

Aim: In this assignment, given an integer k and a set of DNA sequences we ask to find a k -mer pattern that minimizes $\text{Total_Distance}(\text{pattern}, \text{DNA})$ over all strings of that length under Hamming distance. We call such a k -mer a median string for a set of DNA sequences.

Input: An integer k and a set of DNA sequences with the same length.

Example (input.txt):

```
6
TAAGTCTATACCATCGTAGTCTAATTAACGTTATGGTAGGAT
ATCAAGGACGGAATGACCGCAGAGGCGACGTTAATGCGCCGT
CAGAGACGCCCTAAAGATTGCGGTAGGGTCCCGTTGTTAAAG
AGACTTGAGTGCGGTGCTTGATGGGAGTGATTAAGGGCATGT
ATAAGTGTGCTGGGTCTAAGGCATTAAAGCTGAGTCAATAG
TTACATTGCAGATTAACGAGATCTGAAATTAAGGGAGAGATT
CCCAGAGTGCCCTAGTACTTAAGGGCACCCACGCCGCAGGCG
GCCCTACGCCCGTTAATGGTTCGAGTGCTATTCACTAACACA
TTAACGGACGTTTAGTGTGGATTATAGGTGAAGGGTCTGCGC
CACTCCAAGGCAGGGAACATATGTGTTGTTACTATCTTAACG
```

Output: A k -mer v that minimizes $\text{Total_distance}(v, \text{DNA})$ over all k -mers. (If multiple k -mers exist, you may return any of them.)

Example (output.txt): Here is one of the possible answers: TTAACG

Notes:

- Your code must be written by yourself. Sufficient evidence of plagiarism will be treated the same as for plagiarism or cheating.
- Non-compiling submissions will not be evaluated.
- Your code must be complete.
- Do not submit the program binary. You must submit the following items:
 - All of the source files
 - A script to compile the source code and produce the binary (Makefile).
 - A README.txt file that describes how the compilation progress works.
- Submit your answers to Fatma Kahveci at 'fatma.balci@bilkent.edu.tr'.
- Use 'CS481 Assignment-1' in the subject line of your e-mail.
- Zip your files and send them in only one zipped file. File name format='surname_name_hw1.zip'
- C / C++ will be used as programming language.
- All submissions must be made by 23:59, October 10, 2017.
- The fastest wins. **Bonus** will be given for the fastest code.