582670 Algorithms for Bioinformatics, 4 cr — Exam 16.10.2013 — Solutions/grading

All problems were covered at the lectures.

1. **Edit distance computation.** (12 points)

   Give the recurrence for computing the edit distance between two strings and explain why it is correct. Compute the edit distance between the following two strings

   ```
   ACGCA
   ATCT
   ```

   by filling out the dynamic programming matrix and showing the optimal alignment(s).

   Grading:

   - 3p – Recurrence for computing the edit distance
   - 3p – Arguing why it is correct
   - 3p – Correctly filling in the DP matrix
   - 1p – Stating the value of the edit distance from the matrix
   - 2 × 1p – showing the two optimal alignments

2. **Algorithms for finding phylogenetic trees.** (12 points)

   Define the problem of finding a phylogenetic tree. UPGMA and neighbor joining algorithm both solve this problem. Describe briefly the common idea behind the algorithms. How do they differ? In which conditions are the algorithms able to construct the correct tree?

   Grading:

   - 4p – Definition of the problem
   - 3p + 1p – Description of the common idea / differences
   - 2p – Mentioning the two conditions (*additive* or *ultrametric* matrices)
   - 2p – Explaining what additive and ultrametric mean (either through paths in a tree, or by correctly stating the 3-/4-point conditions)

3. **Approximation algorithms for genomic rearrangements.** (12 points)

   Let $d(\pi)$ denote the minimum number of reversals needed to transform a permutation π into the identity permutation. Let $BP(\pi)$ denote the number of breakpoints of a permutation π.

   (a) Argue that $d(\pi) \geq BP(\pi)/2$.

   (b) Prove that if a permutation π has a decreasing strip, then there is a reversal which decreases the number of breakpoints of π by at least 1.

(c) Write a 4-approximation algorithm for computing $d(\pi)$ and argue why its approximation ratio is 4.

Grading:

- 2p – Point (a)
- 5p – Point (b). Many didn't mention that we choose k as the minimum number appearing in *all* decreasing strip, but just chose a decreasing strip and then picked its minimum element. Also, some of you didn't consider the case when the (increasing) strip containing $k-1$ is to the right of the decreasing strip containing k.
- 3p – Writing the correct algorithm.
- 2p – Argumentation that its approximation ratio is 4.

4. **Your choice.** (12 points)

   Choose one of the (non-trivial) problems studied during the course (in study groups, lectures, or/and exercises) not related to the assignments above. Define the problem (input, output), explain how the problem is motivated by molecular biology, and describe an algorithm for the problem by either simulating an example or by giving its pseudocode.

   Grading:

   - 4p – Correct definition
   - 4p – Correct motivation
   - 4p – Correct simulation of pseudo-code