CSE 4065

Computational Genomics

Project 1

Report

Melisa DÖNMEZ -- 150116030

Kevser İLDEŞ -- 150116048

Firstly, for both algorithms, we run corresponding search methods 50 times in code and get best result among all.

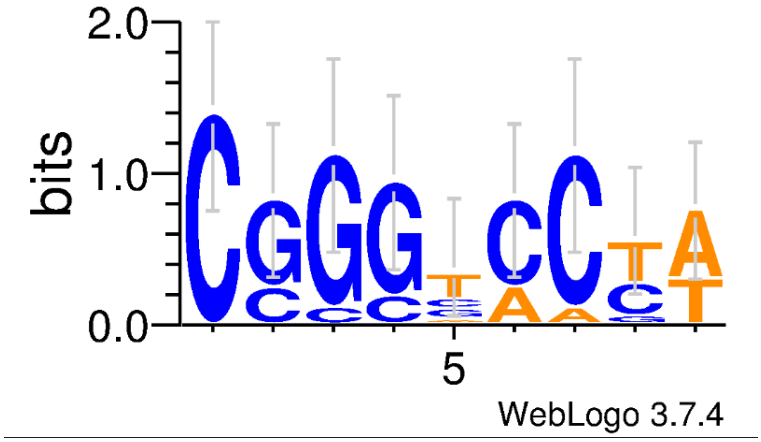
Randomized Motif Search

For randomized motif search some example outputs and consensus string for k=9,10,11 is provided below:

- K=9

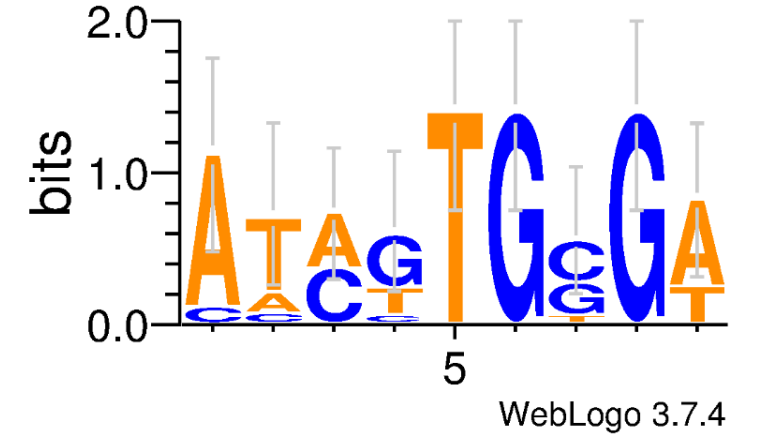
```
Please enter value of k (9,10 or 11): 9

Please enter txt file name: dataset.txt
CGGGGCCCA
CCGCGCCCA
CGGCTACTA
CGGGTAACA
CGGGTCCTT
CGGGTCCTT
CCGGCACGA
CGGGCCCTT
CCGGACCTT
CGCGTCCCA
24
Max score:  34
Average:  28.4
```




```
Please enter value of k (9,10 or 11): 9

Please enter txt file name: dataset.txt
AACGTGCGA
ATAGTGCGA
ATATTGGGA
AAATTGGGA
ACCGTGCGT
ATCGTGCGT
ATCGTGCGT
ATATTGGGA
CTCGTGGGA
ATACTGTGA
21
Max score:  36
Average:  27.6
```



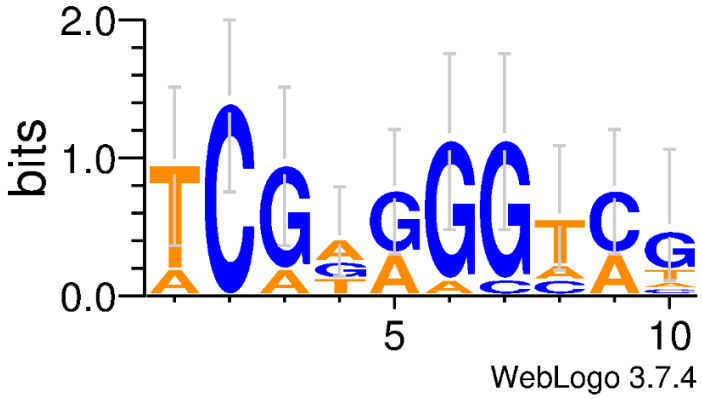
Besides, here is a table for 5 runs of our code for each k value = 9

