CSE 4065 – Computational Genomics

Programming Assignment # 1

Ahmet Hakan Şimşek 150117060

Ahmet Lekesiz 150118509

Mücahit Tanacıoğlu 150115006

In this homework, we are assigned to search motifs and consensus string by implementing randomized motif search and Gibbs sampler algorithms for different k-mer values. We appended 10-mer into Dna string as "AAAAAAAAAA" with up to 4 mutations.

Randomized Motif Search

Implementation
1-First of all, it selects randomly 10 k-mer Motifs from each Dna string
2-From selected k-mers, we construct Profile(Motifs)
3-From created Profile above, we find all possibility of k-mers
4-Change Motifs with highest possibility k-mers
5-Compare score of new motifs and old motifs
6-If new motif's score is better then select them as motif
7-If score have not improve consecutively for 100 times stop algorithm
Go step 2

Gibbs Sampler

Implementation
1-Select randomly 10 k-mer Motifs from each Dna string
2-Remove randomly selected Motif
3-From remaining motifs create Profile(Motifs)
4-For each possible k-mer in removed sequence calculate probability
5-Roll a biased die, select new motif from removed sequence by die
6-If number of iteration with no improvement is 100 then stop algorithm
Go step 2

Usage
python random_search.py --input_file dna_seq.txt -k 11 -m AAAAAAAAAA
python gibb_search.py --input_file dna_seq.txt -k 11 -m AAAAAAAAAA

Randomized Motif Search Tests

9-mer

Inserted mutation:AGTACATACA

Iter 0 : Consensus string:TCATCTCCA number of iteration 5 Min

['TCGTCGACA', 'TCATCCACA', 'TCATACGCA', 'TCTTCTCCA', 'TCATATCCA', 'TCGTCTCCA', 'TAATTTCCA', 'TCATCTCCA', 'TAGTCGCTA', 'TCGTAGACA']

Iter 1 : Consensus string:AAGCGCCAA number of iteration 3 Min

['AACCGCCAA', 'ATCCTTCAA', 'ATGCCCCAG', 'ATGTGCAAT', 'ATCCGACAA', 'AAGTCCAGA', 'AACTCTAGA', 'AACTTCCAA', 'AAGCCTAGG', 'ATGCGCCCT']

Iter 2 : Consensus string:TATTATAAC number of iteration 4 Min

['TAATTGAAC', 'TATTAAAAT', 'TATCAAAAT', 'AAACCTAAC', 'TACGTTAAC', 'AATGCAAAC', 'TGTTAAAAC', 'TCTCTTAAC', 'TATCATTAC', 'TATTCGAAC']

Iter 3 : Consensus string:ACACTCAGT number of iteration 3 Min

['TCACGAAGT', 'ACAGTATGT', 'ATACGCATT', 'CCACCCTGT', 'ACACTCTGT', 'TCGTGCAGT', 'ACATACATT', 'CCACTAAGT', 'GCACTATGT', 'ACACTCACT']

Iter 4 : Consensus string:GAGAGTAAA number of iteration 5 Min

['TATAGCATA', 'GATATTAAA', 'GAGAGCAAG', 'GAGGGTACA', 'TAGAGAAAA', 'TAGAGCGTA', 'TAGATTAAG', 'GTGAGTGAA', 'TTGAGTAGA', 'GAGATCACC']

Iter 5 : Consensus string:TGATAGGTC number of iteration 6 Min

['AGATAGGAC', 'TGATCGGTA', 'TTATCGCTC', 'AGAAAGGTC', 'CCATCAGTC', 'CGATAGCTC', 'TTATAACTC', 'TGCTAGGTA', 'CTATAGCTC', 'TTATCAGAA']

Iter 6 : Consensus string:TTCTGCGTG number of iteration 5 Min

['TTTAGCGTG', 'TTCAGGGTG', 'TCCTATGTG', 'TTCAGTGTG', 'TCGTCGGTT', 'ATGTTTGAG', 'TTCTCCGTG', 'TCGAGCCTG', 'TCGTGCGTG', 'TCCTGGGTG']

Iter 7 : Consensus string:AGGTAAATG number of iteration 4 Min

['GGGTAACAG', 'GGGTCAATG', 'AGGCGAATG', 'AGGGATGTT', 'AGCTAAAAT', 'AGCTAAGTG', 'AGGTATATA', 'AGGTACATG', 'AGGCATTAG', 'AGGGAAATG']

Iter 8 : Consensus string:TATCCAGTC number of iteration 4 Min

['TCTCCAGTC', 'TAGCCAGTG', 'CATCGAGTC', 'TATTCAGTG', 'AAGTCAGTC', 'CATTCAGTC', 'CAGGCAGCC', 'AATCCAGCG', 'TCTCCAGCC', 'CATCCGGCC']

