UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Course: Systems Analysis.

Docent: Carlos Sierra.

Report Bioinformatic workshop

Student:
Devin Santiago Alzate Figueroa

Code:
20231020214

Bogotá, D.C.
September 15 of 2024.

Systemic analysis:

By start, we have the need of build a database with dimenssion 1000<= m <= 2000000, completed with some sequences that have size 5 <= n <= 12, having a invidual elements the characters “A”, “C”, “T” and “G”, that are the most frequeant elements of nulcleotides in a genetic sequence.

- Divide and Conquer:
  This paradigm was used in the procces of build the code, because that help to parameterize the methods and divide the use of them, the uses was there:

  -In the procces of methos codified, stablish a only responsability of each method, like databases and createTxt, that even though made similar tasks, one its in charge of build the structure, and another for print them.

  -For build the sequence, we divide the number parameteriz for each character, also, stablish the number of sequences that we need create for the database and size of that sequence.

Chaos analysis

In the experiments conducted, a noticeable observation is that the occurrence of repetitive sequences is significantly low compared to the total number of sequences generated. This result highlights an interesting aspect of the sequence generation process, where despite the large volume of sequences, the number of identical or highly similar sequences remains minimal.

- Implications for Sequence Generation:

  The low frequency of repetitive sequences may have implications for the design and analysis of experiments involving random sequence generation. It suggests that the generated sequences

cover a broad range of possibilities, minimizing redundancy and enhancing the variability of the databases.

Results

In that part, have some tables, that represent the results of the code, with some variations in the input variables, like size of sequence, size of database, size of motif searched, etc.

Database_size	Motif_size	Motif	Motif_ocurrences	Time_to_find_it
1000	3	TCT	45	13ms
1980	5	CGGCA	10	57ms
5300	23	AGTGCAGCCCAGGGTCTCGGAGG	1	167ms
1000300	61	TGTGTTTACCTCTTGTTCCCTCACTCCGAGCATTTTGGGAGTGACGTTGTACGTTTCCTC	16	11918ms
1050300	48	CCGAGTTGACGCCTGCGATTTCCTTACTCGTACACATTGTGGGGAAAC	33	17105ms

Conclutions

With the workshop of bioinformatic we can see that make a search in a database of biggest size, means have a big time to search in them.

Also, it solidifed the idea of entropy, giving us a prove to at more high was the entropy calculous, more high be the variation in the results, and also, more chaotic.