IT 4133 (P) - Bioinformatics and Computational Biology – 2024
CSH 4123 (P) Bioinformatics -2024
Tutorial – 01

You have up to 1000 DNA reads of equal length (no longer than 50 base pairs), provided in FASTA format. Some reads have a single-nucleotide error.

For each read, one of these conditions holds:

The read was sequenced correctly and appeared in the dataset at least twice (it could be in reverse complement form). However, it is incorrect, appears precisely once, and has a Hamming distance one from one correct read (or its reverse complement) in the dataset.

Task:

Return a list of corrections in the format:

"[incorrect read] -> [correct read]"

Each correction involves only one nucleotide change, and the corrections can be returned in any order.

Sample Data

- TCATC
- TTCAT
- TCATC
- TGAAA
- GAGGA
- TTTCA
- ATCAA
- TTGAT
- TTTCC

Sample Output

- TTCAT->TTGAT
- GAGGA->GATGA
- TTTCC->TTTCA

Write a pseudo code and explain each step and submit it handwritten on or before the 4th of October 2024.