DNA Replication Basics

1. True or False: The Hidden Message Problem is a well-defined computational problem.

   False

   True

2. Compute
   Count(CGCGATACGTTACATACATGATAGACCGCGCGCGATCATATCGCGATTATC, CGCG).

   5

3. What is the most frequent 3-mer of
   TAAACGTGAGAGAAACGTGCTGATTACACTTGTTCGTGTGGTAT?

   GTG

4. What is the reverse complement of TTGTGTC?

   GACACAA

5. Solve the Pattern Matching Problem with Text = GACGATATACGACGATA and Pattern = ATA to find all starting positions of Pattern in Text. Return the starting positions in increasing order (make sure to use 0-based indexing!)

   4 6 14

DNA Replication via GC Skew

1. The position of the E. coli genome at which the skew attains a minimum value is most likely near which of the following?

   the replication terminus

   the origin of replication

   the middle of the reverse strand

   the middle of the forward strand

2. Compute the Hamming distance between CTACAGCAATACGATCATATGCGGATCCGCAGTGGCCGGTAGACACACGT and CTACCCCGCTGCTCAATGACCGGGACTAAAGAGGCGAAGATTATGGTGTG.

   36

3. Identify the value of i for which $Skew_i$(GATACACTTCCCGAGTAGGTACTG) attains a minimum value.

   12

4. Compute $Count_2$(CATGCCATTCGCATTGTCCCAGTGA, CCC).

   15

5. The d-neighborhood of the k-mer Pattern is the collection of all k-mers that are at most Hamming distance d from Pattern. How many 5-mers are in the 2-neighborhood of Pattern = TGCAT? Note that the d-neighborhood of Pattern includes Pattern.

   106

Regulatory Motif Search Basics and Pseudocounts

1. Which type of algorithm enumerates every possible candidate solution to a problem and then checks each one individually?

   Greedy algorithm

   Machine learning algorithm

   Combinatorial algorithm

   Randomized algorithm

   Dynamic programming algorithm

   Brute force search

2. True or false: a motif of lowest score with respect to a collection of strings does not need to appear as a substring of one of the strings.

   True

   False

3. Order the following probability distributions from lowest to highest entropy:

   A: (0.5, 0, 0, 0.5)

   B: (0.25, 0.25, 0.25, 0.25)

   C: (0, 0, 0, 1)

   D: (0.25, 0, 0.5, 0.25)

   B, D, A, C

   B, C, A, D

   C, D, B, A

   C, A, D, B

   D, B, C, A

4. Consider the following profile matrix:

A: 0.4 0.3 0.0 0.1 0.0 0.9

C: 0.2 0.3 0.0 0.4 0.0 0.1

G: 0.1 0.3 1.0 0.1 0.5 0.0

T: 0.3 0.1 0.0 0.4 0.5 0.0

Which of the following strings is a consensus string for this profile matrix? (Select all that apply.)

   AAGAGA   TCGCGA   AGGTCA   **AAGCTA   ACGCGA   ACGTTA**

5. Consider the following motif matrix:

CTCGATGAGTAGGAAAGTAGTTTCACTGGGCGAACCACCCCGGCGCTAATCCTAGTGCCC

GCAATCCTACCCGAGGCCACATATCAGTAGGAACTAGAACCACCACGGGTGGCTAGTTTC

GGTGTTGAACCACGGGGTTAGTTTCATCTATTGTAGGAATCGGCTTCAAATCCTACACAG

Which of the following 7-mers is a median string for this motif matrix? (Select all that apply.)

   AATCCTA   TCTGAAG   AACGCTG   GTCAGCG   **GTAGGAA**   GATGAGT

6. Consider the following profile matrix *Profile*:

A: 0.4 0.3 0.0 0.1 0.0 0.9

C: 0.2 0.3 0.0 0.4 0.0 0.1

G: 0.1 0.3 1.0 0.1 0.5 0.0

T: 0.3 0.1 0.0 0.4 0.5 0.0

Compute Pr(GAGCTA|*Profile*). (Express your answer as a decimal and do not round your answer.)

   0.0054

Regulator Motif Search via Randomization

1. True or False: **RandomizedMotifSearch** performs well when given a uniform profile matrix.

   True

   False

2. True or False: **RandomizedMotifSearch** and **GibbsSampler** are usually run on many choices of initial k-mers.

   True

   False

3. True or False: it is not possible for **RandomizedMotifSearch** to move from a collection of motifs with lower score to a collection of motifs with higher score.

   False

   True

4. Which of the following motif-finding algorithms is guaranteed to find an optimum solution? In other words, which of the following are *not* heuristics? (Select all that apply.)

   GibbsSampler

   BruteForceMotifSearch

   GreedyMotifSearch (without pseudocounts)

   MedianString

5. Assume we are given the following strings *Dna*: TGACGTTC TAAGAGTT GGACGAAA CTGTTCGC.
   Then, assume that **RandomizedMotifSearch** begins by randomly choosing the following 3-mers *Motifs* of *Dna*: TGA GTT GAA TGT

   What are the 3-mers after one iteration of **RandomizedMotifSearch**? In other words, what are the 3-mers *Motifs*(*Profile*(*Motifs*), *Dna*)? Please enter your answer as four space-separated strings.

   TGA TAA GGA TGT