

CSE 4065
Computational Genomics
Programming Assignment 1



ŞÜKRİYE SOYER 150116010

BERRA MERCAN 150117034

Randomized Motif Search

Results of Randomized Motif Search algorithm for k=9:

```
Problems @ Javadoc  
<terminated> RandomMotifS  
GTAGCCTAT  
AGTGGTTAT  
AGAACTTAT  
GTAGGAGAT  
GTAAGTGAT  
GTTAGCGAT  
GTAGCTGAT  
GTAGCTGAT  
ATAGCTGAG  
  
BestScore= 23  
Average Score= 23.8  
Max score= 49  
GTAGCTGAT
```

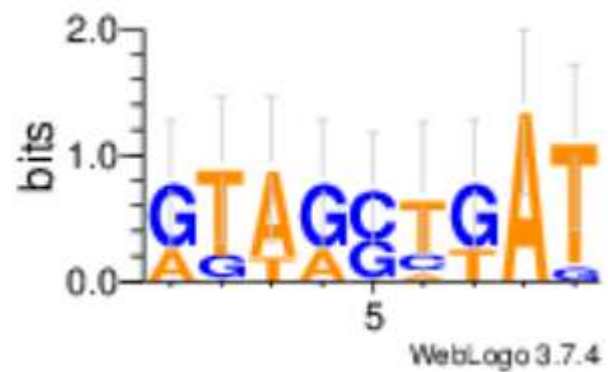


Table for 10 runs of our code for each k value=9:

| K=9 | Best Score | Worst Score | Average Score | Consensus String |
|-----|------------|-------------|---------------|------------------|
| | 28 | 57 | 28.7 | CCATATAGC |
| | 23 | 49 | 23.8 | GTAGCTGAT |
| | 30 | 50 | 30.4 | CAGCCAATT |
| | 28 | 53 | 28.52 | CCAGTATAG |
| | 24 | 54 | 24.9 | GACGTGACT |
| | 25 | 51 | 25.56 | GCGGCACTT |
| | 31 | 49 | 31.44 | AGCCAAGAG |
| | 27 | 54 | 27.74 | GATTGCCGA |
| | 24 | 51 | 24.62 | GAGCATGCG |
| | 27 | 52 | 28.46 | CGCTAGTTG |

Results of Randomized Motif Search algorithm for k=10:

```

Problems @ Javadoc
<terminated> RandomMotifS
|
TGCACTATAA
TAACTATAG
TGAAATCTAG
TAAAGTTTAG
GTAAC TTAT
TTCAGTATAA
ATAACTCCAG
TGGACTATAT
GACAGTTTAG
GTGACTATAG

BestScore= 29
Average Score= 29.68
Max score= 56
TTAACTATAG

```

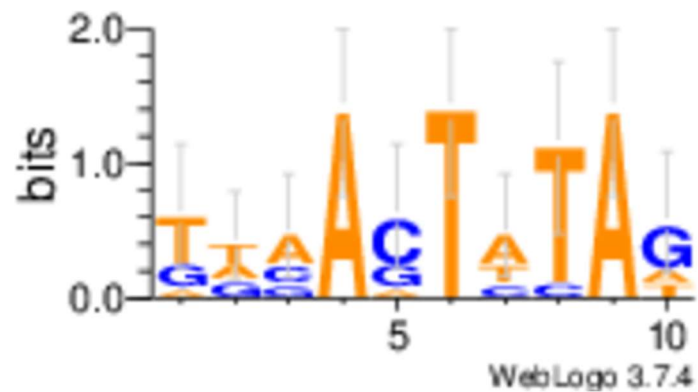


Table for 10 runs of our code for each k value=10:

| K=10 | Best Score | Worst Score | Average Score | Consensus String |
|------|------------|-------------|---------------|------------------|
| | 33 | 55 | 33.44 | GAAGAACTAA |
| | 30 | 62 | 30.98 | TCCTAGGGTA |
| | 33 | 53 | 33.48 | GTACATTAGA |
| | 33 | 56 | 33.52 | ATAAACATCG |
| | 33 | 59 | 34.5 | TGATTTATGA |
| | 39 | 59 | 39.4 | GCATTAGCGC |
| | 32 | 57 | 32.8 | CGTAGCTCCA |
| | 29 | 56 | 29.68 | TTAACTATAG |
| | 38 | 58 | 38.48 | AGGCAACGTC |
| | 38 | 61 | 38.46 | TGAGCATTGA |

