

Diversity and distribution of restriction-modification systems in natural microbial multi-species community of Antarctic Deep Lake

Karyagina A.^{1,2,3*}, Rusinov I.¹, Ershova A.^{1,4}, Alexeevski A.^{1,5}, Spirin S.^{1,5,6}

¹ Belozersky Institute, Lomonosov Moscow State University, Moscow, Russia

² Gamaleya National Research Center of Epidemiology and Microbiology, Moscow, Russia

³ All-Russia Research Institute of Agricultural Biotechnology, Moscow, Russia

⁴ Moyne Institute of Preventive Medicine, Trinity College Dublin, Dublin, Ireland

⁵ Institute of System Studies, Moscow, Russia

⁶ National Research University Higher School of Economics, Moscow, Russia

* akaryagina@gmail.com

Key words: restriction-modification system, Antarctic Deep Lake, horizontal transfer of genes

Motivation and Aim: The distribution of Restriction-Modification (R-M) systems, which are prokaryotic defense systems, in natural multi-species communities has not been yet studied. Thus, distribution of the R-M system genes in stable and closed microbial community of hypersaline Antarctic Deep Lake [1] is of particular interest.

Methods and Algorithms: We performed a bioinformatic analysis of R-M system genes in five metagenomes and in complete genomes of six strains of four archaeal species from Deep Lake. The analysis is based on homology search of metagenomic contigs and genomic sequences over REBASE proteins and on considering R-M system related Pfam domains.

Results: We predicted about 5000 R-M system genes, which can be grouped into 1400 clusters of homologous genes (with > 50 % identity of encoded proteins) and 2300 clusters of nearly identical genes (with > 98 % identity). Only 97 clusters of nearly identical genes are represented in the genomes of the four archaeal species. The lists of R-M system genes found in the complete genomes is significantly different from the lists of genes annotated in REBASE. There are more than 60 putative R-M system genes in these genomes not included in REBASE while REBASE contains 18 genes from the four archaeal genomes that do not meet our criteria. For one of the studied species, *Halorubrum lacusprofundi*, we demonstrate high inter-strain heterogeneity of R-M system composition. Most R-M system genes common for different species are located within rather large highly identical regions. We found that some R-M system genes in complete genomes have metagenomic read coverage significantly different in comparison with housekeeping genes of the same species.

Conclusions: The results for archaeal complete genomes of most abundant species show significant differences with lists of genes annotated in REBASE for the same genomes, which may be explained by difference in the methods of R-M genes identification. The use of an additional criterion, namely the presence of R-M system-related Pfam domains, can make the search for genes of the R-M systems more reliable compared to only homology search vs. REBASE. Some obvious mistakes in REBASE were identified. The microbial community of Antarctic Deep Lake possesses a lot of diverse R-M systems. Only a small part of their variety is encoded in genomes of the four archaeal

species. A significant difference in R-M system related genes between strains for *H. lacusprofundi* is probably a result of an intense gene exchange. Another sign of an active gain and loss of R-M systems is a huge difference in metagenomic read coverage between R-M system genes within one genome, which can interpreted as a presence of strains of the same species but with different R-M system repertoires.

Acknowledgements: The work was supported by the Russian Science Foundation grant 21-14-00135.

References

1. DeMaere M.Z. et al. High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc Natl Acad Sci U S A*. 2013;110(42):16939-16944.