

## Strategies to assemble and analyse metagenomes based on long-read sequencing data

Duration: 6 months

Metagenomics deals with the identification of the different species that can be contained in a single sample taken from the environment. The reconstruction of the genomes that are contained within such samples is challenging due to differences in species abundance and the possible presence of several genotypic versions of the same taxon. Novel sequencing technologies generating kbp-long sequencing reads hold great promise for metagenomics studies since long reads substantially facilitate microbial genome reconstruction. In this Master thesis, public long-read sequencing data sets from metagenomes should be used to benchmark and assess software for metagenome reconstruction and taxonomic analysis. Specifically, the project will comprise the following aspects:

- Assessing metrics and quality of public metagenomic long-read sequencing read datasets
- Sequence correction
- Benchmarking of assembly software
- Assessing the species composition based on assembled and unassembled data
- Comparison and evaluation of metagenome assembly and taxonomic annotation strategies

The computational work will be performed on high-performance Linux systems. Ideally, students should have experience with using Linux commands and working on sequencing data, e.g. as covered in the courses “Essentials for bioinformatics data analysis” and “Sequencing data analysis”. Nevertheless, if you are interested in the project please contact us irrespective of your current skills.

The working group “Bioinformatics” is located at the 6<sup>th</sup> floor of building MUG II. For information about the group, visit <http://seq.boku.ac.at>.

Contact: Ass.Prof. Dr. Juliane Dohm ([dohm@boku.ac.at](mailto:dohm@boku.ac.at))