

Algorithms for Bioinformatics (Autumn 2014)

Exercise 1 (Tue 9.9., 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 juha.karkkainen@cs.helsinki.fi before the exercise session to get credit.

1. Simple things with Python I.

- a) Write a Python procedure that computes the reverse complement of a given DNA sequence.
- b) Find out what are *BioPython* and *FASTA* file format. How would you solve a) with BioPython?

2. Simple things with Python II.

- a) The course book (page 29) describes the naive and simple sorting algorithm called *selection sort*. Implement it with Python.
- b) What build-on procedures does Python offer for sorting? Are they expected to perform better than your implementation of selection sort?

3. Exhaustive enumeration.

Write a Python program that generates all possible DNA sequences of length ℓ .

4. Continuing with Python.

Write a Python program that implements $TotalDistance(v, DNA)$ -function from the lecture.

5. Partial digest.

Consider partial digest

$$L = \{1, 2, 3, 3, 4, 5, 5, 6, 8, 9\}.$$

Solve Partial Digest problem for L (i.e. find X such that $\Delta X = L$).

6. Motif finding.

You are studying a set of genes for which earlier studies indicate that there might be a transcription factor that binds to a motif that consists of two *half-sites*, i.e., having the structure $a_1a_2 \cdots a_m NNNNN b_1b_2 \cdots b_m$, where N is any symbol and B is the reverse complement of A . Modify `BranchAndBoundMedianStringSearch()` pseudocode studied at the lecture so that it finds motifs consisting of half-sites. (You may use `TotalDistance()`, if you define that N matches any symbol, i.e., induces cost 0.)