



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.

```

CCTGCTCAAATGTCTAAAAAGAAATTTGCGGCGGCCTCAAACAGATACTCTCATCGTGACTCCATCAACCACCGCGCC
CTTCGTAAGTGAATAGCCGATTAGGCGTAAGACCGGGTACACTCATCCGACAATTAGGTGTGTGCTACCAACTACA/
TCTAATATGCGGGTCAGCGCTGTATCCTGGGAGCGTAAGTGCTGGTATGCTCAATCGGCCAAGGATGAGTGGTATATA/
CGGCTCGTGCTTCCCCAAGTTTGCCGCTCACGATTTCTTCTGGCTCGGCGTTGCCCTGCACGCCGTAGTCTGGCACC
CGGGGAAAAACCTCCCGGCGAGTGAAGAGTCCGCAATTGCAAGGACTTTGCACCGGAGCATTGTCCCCAATGGTCA/
ACGGCAGTATCCAAGATGAGGGATGGACCCGAGCACAAACGACGATCTCCACCGCAGCGACCCGGCCATTTCCTCGT
ACTCGATAACCTCTATCTCCTTCAACCGCCGGATTGCGCTACGGTAGTTATAAACGCTGCCGGAAAGCGTTGGTCG/
GAGCTTATGCTTACACAGACCGAAATTCTGCAATCTTGGCGAATTGGGATACGCGTGTGCGACTGGGCTAGGACTAA/
AGCACACCCGTAAGATCTCCGTTCCCATCAACACGCCCACGGGATTGGATCATTACATCTCGCTGGTACTACTTTT
CCGCACCGGGAGGGGATCAAGCGCGCGGATAGTGTTCCTCGAAGTTTCAAGGAAAAAGAGAACGTGTACATCAGCGGT

```

```

ATCACTTTGTGCCAAGTTGTGGGCCACTATTAGTCCATTCCCGTTGGCTAAAAAGATGTCATCACCCTTCTGGACAAAACCCCAACTATCAAAATTTCAAG
ACATAGCCAGTGTCATCAACCCAGTGCTTCCAGCTGGCTGCCGCTCCGCGTCACGGCCGGGGCGTGTAACTGTCGGAACCTAATATCATGGCTAACCCGACCGT
ATCCCATTTTAAATACTAATATCAGGTACTGGTGTAGTCTTATTATTAGTTGCCATGGCGACGAGTGGTAATTAGCTCGAGCTACTCAAGGGGTGTGGAACCTC
CTTACTCGAGCGGGTGTACCGAGAGTAACCGGGCAGACTTCCTTAAACCTTGATGCTTGAGCTCATGATCCTCTATCTATGCGCGATTATCTGCATACCATCG
CAGGGCGGTGTGAGCCGCGTACACACCCCTTGCGGAGAGAGACGCTGCTATGCTTCGTTTCGCTTCTAGGAGAATACTACGACGATCGGTCTCCATCACAGTT
TCGTGCGATCCTTACACAATTGCATTAGGCTCCGTTTCGCACCGGCCATTGAGGGGGAGCCCTGTGCAAGCTGTCATTAAAGGCTATGTGGGGCTTTTCGTGGGA
AGGCGACAGGCAGAGCTTAGAGGTTGCTTAGGAGATGAACCACATGCTGACGAGCTGATGGATTAGGCGCGACTGAAGAGACGCCACTCACGCAACCAGCACAGCT
TAAGGGAGGTTTCGTGGAGCGATCACCGAAAGTCTATGCACGGTAGTAAGGGTATTGGTAATGGGTAATTACGATGTATCATCAAGGGATATAGAAATACATAACA
AGCAACGTGATTCCCGATGGCTATTTAGAGAAATTAGCGTCGGGGAAAAATGGTGGCTCGCTCGTTTGGTCAGGTGGTTCCTTCGTTTAATACTGCAGGTCGCCAT
CTTCTGACATATCAAGTCCGAGCAGGTCCGCGGCTAAAGTGTGGCTATGGCTGTGCGATCCCGAGTATTCTCCATGTAGCGCCAAGTATAGAGTAAACATGAGT

```

```

CGCTCGATCCGATTTACTGGGTGCTTGGAGTTCTCTGAGAGCGTTACCAAGCAGCCTTCGGAGTTA
TCTTAAAGCTCATCACTATCAAAAAATTGACACAGGCTCGTTCTTAGACAGATGCCCTCCGAATGG
AAGCCATGGGGTGACAATAGTCCCTAAACGATAGCGAATAGGTCGATCGCATAGGACTTTATTCTT
ACTGTAACATCGGACGGGCCCCTTGTATCCCTATAGAGTCTTGCGCGGAGAGTAACGGCTATTATG
GACTGGAATGTGGGAGATACGAAGTCCCTCCGTTAGACCTGCAGGGCCACTGGGGGAAAATTGAGT
GGAAAAACAAACCGCTCTTACATGAGACCGTAATGCACTCCACTTATACCTGTTCTTTTCTCCACT
TGTTGATGTCATGCGCGGCAGAGCGTACGGAAGCGAGGACCTTATTGTTGGTGTGAAGGTCATGGG
CTCAGCGTGCTAGCCGCAGCTTCGAACAATAAACACTTGGTCCGTCGACGGGCCCCGCTTCTCAAGG
CGCCTTGTCGACTTCAGTGGTCAGGTGAATTAAGTAATGACGGATCGCAGGTGGCAGCGAATGT
ACTCCGGTCGTCAAACCTCATGGTCATAATGTTCCAGATGTGGTCCGCCCCAGGATTGTCTAAGTGG