Iter 9 : Consensus string:GAGTTAAAA number of iteration 4 Min

['GAACTCAGA', 'ATATTAAAA', 'GATCTAAGA', 'GGGTTAAAG', 'CAGCTAAAA', 'GATTGAAGA', 'ATGTTAAAA', 'CTGTTAATA', 'GAATTCATA', 'GCGTGAAGA']

9-mer

Inserted mutation:AAAAAAAAAA

Iter 0 : Consensus string:GCCGTAAAA number of iteration 3 Min

['GTTGTGAAA', 'GTCGTCGAA', 'GCTGTAAGA', 'GTCGGTAGA', 'GCAATAAAA', 'GTTGGCAAA', 'CCTGGGAGA', 'CCCGTAAGA', 'GCCAGAAAA', 'GCCGTAAAT']

Iter 1 : Consensus string:GAGCCCAAG number of iteration 3 Min

['GAGCCATGC', 'GTACCCACG', 'GAACCCGAT', 'GAGACCGCA', 'GAACCAAAG', 'GAGGCCAAC', 'GAGCCCGAG', 'GTACCCTGG', 'GAGCGCGCG', 'GAGCCAACT']

Iter 2 : Consensus string:AAAAAACAA number of iteration 6 Min

['AAAAAACAA', 'AAAAAAAAA', 'AAAAAAAAA', 'GCATGACAT', 'AAAAAACAA', 'AAAAAAAAA', 'AAAAAAAAA', 'ACAGGACAC', 'AAATAAAAA', 'AAATGACAC']

Iter 3 : Consensus string:CGTAAATCG number of iteration 6 Min

['CTAAGACCG', 'CGAAAATCG', 'CATAAATCG', 'CATAGACAG', 'CGTGAATAG', 'CTTAGATGG', 'CGTAAATCG', 'CGAGGTTCG', 'CGAAATCAG', 'CGTAAATGG']

Iter 4 : Consensus string:TTACAAACC number of iteration 3 Min

['TTTCATACC', 'TTTCATGAC', 'GTACTACTC', 'ATACTAACC', 'TTACGACCG', 'ATTCAAAGG', 'GTTCATCCG', 'ATTCAACCG', 'GCACAAACC', 'TTACTAATC']

Iter 5 : Consensus string:ATAAAGAAA number of iteration 4 Min

['AGAAAGAAA', 'ATAAAGAAA', 'ATAAAGGAA', 'TGAAAGAAA', 'ATAAAGCCA', 'TTAACGGGC', 'TTCAAGAAA', 'ATAAAGCAT', 'AGAAAGAAA', 'AGAAAGGGC']

Iter 6 : Consensus string:ATTGTCCAC number of iteration 4 Min

['ATTGTTCAC', 'ATTGTCGTC', 'TTAGTCTTC', 'AAAGATCAA', 'ATCGTACAC', 'TTAGTCCAC', 'AGTGTTCAA', 'ATTGATTCA', 'AATGTCGAA', 'ATCGTCGAC']

Iter 7 : Consensus string:ACAAAAAAA number of iteration 5 Min

['ACAAAAGAC', 'ACAAATAAA', 'ACAAAAAAA', 'ACAAAAAAA', 'ACAAAAAAA', 'AGAAAAAAA', 'AGAAAAAAA', 'ACAAAAAAA', 'ACAAAAAAC', 'ACAAATGAC']

Iter 8 : Consensus string:AAATGCGCG number of iteration 4 Min

['AAATGCGCG', 'AATTCGGCG', 'GAATCCCGT', 'AAAGGCAGA', 'AAATGCTCT', 'GATTGCGCT', 'AATTGATGA', 'AAAGGCCCA', 'GATTGCCGG', 'AATTGCACG']

Iter 9 : Consensus string:AAAAAAAAA number of iteration 8 Min

['AAAAAAGAA', 'AAAAAAAAA', 'AAAAAAAAA', 'AAAAAAAAA', 'AAAAAAAAA', 'AAAAAAAAA', 'AAAAAAAAA', 'AAAAAAAAA', 'AAAAAAAAA', 'ATCTAAGCG']

10-mer

Inserted mutation:AAAAAAAAAA

Iter 0 : Consensus string:TCTCAAAGTG number of iteration 3 Min

['TCACAACCTT', 'TGTCAATTTG', 'TCACTAGGGT', 'TCTCGTAGCG', 'TCACAGATTG', 'TCTCAAATCG', 'TTTCAATCTT', 'TCTCTGTTTG', 'ACTCGAAGCG', 'TCTCGAAGGC']

Iter 1 : Consensus string:CTTGCGCAAG number of iteration 4 Min

['CGTCCAGACG', 'CTACCGGCCG', 'CTTGGGCAAG', 'CTAGCGAAAT', 'TATGCACAAG', 'CTTCTGCACG', 'CGACCGCGAG', 'CGCCGGAACG', 'CTTGGGAAAG', 'TTAGCTGACG']

