

Computational Biology Assignment

By implementing this assignment. You should be able to apply branch and bound median string method on multiple sequences of raw DNA.

In addition, by using already aligned multiple DNA sequences, you should also be able to calculate the probability of a new sequence being added to the multiple aligned DNA. The calculation is done by doing position specific scoring matrix (PSSM).

Instructions

Branch and bound median string

1. The raw DNA input in median string is produced in 2 ways:
 - a. Reading the prepared raw DNA from file called rawDNA.txt where first line is for the values of t sequences, n nucleotides, and L which is length of pattern to be found.
 - b. Generate t sequences where one sequence of length n is formed of random order of nucleotides. And take from user the length L of the pattern to be found.
2. Apply branch and bound method string algorithm
3. Print the multiple alignment starting and their starting indices
4. Return consensus string of length L.

PSSM

1. The multiple aligned DNA sequences is produced in 2 ways:
 - a. Reading the prepared multiple aligned DNA sequences from file called PSSMData.txt where first line is for the values of t sequences, and n nucleotides.
 - b. Generate t sequences where one sequence of length n is formed of random order of nucleotides.
2. Print the multiple aligned DNA sequences.
3. Apply PSSM on the multiple aligned DNA sequences.
4. Print the PSSM matrix
5. Let user enter a new sequence of length n, then print the probability of new entered sequence joining the rest of the multiple aligned DNA sequences.

Deliverables:

1. Source file of your code (.py)

Submission Rules:

1. You will upload a zipped folder that contains your .py file **(Don't include any .exe files in your submission)**.
2. Assignment submission is on Google Classroom **(No submission through mail)**.
3. Assignment is submitted in **teams of maximum 6 students**.
4. Follow this convention for naming your folder:
ID1_ID2_ID3_ID4_ID5_ID6 (i.e
20200111_20200222_20200333_20194323_20220341_2018453)
5. Deadline of the Assignment: **22 May 2023 - 11:59 PM.**
6. **Failing to follow any of the above rules will lead to discard your submission and consider that your team didn't submit.**