Introduction to Computational and Systems Biology

Exercise set 4

Due: December 12, 2019

1. Assume the existence of a linear time algorithm that finds whether a sequence S is a subsequence of T. Give a linear time algorithm to check whether a sequence S is a *circular shift* of another sequence T. For example, *ciseexer* is a circular shift of *exercise*. What happens if one wants to test whether S is a *subsequence* of a circular shift of T?

2. We apply the Sanger sequencing technique to the single stranded DNA sequence ``CATCAGTGCAGTGATC''. What single stranded DNA sequences will you find after the reaction in each of the four test tubes and what will be the picture obtained by running those test tubes through gel electrophoresis? What do you read from that picture?

3. The following two DNA sequences contain the same 4 genes in different arrangements (in different order and on different strands). Find those arrangements and give one way to transform one arrangement into the other one using reversals only, that is, in each step of your strategy, you can only choose a certain substring and invert it. Take into account the orientation (strand) of each gene.

   GGTCAAATTACAGTCTGGCACTAG CCAGACTTGACCCACTAGGTAATT
   CCAGTTTAATGTCAGACCGTGATC GGTCTGAACTGGGTGATCCATTAA

4. Find a solution for a partial digest experiment giving the following results: 3, 4, 6, 10, 12, 15, 16, 20, 22, 25, 28, 31, 32, 35. Note that some of these lengths may have occurred in the experiment more than once, although they are only reported by your bio-lab only once in the list above.

5. For the following matrix, find a permutation of its columns that either transforms it into a matrix with consecutive 1s in each row, or into one having at most two gaps in its blocks of ones (at most two gaps in the whole matrix).

0	1	1	0	1	1	1	0
1	1	0	0	1	0	1	0
0	1	1	0	1	1	0	1
1	0	0	1	0	0	1	0

6. In an experiment on sequencing by hybridization, the following error-free results were return: ACTTA, ACGAT, ATCGT, ATTGT, ATGTT, CGATC, CGATT, CGTAC, CTTAT, GATCG, GATTG, GTACG, GTACT, TACGA, TACTT, TATGT, TCGTA, TGTAC, TGTTT, TTGTA, TTATG. Find at least one sequence that may give this result.

7. Find a shortest common superstring of the following strings: TCAGT, AGTGC, GCA, CATCAG, GTG. **Prove** that no shorter solutions exist.

8. Sort the following permutations using reversals only:

   a. 2 7 3 8 1 5 4 6
   b. -8 2 -5 6 -1 7 -4 3
   c. 3 8 2 7 1 4 6 5
   d. 2 -7 1 8 -4 3 5 -6