

18.417 Introduction to Computational Molecular Biology

Problem Set 1

Issued: September 14, 2004

Lecturer: Ross Lippert

Due: September 28, 2004

These problems relate to the digest problems and basic motif finding.

problem 1 4.2 of JP

problem 2 4.6 of JP: the question doesn't really work with ΔA defined as they have defined it in the chapter. Instead, for this problem let $\Delta A = A \ominus A$ and otherwise keep it the same.

***problem 3** 4.9 of JP: the question is somewhat vague. Instead, I'd like you to solve the following double digest problem:

$$\begin{aligned} frags(A) &= (1, 1, 2, 4, 5, 9, 11, 12, 13, 20, 22) \\ frags(B) &= (2, 2, 5, 7, 7, 11, 15, 23, 28) \\ frags(A + B) &= (1, 1, 2, 2, 2, 2, 2, 2, 2, 4, 5, 5, 5, 9, 9, 12, 15, 20) \end{aligned}$$

You may assume that fragments come from a strand of total length 100 and thus the restriction site locations range from 1 to 99. If you solve this by hand, please turn in a clear description of the reasoning which lead to your solution. If you solve this with a program, please turn enough of the source code to make it clear what approach you took.

problem 4 4.12 of JP

problem 5 4.13 of JP

***problem 6** Attached is an allegedly better version of the branch and bound method for median string searching as discussed in class. This one happens to be from the authors of our text and claims to be the solution of problem 4.17. I don't think it is correct. Am I right, or am I wrong? If wrong, why?¹

¹Obviously, this is an easier problem if I am wrong.

Better Bounded Median String Search

```

1. ImprovedBranchAndBoundMedianString(DNA t n l)
2.   s = (1, 1, ..., 1)
3.   bestDistance = ∞
4.   i = 1
5.   while i > 0
6.     if i < l
7.       prefix = nucleotide string corresponding to (s1, s2, s3, ..., si)
8.       optimisticPrefixDistance = TotalDistance(prefix, DNA)
9.       if (optimisticPrefixDistance < bestSubstring[i])
10.        bestSubstring[i] = optimisticPrefixDistance
11.        if (l - i < i)
12.          optimisticSufxDistance = bestSubstring[l - i]
13.        else
14.          optimisticSufxDistance = 0;
15.        if optimisticPrefixDistance + optimisticSufxDistance ≥ bestDistance
16.          (s, i) = Bypass(s, i, l, 4)
17.        else
18.          (s, i) = NextVertex(s, i, l, 4)
19.      else
20.        word = nucleotide string corresponding to (s1, s2, s3, ..., si)
21.        if TotalDistance(word, DNA) < bestDistance
22.          bestDistance = TotalDistance(word, DNA)
23.          bestWord = word
24.          (s, i) = NextVertex(s, i, l, 4)
25.   return bestWord

```