Report: Workshop 1

Catalina Ariza Ardila 20222020084

Facultad de ingeniería, Universidad Distrital Francisco José de Caldas

Group 020-84, Systems analysis

Carlos Andrés Sierra Virguez

September 15th 2024

BOGOTÁ D.C

Systemic analysis

The system to be analyzed will be the DNA, more specifically, the sequence motifs (specific nucleotide patterns) of the nitrogenous bases in it, whose purpose will be to identify them and, in turn, identify the most diverse ones (those that are less repeated).

The components of the system are the nitrogenous bases: Adenine (A), Guanine (G), Cytosine (C), and Thymine (T). These bases are combined in triplets, where the order matters, meaning that ACG is not the same as CGA, and additionally, the bases can be repeated (AAA).

DNA molecules can have thousands of DNA triplet sequences.

The inputs needed to solve the problem involve the number of nitrogenous base sequences, the size of the molecule we want to analyze, and the size of the sequence motifs we want to identify.

The expected output is a table showing:

- Database Size

- Probability of Bases

- Motif Size

- Motif

- Motif Occurrences

- Time to Find Motif

To obtain these outputs, first, it is necessary to define the number of ACGT sequences and their size; these will be randomly created and stored in a database. It is also important to define the probability of finding a given nitrogenous base in the DNA molecule.

Next, the size of the motif to be found should be defined, and the total occurrences of the motifs should be calculated, as well as the time it took the machine to find them.