

Home Work 5

CSE/BENG/BIMM 181, Winter 2012

Given on Fri, March 9, 2012. Due on Fri. March 16, 2012

State your answers legibly and concisely. Your solutions will be graded on correctness, elegance, **clarity** and originality. Your proofs should **avoid getting bogged down in too much detail**. Your algorithms should be as efficient as possible (inefficient algorithms will result in lower grades). Please note that the work handed in must be your own. No late assignments will be accepted without prior arrangement. Every problem is 20 points.

The Following Problems are from textbook:

Problem 10.6 (20 pts)

How do you calculate the trimming parameter δ in AdditivePhylogeny?

Problem 10.12 (20 pts)

Design a backtracking procedure to reconstruct the optimal assignment of characters in the Sankoff algorithm for the Weighted Small Parsimony problem.

Problem 8.10 (20 pts)

The SBH problem is to reconstruct a DNA sequence from its l -mer composition

- Suppose that instead of a single target DNA fragment, we have two target DNA fragments and we simultaneously analyze both of them with a universal DNA array. Give a precise formulation of the resulting problem (something like the formulation of the SBH problem).
- Give an approach to the above problem which resembles the Hamiltonian Path approach to SBH
- Give an approach to the above problem which resembles the Eulerian Path approach to SBH.

Problem 8.11 (20 pts)

Suppose we have k target DNA fragments, and that we are able to measure the overall multiplicity of each l -mer in these strings. Give an algorithm to reconstruct these k strings from the overall l -mer composition of these strings.

Problem U - Small Parsimony Problem (20 pts)

Register for the contest if you haven't already done so: http://judge.bioinf.spbau.ru/cgi-bin/new-register?contest_id=5

Login for full description and to submit code: http://judge.bioinf.spbau.ru/cgi-bin/new-client?contest_id=5

Given a tree T with every vertex labeled by an m -long string of characters, one can set the length of an edge (v, w) in the tree to the Hamming distance $d_H(v, w)$ between the character strings for v and w . The parsimony score of a tree T is simply the sum of lengths of its edges. In reality, the strings of characters assigned to internal vertices are unknown (only strings at leaves are known)

and the problem is to find strings that minimize the parsimony score.

Tree in this problem is unrooted binary tree.

First line of the input contains one integer: $2 \leq M \leq 1000$ – number of leaves in the tree. Total number of nodes in this tree (by definition of the unordered binary tree) is $2M - 2$. Leaves are numbered from 1 to M , while nodes from $M + 1$ to $2M - 2$ are internal nodes. Next M lines contains ACGT-strings of equal length at most 1000bp for nodes 1 to M respectively. Next $2M - 3$ lines contains two integers each – adjacent nodes (by an edge).

Output minimal possible parsimony score in the first line. Next $M - 2$ lines of the output should contain ACGT-strings for internal nodes from $M + 1$ to $2M - 2$ respectively.

*See example table on website.