

## Computational Tools for the Taxonomic Analysis of Shotgun Metagenome Samples

**Presenter:** [Johannes Dröge](#), Heinrich-Heine-Universität Düsseldorf

**Motivation:** The taxonomic annotation of environmental sequence data provides insight into the structure and functions of microbial communities by assigning genes and their corresponding functions to specific phyla. This information is used to generate taxonomic profiles of metagenome samples for comparisons of different microbial communities and to recover genomes of individual community members from a metagenome. In this demo talk, we present two programs: *Taxator-tk* and *PhyloPythiaS+*. Both implement robust and flexible algorithms to adapt to the heterogeneous nature of metagenome data generated for microbial communities of varying organismal complexity, evolutionary distances from sequenced genomes and sequencing technologies. We consider processing speed an equally important property which is needed for characterizing the large data-sets generated by current and future sequencing technologies.

### Topics:

#### 1) Software *taxator-tk*

We present the theoretical foundations and design of a software package called *taxator-tk*. It performs taxonomic assignment of nucleotide sequences by fast, approximate determination of evolutionary neighbors from sequence similarities. We demonstrate how the phylogenetic approach greatly reduces false-positive assignments with exemplary data sets and that this results in robust annotations regardless of the type of input data. The assignment algorithm works with arbitrarily-sized nucleotide sequences, ranging from ultra-short to genome-long, and assigns nucleotide sequences of Bacteria, Archaea and Eukaryotes.

We show a sample pipeline for the generation of taxonomic profiles of metagenome sequence samples using *taxator-tk* and have a look on the input, intermediate and output data and format. The modular design enables users to create custom pipelines with different applications and homology search programs. The audience is encouraged to ask questions and discuss on various aspects of the software.

#### 2) Software *PhyloPythiaS+*

We shortly present the composition-based taxonomic assignment program *PhyloPythiaS+*. It automatically derives a structured SVM (Support Vector Machine) sample model via marker gene identification and classification. After training of the model with sample-derived and reference nucleotide sequences, it can rapidly assign metagenomic sequences of sufficient length to corresponding sample taxa.

**Organization:** The demo is divided into two parts with a focus on *taxator-tk* (first part).

**Availability:** All software is available at <http://algbio.cs.uni-duesseldorf.de/software/>.