Iter 2 : Consensus string:AAAACAACTC number of iteration 4 Min

['AAAACAACGC', 'TACCCGACTT', 'AAGACAACTT', 'GAAACGACTC', 'TAAGAAGTTC', 'AATCGAACTG', 'CATACGGGTC', 'AAACAAACTT', 'AATAAAATTC', 'TGAAGAGCGC']

Iter 3 : Consensus string:TACCCAACGA number of iteration 4 Min

['AGCTCAACTC', 'CACCCAACCC', 'AACTCAAGGT', 'TAGCCTACTC', 'TGCCCAGCGC', 'TACTCAACCT', 'AACCCAACGA', 'AACCCAAGCA', 'TACCCCAGTA', 'TGCTCCACGA']

Iter 4 : Consensus string:CAAGTATGCA number of iteration 3 Min

['CAAATAGGAA', 'AAAGTAGTCA', 'CATGTGTGCA', 'CCAGTCTACT', 'CCAGAGTGCA', 'CGAGTTTGCT', 'TGAATATACA', 'CAAGCAGACA', 'CCAGTATGCT', 'CCCGTATTGA']

Iter 5 : Consensus string:AAAACCACGC number of iteration 3 Min

['AAAACAACGC', 'AATTCCACGC', 'ATGTTCCCTC', 'AAAAACACGC', 'TAGACCACGT', 'AAAACTCCCC', 'AAGTACACGC', 'TAAATCTCTC', 'AAAATTACGC', 'TTGACTTCTC']

Iter 6 : Consensus string:AGAACTCCAC number of iteration 3 Min

['AAAGCTCAAC', 'AGGGCTGTAC', 'GGGGGTCCAA', 'CAGTCTACTC', 'CGGACTCCTC', 'AAAACTCCCC', 'AAAAGTACAC', 'AGAAGTGCCC', 'CGGTCTCCGC', 'AGAGCGCAAC']

Iter 7 : Consensus string:CTGTACAGAG number of iteration 4 Min

['CTGTACCTGG', 'CTATAGAAAC', 'CTGTACCGTG', 'CTCGCCAATG', 'CTTGCCAGAG', 'CTCAACAGAG', 'CTCTCCCGGG', 'CTCTACCGGC', 'CTGTATAGAG', 'CTGAATAAAG']

Iter 8 : Consensus string:AAAAAAAATA number of iteration 4 Min

['AAAAAAAATA', 'AAAAAAAATA', 'AAAAAAAATA', 'AAAAAAAATA', 'AAAAAAAATA', 'AAAAAAAATA', 'AAAAAAAATA', 'AAAAAAAATG', 'AAAAAAAATA', 'TACAAAGAAG']

Iter 9 : Consensus string:CGTGCGAAAA number of iteration 4 Min

['CGTACGCAAA', 'AGTACCAAAA', 'CGTGCTACAG', 'CTAGCGAAAT', 'CTTGCCAGAG', 'CGAGGGAAAA', 'CGAGCAAAAA', 'ATTGCAAAAT', 'CTTGGGAAAG', 'AGAGCGCAAC']

11-mer

Inserted mutation:AAAAAAAAAA

Iter 0 : Consensus string:TCTATTGCTAG number of iteration 5 Min

['TCTACTGCAAG', 'TCTATTTGGAG', 'TCTAGTTCTTA', 'TCAATCGCAAT', 'TCAATTGCTAT', 'TCAATTGGTTG', 'TAAATTTCGAA', 'TCTCCCTCTAG', 'TCTAACGTTAG', 'TCTAGTGGGTT']

Iter 1 : Consensus string:TAAGAAAAAAC number of iteration 5 Min

['GTCTCCAACAC', 'TGAGCAAAAAC', 'TAAGAAAAAAA', 'TACGAAAACAC', 'CTACAAAACTC', 'CAACAAAAATC', 'CACCAAAAAAA', 'TTCCCAAAAAA', 'TTATCAAAAAC', 'TGCGTCACATA']

Iter 2 : Consensus string:TAAAAAAAAAA number of iteration 5 Min

['TAAAAAAAAAA', 'CACAAAAAAAA', 'TAAAAAAAAAA', 'CAAAAAAAAAA', 'TAAAAAAAAAA', 'GAAAAAAAAAA', 'CAAAAAAAAAA', 'CAAAAAAGAAA', 'TAAAAAAAAAA', 'GACAGAGCCTA']

Iter 3 : Consensus string:AAAAAAAAAAA number of iteration 9 Min

['AAAAAAAAAAA', 'AAAAAAAAAGA', 'AAAAAAAAAAA', 'AAAAAAAAAGT', 'AAAAAAAAAAA', 'AAAAAAAAAAA', 'AAAAAAAAAAA', 'AAAAGAAAAGT', 'AAAGAACAAAA', 'GACGGTCAAGT']

