

Исследование разнообразия вирусов в экосистеме озера Байкал с помощью метагеномного анализа

Букин Ю.С.^{1, 2*}, Антипова А.Ю.²

¹ Лимнологический институт СО РАН, Иркутск, Россия

² Иркутский государственный университет, Иркутск, Россия

* bukinyira@mail.ru

Ключевые слова: метагеномика; вирусы; метавирусный анализ; высокопроизводительное секвенирование; озеро Байкал

Введение: Вирусы являются важным компонентом водных экосистем. Концентрация вирусоподобных частиц в пресноводных водоемах достигает 10^{11} на миллилитр. Большинство вирусов являются бактериофагами, поражающими различные виды прокариот. Часть вирусов поражает автотрофный и гетеротрофный планктон, регулируя его численность и динамику. Вирусы также могут поражать любые водные организмы и ассоциированный с ними микробиом. Одним из наиболее перспективных методов исследования вирусов является шотган метагеномный анализ. Данный метод основан на массовом параллельном секвенировании ДНК или РНК из природной смеси организмов с последующей полногеномной сборкой и идентификацией фрагментов геномов или полных геномов вирусов. Оценка покрытия фрагментов вирусных геномов позволяет получить данные об их представленности в исследуемых образцах. К предварительной подготовке генетического материала для исследования вирусов в водных экосистемах существует два подхода. Первый подход предполагает обогащение образца воды путем фильтрации и центрифугирования вирусными частицами. В результате мы получаем массив вирусных данных. Второй подход предполагает расшифровку всего генетического материала ДНК или РНК (метагеномный подход) с дальнейшим поиском в общих данных вирусных фрагментов. Такая методика чаще всего применяется для исследования вирусов, ассоциированных с различными водными многоклеточными организмами.

Цель нашего исследования заключалась в анализе таксономического и функционального разнообразия ДНК вирусов в воде озера Байкал и разнообразия ДНК и РНК вирусов, ассоциированных с байкальскими ракообразными амфиподами.

Методы и алгоритмы: Первичные ДНК данные для образцов воды были получены путем обогащения материала вирусными частицами. Для анализа ДНК и РНК вирусов, ассоциированных с байкальскими амфиподами, использовались данные транскриптомного анализа. Данные транскриптомики позволяют получить представления о разнообразии РНК вирусов и экспрессирующихся фрагментов ДНК вирусов. Расшифровку генетического материала проводили методом дробовика по технологии Illumina MiSeq и Illumina HiSeq.

De novo геномные сборки проводили метагеномным сборщиком metaSPAdes. Идентификация вирусных скаффолдов (фрагментов вирусных геномов) проводилась комбинацией методов, включая методы, основанные на машинном обучении (скрытые Марковские модели) (программы Virsorter, VIBRANT) и методы, основанные на BLAST подобных алгоритмах выравнивания (DIAMOND,

Kaiju). Основой для идентификации вирусов послужили полные вирусные протомы базы данных NCBI RefSeq. Спектр потенциальных хозяев вирусов определялся по базе данных Virus-Host Database. Полное описание методов можно найти в наших публикациях [1–5].

Результаты: Полное описание результатов по исследованию разнообразия вирусов в образцах воды озера Байкал можно найти в наших публикациях [1, 2]. Эти исследования показали, что в Байкале среди ДНК вирусов доминируют разнообразные бактериофаги, также присутствуют вирусы одноклеточных водорослей. В свободном виде обнаружены вирусные частицы позвоночных и беспозвоночных животных.

В транскриптомах амфипод в большом количестве были представлены экспрессирующиеся гены различных ДНК бактериофагов. Очевидно, что это вирусы, ассоциированные с микробиомом амфипод, и частично вирусы из воды. Также были обнаружены фрагменты РНК вирусов, предсказанными хозяевами для которых были различные беспозвоночные животные, в том числе ракообразные.

Финансирование: Работа поддержана темой бюджетного финансирования ЛИН СО РАН «Генетика сообществ байкальских организмов: структура генофонда, стратегии консервации» 0279-2021-0010 (121032300196-8).

