Assignment-2 : Motif-discovery

Given a set of DNA sequences (dna.csv), write a program to find the MOTIF (**size of the lmer is 6**). You can implement using **ANY ONE** of the approaches indicated below:

(i) Brute Force algorithm,
(ii) Median string algorithm
(iii) Greedy algorithm
(iv) Gibbs sampling algorithm.

(a) Estimate the **time complexity** (by setting up a counter in your program) and print it. In the end, we will analyze how much time is taken by different algorithms.
(b) In the case of Greedy and Gibbs sampling approaches, show how the score (computed based on the profile matrix) varies as the number of steps or as computation progresses. Does it converge to a constant value? Comment on it.
(c) In the case of Brute-force provide the score (profile matrix based) computed for all possible motifs and show how you picked up the best motif.
(d) In the case of Median string algorithm, show how the score (Hamming distance based) computed for each of the consensus/reference motifs (4**6) with the motifs in the DNA sequences ((N-L-1)*T) varies with the number of steps.

Provide a document (README.txt) of the steps involved in your implementation.
Use comments appropriately in your code to make it more readable.
Provide details on how to run it and software requirements.
You may not be using any of the built-in function or libraries for doing this task

PS: You can work in a group of 3.
Announce your group in the shared document below.

https://docs.google.com/document/d/1iqPxCTkkKcn_hrfc7TlkKLQLQ4_c5S03IuIIfVf5Ihw/edit?usp=sharing

The input file, dna.csv is available in the following URL:

https://drive.google.com/file/d/1ULv3lCscn3EfzuSG4a_zEdIdU035VDkT/view?usp=sharing