Iter 4 : Consensus string:AAAAAAAAAAA number of iteration 5 Min

['AAAAAAAAAAA', 'CACAAAAAAAA', 'AAAAAAAAAAA', 'CAAAAAAAAAA', 'AAAAAAAAAAA', 'AAAAAAAAAAA', 'AAAAAAAAAAA', 'AAAAAAGAAAA', 'AAGAACAAAAA', 'GATACGCGTAA']

Iter 5 : Consensus string:CAAAAAAAAAA number of iteration 5 Min

['CAAAAAAAAAA', 'TGAAACAAAAA', 'TAAAAAAAAAA', 'CAAAAAAAAAA', 'TAAAAAAAAAA', 'CAAAAAAAAAA', 'CAAAAAAAAAA', 'CGAAAGAAAGA', 'TAAAAAAAAAA', 'CGAACCTTAGA']

Iter 6 : Consensus string:AAAAAAAAAAG number of iteration 5 Min

['AAAAAAAAAAA', 'AGAAGTTTATG', 'AAAAAAAAAAA', 'AAAAAAAAAAG', 'AAAAAAAAAAA', 'AAAAAAAAAAG', 'AAAAAAAAAAA', 'AAAAGTTTAAT', 'AAAAAAAAAAG', 'AACGGTATTAG']

Iter 7 : Consensus string:TCAATTAAGAG number of iteration 5 Min

['TCGTATAAGAT', 'TCTATTTGGAG', 'TCACTAAAAAG', 'TCGATTAGGAT', 'TCAATTGCTAT', 'TCGAAAGAAAT', 'TTAATAAAGAT', 'GCACTTGATAG', 'TCACTTTACCG', 'TCATTTTAAAG']

Iter 8 : Consensus string:AAAAAAAAACA number of iteration 6 Min

['AAAAAAAAACA', 'AAAAAAAAAGA', 'AAAAAAACCCA', 'AGAAACGGTCA', 'AAAAAAAAACA', 'AAAAAAAAAGA', 'AACAAAAATGA', 'AAAAAAAAAGA', 'AAAAAAAAACA', 'GTCAAGAGCCA']

Iter 9 : Consensus string:CAAAAACAAAA number of iteration 5 Min

['CAACAATAAAA', 'TTGAAACAAAA', 'AAGAGACAAAC', 'CAACAACAAAA', 'AAAAAACGAAA', 'CAACAACAAAA', 'CTGAAATAAAA', 'CTAAAACAAAA', 'AAAGAACAAAA', 'TAGAATTAGAT']

Gibs Sampler Tests

9-mer

Inserted mutation:AAAAAAAAAA

Iter 0 : Consensus string:TGTTTAGTT number of iteration 164 Min

['AGCTTAGAT', 'GAAGTAATT', 'GAGGTAGAT', 'CCCGGAAAA', 'CTCGCAGTA', 'GCTTCACTA', 'TATACAGAC', 'ATTTTGCAT', 'ATATCAGTT', 'AATGTAATA']

Iter 1 : Consensus string:CCCCCCGAA number of iteration 174 Min

['ATTGTTGCA', 'AGCGTAGCA', 'GTGGAAACA', 'AGGGGCCCA', 'GCTCTTCCG', 'ATGGTATCC', 'ATTGGGCCC', 'ATCGGGGCC', 'AGGGCAGCA', 'TTTGGGTCC']

Iter 2 : Consensus string:AAAAAAAAAA number of iteration 359 Min

['AAAAAAAAAA', 'ATAAAAAAA', 'AAAAAAAAAA', 'AAAAAAAAAA', 'AAAGAAAAC', 'ATAAAAAAA', 'TAATAAAAA', 'AAAAAATAA', 'AAAAAAAAAA', 'AAACAAATA']

Iter 3 : Consensus string:AAAAAAAAAA number of iteration 253 Min

['AAAAAAAAAA', 'AAAAAAAAT', 'AAAATAAAA', 'AAAAAAAAAA', 'ACAAAAAAA', 'AAAAAAAAAA', 'AGAAAAAAA', 'AAAATAAAA', 'AAAAAAAAAA', 'AAGAAAGAA']

Iter 4 : Consensus string:AAAAAAAAAA number of iteration 242 Min

['AAAAAAAAAA', 'AAATGAAAA', 'TAAAAAAAA', 'TAGTAAAAC', 'AAAAAATAA', 'AAAAAAAAC', 'AAAAAAAAC', 'AAAAAAAAC', 'TAAAAAAAA', 'AAGAAAGAA']

