

Introduction to Computational and Systems Biology

Exercise Set 3

Due: December 11, 2019

1. Suppose we have the following fragments:

$f_1 = \text{ATCCGTTGAAGCCGCGGGC}$
 $f_2 = \text{TAACTCGAGG}$
 $f_3 = \text{TAAAGTACTGCCCCG}$
 $f_4 = \text{ATCTGTGTCTGGG}$
 $f_5 = \text{CGACTCCCGACACA}$
 $f_6 = \text{CACAGATCCGTTGAAGCCGCGGG}$
 $f_7 = \text{CTCGAGTTAAGTA}$
 $f_8 = \text{CGCGGGCAGTACTT}$

We know that the total length of the target molecule is about 50 bp and may be ready to accept a solution of length between 45 and 55 bp. Assemble these fragments and obtain a consensus sequence. Be prepared to deal with errors. You may also have to use the reverse complement of some of the fragments.

2. Let $F = \{\text{ATC}, \text{TCG}, \text{AACG}\}$. Find the best layout for this collection according to the sequence reconstruction model, with error level $e=0.1$. The same problem for $e=0.25$. Be sure to consider also reverse complements.
3. a) You want to use the sequencing by hybridization method (SBH) to sequence a DNA fragment. For this, you are using a DNA array containing all DNA sequences of length 3 and test which of these sequences bind to your target. As a result, you find out that the target sequence has the following substrings of length 3:
 $\{\text{ATG}, \text{CTG}, \text{GCT}, \text{GGT}, \text{GTG}, \text{TGA}, \text{TGC}, \text{TGG}\}$
Find at least 2 DNA sequences validating this data.
b) How many solutions do you have if, using a DNA array containing all sequences of length 4, you obtain that the target sequence has the following substrings of length 4:
 $\{\text{ATGG}, \text{CTGA}, \text{GCTG}, \text{GGTG}, \text{GTGC}, \text{TGCT}, \text{TGGT}\}$?
4. You are assembling a DNA sequence containing a repeat of the form XXX. Having given the fragments *ATG*, *CTTGAT*, *TGT*, *TGTCA*, *TCAGAT*, *TGTA ACT*, find at least two such DNA sequences knowing that
 - No fragment is included into some other
 - The fragments provide “good linkage”, in the sense that all fragments (except the one covering the ends of the sequence) overlap with at least one fragment at left and with another at right.
5. a) Assemble the following error-free fragments using the shotgun approach: *ATGTG*, *GCCGCA*, *GTGCCG*, *TGTGCC*.
b) The same problem as above, replacing the second fragment above with *CCCGCA*. Assemble the fragments using the shotgun approach. Assemble the fragments using also the SBH-style shotgun approach. Compare the results and also with the result obtained at a), knowing that fragment *CCCGCA* had one substitution error – the correct one was the second fragment in a).