Metagenomic Systems Biology of the Human Microbiome

The human microbiome is an integrated part of the human body, outnumbering the human cells by approximately a factor 10. These microorganisms are very important for human health, hence knowledge about this, "our other genome", has been growing rapidly in recent years. This is mainly due to the advances in next generation sequencing, which has allowed for large-scale metagenomics studies of different niches of the human microbiota. Especially the gut microbiota has been studied intensively. However, most studies have been purely descriptive, thus there is still a lot to learn regarding the interplay between species in the microbiota and also between the host and the inhabiting microorganisms. Additionally, the non-bacterial part of the microbiota, which includes bacteriophages, plasmids and micro-eukaryotes, is not very well described.

In this thesis, metagenomics data from the human gut, nose and oral cavity has been analyzed. The central method has been a co-abundance clustering method, which separates genes from metagenomics data under the assumption that genes originating from the same DNA (e.g. a bacterial genome, a phage or a plasmid) will co-vary across samples. Thus, co-abundance gene groups (CAGs) are obtained, which represent bacterial genomes, phages, plasmid or other genetic elements in the system. The ability to reassemble the metagenome in this way opens up new possibilities for analyzing the functional potential of species in the microbiota as well as the interactions in the system. Applying the CAG clustering method to data from the human gut microbiome, we identified dependency-associations between plasmids, phages and clone-specific gene sets to their bacterial host. Connections between CRISPR-elements and phages were also observed. Additionally, the persistence of some bacterial species in the human gut could be predicted based on absence or presence of specific genetic modules.

Based on the same CAG clustering of the human gut microbiome data, the link between bile acid degradation of bacteria in the gut and obesity was investigated. There seemed to be a slight correlation between the two. However, this remains to be a hypothesis for further studies. Furthermore, the prevalence of the parasite Blastocystis in the human gut microbiome data was analyzed. This is the first time a metagenomics approach has been applied to this problem and the results were similar to previous Blastocystis prevalence studies. Moreover, it was found that individuals with a Bacteroides-driven enterotype were less prone to harbor the Blastocystis parasite.

Finally, the CAG clustering method was applied to metagenomics data from the human nose- and oral-cavity. It was concluded that this method needs further improvement in order for it to be directly transferable to other datasets. In summary this thesis presents co-abundance gene groups (CAG) clustering as a valuable tool for analyzing human microbiome data. Furthermore, results based on this method regarding important topics in relation to the human gut microbiota are described, including the interplay between bacterial species and other genetic elements in the system, factors that might influence development of obesity and prevalence studies of eukaryotes. Studies of other areas of the human microbiome might also benefit from CAG based analyses once the tool has been optimized.