

Exploration of environmental metagenomes and metatranscriptomes: current possibilities and limitations in data analysis

Petr Baldrian

Laboratory of Environmental Microbiology, Institute of Microbiology of the ASCR, Prague, Czech Republic
<http://www.biomed.cas.cz/mbu/lbwrf>

Environmental metagenomes and metatranscriptomes are extremely complex, considering that one gram of soil may harbor tens of thousand species of bacteria and thousands of species of eukaryotic microorganisms. Their exploration thus currently relies on methods delivering relatively long sequence reads, i.e. these obtained with the Roche or Pacific Biosciences instruments. Shotgun approaches are combined with sequencing of PCR amplicons of genes with sufficient taxonomic resolution (rDNA) or, less frequently, functional genes. Our recent experience shows, that a description of total (DNA sequencing) and active (sequencing of cDNA derived from environmental RNA) soil microbial communities or the pools of functional genes and their transcripts (mRNAs) can be sufficiently well characterised using amplicon sequencing (1). The analysis of metagenomes is much more challenging since the sequence identity has to be determined and the assignment of functions and microbial producers to such sequences is not trivial. Current possibilities of metagenomic data analysis would benefit mainly from the tools allowing to search not only in GenBank (as most of the current tools do) but also in the full genomes of individual microorganisms, or, as a best option, in a database covering all these genomes. Furthermore, amplicon sequencing, that now relies on the construction of consensus sequences representing putative microbial species (OTUs, operational taxonomic units) would greatly advance if an automatic tool of consensus construction of all identified similarity clusters is developed. As our first results in the field of environmental metaproteomics show, even more sophisticated tools would be needed if metaproteomic data, typically short sequences of amino acids, need to be compared with nucleotide sequences obtained using DNA or cDNA sequencing.

References

1. Baldrian, P., Kolarik, M., Stursova, M., Kopecky, J., Valaskova, V., Vetrovsky, T., Zifcakova, L., Snajdr, J., Ridl, J., Vitek, C., Voriskova, J. (2011) Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME Journal in press, doi:10.1038/ismej.2011.95.