

Идентификация и характеристика новых некультивируемых линий бактерий и архей из глубинной подземной биосферы

Кадников В.^{1*}, Белецкий А.¹, Марданов А.¹, Карначук О.², Равин Н.¹

¹ Институт биоинженерии, ФИЦ Биотехнологии РАН, Москва, Россия

² Томский государственный университет, Томск, Россия

* vkadnikov@bk.ru

Ключевые слова: некультивируемые микроорганизмы; геном; экология микроорганизмов; микробное сообщество; метабеномика

Мотивация и цель: Микроорганизмы представляют собой наиболее многочисленную форму жизни, однако наши знания об их разнообразии и осуществляемых ими процессах весьма ограничены, поскольку обычно не более 1 % микроорганизмов из природных сообществ удается культивировать и охарактеризовать традиционными методами. О существовании остальных 99 % «некультивируемых» стало известно лишь с появлением молекулярных методов анализа. Так, если в «домолекулярный» период было описано всего 12 филумов прокариот, то в настоящее время известно уже 175 филумов бактерий и 19 филумов архей (по последней версии Genome taxonomy database), более половины, из которых не имеет культивируемых представителей.

Хотя микроорганизмы «некультивируемых» групп широко распространены в различных экосистемах и могут играть ключевые роли в важных биогеохимических процессах, о них практически ничего (метаболизм, роль в биосфере и т. д.) не известно. Основным инструментом изучения некультивируемых микроорганизмов является метабеномный анализ, позволяющий не только охарактеризовать их на геномном уровне, но и описать генетическое разнообразие, метаболизм микробного сообщества и осуществляемые им биогеохимические процессы.

Одним из наименее изученных микробных местообитаний на Земле являются глубинные подземные экосистемы, которые могут оставаться изолированными от поверхности на протяжении миллионов лет. Большинство микроорганизмов глубинной подземной биосферы принадлежит к некультивируемым линиям, многие из которых специфичны для подземных местообитаний. Целью данной работы была идентификация и характеристика микроорганизмов подземной биосферы с помощью метабеномного анализа.

Методы и алгоритмы: Объектами исследования были глубинные подземные воды Томской области (2–3 км), длительное время (миллионы лет) изолированные от поверхности и грязевые вулканы Керченско–Таманского региона. Была определена температура воды, pH, Eh и ее химический состав, а также состав растворенного в ней газа. Также определили изотопный состав метана с целью уточнения его биогенного или абиогенного происхождения.

Характеристику состава сообществ проводили с помощью высокопроизводительного секвенирования ПЦР фрагментов генов 16S рРНК,

амплифицированных из метагеномной ДНК с «универсальными» праймерами 341F и 806R. Секвенирование проводили на Illumina MiSeq (2x 300 нт). Анализ состава сообществ по последовательностям генов 16S рРНК проводили с использованием пакета программ QIIME.

Для метагеномного анализа использовали технологии секвенирования Illumina и Oxford Nanopore. Для получения основного массива данных использовали Illumina (HiSeq2500). Сборку полученных последовательностей в контиги проводили с помощью metaSPAdes и/или MEGAHIT. Чтения MinION вначале собирали в контиги *de novo* с использованием Flye v. 2.7. Затем последовательности контигов корректировали с помощью Pilon v.1.2.2 в результате картирования на них чтений Illumina с использованием Bowtie 2. Вторым вариантом была совместная сборка чтений Illumina и MinION в контиги с помощью metaSPAdes hybrid assembler. Наконец, собранные только из чтений Illumina контиги определенного MAG (metagenome-assembled genome) объединялись в более протяженные или даже в полный кольцевой геном с использованием длинных чтений MinION. Для этого чтения MinION картировали на входящие в MAG контиги с помощью BWA v.0.7.15. Затем с помощью программы Npscarf v.1.0 формировали скаффолды, пробелы между контигами заполняли с использованием последовательностей Illumina из графа сборки metaSPAdes. Выбор конкретного варианта сборки метагенома определялся его сложностью, составом, объемом полученных последовательностей и т. д. Для биннинга метагеномных контигов в MAG использовали программы MetaBAT и/или CONCOCT. Собранные MAG таксономически классифицировали с использованием Genome Taxonomy Database Toolkit (GTDB-Tk). Доли отдельных микроорганизмов в сообществе определяли по доле прочтений метагеномной ДНК, вошедших в соответствующий MAG. Поиск генов и аннотацию MAG проводили с использованием NCBI Prokaryotic Genome Annotation Pipeline и/или RAST server. Для филогенетического анализа с помощью GTDB-Tk v.0.3.2 в MAG были найдены однокопийные маркерные гены и построено множественного выравнивания их конкатенированных последовательностей из данного MAG и всех видов из Genome Taxonomy Database. Выбранную часть множественного выравнивания, созданного в GTDB-Tk, использовали для построения филогенетического дерева с помощью PhyML v.3.3.

