

Master Thesis

Comparison of Microbiome Analysis Tools

Taxonomic and phylogenetic fingerprinting based on sequence analysis of gene fragments from the small-subunit (SSU) 16S rRNA gene has become the standard technique in bacterial and archeal classification [1]. At the moment a variety of tools and techniques is available for the analysis of 16S amplicon sequencing data [2]. To get an overview about usage and range of functions the most popular analysis tools, mothur [3], QIIME [4], UniFrac [5] and the RDP-Classifer [6] within SnoWMAAn [7] should be evaluated and finally compared.

The overall goal of this project is to compare different microbiome analysis tools in respect to usability, range of functions and final results. The following thesis objectives are specified:

1. Evaluate mothur, QIIME, UniFrac and the RDP-Classifer of SnoWMAAn.
2. Find an existing well described 16S amplicon sequencing data set or create your own mock-community
3. Analyze the selected/created 16S amplicon sequencing data-set with each tool.
4. Compare analysis principles of the different tools.
5. Compare different tools in respect to usability.
6. Compare different tools in respect to available features.
7. Ultimately, compare final results with regards to number of analyzed sequences, number of determined OTUs, α - and β -diversity measures, taxonomic distribution and principal component analysis (PCA).

References:

1. Halwachs B, Gorkiewicz G, Thallinger G: **High-Throughput Characterization and Comparison of Microbial Communities**. In *Computational Medicine*. Edited by Trajanoski Z. Springer Vienna; 2012:37-57.
2. Morgan XC, Huttenhower C: **Chapter 12: Human microbiome analysis**. *PLoS Comput Biol* 2012, **8**: e1002808.
3. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.*: **Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities**. *Appl Environ Microbiol* 2009, **75**: 7537-7541.
4. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.*: **QIIME allows analysis of high-throughput community sequencing data**. *Nat Methods* 2010, **7**: 335-336.
5. Lozupone C, Knight R: **UniFrac: a new phylogenetic method for comparing microbial communities**. *Appl Environ Microbiol* 2005, **71**: 8228-8235.
6. Cole JR, Wang Q, Chai B, Cardenas E, Farris RJ, Kulam-Syed-Mohideen AS *et al.*: **The Ribosomal Database Project**. *10 February, 2008-13 February, 2008*; 2008.
7. Halwachs B, Hoftberger J, Stocker G, Snajder R, Gorkiewicz G, Thallinger GG: **High-Throughput Characterization and Comparison of Microbial Communities**. *Biomed Tech (Berl)* 2013.