BM4322 Genomic Signal Processing

2021 Batch - Academic Year 2025/2026 - Semester 7
Assignment (Individual) - 50% of Final Grade

According to Liu et al. 2011[1], the σ_2 subunit of bacterial RNA polymerase is capable of detecting promoters of the form WAWWWT where W represents a T or A. These promoters typically occur 10 bases upstream. Based on this postulate perform the following computational operations on the sequenced DNA of the given organism.

1. Locate 1100 genes of the organism that have been *predicted based upon homology*. Select the region from 15 to 5 bases upstream for each gene. Using 100 of these selected regions manually extract the 6 bases that most likely form the promoter. If the region does not have at least 6 consecative Ws reject it. Construct the position probability matrix (PPM) of the promoter from this data taking suitable heuristic probability values for bases C and G.

2. For the remaining 1000 regions perform a statistical alignment using the PPM of (1) and determine the presence or absence of promoters within them.

3. Obtain the 1000 samples of every other student of the class and repeat (2) with the PPM of (1)

The genome assigned to each student is as follows:

Student	Genome
210079K	GCA_ 001457635.1
210179R	GCA_ 019048645.1
210504L	GCA_ 900636475.1
210657G	GCA_ 900637025.1
210707L	GCA_ 900475505.1
210732H	GCA_ 019046945.1

Due: 2025.10.17

This assignment accounts for 50% of the module assessment

Upeka Premaratne (upeka@uom.lk)
0719538433 voice and WhatsApp
2025.09.19

---

[1]Liu, X., Bushnell, D. A., & Kornberg, R. D. (2011). Lock and key to transcription: σ-DNA interaction. Cell, 147 (6), 1218-1219