T.C.

MARMARA UNIVERSITY

FACULTY of ENGINEERING

COMPUTER ENGINEERING DEPARTMENT

CSE4088

Introduction to Computational Genomics

Assignment 1

150116034 – Enes Garip

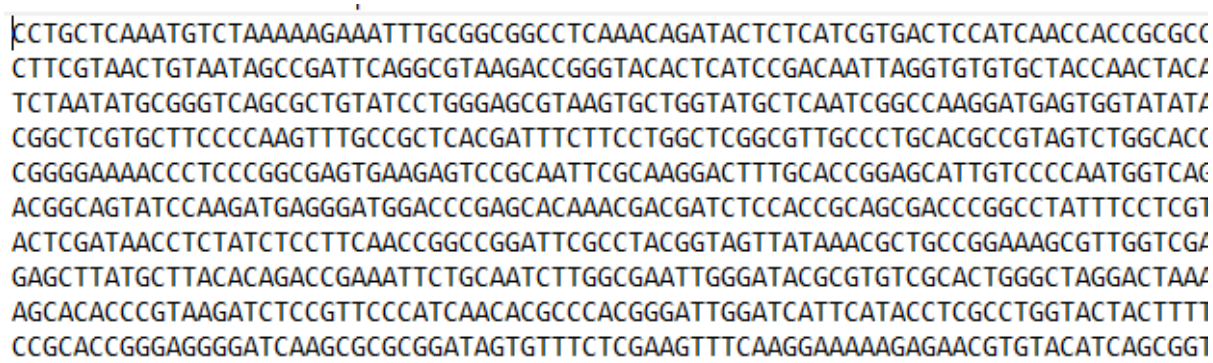
150117062 – Ahmet Tunahan Cinsoy

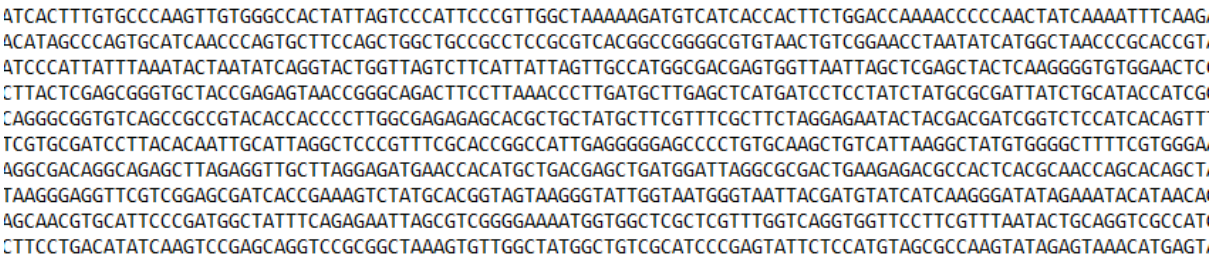
150116013 – Muhammet Yasin Tufan

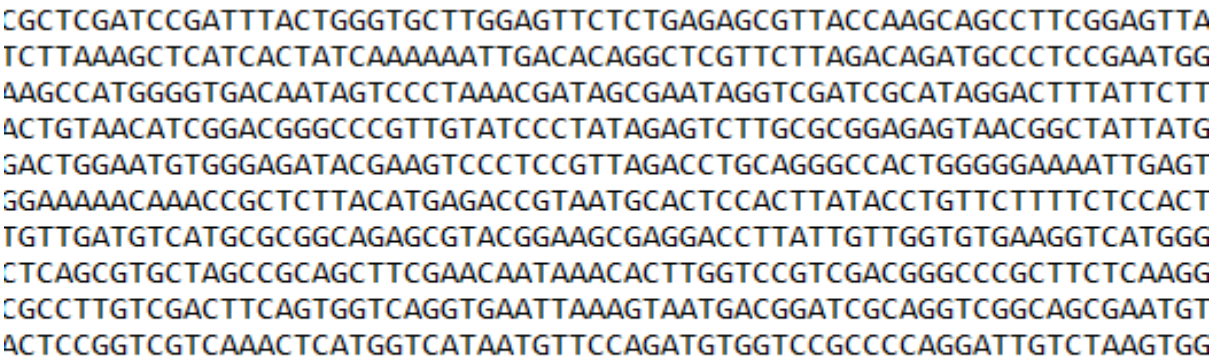
Input File Format

We produce randomly generated “DNA.txt” file as mentioned in the documentation.

You can see a part of the generated “DNA.txt” files as listed below.