K = 9	Best Score	Worst Score	Average Score	Consensus String
	23	34	28.16	TGTTAGTGC
	24	34	28.6	TATTCACTC
	21	33	28.24	TCTTTAGCG
	23	34	28.3	ACTCTGCGT
	22	35	27.46	GGTTAGTGC

- K=10

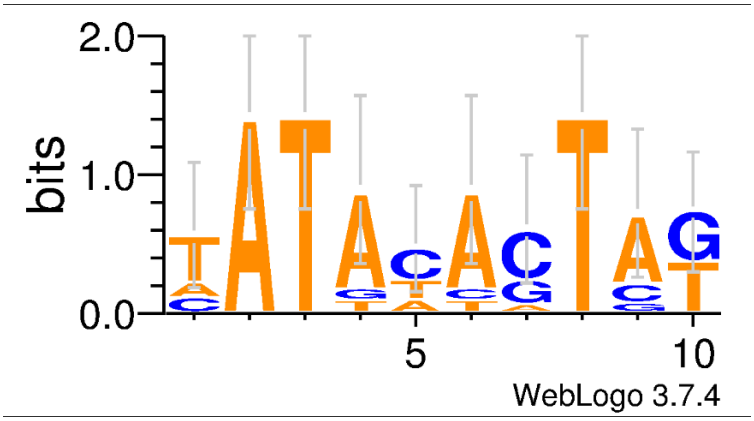
```
Please enter value of k (9,10 or 11): 10

Please enter txt file name: dataset.txt
TCGGGGGTCA
TCAAAGGCCT
ACGTAGGTCT
TCGTGGGTCC
TCGGGGGACG
TCGAGGCTAG
ACGAAGGTAG
TCAAAAGCAG
TCGTGGGACG
TCGGGGGTAG
28
Max score:  40
Average:  33.82
```




```
Please enter value of k (9,10 or 11): 10

Please enter txt file name: dataset.txt
TATAAACTAG
TATTCACTCG
AATATCGTAT
CATGCACTAG
TATAAAATAG
TATATACTAT
TATACAGTAG
CATACTCTGT
TATACAGTCT
AATATACTAT
25
Max score:  42
Average:  33.7
```



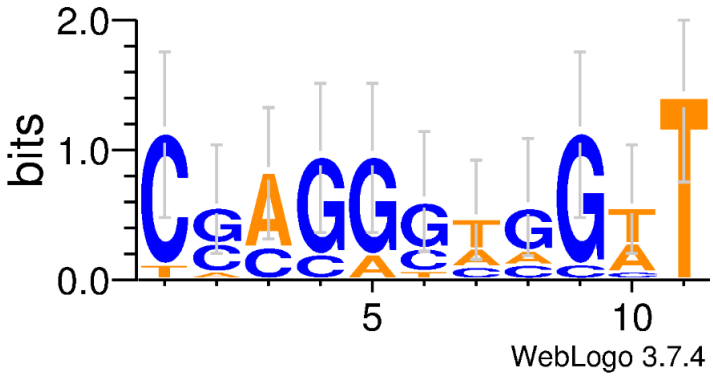
Besides, here is a table for 5 runs of our code for each k value = 10

K = 10	Best Score	Worst Score	Average Score	Consensus String
	27	40	33.12	GCCTTTTAGC
	26	37	32.92	ATTCGGTGAG
	24	43	33.9	TGACCCGGGA
	28	40	33.5	TGTTGTCTCG
	26	44	33.18	TATACACTCA

