ITX Integrative Transcriptomics
Dr. K. Nieselt, Faculty of Science

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Lecture: Bioinformatics I — WS 2015/16

Assignment No. 2 (10 points)

Hand out: Monday, October 19
Hand in due: Monday, October 26, 10:15
Direct inquiries to `alexander.seitz@uni-tuebingen.de`

Theoretical Assignments

1. **Hirschberg algorithm for alignment in linear space** (3P, max. 3h)

   Show all the steps of the Hirschberg algorithm for the sequences

   ```
   X = TGGA
   Y = TTGAGA
   ```

   Use the scoring function

   $$\begin{aligned} s(a,a) &= 4 \\ s(a,b) &= -3 \text{ for } a \neq b \ (a,b \in \Sigma) \\ \text{linear gap score penalty } d &= 3 \end{aligned}$$

Practical Assignments

2. **Do scoring matrices have expected score smaller than zero?** (1P, max. 1.5h)

   Calculate the expected score for each of the provided typically used substitution matrices. Assume uniform distribution of amino acids. Which of the provided matrices is not a useful substitution matrix?

   Do not perform these calculations by hand! Instead use or write an appropriate program for this task. Document your efforts and summarize your results.

3. **Needleman-Wunsch algorithm for affine gap scores** (6P, max 6h)

In the last assignment, you implemented a modular version of the Needleman-Wunsch algorithm for linear gap penalties. In the following, you are going to extend this program to allow affine gap penalties. If you don't want to use your previous solution, use the provided official solution for the linear gap Needleman-Wunsch algorithm.

Extend the existing program to allow affine gap penalties. Add command line options to

- run the program with linear gap penalty (as before)
- run the program with affine gap penalty and
- specify the *gap open* and *gap extension* penalties.

Please hand in your program as an **executable** `jar` file, that in addition contains your Java source files. Apply the program to compare `test1.fasta` with `test2.fasta`, as well as `PAX6Human.fasta` with `PAX6Mouse.fasta`. Run the program with

(a) `match` = 3; `mismatch` = -3; `gapopen` = 4 and `gapextend` = 1.

(b) `match` = 3; `mismatch` = -3; and the linear gap cost = 4.

Briefly discuss the different results (linear vs. affine).

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.