Quiz 9. Heuristic Alignment algorithms

1. If we are searching for the similarity between a query sequence Q of length 100 and the set of 30 sequences of approximately the same length each, and we want to find all the sequences in the set which approximately match sequence Q with maximum 9 differences (insertions, deletions, substitutions)

How can we speed up the search by:

a. Filtration

BYP filtration algorithm:
Partition Q into k+1=10 equal substrings
Test if any of these substrings occurs exactly as a substring of the database strings
Filter out the database strings which do not contain an exact copy of at least 1 of the partitions

b. Decreasing the number of DP table cells to be computed

When comparing Q with each of the database strings, compute the values of the cells in the strip of width 9*2+1=19 around the main diagonal. If we have computed the resulting distance without exceeding this strip, report the database string. If all the values in some row exceeded 9, then abort the computation.