```

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:
    ATGTTGCTA
    TCCCTGCTTA
    TCGTGGCATT
    TCAATCTCTC
    TGGGTCTCAA
    GCGGTGTGTG
    TTCCTGTCTT
    TGATTATATA
    TCGGTTGATA
    TGCTTGAATA

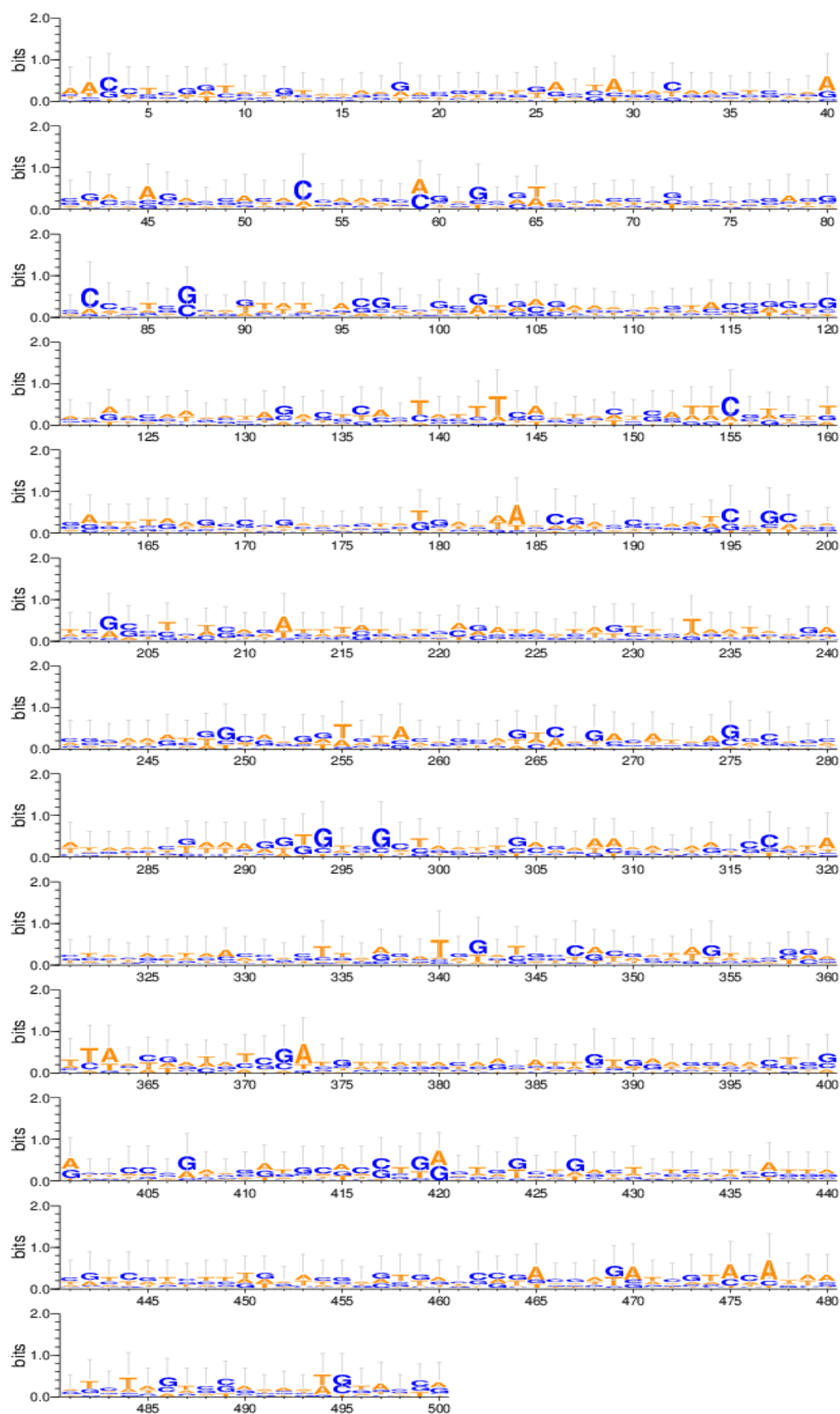
----- RANDOMIZED MOTIF SEARCH -----
Randomized Motif Search Score: 30
Randomized Motif Search Consensus: ACAGTCTGTTA
    ATAGGATGTTA
    ACAGTCAGTTC
    CCAGTGTGTTA
    AAAGGGTTTTA
    ATAATATGTTC
    AGATGCTGTAC
    TCATTCTGTAC
    GCAATCTTGGC
    TTAGACAGTTA
    AGAGGGTGTTA

----- 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



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 understand 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.