DOI 10.18699/BGRS/SB-2020-006

Metavirome analysis of Baikal sponges

Tatyana Vladimirovna Butina
Laboratory of Analytical and Bioorganic Chemistry
LIN SB RAS
Irkutsk, Russia
tvbutina@mail.ru

Yurij Sergeevich Bukin
Laboratory of Genosystematics
LIN SB RAS
Irkutsk, Russia
bukinyura@mail.ru

Igor Veniaminovich Khanaev
Laboratory of Ichthyology
LIN SB RAS
Irkutsk, Russia
igkhan@lin.irk.ru

***Abstract* — Sponges are the oldest multicellular invertebrates (phylum Porifera); they are ecologically important components of marine and freshwater benthic environments. Associated communities of sponges include a variety of microorganisms: fungi, algae, archaea, bacteria and viruses. The aim of this study was to elucidate the genetic diversity of viruses in the community of Baikal endemic sponges *Baikalospongia bacillifera* using metagenomic approach. We have shown for the first time a high genetic, potential taxonomic and functional diversity of dsDNA viruses in these Baikal sponges. Identified viral sequences belonged to 28 viral families that infect a wide range of organisms. The bacteriophages of the *Myoviridae*, *Siphoviridae* and *Podoviridae* families dominated in the samples. The viruses of the *Phycodnaviridae*, *Poxviridae* *Mimiviridae*, *Herpesviridae*, *Baculoviridae*, *Polydnaviridae* and *Iridoviridae* families were also the most numerous. Viral communities of visually healthy and diseased Baikal sponges were significantly different. Analysis of viral sequences has indicated 22 functional categories of proteins and revealed a wide variety of structural viral proteins and enzymes. Among those the genes of proteins involved in the metabolism of host cells were also identified. Thus, the role of viruses in sponges may be both in the regulation of the number and diversity and in the maintenance of the vital activity of their hosts and the associated community as a whole.**

Keywords* — *viruses, viral diversity, metagenomics, virome, sponge holobiont, freshwater sponges, Lake Baikal

Introduction

Sponges are the oldest multicellular invertebrats (phylum Porifera) that represent complex symbioses (holobionts) in marine and freshwater ecosystems. These unusual animals are active biofilters and, due to their mass distribution, they play an important role in the ecosystem of Lake Baikal.

The community of sponges includes various microorganisms – fungi, algae, archaea, bacteria and viruses; they also serve as food and habitat for many invertebrates. The number, diversity and the role of viruses in the sponge holobionts may be very significant; however, viruses in the symbiotic associations of sponges remain poorly understood compared to other microorganisms. The first studies of the genetic diversity of sponge holobiont viruses have recently been published and concern some species of sponges inhabiting reefs [1, 2] and Lake Baikal (branched sponge *Lubomirskia baikalensis*) [3]. The results of these studies demonstrate the high diversity and important role of viruses in the composition of marine and freshwater sponge communities.

Against the backdrop of global climate change and increasing anthropogenic pressure in Lake Baikal, the mass destruction and death of sponges is registered, starting in 2011. The causes of the disease are still unknown.

The aim of this work was to study the genetic diversity of viruses in the associated community of the Baikal endemic sponges *Baikalospongia bacillifera* (diseased and visually healthy) using the metagenomic approach. Sponges *B. bacillifera* – one of the most abundant in Lake Baikal, they have a massive globular shape.

I. Materials, and Methods

The sponges *B. bacillifera* were sampled in the southern basin of Lake Baikal (near Bolshiye Koty, 51°54'07.5"N, 105°06'12.0"E) at depths of 16-20 m in May 2018 using lightweight diving equipment. Two individuals of *B. bacillifera* of 5-7 cm^3 in volume were collected: one looked healthy (Sv2478), and another had necrosis lesions (Sv2475). The sponge samples were twice washed in sterile Baikal water and thoroughly homogenized. Then homogenates were frozen in nitrogen and transported to the laboratory. The samples were gently thawed, twice diluted with SM buffer (0.2 M NaCl; 10 mM $MgSO_4$; 50 mM Tris HCl, pH 7.5), shaken with a Heidolph Multi Reax Vortex Mixer (10,000 rpm, 30 min), and were centrifuged 400 g for 15 min followed by 16,000 g for 30 min. The aqueous fraction was passed through a syringe filter with a pore size of 0.2 μm (Sartorius) and treated with DNase I and RNase A enzymes (Thermo Fisher Scientific) to remove contaminating nucleic acids. Viral DNA was extracted by ZR Viral DNA kit (Zymo Research).

The preparation and sequencing of DNA libraries were performed in The Center of Shared Scientific Equipment "Persistence of microorganisms" of Institute for Cellular and Intracellular Symbiosis UB RAS, Russia. The paired-end libraries were prepared using a NEBNext Ultra II FS DNA Library Prep Kit for Illumina (NEB) according to the manufacturer's protocol. The validation of DNA libraries was verified by Agilent 2100 Bioanalyzer (Agilent Technologies). Sequencing of the libraries was conducted on a MiSeq genome sequencer using MiSeq Reagent Kit v3 (2x300cycles, Illumina).

The primary processing (quality control and trimming) of the metavirome datasets (paired reads of 2x300 bp) was performed using the R package "ShortReads" [4]. The first (up to 15) and last (up to 30) nucleotides with low quality were removed. The sequences of less than 80 nucleotides were excluded from datasets.

