

Разнообразие вирусов в образцах байкальских брюхоногих моллюсков (*Benedictia baicalensis*)

Бутина Т.^{1*}, Букин Ю.¹, Бондарюк А.², Петрушин И.³, Ханаев И.¹

¹ Лимнологический институт СО РАН, Иркутск, Россия

² Иркутский научно-исследовательский противочумный институт Сибири и Дальнего Востока, Иркутск, Россия

³ Иркутский государственный университет, Иркутск, Россия

* tvbutina@mail.ru

Ключевые слова: вирусы; вирусное разнообразие; метатранскриптомный анализ; пресноводные брюхоногие моллюски; озеро Байкал

Мотивация и цель: Первые сведения о наличии вирусов у моллюсков появились в начале 1970-х гг. на основе гистологических и цитологических наблюдений с использованием световой и электронной микроскопии [1]. В настоящее время в связи с проблемой появления новых вирусных заболеваний и расширением возможностей молекулярной вирусологии реализуются проекты по массовому параллельному секвенированию (в основном метатранскриптомные исследования) различных видов беспозвоночных, в том числе моллюсков [2, 3]. Проведенные исследования выявили огромное разнообразие по большей части неизвестных вирусов и заполнили пробелы в истории эволюции вирусов. Большинство исследований сосредоточено на морских, преимущественно двустворчатых моллюсках, и значительно реже на брюхоногих моллюсках. Гораздо меньше внимания уделяется пресноводным моллюскам. В нашей работе мы провели оценку вирусного разнообразия в образцах байкальских эндемичных брюхоногих моллюсков *Benedictia baicalensis* с помощью метатранскриптомного анализа.

Методы и алгоритмы: Моллюски рода *Benedictia* (*B. baicalensis*) были отобраны на двух станциях Южного Байкала: в районе пос. Листвянка и пос. Большие Коты; а также в Северном Байкале вблизи Ушканьих островов. Отбор проб осуществлялся водолазами, как описано ранее [4].

Суммарную РНК выделяли из образцов *B. baicalensis* с использованием TRI Reagent (Molecular Research Center, США) и набора Direct-zol RNA MiniPrep (Zymo Research, США). Библиотеки готовили с использованием набора MGI Easy RNA Library Prep Kit (MGI Tech Co., Ltd., Китай) в соответствии с протоколами производителя. Библиотеки кДНК секвенировали с обоих концов на платформе DNBSEQTM с использованием секвенатора DNBSEQ-400 и набора для высокопроизводительного секвенирования DNBSEQ-G400RS (MGI Tech Co., Ltd., Китай). Длина секвенированных фрагментов составляла 150 оснований.

Проверку качества наборов данных (парных прочтений) осуществляли с помощью программы FASTQC v.0.11.9. Фильтрацию прочтений по качеству проводили с помощью программы Trimmomatic v.0.32 (MAXINFO:40:0.05 AVGQUAL:15 MINLEN:90). После обработки наборы данных содержали от 63 до 67 млн пар прочтений на образец.

Метатранскриптомные данные были агрегированы в один массив для комбинированной («микс») сборки *de novo* [5] с помощью программы MEGAHIT v.1.2.9. С использованием программы VirSorter2 v.2.2.4 были идентифицированы вирусные скаффолды и открытые рамки считывания (ORFs). Результаты идентификации вирусов проверялись программой CheckV. Все скаффолды, идентифицированные VirSorter2 как вирусные, сравнивались с хромосомными геномами моллюсков и протеомными данными из базы данных NCBI с использованием алгоритма BLASTn и инструмента DIAMOND v2.0.14.152. Вирусные скаффолды также проверялись на наличие мобильных генетических элементов с помощью программы DFAMSCAN.

Таксономическую идентификацию вирусных скаффолдов осуществляли путем сравнения предсказанных вирусных белков с базой данных полных вирусных протеомов NCBI RefSeq с помощью алгоритма DIAMOND. Программы Bowtie2 v.1.3.1 и SAMtools v.1.7 использовались для определения количества ридов из каждого образца, входящих в состав вирусных скаффолдов. Предсказанные вирусные ORFs сопоставлялись с функциональными мотивами из базы данных Pfam с использованием программного обеспечения PfamScan v.1.5 и базы данных консервативных доменов CDD (Conserved Domain Database).

Филогенетическую реконструкцию на основе генов *RdRp* проводили с помощью IQTREE v.1.6.12. В наш анализ были включены BLAST-хиты из базы NCBI nr и известные вирусы из базы данных RefSeq с наибольшей степенью сходства аминокислотных последовательностей с идентифицированными нами вирусами. Выравнивание аминокислотных последовательностей проводили с помощью MAFFT v.7.490. Наиболее подходящая модель замены аминокислот (LG+G4+Inv+F) была выбрана на основе значений BIC критерия, с помощью ModelFinder, реализованного в программе IQTREE.