Studying the of viral diversity in the Lake Baikal ecosystem using metagenomic analysis

Bukin Yu.S.^{1, 2*}, Antipova A.Yu.²

¹ Limnological Institute, SB RAS, Irkutsk, Russia

² Irkutsk State University, Irkutsk, Russia

* bukinyira@mail.ru

Key words: metagenomics; viruses; metavirome analysis; high-throughput sequencing; Lake Baikal

Motivation and Aim: Viruses are an important component of aquatic ecosystems. The concentration of virus-like particles in freshwater bodies reaches up to 10¹¹ per milliliter. Most viruses are bacteriophages that infect various types of prokaryotes. Some viruses infect autotrophic and heterotrophic plankton, regulating their numbers and dynamics. Viruses can also infect any aquatic organism and its associated microbiome.

One of the most promising methods for studying viruses is shotgun metagenomic analysis. This method is based on massively parallel sequencing of DNA or RNA from a natural mixture of organisms, followed by whole-genome assembly and identification of genome fragments or complete genomes of viruses. Assessing the coverage of viral genome fragments allows us to obtain data on their representation in the studied samples. There are two approaches to the preliminary preparation of genetic material for the study of viruses in aquatic ecosystems. The first approach involves enriching a water sample by viral particles with filtration and centrifugation. As a result, we obtain an array of virome data. The second approach involves decoding the entire genetic material of DNA or RNA (metagenomic approach) with a further search for viral fragments in the general data. This technique is most often used to study viruses associated with various aquatic multicellular organisms.

The purpose of our study was to analyze the taxonomic and functional diversity of DNA viruses in the water of Lake Baikal and the diversity of DNA and RNA viruses associated with Baikal amphipod.

Methods and Algorithms: Primary DNA data for water samples was obtained by enriching the material with viral particles. Transcriptomic analysis data were used to analyze DNA and RNA viruses associated with Baikal amphipods. Transcriptomics data provide insight into the diversity of RNA viruses and expressed DNA viruses fragments. Genetic material was sequenced using the shotgun method using Illumina MiSeq and Illumina HiSeq technology.

De novo genome assemblies were performed with the metaSPAdes metagenomic assembler. Identification of viral scaffolds (fragments of viral genomes) was carried out by a combination of methods, including methods based on machine learning (hidden Markov models) (“Virsorter” and “VIBRANT” programs) and methods based on BLAST-like alignment algorithms (“DIAMOND”, “Kaiju”) The basis for virus identification was the complete viral proteomes of the NCBI RefSeq database. The range of potential virus hosts was determined using the Virus-Host Database. A complete description of the methods can be found in our publications [1–5].

Results: A complete description of the results of studying the diversity of viruses in water samples from Lake Baikal can be found in our publications [1, 2]. These studies showed that in Baikal, various bacteriophages dominate among DNA viruses; viruses of unicellular algae are also present. Viral particles of vertebrate and invertebrate animals were found in free form.

Expressed genes of various DNA bacteriophages were abundantly represented in amphipod transcriptomes. Obviously, these are viruses associated with the microbiome of amphipods and partly viruses from water. Fragments of RNA viruses were also discovered, the predicted hosts for which were various invertebrate animals, including crustaceans.

Funding: Work supported by budget funding of Limnological Institute SB RAS “Genetics of communities of Baikal organisms: structure of the gene pool, conservation strategies” 0279-2021-0010 (121032300196-8).

Список литературы/References

1. Butina T.V., Bukin Y.S., Krasnopeev A.S., Belykh O.I., Tupikin A.E., Kabilov M.R., Belikov S.I. Estimate of the diversity of viral and bacterial assemblage in the coastal water of Lake Baikal. *FEMS Microbiol Lett.* 2019;366(9):fnz094. doi 10.1093/femsle/fnz094
2. Butina T.V., Bukin Y.S., Petrushin I.S., Tupikin A.E., Kabilov M.R., Belikov S.I. Extended evaluation of viral diversity in lake baikal through metagenomics. *Microorganisms.* 2021;9(4):760. doi 10.3390/microorganisms9040760
3. Butina T.V., Petrushin I.S., Khanaev I.V., Bukin Y.S. Metagenomic assessment of DNA viral diversity in freshwater sponges, *Baikalospongia bacillifera*. *Microorganisms.* 2022;10(2):480. doi 10.3390/microorganisms10020480
4. Alexyuk M.S., Bukin Y.S., Butina T.V., Alexyuk P.G., Berezin V.E., Bogoyavlenskiy A.P. Viromes of coastal waters of the North Caspian Sea: Initial assessment of diversity and functional potential. *Diversity.* 2023;15(7):813. doi 10.3390/d15070813
5. 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(2):418-433. doi 10.17537/2023.18.418