- K=11

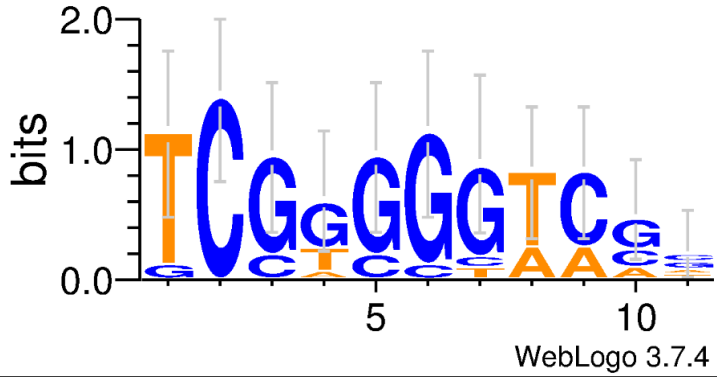
```
Please enter value of k (9,10 or 11): 11

Please enter txt file name: dataset.txt
CGAGAGTGGTT
CCACGGTCGAT
CGAGGGAGGTT
TCCGGGCGGAT
CCCGGGTCCTT
CGAGGCTAGAT
CAACGGAGGCT
CCAGGTTGGAT
CGCGACCGGTT
CGAGGCAAGTT
32
Max score:  49
Average:  38.74
```




```
Please enter value of k (9,10 or 11): 11

Please enter txt file name: dataset.txt
TCGGGGGTCAT
TCCACGGTCGA
GCCGGGGTCGG
TCGTGGGTCCG
TCGGGGGACGC
TCGTGCGTAAA
TCGGCGTTCCT
TCGGGGCAACG
TCGTGGGACGC
TCGGGGGTAGC
30
Max score:  47
Average:  39.56
```



Besides, here is a table for 5 runs of our code for each k value = 11

K = 11	Best Score	Worst Score	Average Score	Consensus String
	32	47	39.86	GTTTAATTCGA
	33	46	39.96	GAGGCGAGTGG
	30	44	38.54	CCCCCCGTGCA
	31	43	39.3	TATTCGTCCAA
	32	46	39.32	GCGAGGTTCAA

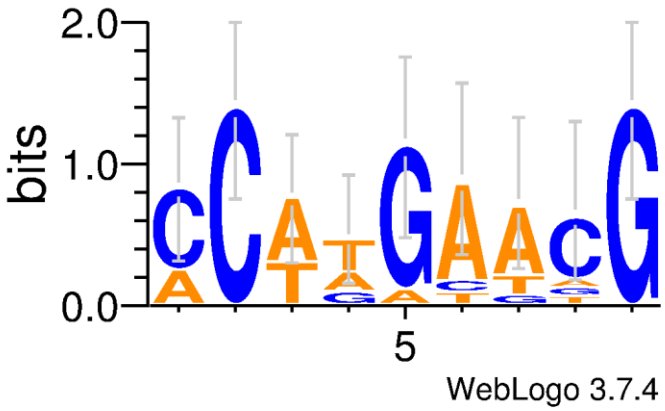
Gibbs Sampler

For gibbs sampler algorithm some example outputs and consensus string for k=9,10,11 are provided below:

- K = 9

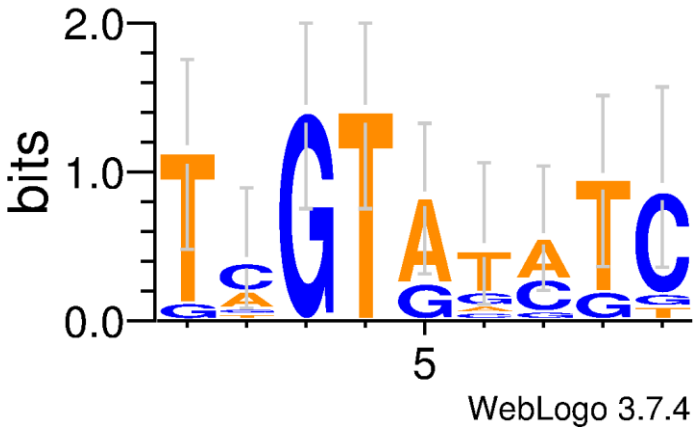
```
Please enter value of k (9,10 or 11): 9
Please enter txt file name: dataset.txt
Motifs
--------------
CCAGGAACG
CCATGAACG
CCTTGAACG
ACATGATCG
CCATGAGCG
CCTTGCACG
ACTAGAACG
CCAGGTTGG
CCTAGAAAG
ACAAAAATG

Score Result
--------------
Best Score :  21
Max Score:  41
Average Score:  31.52
```



```
Please enter value of k (9,10 or 11): 9
Please enter txt file name: dataset.txt
Motifs
--------------
TTGTGACTC
TGGTATATG
TAGTATCGC
TCGTATATC
TCGTAGCTC
TCGTATATC
TCGTAGGTT
TAGTATAGC
GAGTGCATC
TCGTGTCTC

Score Result
--------------
Best Score :  22
Max Score:  39
Average Score:  31.3
```



