

Algorithms for Bioinformatics (Autumn 2015)

Exercise 4 (Tue 29.9., 10-12, B222)

If you cannot make it to the exercise session, please e-mail your solutions and the reason why you cannot attend to daniel.valenzuela@cs.helsinki.fi before the exercise session to get credit.

Some of the problems below are programming exercises on the Rosalind platform at <http://rosalind.info/problems/list-view/?location=bioinformatics-textbook-track>

1. Consider the alignment below:

```
ACGATGAT--CT
A-GA-CATAAAT
```

What is the alignment score in the edit distance model? What is the global alignment score with the mismatch and indel penalties -1 and match premium $+1$? What is the best local alignment score inside the given global alignment?

2. Compute the edit distance between **ACGTA** and **AGAA** by filling the dynamic programming matrix, and give all the optimal alignment(s).
3. Solve the Rosalind problem BA5G: *Compute the Edit Distance Between Two Strings*.
4. Solve the Rosalind problem BA5E: *Find a Highest-Scoring Alignment of Two Strings*.
5. We are interested in *overlap alignments* of strings A and B such that suffix of A is aligned against prefix of B . For example, an overlap alignment of **ACGATGAT** and **GACATAAAT** is

```
ACGATGAT
   GA-CATAAAT
```

Derive a variant of global alignment equations that gives the best scoring overlap alignment of A and B .

6. Derive a variant of edit distance equations that gives the overlap alignment of A and B with minimum cost, with the restriction that overlap should be at least of length ℓ . (Why is such restriction required?)