Iter 5 : Consensus string:AAAAAAAAAA number of iteration 179 Min

['TAAAAAAAC', 'GAAAAACAA', 'GAAAAAAAA', 'GAAAAAAAA', 'GAAAGAAAA', 'TGAATAAAA', 'ATAAAAAAA', 'GAATAAAAA', 'TTAAAAAAA', 'TATAAACAA']

Iter 6 : Consensus string:AAAAAAAAAA number of iteration 229 Min

['AATAAAAAA', 'AATAAAAAA', 'AAAAAATGA', 'AGAAAAAAA', 'AAAAAATT', 'AAAAGAAAC', 'AAAAAAAAAA', 'CAAGTAAAT', 'AAAAATAAC', 'AACAATAAT']

Iter 7 : Consensus string:AAAAAAAAAA number of iteration 201 Min

['AAAAAAAAAA', 'AAAAACAAA', 'AAAAAAAAAA', 'GAAAAAAAA', 'ATCGAAAAA', 'AAAAAAAAG', 'AACAAAAAA', 'AATAAAAAA', 'CAAAAAAAA', 'ATAAACAAA']

Iter 8 : Consensus string:AAAAAAAAAA number of iteration 196 Min

['AAAAAAACA', 'AAATGAAAA', 'AAAAAAATG', 'GAAAAAAAA', 'AAAAAAAAAA', 'AAAAAAATA', 'AAAAAAAAAA', 'GAAAAAATA', 'ACAAAAAAA', 'ATAAACAAA']

Iter 9 : Consensus string:AAAAAAAAAA number of iteration 155 Min

['AAAAAAAAT', 'AGAAAAAAA', 'AAAAAAAAC', 'AACGAAGAA', 'AAAAAAAAAA', 'AAAAAAAAAA', 'AACAAAAAA', 'AAAACAGAA', 'ACAAAAAAA', 'TAAACACAA']

9-mer

Inserted mutation:ACGTAACAGT

Iter 0 : Consensus string:CTGGTCTTG number of iteration 201 Min

['GTACGTTGT', 'CTACATTAC', 'GCGCGGTAC', 'CGACGATTT', 'GCACGTTCT', 'CGGCGCTGG', 'TTAAGTTAA', 'ATACATTGT', 'GATCGTCGT', 'CTACGAGAT']

Iter 1 : Consensus string:CGATAGTTT number of iteration 147 Min

['TCGGACCCG', 'TCGCAACTC', 'GGGGTACTA', 'TTATAGGGT', 'TTCGTGAGG', 'GCTCACCCG', 'TTCCTGATG', 'TTGTTGCCG', 'TTCCTCCAA', 'TTACAGGTC']

Iter 2 : Consensus string:GGTATTACG number of iteration 122 Min

['TAGGCCATG', 'AAGTTGCTA', 'TACGTCATG', 'GAGGTAATA', 'AATGGCACG', 'AAGGGAATG', 'TGTGTGACA', 'TATCGCCTG', 'GATGGACTC', 'AAAGGGCTA']

Iter 3 : Consensus string:TAACAGTTA number of iteration 261 Min

['TAACAGTTG', 'TAACAGTTG', 'TAACAGTAG', 'TAACAGTTA', 'AAACAGTAA', 'TTAGAGTAA', 'TAACAGTTA', 'TCACTTTGA', 'TTACAGTTG', 'TAACAGCTG']

Iter 4 : Consensus string:TTTGAATCC number of iteration 153 Min

['AGAAGAATG', 'CGAGCAATG', 'CGAGCAGTG', 'ACAGCACTG', 'GGTGCAATT', 'CCAGCTCTT', 'GAAGAAAGT', 'GTAACTGTG', 'GGAACAATT', 'TAAGCAATA']

Iter 5 : Consensus string:GTAACAGTT number of iteration 163 Min

['GTAACAGTT', 'GTAACAGTT', 'GTAACAGTT', 'GTAAGAGTG', 'GAAGCAGTG', 'GTGACAGTT', 'GTAACAGTT', 'GTAACAGTG', 'GTAACAGTG', 'GTAGCAGAG']

Iter 6 : Consensus string:GTGAAAGTG number of iteration 111 Min

['CACCTGCTC', 'CCTATGGAA', 'GTGTTGCTG', 'CTCTTGCTA', 'CCGCTGCGA', 'CTGCTTATC', 'GCTCTGGAG', 'GTGTTGCGG', 'ATCCCTATT', 'AAGTTGGAC']

Iter 7 : Consensus string:GGATAACGT number of iteration 214 Min

['TGAACCTAT', 'GCCGTCGTA', 'GGCGGCGTA', 'GTAATACTG', 'TTCCCCATA', 'TTAACAGTG', 'TTCATATAA', 'GTAACAGTT', 'GTAACAGTT', 'AAAAAACTA']