Результаты: Мы изучили микробные сообщества глубинных термальных водоносных горизонтов в мезозойских осадочных породах Западно-Сибирского региона с помощью метагеномного подхода. Были исследованы географически удаленные водоносные горизонты, доступ к которым осуществлялся через нефтепоисковые скважины глубиной 2–3 км. Выявлено три типа микробных сообществ. Сообщества первого типа (скважины 3Р и Vi-1) состояли в основном из хемолитоавтотрофов, – метаногенных архей (*Methanothermobacter*) и сульфат-редуцирующих *Bacillota* и/или *Desulfobacterota*. Второй тип сообществ (скважина 1-Р) содержал сульфат-редуцирующие *Bacillota* и *Desulfobacterota*, а также различные некультивируемые линии филумов *Chloroflexota*, *Ignavibacteria*, *Aminicenantia* и *Riflеbacteria*, археи отсутствовали. Анализ геномов *Chloroflexota*, *Ignavibacteria*, *Riflеbacteria* и *Aminicenantia* показал, что они могут сбраживать углеводы, а некоторые способны к аэробному и/или анаэробному дыханию. Вероятно, захороненное органическое вещество мезозойских отложений является субстратом для их роста, а образуемые водород и ацетат используются сульфат-

редукторами. Микробное сообщество третьего типа (скважина 5P) включало примерно равные доли бактерий и метаногенных архей. Среди бактерий найдены органотрофные представители *Bacillota*, *Chloroflexota*, *Ignavibacteria*, *Armatimonadota*, *Pseudomonadota*, *Bacteroidetota* и *Spirochaetota*. Сульфат-редукторы составляли небольшую долю и были представлены ‘*Ca. Karabacteria*’ и *Thermodesulfovibrio*. Анализ метагеномов позволил получить несколько сот геномов микроорганизмов, в том числе представителей «некультивируемых» классов/филумов *Aminicenantia*, *Bipolaricaulota*, *Patescibacteria*, *Atribacteriota*, *Armatimonadota*, *Rifl bacteria*, WOR-3 и BRC1.

Исследовано три грязевых вулкана Керченский полуостров – сопка Андрусова (KV1), сопка Тищенко (KV2) и вулкан «принц Ольденбургский» (KV3). Вулканы представляют собой небольшие (до нескольких метров) конусообразные структуры, из которых на поверхность постоянно поступают грязевые массы и газы (в основном метан). Микробные сообщества образцов флюида, отобранных из грязевых вулканов, в целом были сходны между собой, хотя имелись важные отличия, как между вулканами, так и между образцами, отобранными на разной глубине. Археи составляли от 9 до 55 % (доля возрастает с глубиной) и в основном были представлены организмами цикла метана, хотя в верхних горизонтах вулканов KV1 и KV2 также в значительных количествах встречались *Bathyarchaeia*. Большинство архей относилось к линиям, осуществляющим анаэробное окисление метана (ANME). Археи группы ANME-1 были обнаружены в двух верхних горизонтах (20 и 70 см) образцов KV1 и KV2, причем в KV1 были найдены ANME-1a (1–3 %), а в KV2 – ANME-1b (около 2 %). Группа ANME-2a-2b была найдена только в KV2, где ее доля возрастала с глубиной с 2 до 14%. Наиболее многочисленной группой ANME в образцах KV1 и KV2 была *Methanoperedenaceae* (осуществляют окисление метана, сопряженное с восстановлением нитрата или марганца), причем их доля возрастала с глубиной до 51 % в KV1 и до 14 % в KV2. В вулкане KV3 среди метан-окисляющих архей на всех глубинах доминировала группа ANME-3, на которую приходилось 32–47 % всего сообщества. ANME-3 также были найдены в глубинных слоях в KV1 (1.2 %) и KV2 (7.4 %). Доля метаногенных архей была намного ниже, чем ANME, в значительных количествах метаногены присутствовали только в KV3, в котором *Methanocaldococcus* составлял 3–6 % сообщества. Среди бактерий наиболее многочисленными были филумы *Desulfobacterota*, *Pseudomonadota*, *Bacillota*, *Bacteroidota* и *Campylobacterota*. *Desulfobacterota* в основном относились к сульфат-редуцирующим линиям (*Desulfuromonadales* и др.), причем их доля была максимальной в верхних горизонтах KV1 (28 %). Сульфат-редукторы филумов *Bacillota* (*Desulfotomaculalia*) и *Nitrospirota* (*Thermodesulfovibrionia*) также присутствовали, но в меньших количествах. Среди протеобактерий наоборот, преобладали серо-окисляющие линии, в частности, рода *Thiomicrospira* (до 8 %), *Thiopfundum*, *Thioalkalivibrio* и др. Также среди потенциальных сероокислителей были найдены *Campylobacterota* (2–9 % сообществ), родов *Sulfurimonas*, *Sulfurospirillum*, *Sulfuricurvum* и *Sulfurovum*. В целом, выявленный нами состав микробных сообществ грязевых вулканов показал преобладание хемолитоавтотрофных микроорганизмов – участников циклов серы и метана, причем среди последних преобладали ANME археи.