Bioinformatic analysis of data was performed as described previously [3]. Briefly, taxonomic identification of viral sequences was performed using the BLASTn algorithm [5] against NCBI RefSeq viral complete genomes database (September 2018 release). The sequence reads were considered 'identified' if they had a relative in the reference database with an e-value of $\leq 10^{-5}$ and bit score ≥ 50. For the functional annotation of viral sequences, we used the local Blastx application [5] and COG database [6]. The reliability of the difference between two *B. bacillifera* viral communities

was estimated using the chi-square test. Statistical calculations were performed using the R packages “vegan” and “pvclust”.

Results and discussion

The raw data contained 3 842 088 and 5 035 528 pair sequence reads for the samples (Sv2475) and (Sv2478), respectively. After quality processing of data, we have obtained 3 375 063 and 4 063 311 reads, ranging from 80 to 256 bp. Of them, 97 557 and 88 517 sequences were identified as viral using the NCBI RefSeq viral genomes database, accounting for ca. 2.9 and 2.2% of datasets.

The families *Myoviridae*, *Phycodnaviridae*, *Siphoviridae*, *Poxviridae*, *Podoviridae*, *Mimiviridae*, *Herpesviridae*, *Baculoviridae,* and *Iridoviridae* were the most numerous, represented more than 1% of the sequences and in total accounted for more than 70% of the identified virome sequences (Fig. 1). We did not classify the significant parts of viral reads (21.4 and 23.9% in the samples Sv2475 and Sv2478, respectively) at the family rank. The diversity, richness and difference of two viral communities were estimated using Shannon, Simpson, ACE and Chao1 indices [7] (Table 1), rarefaction technique and chi-square test. The rarefaction curves for the both samples reached a plateau (data not shown).

TABLE 1 - BIODIVERSITY AND RICHNESS INDICES FOR THE VIROME DATASETS

Samples	Chao1/ACE	Shannon index	Simpson index
*B. bacillifera*_Sv2475	986/986	5.26	0.98
*B. bacillifera*_Sv2478	973/973	5.36	0.98

A comparative analysis of the diseased and visually healthy sponges did not reveal significant differences at the family level, the composition of the families was the same, but their percentages varied. In general, the viral communities of visually healthy and diseased Baikal sponges were significantly different (p-value < 2.2e-16). This indicates the changes in the structure of microbial communities in the affected sponges and the propagation of microorganisms and viruses that are not characteristic for the normal state of the sponges.

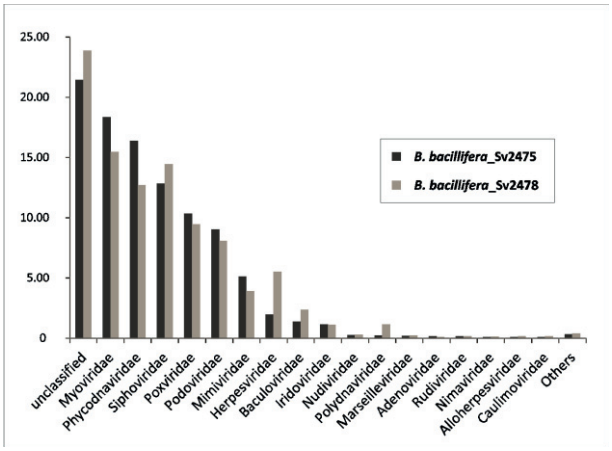

Fig. 1. The proportion of identified DNA viral families and viruses that were unclassified at the family rank

The comparison of revealed viral reads with the COG database has indicated the 22 functional categories of proteins and enzymes. The most representative (more than 5%) were proteins of replication, recombination and repair, nucleotide transport and metabolism, and mobile genomic elements (prophages and transposons).

Structural viral proteins, as well as the enzymes involved in the synthesis, modification, and packaging of DNA/RNA (methylases, reductases, terminates, helicases, polymerases, etc.) were the most numerous in *B. bacillifera*. The proteins of the host metabolism (auxiliary metabolic genes, AMGs) were also identified. The viruses, in particular bacteriophages, may synthesize the proteins necessary to maintain the vital activity of bacteria under unfavorable conditions. This, in turn, contributes to greater production of viral particles. Thus, viruses can play an important role in sponges by regulating the number and variety of micro- and macroorganisms, as well as participating in the metabolism and maintenance of the vital activity of their hosts.

References

[1] P.W. Laffy, E.M. Wood-Charlson, D. Turaev, K.D. Weynberg, E.S. Botté et al., “HoloVir: A workflow for investigating the diversity and function of viruses in invertebrate holobionts”, Front. Microbiol., vol. 7, 822, 2016.

[2] P.W. Laffy, E.M. Wood-Charlson, D. Turaev, S. Jutz, C. Pascelli et al., “Reef invertebrate viromics: diversity, host specificity and functional capacity”, Environ. Microbiol., vol. 20, pp 2125-2141, 2018.

[3] T.V Butina, Yu.S. Bukin, I.V. Khanaev et al., “Metagenomic analysis of viral communities in diseased Baikal sponge *Lubomirskia baikalensis*”, Limnology and Freshwater Biology, vol. 1. pp. 155-162, 2019.

[4] M. Morgan, S. Anders, M. Lawrence, P. Aboyoun, H. Pages, R. Gentleman, “ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data”, Bioinformatics, vol. 25, pp. 2607–2608, 2009.

[5] S.F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, “Basic local alignment search tool”, J. Mol. Biol., vol. 215, pp. 403–410, 1990.

[6] R.L. Tatusov, M.Y. Galperin, D.A. Natale, E.V. Koonin, “The COG database: a tool for genome-scale analysis of protein functions and evolution”, Nucleic Acids Res., vol.28, pp. 33-36, 2000.

[7] M.O. Hill, “Diversity and evenness: a unifying notation and its consequences”, Ecology, vol. 54, pp. 427–432, 1973.