Results of Randomized Motif Search algorithm for k=11:

```

Problems @ Javadoc
<terminated> RandomMotifSe

TCTGCACTATA
GAGACGCATTA
TAGGCGCACTC
TCTACGCATTA
TCCGCGCTGTC
TCGGCCCTCTA
TAAGCGCATGC
TAAGCGCATGC
TAGGCGCTATA
GAGGCTCGATC

BestScore= 34
Average Score= 34.62
Max score= 58
TAGGCGCATTA

```

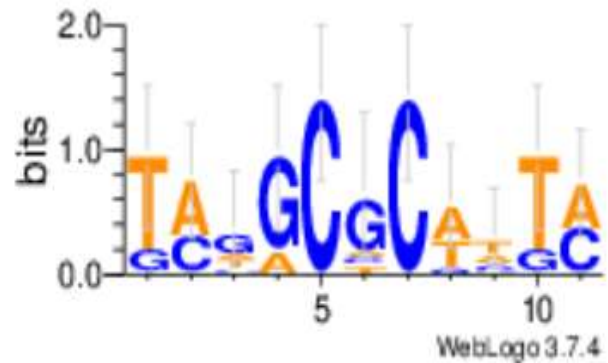


Table for 10 runs of our code for each k value=11:

| K=11 | Best Score | Worst Score | Average Score | Consensus String |
|------|------------|-------------|---------------|------------------|
| | 37 | 65 | 37.56 | AGTTAGACTGA |
| | 38 | 67 | 38.58 | TTCTTCCATTC |
| | 34 | 58 | 34.62 | TAGGCGCATTA |
| | 39 | 59 | 39.4 | ATAGCTAGCCA |
| | 39 | 65 | 39.52 | TGTCAGACCGA |
| | 37 | 58 | 37.42 | TTTGATTGCCT |
| | 41 | 67 | 41.52 | GCATTAGACGG |
| | 35 | 65 | 35.6 | TCGGACCATGC |
| | 38 | 61 | 38.52 | TTCGACTATCC |
| | 37 | 60 | 37.46 | GTCTTACGTGT |

Gibbs Sampler

Results of Gibbs Sampler algorithm for k=9:

```
Problems @ Javadoc  
<terminated> GibbsSampler [J  
TGCTCTGCA  
TGCCAGGCT  
TGCCGTGCT  
CGCCATGAG  
TGTCATGCA  
TGCCATGCG  
TGCCATGCC  
TGCCATGCA  
TGTCGTGCG  
TGTCGTGCA  
  
BestScore= 17  
Average Score= 22.84  
Max score= 53  
TGCCATGCA
```

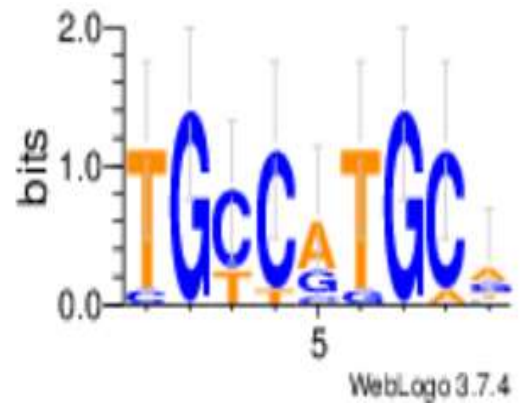


Table for 10 runs of our code for each k value=9:

| K=9 | Best Score | Worst Score | Average Score | Consensus String |
|-----|------------|-------------|---------------|------------------|
| | 22 | 52 | 29.94 | CGAATTCAG |
| | 22 | 47 | 26.88 | CATGTGAGA |
| | 22 | 52 | 26.98 | TCCAAAACC |
| | 21 | 53 | 28.02 | GTCGGGAAC |
| | 19 | 55 | 28.88 | TGACGTGAC |
| | 27 | 54 | 32.36 | GCGTTATAT |
| | 19 | 52 | 27.14 | CCAATTGTA |
| | 17 | 53 | 22.84 | TGCCATGCA |
| | 25 | 48 | 29.04 | CAATTTAAG |
| | 22 | 51 | 26.46 | GACACGTGA |

Results of Gibbs Sampler algorithm for k=10:

```

Problems Javadoc
<terminated> GibbsSampler [Java]
TAGCGGAAGT
CAGCGTGAGT
TGGCGTTACT
TAGCGAGAAT
TAGCGTGGGC
CGGAGTAAGT
TAGCGTAATT
TAGCTAAGGT
TAGCGTAGGT
TAGCGTCAGT

BestScore= 21
Average Score= 28.84
Max score= 59
TAGCGTAAGT