Финансирование: Данная работа была выполнена при поддержке Российского научного фонда № 22-14-00178.

Identification and characterization of new uncultured lineages of bacteria and archaea from the deep subsurface biosphere

Kadnikov V.^{1*}, Beletsky A.¹, Mardanov A.¹, Karnachuk O.², Ravin N.¹

¹ Institute of Bioengineering, Federal Research Center of Biotechnology RAS, Moscow, Russia

² Tomsk State University, Tomsk, Russia

* vkadnikov@bk.ru

Key words: uncultured microorganisms; genome; microbial ecology; microbial community; metagenomics

Motivation and Aim: Microorganisms are the most abundant form of life, however, our knowledge of their diversity and the processes they carry out is very limited, since usually no more than 1 % of microorganisms from natural communities can be cultivated and characterized by traditional methods. The existence of the remaining 99 % “uncultivable” became known only with the advent of molecular methods of analysis. Thus, if in the “pre-molecular” period only 12 phyla of prokaryotes were described, then at present 175 phyla of bacteria and 19 phyla of archaea are known (according to the latest version of the Genome taxonomy database), more than half of which have no cultivated representatives.

Although microorganisms of “uncultivated” groups are widespread in various ecosystems and can play key roles in important biogeochemical processes, practically nothing is known about them (metabolism, role in the biosphere, etc.). The main tool for studying uncultivated microorganisms is metagenomic analysis, which allows not only to characterize them at the genomic level, but also to describe genetic diversity, the metabolism of the microbial community and the biogeochemical processes carried out by it.

One of the least studied microbial habitats on Earth are deep subsurface ecosystems, which can remain isolated from the surface for millions of years. Most microorganisms deep in the underground biosphere belong to uncultivated lineages, many of which are specific to underground habitats. The goal of this work was to identify and characterize microorganisms in the subsurface biosphere using metagenomic analysis.

Methods and Algorithms: The objects of study were deep underground waters of the Tomsk region (2–3 km), isolated from the surface for a long time (millions of years), and mud volcanoes of the Kerch-Taman region. The water temperature, pH, Eh and its chemical composition, as well as the composition of the gas dissolved were determined. The isotopic composition of methane was also determined in order to clarify its biogenic or abiogenic origin.

Characterization of the community composition was carried out using high-throughput sequencing of 16S rRNA gene fragments amplified from metagenomic DNA with “universal” primers 341F and 806R. Sequencing was performed on Illumina MiSeq (2x300 nt). Analysis of community composition based on 16S rRNA gene sequences was carried out using the QIIME software package.

Illumina and Oxford Nanopore sequencing technologies were used for metagenomic analysis. Illumina (HiSeq2500) was used to obtain the main dataset. The obtained sequences were assembled into contigs using metaSPAdes and/or MEGAHIT. MinION reads were first assembled into contigs *de novo* using Flye v. 2.7. The contig sequences were then corrected using Pilon v.1.2.2 as a result of mapping Illumina reads to them

using Bowtie 2. The second option was to jointly assemble Illumina and MinION reads into contigs using the metaSPAdes hybrid assembler. Finally, contigs of a specific MAG (metagenome-assembled genome) assembled only from Illumina reads were assembled into longer ones or even into a complete circular genome using MinION long reads. For this purpose, MinION reads were mapped to MAG contigs using BWA v.0.7.15. Scaffolds were then formed using the Npscarf v.1.0 program, and the gaps between contigs were filled using Illumina sequences from the metaSPAdes assembly graph. The choice of a specific metagenome assembly option was determined by its complexity, composition, volume of obtained sequences, etc. MetaBAT and/or CONCOCT programs were used for binning metagenomic contigs into MAGs. The assembled MAGs were taxonomically classified using the Genome Taxonomy Database Toolkit (GTDB-Tk). The relative abundancies of individual strains in the community were determined by the proportion of metagenomic DNA reads included in the corresponding MAG. Gene searching and MAG annotation were performed using the NCBI Prokaryotic Genome Annotation Pipeline and/or RAST server. For phylogenetic analysis using GTDB-Tk v.0.3.2, single-copy marker genes were found in MAGs and a multiple alignment of their concatenated sequences from these MAGs and all species from the Genome Taxonomy Database was constructed. A selected portion of the multiple alignment generated in GTDB-Tk was used to construct a phylogenetic tree using PhyML v.3.3.

