Beyond 16S rRNA Profiling - Efficient Algorithms and Software for Taxonomic and Functional Analysis of Shotgun Metagenomics Data

Prof. Dr. Daniel Huson
Chair of Algorithms in Bioinformatics,
Center for Bioinformatics,
University of Tuebingen
Visiting Professor,
Life Sciences Institute,
National University of Singapore

Abstract
There is currently much interest in microbiome analysis, the study of microorganisms in a particular environment, such as the human body, water or soil. Much work has been based on 16S rRNA profiling, which allows one to identify bacteria down to the taxonomic rank of genus. The resulting datasets are small and computational analysis is straightforward and takes little time. The cost of sequencing is now so low that using shotgun sequencing to collect tens of millions of metagenome sequencing reads per sample is easily affordable, even for large studies involving many tens or hundreds of samples. The analysis of such datasets will allow one to address more detailed taxonomic and functional questions. This development raises a number of difficult computational questions, namely how to efficiently align hundreds of millions of sequencing reads against a protein reference database such as NR? How to efficiently store and provide access to the results? How to enable interactive exploration and analysis of large numbers of very large samples? In this talk, we present three new tools that we have developed to address these issues. Our new alignment tool DIAMOND allows one to compare Illumina reads against the NR database at 20,000 times the speed of BLASTX, without loss of sensitivity, thus making functional analysis of very large datasets possible. We are soon to release version 6 of MEGAN, our interactive metagenome analysis tool, which will have many new features that explicitly support the analysis of large numbers of large datasets. In addition, our new tool MeganServer allows one to host all files on a server and to access them remotely from within MEGAN. We also describe recent work on gene-centric assembly. To illustrate the use of these new tools, we will report on two projects that we are involved in, namely one that focuses on impact of antibiotics on resistance genes in the human gut and one that analyzes halogenases and dehalogenases in soil.

Selected Publications for Reference