Результаты: В метатранскриптомных наборах данных из образцов байкальских моллюсков было выявлено 64 вирусных скаффолда; из них только один был идентифицирован как ДНК-вирус, родственник одноцепочечным ДНК CRESS-вирусам из пресноводных водоемов Новой Зеландии (озера Сара) и США. Остальные 63 скаффолда (от 24 до 33 на образец) были идентифицированы как РНК-вирусы, сходные с таковыми из разных семейств вирусов (*Picornaviridae*, *Marnaviridae*, *Dicistroviridae*, *Solemoviridae*, *Partitiviridae*, *Tombusviridae*, *Nodaviridae*, *Narnaviridae*, *Caulimoviridae* и *Qinviridae*). В целом четыре исследованные пробы были схожими, несмотря на большие расстояния между станциями отбора проб. Это можно объяснить активными процессами перемешивания вод озера Байкал, в том числе за счет течений и ветров [6], а также длительной циркуляцией разнообразных вирусов в экосистеме озера.

Среди идентифицированных вирусных геномов (скаффолдов) оказались пикорноподобные вирусы, родственные ранее описанным вирусам пресноводных моллюсков, *Biomphalaria virus 1* и *Biomphalaria virus 3*. Как известно, пикорноподобные вирусы могут вызывать массовую смертность морских моллюсков [7]. В целом сходство белков вирусов из *B. baicalensis* с известными вирусами из базы данных NCBI было низким и в единичных случаях превышало 70 %. Только один скаффолд имел 99.3 % сходства (при выравнивании 99 % последовательности) с неструктурированным полипротеином вируса некроза тигрового фугу (*Tiger puffer nervous necrosis virus*, семейство *Nodaviridae*). Как оказалось, резервуарами «вирусов нервного некроза» могут быть различные двусторчатые, брюхоногие и

головоногие моллюски, а также другие виды морских беспозвоночных (крабы, креветки, артемии и коловратки), обнаруженные в водах Южной Кореи, Японии и Китая, в европейской части Атлантического океана и Средиземном море [8]. Скорее всего, байкальские моллюски также могут выступать резервуарами для нодавирусов, как и других вирусов, и влиять на их эпидемиологию.

На филогенетическом древе последовательности *RdRp* из образцов байкальских брюхоногих моллюсков вошли в состав пяти кластеров, соответствующих РНК-вирусам порядков *Picornavirales*, *Tolivirales*, *Nodamuvirales*, *Durnavirales* и *Sobelivirales*. Полученные результаты демонстрируют отдаленное родство выявленных нами вирусных геномов (геномных фрагментов), а также разнообразие РНК-вирусов с широким спектром хозяев (моллюски, насекомые, растения, водоросли, простейшие, грибы и позвоночные животные). Наши данные выявили также значительные генетические расстояния между генами *RdRp* байкальских вирусов и вирусов из других организмов и биотопов, что указывает на уникальность обнаруженных нами вирусов.

Выводы: Проведенное метатранскриптомное исследование образцов байкальских моллюсков позволило раскрыть новый спектр уникальных вирусов в экосистеме озера Байкал. Выявленное высокое разнообразие вирусов в байкальских эндемичных брюхоногих моллюсках *B. baicalensis* включало неизвестных представителей разных отрядов и семейств, поражающих широкий круг хозяев.

B. baicalensis обитают на каменистых и смешанных песчано-каменистых биотопах и являются всеядным видом, питающимся субстратом [9]. Это способствует передаче и биоаккумуляции различных вирусов, поражающих широкий круг придонных и планктонных обитателей озера Байкал. Таким образом, для установления фактического круга хозяев новых вирусов и выявления истинного виroma байкальских моллюсков необходимо продолжать начатые нами исследования.

Финансирование: Исследование поддержано Российским научным фондом (№ 22-24-01120) и Министерством науки и высшего образования РФ (№ 0279-2021-0005).

