

Lecture: Bioinformatics I

WS 2015/16

Assignment No. 4

(10 points)

Hand out:

Monday, November 2

Hand in due:

Monday, November 9, 10:15

Direct inquiries to alexander.seitz@uni-tuebingen.de

Theoretical Assignments

1. Optimal multiple alignment

(2P, max. 1h)

Execute the dynamic programming algorithm for an optimal multiple sequence alignment (MSA) on the following three sequences:

A: CTT

B: TC

C: CCT

Use the sum-of-pairs (SP) score with $s(a, a) = 2$, $s(a, b) = -2$ for $a \neq b$, gap penalty $d = 2$.

2. Progressive alignment

(2P, max. 1h)

Perform a progressive alignment by hand of the sequences

A: CTT

B: TC

C: CCT

D: CACT

Use the parameters of the previous task. Compute first a distance matrix from all aligned pairs of sequences, then a guide tree (with a method of your choice). For aligning sub-alignments, use the “complete alignment” approach.

Practical Assignments

3. Comparing multiple alignments

(6P, max. 6-8h)

Many programs exist for multiple sequence alignments. Your task is to choose three of the following five programs and compare their results.

- ClustalW (<http://www.clustal.org/>)
- Clustal Omega (<http://www.clustal.org/>)
- Muscle (<http://www.drive5.com/muscle/downloads.htm>)
- T-Coffee (<http://www.tcoffee.org/Projects/tcoffee/>)
- MAFFT (<http://mafft.cbrc.jp/alignment/software/source.html>)

We provide test sets from BALiBASE and Pfam. In the zip folder you will find the original (non-aligned) sequences as well as the reference alignment as provided by BALiBASE and/or Pfam.

Use the non-aligned sequences to align them with the programs mentioned above. Then, compare the results to the respective reference alignment. Compare the programs with respect to the following three measures:

- total column score, which is the percentage of alignment columns in the reference alignment that were accurately reconstructed
- sum of pairs score, which is here (!) the percentage of pairs of aligned residues in the reference MSA that are similarly aligned in the reconstructed MSA
- computation time

The first two scores are implemented by the developers of BALiBASE, and available as C code.

If you do not want to use the code from BALiBASE, you can also use or implement other measures to compare the results. Be sure to state in your findings, why you used these scores.

We ask you to hand in your results with a written text (about 1 page) describing your methods, your approach, your findings and your conclusion (be as scientific as you can, i.e. your text should be similar to a typical paper-style text!). Also make clever use of a figure and/or table to illustrate your results.

Please read the questions carefully. If there are any questions, you may ask them during the tutorial session or via e-mail to your tutor. You will usually get an answer in time, but late e-mails (e.g. on Monday morning before class) might not be answered in time. Please send all your electronic solutions to `alexander.seitz@uni-tuebingen.de` or `alexander.peltzer@uni-tuebingen.de` (depending on your tutor). Please pack both your source code as well as the theoretical part into one single archive file. Source code should compile correctly. Make sure, that you export the source code and not only the binaries. Handwritten assignment solutions (e.g. for the theoretical part) can be turned in during the lecture.