

Interactive comment on “Diversity and seasonal dynamics of airborne *Archaea*” by J. Fröhlich-Nowoisky et al.

R. Conrad

conrad@mpi-marburg.mpg.de

Received and published: 3 July 2014

Reviewers #2 and #3 both criticize the use of Sanger sequencing instead of next generation sequencing (NGS), which would have allowed a much deeper assessment of the community structure. Especially Illumina sequencing nowadays allows obtaining millions of sequences at a very reasonable price. I completely agree with these reviewers about the advantages of NGS. Nevertheless, there were good reasons why we applied Sanger sequencing instead.

1. The main reason was that we started the project with first air samples in 2006. At this time, NGS has just been invented (Margulies et al. Nature 437 (2005) 376). After that it still took several years until tagged sequencing (which would have been necessary for our study) had been established and the evaluation pipelines for the large number

C3198

of sequences were available. Illumina sequencing, in particular, delivered until recently only very short reads (<100 bp) of poor quality. If we would start the project today, we would certainly consider NGS as a possible method. However, at the beginning of our project, we could only use Sanger sequencing.

2. NGS is very powerful when targeting samples with a high potential richness of different microbes. Estimation of true diversity would require sequencing up to saturation. For a soil community with 10⁹ individual microbes per gram this would require about 10⁹ sequences. Quite obviously, only NGS could come near to this requirement. However, the archaeal community of air samples is not as rich. We are talking of the order of 10 individual archaea per m³ air. Admittedly, this would also add up to 10⁴ archaea in 1000 m³, but the situation is by orders of magnitude more relaxed than when analyzing a soil microbial community. Literature shows that most soil studies using NGS are satisfied with analyzing a few thousand sequences. Therefore, the relatively low sequencing depth in our study is in proportion.

3. An important objective was to obtain phylogenetic information out of the sequences retrieved from air samples. Therefore, we aimed at relatively long sequences (ca. 600 bp). Not even the most modern Illumina technology allows such read lengths (250 bp with paired end sequencing). The safest way for obtaining phylogenetic information is aligning the sequences with others that represent the relevant phylogenetic breadth and constructing trees. This procedure is straight forward using Sanger sequences. With thousands of NGS sequences, however, such procedure is not possible. Instead the sequences are compared (e.g. BLAST) to data bases, which sometimes are of insufficient quality. Since many of our archaeal sequences belong to the crenarchaeotal/ thaumarchaeotal branch of the archaeal domain, which hardly contain any known species (in 2006 there were only 2-3 species), BLAST analyses would not give satisfactory results. In such case, Sanger sequencing is still the gold standard.

The relatively large frequency of archaeal sequences belonging to the thaumarchaeotal group is the reason why the amoA gene (coding for a subunit of the ammonia monoxy-

C3199

genase, the key enzyme of ammonia oxidizing nitrifiers), which characteristically occurs in Thaumarchaeota, was also sequenced. Targeting this gene provides additional phylogenetic information for this not well characterized archaeal group.

Interactive comment on Biogeosciences Discuss., 11, 6945, 2014.

C3200