Automated analysis of tagged amplicon sequences from Next Generation Sequencing

Thomas Berner[1], Holger Heuer[2] and Jens Keilwagen[1]
[1]Julius Kühn-Institut, Institute for Biosafety in Plant Biotechnology, Quedlinburg
[2]Julius Kühn-Institut, Institute for Epidemiology and Pathogen Diagnostics, Braunschweig
Email of corresponding author: thomas.berner@jki.bund.de

Repeated tasks can be handled with workflow management systems that allow for defining, managing and running series of task to produce specific results. Using the open-source software Galaxy, we provide a workflow management system for the JKI.
Based on this system, we provide a workflow for analyzing tagged amplicon sequencing data to predict operational taxonomic units (OTU) and analyse their phylogenetic affiliation. This can be used to compare communities of bacteria, fungi or nematodes between experimental samples.

In a typical experiment, soil or rhizosphere was sampled with replicates from two or more treatments. Total DNA of each sample was extracted and used as target for amplifying small subunit (SSU) rRNA genes (bacteria, nematodes) or internal transcribed space (ITS) fragments (fungi) by PCR. For Next Generation Sequencing, a library was constructed with fusion primers containing sequencing adapters and a sample-specific eight-base barcode.

We used the 454 Genome Sequencer FLX platform (Roche – 454 Life Sciences, Branford, CT) at Biocant (Cantanhede, Portugal) for library construction and sequencing. Typically 2000-10000 sequences were obtained for each sample.

The workflow for SSU rRNA genes compares each sequence to a ribosomal database by BLAST. In addition, the RDP naive Bayesian classifier is used for taxonomic classification. All sequences are clustered into OTU based on pairwise similarities. For each OTU, the workflow reports its taxonomic classification as well as the number of observed sequences for each sample.

This OTU report is the basis for further analysis to statistically compare communities and find taxa, which respond to the experimental treatment.