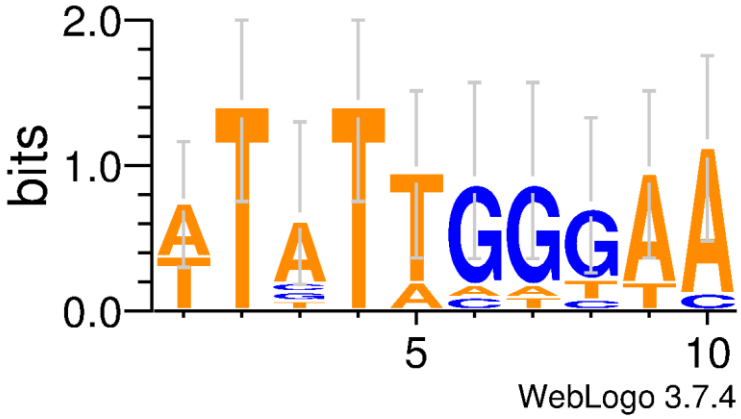
Besides, here is a table for 5 runs of our code for each k value = 9

K = 9	Best Score	Worst Score	Average Score	Consensus String
	21	41	31.52	CCATGAACG
	22	39	31.3	TCGTATATC
	23	41	31.7	GAGTGCATT
	23	40	31.50	ATATACAAT
	24	41	31.52	TATCGGGGG

- K = 10

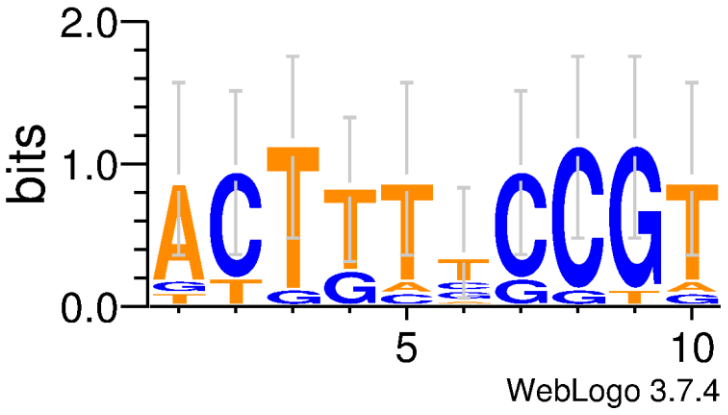
```
Please enter value of k (9,10 or 11): 10
Please enter txt file name: dataset.txt
Motifs
--------------
TTTTAGAGAC
TTATTGGCTA
ATATTGGGAA
TTATTGGTAA
TTATTGGTAA
ATATTGGGAA
ATATTGGGAA
ATATTGGGAA
ATCTTAGGAA
TTGTACTGTA

Score Result
--------------
Best Score :  20
Max Score:  44
Average Score:  35.04
```




```
Please enter value of k (9,10 or 11): 10
Please enter txt file name: dataset.txt
Motifs
--------------
TTTGTCCCTT
ACTTCTGCGA
ACTGTGCCGT
GCTTTAGGGT
ACTTTTCCGT
ACTTTTCCGT
ACTTTTCCGT
ACTTTTCCGT
ATGTTCCCGT
ACTGAGCCGG

Score Result
--------------
Best Score :  21
Max Score:  41
Average Score:  34.38
```