Results: We studied microbial communities of deep thermal aquifers in Mesozoic sedimentary rocks of the West Siberian region using a metagenomic approach. Geographically distant aquifers, accessed through 2–3 km deep oil exploration wells, were explored. Three types of microbial communities were identified. Communities of the first type (wells 3P and Vi-1) consisted mainly of chemolithoautotrophs, methanogenic archaea (*Methanothermobacter*) and sulfate-reducing *Bacillota* and/or *Desulfobacterota*. The second type of community (well 1-R) contained sulfate-reducing *Bacillota* and *Desulfobacterota*, as well as various uncultivated lineages of the *Chloroflexota*, *Ignavibacteria*, *Aminicenantia* and *Riflibacteria*; archaea were absent. Analysis of the genomes of *Chloroflexota*, *Ignavibacteria*, *Riflibacteria* and *Aminicenantia* showed that they can ferment carbohydrates, and some are capable of aerobic and/or anaerobic respiration. It is likely that the buried organic matter of Mesozoic sediments is a substrate for their growth, and the hydrogen and acetate produced are used by sulfate reducers. The microbial community of the third type (well 5P) included approximately equal proportions of bacteria and methanogenic archaea. Among the bacteria, organotrophic members of *Bacillota*, *Chloroflexota*, *Ignavibacteria*, *Armatimonadota*, *Pseudomonadota*, *Bacteroidetota* and *Spirochaetota* were found. Sulfate reducers accounted for a small proportion and were represented by ‘*Ca. Kapabacteria*’ and *Thermodesulfovibrio*. Analysis of metagenomes made it possible to obtain several hundred genomes of microorganisms, including representatives of the “uncultured” classes/phyla *Aminicenantia*, *Bipolaricaulota*, *Patescibacteria*, *Atribacteriota*, *Armatimonadota*, *Riflibacteria*, *WOR-3* and *BRC1*.

Three mud volcanoes of the Kerch Peninsula were studied – Andrusov Hill (KV1), Tishchenko Hill (KV2) and Prince of Oldenburg Volcano (KV3). Volcanoes are small (up to several meters) cone-shaped structures, from which mud masses and gases (mainly methane) constantly flow to the surface. The microbial communities of fluid samples collected from mud volcanoes were generally similar to each other, although there were important differences both between volcanoes and between samples collected at different depths. Archaea constituted from 9 to 55 % (the proportion increases with

depth) and were mainly represented by organisms of the methane cycle, although Bathyarchaea were also found in significant quantities in the upper horizons of volcanoes KV1 and KV2. Most archaea belonged to lineages performing anaerobic methane oxidation (ANME). Archaea of the ANME-1 group were found in the upper two horizons (20 and 70 cm) of samples KV1 and KV2, with ANME-1a (1–3 %) found in KV1 and ANME-1b (about 2 %) in KV2. The ANME-2a-2b group was found only in KV2, where its proportion increased with depth from 2 to 14 %. The most numerous ANME group in samples KV1 and KV2 were Methanoperedenaceae (carry out methane oxidation coupled with the reduction of nitrate or manganese), and their share increased with depth to 51 % in KV1 and up to 14 % in KV2. In the KV3 volcano, the ANME-3 group dominated the methane-oxidizing archaea at all depths, accounting for 32–47 % of the total community. ANME-3 was also found in deep layers in KV1 (1.2 %) and KV2 (7.4 %). The proportion of methanogenic archaea was much lower than ANME, with methanogens present in significant quantities only in KV3, in which Methanocalculus comprised 3–6 % of the community. Among the bacteria, the most abundant phyla were Desulfobacterota, Pseudomonadota, Bacillota, Bacteroidota and Campylobacterota. Desulfobacterota mainly belonged to sulfate-reducing lineages (Desulfuromonadales and others), and their proportion was maximum in the upper horizons of KV1 (28 %). Sulfate reducers of the phyla Bacillota (Desulfotomaculia) and Nitrospirota (Thermodesulfobionia) were also present, but in smaller quantities. Among the Pseudomonadota, on the contrary, sulfur-oxidizing lineages predominated, in particular, the genus Thiomicrospira (up to 8 %), Thioprofundum, Thioalkalivibrio, etc. Also among potential sulfur-oxidizers were found Campylobacterota (2–9 % of communities), the genera Sulfurimonas, Sulfurospirillum, Sulfuricurvum and Sulfurovum. In general, the composition of the microbial communities of mud volcanoes that we identified showed the predominance of chemolithoautotrophic microorganisms – participants in the sulfur and methane cycles, and ANME archaea predominated among the latter.

Funding: The work was supported by the Russian Science Foundation (No. 22-14-00178).