Viral diversity in samples of Baikal gastropods (*Benedictia baicalensis*)

Butina T.^{1*}, Bukin Y.¹, Bondaryuk A.², Petrushin I.³, Khanaev I.¹

¹ Limnological Institute, SB RAS, Irkutsk, Russia

² Irkutsk Anti-Plague Research Institute of Siberia and Far East, Irkutsk, Russia

³ Irkutsk State University, Irkutsk, Russia

* tvbutina@mail.ru

Key words: viruses; viral diversity; metatranscriptomic analysis; freshwater gastropods; Lake Baikal

Motivation and Aim: The first evidence of viruses in mollusks appeared in the early 1970s on the basis of histological and cytological observations using light and electron microscopy [1]. Currently, due to the problem of the emergence of new viral diseases and the expansion of the capabilities of molecular virology, the different projects are being implemented for massive parallel sequencing (mainly metatranscriptomic studies) of various species of invertebrate, including mollusks [2, 3]. These studies have revealed

a huge diversity of mostly previously unknown viruses and filled gaps in the evolutionary history of viruses. Most studies have primarily focused on marine mollusks, mostly bivalves and less often gastropods. Much less attention has been paid to freshwater mollusks. In our study, we estimated the viral diversity in samples of the Baikal endemic gastropod *Benedictia baicalensis* using metatranscriptomic analysis (total RNA-sequencing).

Methods and Algorithms: Mollusks of the genus *Benedictia* (*Benedictia baicalensis*) were taken at two stations of Southern Baikal: in Listvennichny Bay (near the village of Listvyanka) and near the village of Bolshie Koty; and near the Ushkany Islands in Northern Baikal. Sampling was carried out by divers [4].

Total RNA was isolated from samples *B. baicalensis* (pools of five) using TRI Reagent (Molecular Research Center, USA) and a Direct-zol RNA MiniPrep kit (Zymo Research, USA). Libraries were prepared using the MGI Easy RNA Library Prep Kit (MGI Tech Co., Ltd., China) in accordance with the manufacturer's protocols. The cDNA libraries were sequenced from both ends on a DNBSEQTM platform using a DNBSEQ-400 NGS sequencer and a DNBSEQ-G400RS High-throughput Sequencing Kit (MGI Tech Co., Ltd., Wuhan, China). The length of the sequenced fragments was 150 bases.

The quality visualization of datasets (paired reads) was carried out using the FASTQC v.0.11.9 program. Trimming of reads by quality was carried out with the Trimmomatic v.0.32 program (MAXINFO:40:0.05 AVGQUAL:15 MINLEN:90). After trimming, the datasets finally contained 63–67 million read pairs.

All metatranscriptomic data were aggregated into one array for *de novo* cross-assembly [5] using the MEGAHIT v.1.2.9 NGS assembler. We identified the viral scaffolds and open reading frames (ORFs) within them using the VirSorter2 v.2.2.4 tool. The results of virus identification were checked by the CheckV program. All scaffolds identified as viral by VirSorter2 were compared with chromosomal genome and proteomic data of mollusks from the NCBI database using BLASTn algorithm and DIAMOND v2.0.14.152 tool. The viral scaffolds were also checked for mobile genetic elements using the DFAMSCAN program.

Taxonomic identification for the viral scaffolds was carried out by comparisons of predicted viral proteins in scaffolds with the NCBI RefSeq complete viral proteome database and DIAMOND algorithm. The Bowtie2 v.1.3.1 (with default options) and SAMtools v. 1.7 results were used to determine the number of reads mapped on each predicted viral scaffold from each sample. The predicted viral proteins (ORFs) were matched with functional motifs in the Pfam database using PfamScan v.1.5 software and the Conserved Domain Database.

Phylogenetic reconstruction based on *RdRp* genes was performed with IQTREE v.1.6.12. Our analysis included NCBI nr BLAST-hits and known, well-studied viruses from the RefSeq database with a significant degree of amino acid sequence similarity to the viruses we identified. Aligning of amino acid sequences was carried out in MAFFT v.7.490. The best-fitting amino acid substitution model (LG+G4+Inv+F) was selected based on the BIC criterion values calculated by ModelFinder implemented in IQTREE.

Results: In total in metatranscriptomic data from the samples of Baikal mollusks we revealed 64 viral scaffolds; of these, only one was identified as a DNA virus related to ssDNA CRESS viruses from a New Zealand freshwater Lake Sarah and USA freshwater springs. The rest 63 viral scaffolds (from 24 to 33 per sample) were identified as RNA viruses similar to those from different viral families (*Picornaviridae*, *Marnaviridae*, *Dicistroviridae*, *Solemoviridae*, *Partitiviridae*, *Tombusviridae*, *Nodaviridae*,

Narnaviridae, *Caulimoviridae* and *Qinviridae*). Overall, the four examined pool samples were similar despite the large distances between sampling stations. This can be explained by the active mixing processes of the Lake Baikal waters, including those due to the currents and winds [6], as well as the long-term circulation of a wide variety of viruses in the lake's ecosystem.

