

Lecture: Bioinformatics I

WS 2015/16

Assignment No. 10

(10 points)

Hand out:

Monday, January 18

Hand in due:

Monday, January 25, 10:15

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

Theoretical Assignments

1. Assembly using *de Bruijn* graph (3P, 2h)

Construct the *de Bruijn* graph for the following set of sequencing reads $\mathcal{F} = \{f_1, \dots, f_5\}$ and a k -mer size of $k = 4$ (you do not need to add the reverse complements of the reads):

$f_1 = \text{ACCGT}$

$f_2 = \text{CGTAACGTT}$

$f_3 = \text{ACGTTA}$

$f_4 = \text{GTAA}$

$f_5 = \text{TAAACTG}$

Using your graph, compute an Euler path and the resulting superstring. Do you encounter any problems? Explain.

Now construct the *de Bruijn* graph for $k = 5$. Compare the two graphs, the resulting superstrings and discuss the differences. What can you say about the choice of k .

Practical Assignments

2. Assemble the human mitochondrial genome (7P, 4h)

Assemble the human mitochondrial genome using Velvet (<http://www.ebi.ac.uk/~zerbino/velvet/>) and SOAPdenovo (<http://soap.genomics.org.cn/soapdenovo.html>). To do so, download the reads from the additional material on the lecture's website. Then use the assemblers to produce contigs from it. Experiment with different k -mer sizes. What are your conclusions? Can you derive a relationship between the chosen k -mer and the number of resulting contigs? What is the optimal result?

Which of the two assemblers performed better? Think about statistics with which you can compare the results.

Note: In order to use SOAPdenovo you should have at least 5GB of physical memory. If you do not have such a system or you run into memory problems during execution, please contact your tutor.

Hand in a short scientific paper, including a brief description of the methods, the results and discussion. Remember to cite all used sources and provide all references in proper form.

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.