

Exercise 5: Analysis of Microbiome Data

This exercise deals with the analysis of microbiome data from a diarrhea study (Gorkiewicz G, Thallinger GG, *et al.*, 2013). The goal of the study was to investigate the influence of diarrhea on the composition of the microbial community in the human gut. To this end, stool samples from four healthy individuals at four different time points have been obtained:

Table 0.1: Description of the time points within the diarrhea study.

P	Time point	Description
1	day -7	taken 7 days before diarrhea was induced. All patients have been set on the same diet.
2	day 0	taken before diarrhea was induced
3	day 3	sample at the end of the diarrhea
4	day 10	last sample, taken 10 days after diarrhea induction

The 16S DNA was extracted from the samples and sequenced using Roche/454 technology. A basic analysis of the 16S sequence data has been performed in the microbiome analysis pipeline **SnoWMA**n (<http://snowman.genome.tugraz.at>). The result of this analysis is available in form of a tab separated file summary file (*global-summary-dist-0.03.txt*) from the course web site. The structure of the file is described in the document (*SnoWMA*n *Summary.txt*).

5.1 Summary Statistics

Read the summary file in a data.frame and report the total number of OTUs and include a table with the number of OTUs and total number of reads per sample. Create a barchart with the total number of sequences per sample.

5.2 Taxonomic Classification

Aggregate the OTUs for each sample on the "Phylum" level and write the resulting data.frame to a tab separated file. Report the number of unique Phyla per sample in form of a table and create a stacked barchart with the relative phylum distribution per sample.

5.3 Alpha and Beta - Diversity

Create a per sample table with the following alpha diversity measures: Chao1, ACE, Richness, Evenness and Shannon. Create a heatmap of the sample beta diversity using the Bray-Curtis distance.

- Compare and discuss the alpha diversity of the different patients for each of the time points.
- Compare and discuss the beta diversity of one sample (patient) over all time points.

- Compare and discuss the beta diversity of all four samples (patients) at the same time point.

5.4 Differentially Abundant Taxa (optional)

Identify taxa at the phylum level which are differentially abundant between time-points 2 and 3. Write the result to a tab delimited file. Include a table with all phyla, their name, fold-change and the false discovery rate. How many of the statistically significant phyla are up-regulated, how many are downregulated ? Give your answer for both an absolute fold-change of 2 and a false discovery rate less than 0.05.

5.5 Sweave Report

Create your report using a Sweave file which contains text, R commands and results in a single file.