Among the viral genomes (or scaffolds) identified were picorna-like viruses closely related to the previously described freshwater shellfish viruses *Biomphalaria virus 1* and *Biomphalaria virus 3*. As known, picorna-like viruses may also be responsible for mass mortality events in marine mollusks [7].

In general, the similarity of proteins of *B. baicalensis* viruses to known viruses from the NCBI database was low and in single cases exceeded 70 %. Only one scaffold showed 99.3 % identity and 99 % query coverage with unstructured polyprotein of *Tiger puffer nervous necrosis virus* (TPNNV, family *Nodaviridae*). Reservoirs of 'nervous necrosis viruses' were reported in different mollusks (bivalves, gastropods and cephalopods) and other marine invertebrate species (crabs, shrimps, artemias and rotifers) in waters of South Korea, Japan and China, in the European Atlantic and Mediterranean waters [8]. Baikal mollusks may also act as reservoirs for nodaviruses and other viruses and influence their epidemiology.

RdRp sequences from Baikal gastropods were distributed into five clusters corresponding to the RNA viral orders *Picornavirales*, *Tolivirales*, *Nodamuvirales*, *Durnavirales* and *Sobelivirales*, demonstrating the distant relationship of the identified viral genomes (genomic fragments) and the diversity of RNA viruses with a wide range of hosts (mollusks, insects, plants, algae, protozoa, fungi and vertebrates). Our data also revealed significant distances between the *RdRps* of Baikal viruses and those from other organisms and biotopes, indicating the novelty of the viruses we discovered.

Conclusion: The metatranscriptomic study of samples of Baikal mollusks made it possible to reveal a new range of unique viruses in the ecosystem of Lake Baikal. The detected high diversity of viruses in the Baikal endemic gastropod *B. baicalensis* included unknown representatives of different orders and families infecting a wide range of hosts.

B. baicalensis lives on rocky and mixed sandy-stony biotopes and is an omnivorous species that grazes on the substrate [9]. This contributes to the transmission and bioaccumulation of various viruses that infect a wide range of benthic and planktonic inhabitants of Lake Baikal. Thus, additional research is required to establish the actual host range for novel viruses and uncover the true virome of Baikal mollusks.

Funding: The study is supported by the Russian Science Foundation (No. 22-24-01120) and the Ministry of Science and Higher Education of the Russian Federation (No. 0279-2021-0005).

Список литературы/References

1. Farley C.A. Viruses and virus-like lesions in marine mollusks. *Mar Fish Rev.* 1978;40:18-20
2. Shi M. et al. Redefining the invertebrate RNA virosphere. *Nature.* 2016;540:539-543. doi 10.1038/nature20167
3. Zhang Y.Y., Chen Y., Wei X., Cui J. Viromes in marine ecosystems reveal remarkable invertebrate RNA virus diversity. *Sci China Life Sci.* 2022;65:426-437. doi 10.1007/s11427-020-1936-2
4. Butina T.V. et al. Viral diversity in samples of freshwater gastropods *Benedictia baicalensis* (caenogastropoda: benedictiidae) revealed by total RNA-sequencing. *Int J Mol Sci.* 2023;24:17022. doi 10.17537/2023.18.418
5. Букин Ю.С., Бондарюк А.Н., Бутина Т.В. Анализ эффективности микс-сборки метатранскриптомных наборов данных в исследовании вирусных сообществ. *Математическая биология и биоинформатика.* 2023;18(2):418-433. doi 10.17537/2023.18.418

- [Bukin Y.S., Bondaryuk A.N., Butina T.V. Performance analysis of cross-assembly of metatranscriptomic datasets in viral community studies. *Math Biol Bioinf.* 2023;18:418-433 (in Russian)]
6. Байкал: атлас. Москва: Роскартография, 1993
[Baikal: atlas. Moscow: Roscartography, 1993 (in Russian)]
Renault T., Novoa B. Viruses infecting bivalve molluscs. *Aquat Living Resour.* 2004;17:397-409. doi 10.1051/alr:2004049
 7. Johnstone C., Pérez M., Arizcun M., García-Ruiz C., Chaves-Pozo E. Reservoirs of red-spotted grouper nervous necrosis virus (RGNNV) in squid and shrimp species of Northern Alboran Sea. *Viruses.* 2022;14:328. doi 10.3390/v14020328
 8. Ситникова Т.Я., Репсторф П. Эти моллюски живут только в Байкале. *Наука из первых рук.* 2004;1(2):84-99
[Sitnikova T.Y., Roepstorf P. Mollusks that Live Exclusively in Lake Baikal. *Nauka iz Pervykh Ruk.* 2004;1(2):84-99 (in Russian)]