```

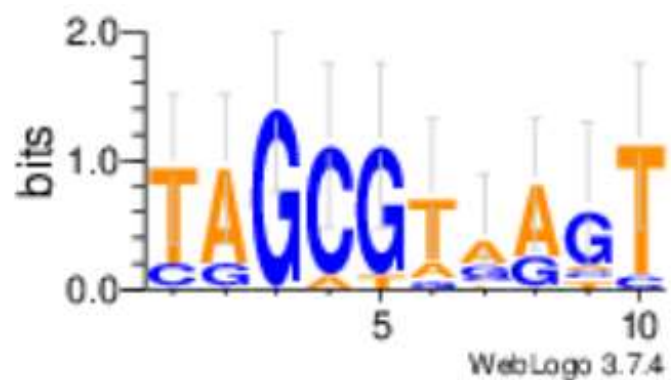


Table for 10 runs of our code for each k value=10:

| K=10 | Best Score | Worst Score | Average Score | Consensus String |
|------|------------|-------------|---------------|------------------|
| | 27 | 57 | 33.04 | CAACTGTAGC |
| | 27 | 56 | 31.58 | CAGAATGCGG |
| | 28 | 55 | 33.98 | TCGCGAGATT |
| | 21 | 59 | 28.84 | TAGCGTAAGT |
| | 23 | 54 | 30.08 | ACCAATAGGC |
| | 22 | 55 | 27.96 | AATTTAGCCC |
| | 32 | 65 | 36.6 | CTGGGATCGT |
| | 30 | 56 | 34.4 | TCAATCGTTT |
| | 24 | 58 | 32.32 | CTAGGCGAGT |
| | 24 | 56 | 31.46 | AAATGTCGCT |

Results of Gibbs Sampler algorithm for k=11:

```

Problems @ Javadoc
<terminated> GibbsSampler [Jav
GGAGCTGTATA
CGACATGTGAG
CGACATATGAG
TGA CTTGTAAG
GGACATGTAAA
GAAGATGTAAG
GGACACGTGAG
GGACACATAAA
TGACAGTTTAG
GGAAACTTGAG

BestScore= 27
Average Score= 36.66
Max score= 67
GGACATGTAAG

```

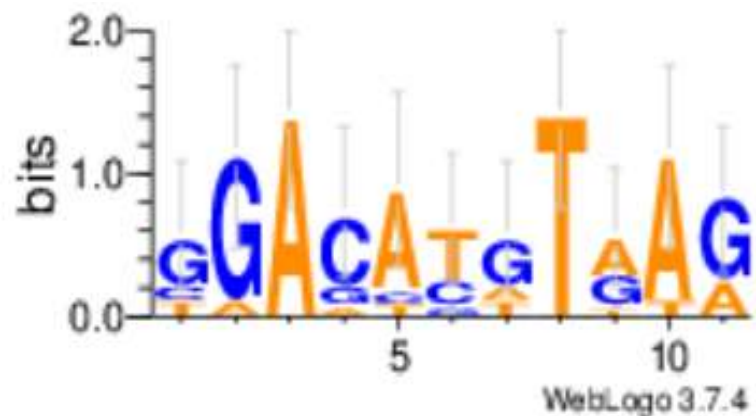


Table for 10 runs of our code for each k value=11:

| K=11 | Best Score | Worst Score | Average Score | Consensus String |
|------|------------|-------------|---------------|------------------|
| | 27 | 67 | 36.66 | GGACATGTAAG |
| | 31 | 65 | 39.34 | CCATCCAATTC |
| | 32 | 64 | 39.24 | GCTGAGGGCAG |
| | 31 | 65 | 36.28 | AGCAATACGCG |
| | 29 | 63 | 36.98 | GGAGAGTCATA |
| | 31 | 61 | 36.12 | TAAACAACTAG |
| | 32 | 65 | 38.26 | GATTCATACCA |
| | 36 | 58 | 41.24 | CGTACATAAGCT |
| | 28 | 62 | 34.64 | TAGCTAAAGTTT |
| | 30 | 64 | 36.54 | GCGCATTCGTT |

Conclusion

As a result, we noticed that the Gibbs algorithm works better than Random Motif Search. When the k value increases, the score increases in both algorithms. Both of algorithms run time is close each other.