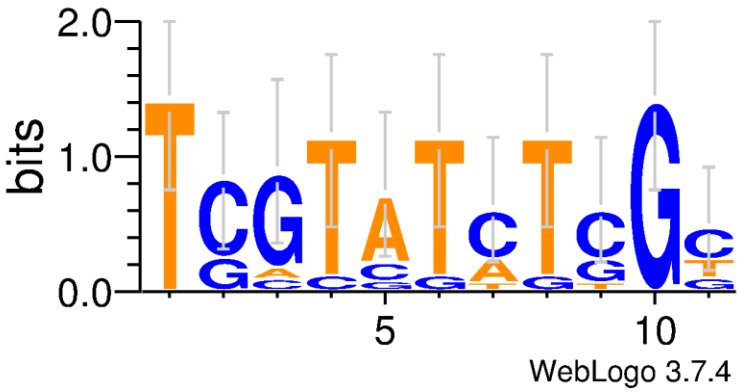
Besides, here is a table for 5 runs of our code for each k value = 10

K = 10	Best Score	Worst Score	Average Score	Consensus String
	20	44	35.04	ATATTGGGAA
	21	41	34.38	ACTTTTCCGT
	25	44	35.48	TATACTGTGA
	26	45	35.38	TAGTCGAATA
	27	41	35.28	CGTTTCCACA

- K = 11

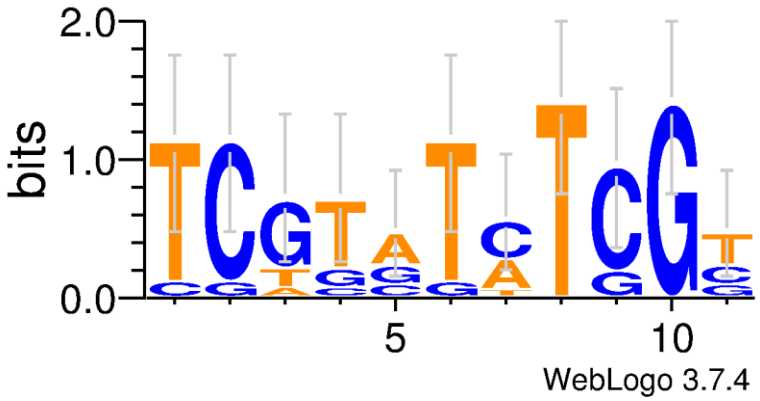
```
Please enter value of k (9,10 or 11): 11
Please enter txt file name: dataset.txt
Motifs
--------------
TGGTCGCTCGC
TGGTATATGGG
TCGTATTTGGC
TCGTATATCGC
TCATCTCTCGC
TCGTATATCGT
TGGTATCGTGC
TCGCATCTCGG
TCCTATCTGGT
TCGTGTCTCGT

Score Result
--------------
Best Score :  24
Max Score:  51
Average Score:  38.94
```



```
Please enter value of k (9,10 or 11): 11
Please enter txt file name: dataset.txt
Motifs
--------------
TCGTCGCTCGT
TGGTATATGGG
TCGTATTTGGC
TCGTATATCGC
TCATCTCTCGC
TCGTATATCGT
CCTGGTATCGT
TCGCATCTCGG
TCTGGTCTCGT
TCGTGTCTCGT

Score Result
--------------
Best Score :  26
Max Score:  45
Average Score:  38.42
```



Besides, here is a table for 5 runs of our code for each k value = 11

K = 11	Best Score	Worst Score	Average Score	Consensus String
	24	51	38.94	TCGTATCTCGC
	26	45	38.42	TCGTATCTCGT
	29	48	38.6	AGAACGTGCCA
	30	48	39.38	TAACCCGTCTC
	31	48	39.52	ATACTGTTATC

Conclusion

As a result, we noticed that when k values increased, the scored increased for both algorithms. We can understand that these algorithms create consensus strings closer to the original string at small k values. If we compare the two algorithms, we can say that Gibbs Sampler algorithm works better. We got our best score with Gibbs Sampler that is 20. Also, best score in the randomized motif search is 21.