Geophysical Research Abstracts
Vol. 18, EGU2016-17511, 2016
EGU General Assembly 2016

Fast, accurate and easy-to-pipeline methods for amplicon sequence processing

Livio Antonielli and Angela Sessitsch

AIT Austrian Institute of Technology GmbH, Austria, Health & Environment Department, Bioresources Unit (livio.antonielli@ait.ac.at)

Next generation sequencing (NGS) technologies established since years as an essential resource in microbiology. While on the one hand metagenomic studies can benefit from the continuously increasing throughput of the Illumina (Solexa) technology, on the other hand the spreading of third generation sequencing technologies (PacBio, Oxford Nanopore) are getting whole genome sequencing beyond the assembly of fragmented draft genomes, making it now possible to finish bacterial genomes even without short read correction.
Besides (meta)genomic analysis next-gen amplicon sequencing is still fundamental for microbial studies. Amplicon sequencing of the 16S rRNA gene and ITS (Internal Transcribed Spacer) remains a well-established widespread method for a multitude of different purposes concerning the identification and comparison of archaeal/bacterial (16S rRNA gene) and fungal (ITS) communities occurring in diverse environments. Numerous different pipelines have been developed in order to process NGS-derived amplicon sequences, among which Mothur, QIIME and USEARCH are the most well-known and cited ones.
The entire process from initial raw sequence data through read error correction, paired-end read assembly, primer stripping, quality filtering, clustering, OTU taxonomic classification and BIOM table rarefaction as well as alternative “normalization” methods will be addressed. An effective and accurate strategy will be presented using the state-of-the-art bioinformatic tools and the example of a straightforward one-script pipeline for 16S rRNA gene or ITS MiSeq amplicon sequencing will be provided. Finally, instructions on how to automatically retrieve nucleotide sequences from NCBI and therefore apply the pipeline to targets other than 16S rRNA gene (Greengenes, SILVA) and ITS (UNITE) will be discussed.