



Lecture: Grundlagen der Bioinformatik

SoSe 2022

Assignment 4

(20 points)

Hand out:

Hand in due:

Direct inquiries via the ILIAS forum or to your respective tutor at:

Mathias Witte Paz: iizwi01@uni-tuebingen.de

theresa-anisja.harbig@uni-tuebingen.de

meret.haeusler@student.uni-tuebingen.de

jules.kreuer@student.uni-tuebingen.de

simon.heumos@qbic.uni-tuebingen.de

Thursday, May 19

Thursday, May 26, 18:00

Theoretical Assignments

1. Unlabeled branching pattern

(2P)

For unrooted binary trees, how many leaves do there have to be at least to obtain more than one unlabeled branching pattern? Provide sketches of the topologies to support your answer.

2. Proof of cluster distance computation

(4P)

Show that, if the distances between two clusters C_i and C_j are defined by the following equation

$$d(i, j) = \frac{1}{|C_i||C_j|} \sum_{x \in C_i, y \in C_j} d(x, y)$$

and if $C_k = C_i \cup C_j$, then $d(k, l)$ for any l is given by

$$d(k, l) = \frac{d(i, l)|C_i| + d(j, l)|C_j|}{|C_i| + |C_j|}$$

3. Manually reconstruct the NJ and UPGMA tree*

(6P)

Reconstruct the Neighbor-Joining and UPGMA tree for the following input distance matrix of 4 taxa A, B, C, D . Show **all intermediate steps** (you may hand-in a scan of your handwritten solution). Afterwards, compare the results of the two approaches. Which tree if any perfectly represents the input data? (* This task has a runtime of two weeks, you can either hand in the solution until May 26 or June 2).

	A	B	C	D
A	0	8	7	4
B		0	5	6
C			0	5
D				0

4. Marker genes for the prokaryotic tree of life (2P)

In the lecture, two universal marker genes (16S rRNA and aminoacyl tRNA synthetase) were discussed since they are used in the phylogenetic reconstruction of prokaryotes. Try to find two other marker genes in prokaryotes that are also used for this purpose. Write 2 sentences (around 40 words, excl. citations) for this subtask.

Practical Assignments

5. Phylogenetic analysis of 16S rRNA sequences (6P)

For this exercise we ask you to use the phylogenetic software package MEGA11 <http://www.megasoftware.net/> (note that this software package needs a registration, but it is free for academic users). Central to this task is a multiple sequence alignment of 16S rRNA sequences from various organisms (16SrRNA.fasta).

- (a) Choose 2 different phylogenetic reconstruction methods from MEGA11 to compute phylogenetic trees of the provided 16S rRNA alignment.
- (b) Summarize and discuss your results: Report your results and provide figures of the reconstructed trees with an informative caption as well as the name of the methods and the precise parameters used for (a). In your text, make sure that you correctly refer to the included figures. Discuss: Do the different methods agree on the tree? Maybe only in parts? For this also state the possible reasons why the two different reconstruction methods result in possibly different phylogenies even when applied to the same data. In total, write about 400 words.

Remember to cite your sources and the used software correctly.

Please read the questions carefully. If there are any questions, you may ask them during the tutorial session or in the forum of ILIAS. You will usually get an answer in time, but late e-mails (e.g. the evening of the hand-in) might not be answered in time. Please upload all your solutions to ILIAS. Don't forget to put your names on every sheet **and** in your source code files. Please pack both your source code as well as the theoretical part into one single archive file and give it a name using this scheme: <name1>_<name2>_<Assignment>_<#>.zip. The program should run without any modification needed.