Iter 8 : Consensus string:CGTAACAGT number of iteration 178 Min

['CGTAACAGT', 'CGTAACAGT', 'CGTAACAGT', 'CGTACCAGA', 'CGTAACAGT', 'CGTAACAGT', 'CGTAACAGT', 'CGTTACAGA', 'CGTAACAGT', 'CGTACCAGC']

Iter 9 : Consensus string:AACAGTGTT number of iteration 412 Min

['AACAGTCAT', 'AACAGTCGT', 'AACAGTGGG', 'ACCAGGAGG', 'ACCAGTGGC', 'AACAGTAGG', 'AACATTCGA', 'AACAGTACA', 'AACAAATAG', 'AAAACTACA']

10-mer

Inserted mutation:AAAAAAAAAA

Iter 0 : Consensus string:AATAAAAAAA number of iteration 176 Min

['CCAGCTCGAA', 'CACGAAGTAG', 'CAGGGCAAAG', 'CCCTTTGCAG', 'CCATTTGGAG', 'AGAGATTATC', 'GCTGATAGAC', 'CGTTCAATAA', 'GCCCCTAGTG', 'CTCCCCCAAG']

Iter 1 : Consensus string:AAAAAAAAAA number of iteration 377 Min

['ATAAAAAAAA', 'AAAAAAAAAA', 'AAAAAAAAAA', 'AAAAAACAAT', 'AAAAAGAAAT', 'CAAAAAAAGG', 'AAAAAAAAAA', 'AAAAAAAAAT', 'ACAAAAAAGT', 'CAAAACGATT']

Iter 2 : Consensus string:AAAAAAAAAA number of iteration 261 Min

['ATAAAAAAAA', 'AGAAAAAAAA', 'AAAAGAAGAA', 'AAAAAAACAA', 'AAAAAAACCG', 'AAGAAAATAA', 'AAAAAAAGCA', 'AAAGAAAAAA', 'AAAAAAAACA', 'ACAAAAGGAA']

Iter 3 : Consensus string:AAAAGAAAAA number of iteration 294 Min

['AAGAAAAAAG', 'AAAAAAAAAA', 'AAGAAAAAAG', 'AAAAAACAAT', 'AAAAAAAAAA', 'AAAAAAAAAC', 'CAAAAAAAAA', 'ATAAAAGAAA', 'AAAAGAAAAA', 'CAAAAGGAAC']

Iter 4 : Consensus string:AAAAAAAAAA number of iteration 244 Min

['AAAAAAAAAA', 'AAAAAAAATA', 'AAAAAAAAAT', 'AAAAAAACAA', 'AAAAAAAAAA', 'TGAAAAAAAA', 'TAAGAAAAAA', 'TAAAAAAAAA', 'AAAAGAAAAA', 'TAAGCAGAAA']

Iter 5 : Consensus string:AAAAAAAAAA number of iteration 159 Min

['CATAAAAGAT', 'CAAAAAAAAA', 'AAAAAAATTA', 'AAAACAAAAA', 'AAAAAAAAAA', 'AAAAAAAAAA', 'AAAAAAAAAA', 'AAAAATAAAT', 'AAGAAAAAAA', 'ACAAAAGGAA']

Iter 6 : Consensus string:AAAAAAAAAA number of iteration 231 Min

['AAAAAAAAAA', 'AAAAAACATA', 'AAAAAAACTC', 'AACAAAAAAT', 'AAAAGAAATA', 'GAAAAAAAAA', 'AAAAAAAAAA', 'AAAAAAAATA', 'AAAAAATTAA', 'AAAGGAACCA']

Iter 7 : Consensus string:AAAAAAAAAA number of iteration 328 Min

['AAAAAAAAAT', 'AAAAAAAAAA', 'CAAAAAAAAA', 'AAAAAAACAA', 'CAAAAAGAAA', 'GAAAAAAAAT', 'CCAAAAAAAA', 'AAAAAAAAAT', 'AAAAAAAAAA', 'ACAAAAGGAA']

Iter 8 : Consensus string:AAAAAAAAAA number of iteration 328 Min

['AAAAAAAAAT', 'CAAAAAAAAA', 'CATAAAAAAA', 'AAACAAAAAA', 'AAAAAAAAAA', 'TGAAAAAAAA', 'AAAAAAAAAA', 'CAAAAAAAAA', 'AAAGAAAAAA', 'CAAGACTATT']

Iter 9 : Consensus string:AAAAAAAAAA number of iteration 163 Min

['GAAAAAAAAA', 'AAAAAAAAAA', 'AAAAGAAGAA', 'GAGAAAAAAA', 'AAAAAAAAAA', 'AAAAAAAAAA', 'AAAAGAAAAA', 'AAAAAAAAAA', 'AAAAAAAAAA', 'ACAAAAGGAA']

