sequences and then the distance score is set to be $1-y/x$, where x and y are the number of non-gap positions and the number of identical positions, respectively. The guide tree is then built according to the distance matrix (and not the raw alignment scores) using the neighbour joining algorithm presented in the lectures. The last step involves doing profile-profile alignments and may be computed using dynamic programming to with some PSP score.



Using the PAM250 scoring matrix with a gap penalty of -10, one can compute the SP score of the MSA by hands using the formula presented in the course. It is however more convenient to write a code (see `MSA_SP_score.py`) to do this task. We find a SP score of 101.