

CAP5510 Introduction to Bioinformatics

Fall 2009

Homework #3

(Assigned Oct.26 Due: Nov. 2)

1. Consider the partial digest $L=\{1,1,1,2,2,3,3,3,4,4,5,5,6,6,6,9,9,10,11,12,15\}$
Solve the partial digest problem L (i.e., find X such that $\Delta X=L$.) [25 points]
2. Do Problem 4.16 and Problem 4.17 (from text by Jones=Pevzner, pp.122-123 [25 points]
3. Explain the basic idea behind SBH (Sequencing by Hybridization). What are its limitations for realistic Genome sequencing? Use the Eulerian path approach to solve the SBH problem for the following spectrum:
$$\mathbf{S} = \{ATG, GGG, GGT, GTA, GTG, TAT, TGG\}$$
Label edges and vertices of the graph, and give all possible sequences s such that $Spectrum(s, 3) = \mathbf{S}$ [25 points]
4. DNA sequencing reads contain errors that lead to complications in fragment assembly. Fragment assembly with sequencing errors motivates the **Shortest k-approximate Superstring** problem: Given a set of strings $S = (S_1, S_2, \dots, S_t)$ find a shortest string s such that each string in S matches some substring of s with at most k errors. [25 points]