11-mer

Inserted mutation:AAAAAAAAAA

Iter 0 : Consensus string:CAAAAAAAAAA number of iteration 381 Min

['AAAAAAAAAAT', 'CAAAAACGAAC', 'AAAACAAAAAG', 'AAGTCAAAATC', 'ATAAATCAAAG', 'AAAAAAAAAAA', 'AAAACTAGAAT', 'AAAACAAAAAC', 'GAAGAATCAAA', 'AAGGGAAAAAG']

Iter 1 : Consensus string:CAAAAAAAAAA number of iteration 316 Min

['AAAAAAAAAAT', 'AAAAACAAAAT', 'CAATAAAAAAA', 'ACAACCGACAT', 'AAAACCTAAAT', 'CAAAAAAAAAA', 'AAAAAAAAAAT', 'CAAAAAAAAAA', 'AAAAAAAAAAT', 'ATCACCTAAAC']

Iter 2 : Consensus string:AAAAAAAAAAT number of iteration 233 Min

['AAAAAAAAAAT', 'AAAAACAAAAT', 'AATAAAAAAAA', 'AATAAAAAAGC', 'TTAAAAATATC', 'AAAAAAAAAAG', 'AAAAAAAAAAT', 'AAAACAAAAAC', 'AAAAAAAAAAT', 'TATACCAATTC']

Iter 3 : Consensus string:CAAAAAAAAAA number of iteration 235 Min

['GAAAAAAAAAA', 'ACGAAAACAA', 'ATAAAAAAAAA', 'GAAATAAAAAA', 'ACTAAAAAAAA', 'GCAAAAAAAAA', 'GCCAAAAAAAA', 'GGAAAAAAAAA', 'GACAAAAAAAA', 'GTCACAATTAA']

Iter 4 : Consensus string:AAAAAAAAAAG number of iteration 252 Min

['AAAAAAAATAT', 'AAAAACAAAAT', 'AAAAAAAAATA', 'AAAGAAACAAC', 'AAAAAAAAGAC', 'AAAAAAAAAAT', 'AAAAAAAAAAT', 'AAAAAAAAAAC', 'AAAAAAAAAAT', 'AAAAAGTCCCC']

Iter 5 : Consensus string:CAAAAAAAAAA number of iteration 254 Min

['AAAGAAAAAAT', 'AAAGGAAAAAT', 'AAAAAAAAAAT', 'AAAAAAAAATT', 'GAAAAAAATAA', 'AAAAAAAAAAT', 'AAAAAAAAAAT', 'AAAAAAAAAAT', 'ACAAAAAAAAA', 'GAAGGGAAAAA']

Iter 6 : Consensus string:AAAAAAAAAAT number of iteration 244 Min

['AAAAAAAAAAC', 'AAAAAAAAAAC', 'AAAACAAAAAT', 'AAAAAAAAATT', 'AAAAAAAAGAC', 'AAAAAAAAAAT', 'AAAAAAAAATT', 'AAAAAAAAAAT', 'AAAAAAAAAAT', 'AGGGAAAAAGT']