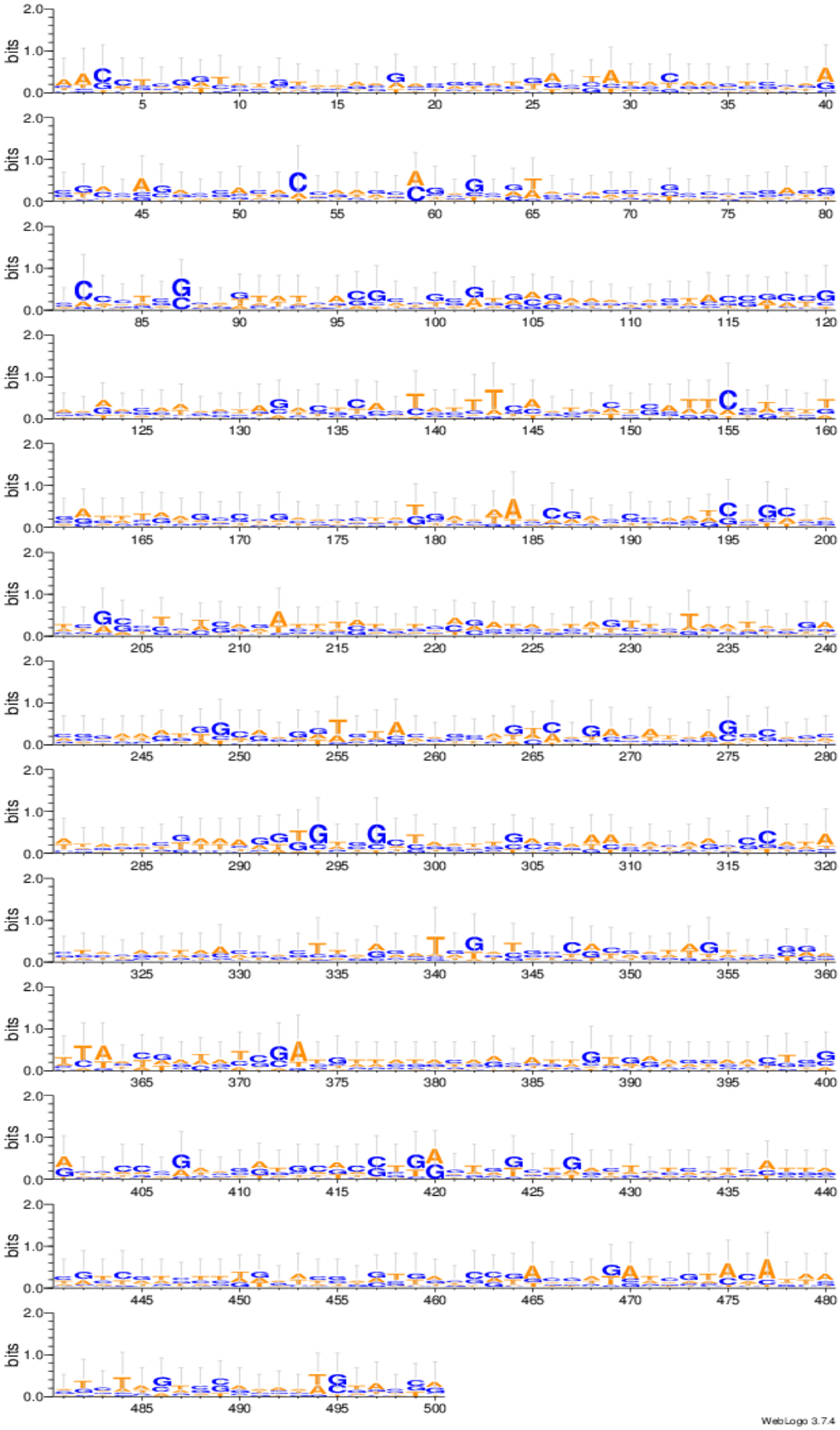
Output Format

We take three different k values as mentioned in the documentation then we calculate the mutated strings, scores, consensus strings for Randomized Motif Search and Gibbs Algorithm. You can see sample output below.

```
Enter k value: 11
Original String:TCGTTGTCTA
Mutated Strings:
                ATGTTCGCTA
                TCCCTGCTTA
                TCGTGGCATT
                TCAATCTCTC
                TGGGTCTCAA
                GCGGTGTGTG
                TTCCTGTCTT
                TGATTATATA
                TCGGTTGATA
                TGCTTGAATA

--------- RANDOMIZED MOTIF SEARCH ---------
Randomized Motif Search Score: 30
Randomized Motif Search Consensus: ACAGTCTGTTA
                ATAGGATGTTA
                ACAGTCAGTTC
                CCAGTGTGTTA
                AAAGGGTTTTA
                ATAATATGTTC
                AGATGCTGTAC
                TCATTCTGTAC
                GCAATCTTGGC
                TTAGACAGTTA
                AGAGGGTGTTA

-------------- GIBBS SAMPLER --------------
Gibbs Sampler Score: 31
Gibbs Sampler Consensus: ATACGCAAGGA
                ATCTGCAAGGC
                AAACGCATGGA
                AAACGCCTGGA
                ACATCCCAGGA
                ATTCGCAAGGA
                ACACGGGAGGC
                ATATGCAGGGC
                ATACGCGTGTC
                GTACCCGAGGA
                GTATGCTTGGA

Process finished with exit code 0
```

Table of Score Values

k	Run	1	2	3	4	5
9	RMS	21	19	22	23	21
	GS	17	26	20	17	16
10	RMS	27	27	27	28	22
	GS	26	26	14	25	32
11	RMS	32	31	31	33	33
	GS	28	27	34	32	31


bits
2.0
1.0
0.0
5 10 15 20 25 30 35 40
45 50 55 60 65 70 75 80
85 90 95 100 105 110 115 120
125 130 135 140 145 150 155 160
165 170 175 180 185 190 195 200
205 210 215 220 225 230 235 240
245 250 255 260 265 270 275 280
285 290 295 300 305 310 315 320
325 330 335 340 345 350 355 360
365 370 375 380 385 390 395 400
405 410 415 420 425 430 435 440
445 450 455 460 465 470 475 480
485 490 495 500
WebLogo 3.7.4

Implementation

First, we randomly generate a DNA String and write that string into the “DNA.txt” file. After that, we implement Randomized Motif Search algorithm. In that algorithm, we try to choose a random motif and keep that data into the motifs matrix.

After that, we implement Gibbs Sampler algorithm. In that algorithm, we erased a line that selected randomly from the matrix. Then, we generate profile matrix. At the end, we calculate the result.

Conclusion

In conclusion, we can understands from the results that Gibbs Algorithm most of the times gives better results if you compare with Randomized Motif Search algorithm. In addition to that, we observe that if the k value increase, score of the algorithms grows with the k values.