Individual Project

Topic:

Implement three algorithms to solve multiple sequence alignment (MSA) problems.

Requirements:

- Implement dynamic programming (DP) algorithm to find the optimal solution.
- Implement A-star (A*) algorithm to find the optimal solution.
- Implement genetic algorithm to find the optimal/suboptimal solution.

You need to find the best alignment that yields the least cost. Cost matrix is simple and revealed below.

	MATCH	**MISMATCH**	**GAP**
COST	0	3	2

For multiple sequence alignment with size larger than 2, the cost is computed using sum-of-pairs (sum up all pairwise cost). Note that **GAP-GAP** alignment is considered as a **MATCH** with cost 0.

For example, given

$$ABCD, ACD, BCD$$

the cost of the alignment

$$ABCD$$
$$A-CD$$
$$B-CD$$

is `(0+3+3)+(0+2+2)+0+0=10`.

Dataset:

There are `5` queries for pairwise alignment and `2` queries for three-sequence alignment in *MSA_query.txt*. Database is stored in *MSA_database.txt* with `100` sequences.

Tips:

- Please describe and analyze your implementation, results, running time and time complexity thoroughly in the report.
- Do not cheat.