Iter 7 : Consensus string:CAAAAAAAAAA number of iteration 263 Min
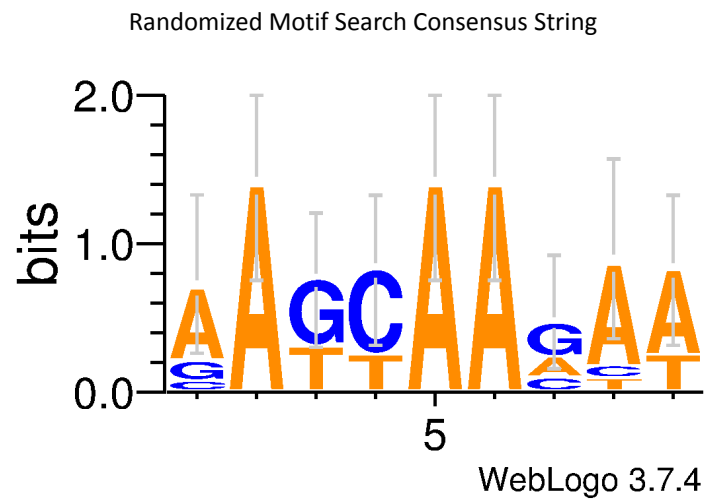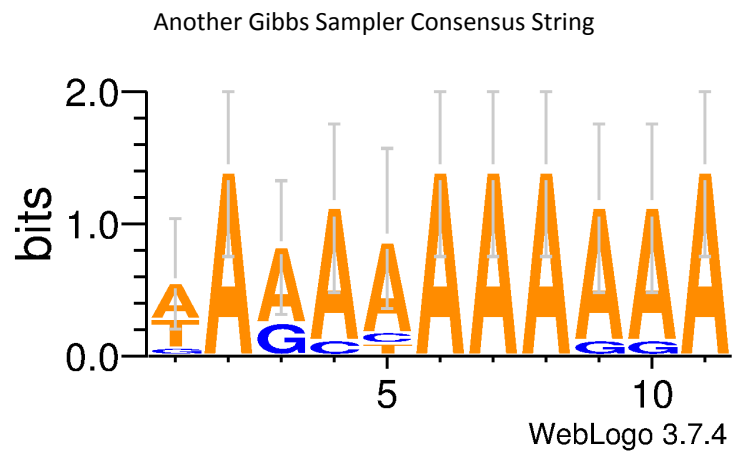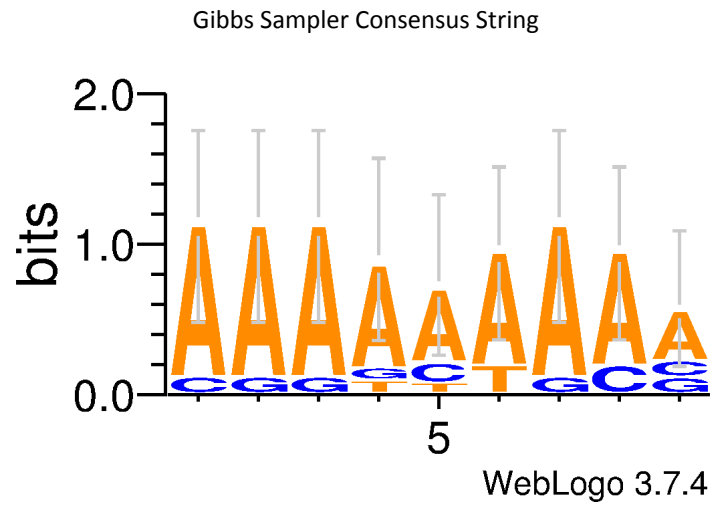
['GAAAAAAAAAA', 'ACCAAAGGAAA', 'AATAAAAAAAA', 'GCAAAAAAAAA', 'AAAAATAAAAA', 'GCAAAAAAAAA', 'TAAAAAACAAA', 'GACAAAAAAAA', 'ACAGGAAAAAA', 'ACAGAATAAGT']

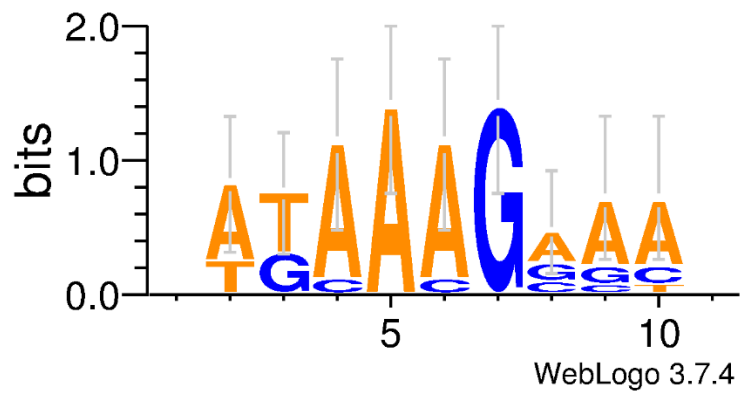Iter 8 : Consensus string:AAAAAAAAAAA number of iteration 184 Min

['AAAAAAAAAAC', 'GCAAAAAAAAA', 'AAAAAAAAATA', 'GCAAAAAAAAA', 'AAAAAAATAAC', 'AAAAAAAAAAA', 'AAAAAAAAAAA', 'AAAAAAAAAAT', 'ACAAAAAAAAA', 'AAGGGAAAAAG']

Iter 9 : Consensus string:AAAAAAAAAAA number of iteration 154 Min

['AAAAAAAGAAA', 'CAAAAAAAAAA', 'AAAAAAAAATA', 'AAAAAAAACAA', 'AGCCAAAAAAA', 'AAAAAAAAATA', 'TAAAAAAAAAA', 'TAAAAAAAAAA', 'CAAAAAAAAAA', 'CAATTAAACAT']

Gibbs Sampler Consensus String

WebLogo 3.7.4

Another Gibbs Sampler Consensus String

WebLogo 3.7.4

Randomized Motif Search Consensus String

WebLogo 3.7.4

Another Randomized Motif Search Consensus String

WebLogo 3.7.4

Conclusion

As we have seen tests above, higher k-mer values are more accurate to find implanted motif. We think first random selection is most important part of both algorithms and in this step of algorithm with higher k-mer value we have higher chance to find implanted motif in first step also with higher k-mer value accuracy will be higher and we can conclude from test results gibbs sampler algorithm gives more accurate results than randomized motif search.