Министерство науки и высшего образования Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ

ИРКУТСИЙ НАУЧНЫЙ ЦЕНТР СИБИРСКОГО ОТДЕЛЕНИЯ  
РОССИЙСКОЙ АКАДЕМИИ НАУК

(ИНЦ СО РАН)

УДК 574.5+004.5

Рег. № НИОКТР АААА-А17-117041250054-8

СОГЛАСОВАНО УТВЕРЖДАЮ

Заведующий Отделом интеграционной Директор ИНЦ СО РАН, д.м.н.

программы, чл.-корр. РАН

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д.П. Гладкочуб \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_К.А. Апарцин

«\_\_\_\_\_» декабря 2018 г. «\_\_\_\_\_» декабря 2018 г.

ОТЧЕТ

О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

ФУНДАМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ И ПРОРЫВНЫЕ ТЕХНОЛОГИИ КАК ОСНОВА ОПЕРЕЖАЮЩЕГО РАЗВИТИЯ БАЙКАЛЬСКОГО РЕГИОНА И ЕГО МЕЖРЕГИОНАЛЬНЫХ СВЯЗЕЙ

(промежуточный)

0341-2018-0001

Книга 11

Направление 4 – Научное обоснование охраны озера Байкал и социально-экономического развития Байкальской природной территории

Блок 4.2. Применение методов NGS-BD (Next Generation Sequencing – Big Data)

для решения вопросов экологии

Руководитель НИР,

научный руководитель

ИНЦ СО РАН,

академик РАН \_\_\_\_\_\_\_\_\_\_\_ И.В. Бычков

Иркутск 2018СПИСОК ИСПОЛНИТЕЛЕЙ

|  |  |  |
| --- | --- | --- |
| Руководители НИР,  д.б.н., проф., зав. отделом | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | Е. В. Лихошвай  (введение, разделы 4, 6-8,  заключение) |
| д.б.н., зав. лаб. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | Т. И. Земская  (введение, разделы 3, 5,  заключение) |
| Исполнители: |  |  |
| с.н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | Ю. П. Галачьянц  (разделы 6, 7) |
| с.н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | Ю. Р. Захарова  (разделы 4, 6, 7) |
| с.н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | Д. П. Петрова  (разделы 4, 6, 7) |
| с.н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | А. В. Ломакина  (разделы 5) |
| с.н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | Ю. С. Букин  (разделы 3, 6, 7) |
| с.н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | Е. А. Черкашин  (разделы 1-3) |
| в.н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | А. О. Шигаров  (разделы 1, 2) |
| с.н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | С. М. Шишлянников  (раздел 5) |
| н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | И. С. Михайлов  (разделы 3, 4, 6, 7) |
| н.с. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | Ф. С. Малков  (разделы 1, 2) |
|  |  |  |
| Нормоконтролер | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ подпись, дата | О. М. Чагина |

**РЕФЕРАТ**

Отчет 72 с., 35 рис., 9 таблиц, 34 источника, 3 прил.

БОЛЬШОЙ ОБЪЕМ ДАННЫХ, ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ, ЭКОИНФОРМАТИКА, ВОДНЫЕ ЭКОСИСТЕМЫ, NGS, МИКРОБИОМ, ДИНАМИЧЕСКАЯ СИСТЕМА

Объект исследования и разработки: большие данные, секвенирования нового поколения (NGS), микробиом оз. Байкал, экосистема.

Цель всего блока проекта – характеристика микробиома оз. Байкал на основе методов секвенирования нового поколения (NGS) и результатов обработки полученной информации аналитическими технологиями Больших Данных (Big Data) для решения задач экологии.

Цель исследования 2018 г. – реализация системного подхода к исследованию оз. Байкал на основе данных NGS и современных мировых стандартов и инструментальных средств.

Методы исследования: выделение и NGS секвенирование ДНК, анализ данных NGS, построение и анализ корреляционных зависимостей между характеристиками микробиома, идентификация соответствующей динамической модели; системный анализ информационных потоков; разработка графовых моделей и структуры данных для облачного хранения.

Результаты исследования: Разработан алгоритм построения концептуальной модели взаимоотношений организмов в сообществах. Показано, что петли V2–V3 16S рРНК обладают наибольшей разрешающей способностью на уровне таксонов низшего ранга (рода, вида). NGS фрагментов генов 16S и 18S рРНК проб фотического слоя 2017 г. оз. Байкал выявил высокое богатство бактерий и микроэукариот, превосходящее полученные ранее данные; демонстрация сезонных изменений OTU (Operational Taxonomic Unit) свидетельствует о перспективности метода NGS для мониторинга. Выполнено прототипирование информационно-аналитической системы поддержки исследования микробиома оз. Байкал, включая схему представления и хранения метагеномных данных и OTU, алгоритм разбора и загрузки OTU из формата пакета программного обеспечения MOTHUR в базу метагеномных данных, веб-ориентированный RESTful интерфейс прикладного программирования для доступа к данным на основе архитектуры взаимодействия компонентов распределённого приложения в сети.

Рекомендации по внедрению результатов научных исследований и разработок будут сделаны после завершения работ.

Область применения: Полученные результаты важны при разработке методик мониторинга микробных сообществ в любых системах с применением NGS.

**СОДЕРЖАНИЕ**

|  |  |  |
| --- | --- | --- |
|  | ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ …………………………………………………. | 5 |
|  | ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ …………………………….. | 6 |
|  | ВВЕДЕНИЕ ………………………………………………………………………... | 7 |
| 1 | Управление данными информационно-аналитической системы поддержки исследования микробиома озера Байкал ………………………………………... | 10 |
| 2 | Визуальное моделирование проблемно-ориентированных пакетов …………... | 18 |
| 3 | Алгоритм построения сетей взаимодействия видов в сообществах микроорганизмов на основе результатов мониторинговых экологических исследований с привлечением геномных и метагеномных данных ………………………. | 25 |
| 4 | Анализ проб фотического слоя озера Байкал 2018 г методами микроскопии и микробиологии и подготовка проб для секвенирования NGS ………………… | 29 |
| 5 | Реконструкция полных геномов выделенных в культуры перспективных для практического применения штаммов бактерий ………………………………… | 45 |
| 6 | Структура бактериальных сообществ фотического слоя озера Байкал в весенний, летний и осенний периоды 2017 г. по данным NGS …………………... | 49 |
| 7 | Сукцессия сообществ бактерий и микроэукариот в южной котловине озера Байкал в 2017 г. по данным NGS ………………………………………………… | 54 |
|  | ЗАКЛЮЧЕНИЕ …………………………………………………………………… | 64 |
|  | СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ …………………………….. | 66 |
|  | ПРИЛОЖЕНИЕ А Публикации по теме проекта за 2018 год …………………. | 69 |
|  | ПРИЛОЖЕНИЕ Б Информация о количестве статей в виде таблицы ………... | 71 |
|  | ПРИЛОЖЕНИЕ В Этапы выполнения проекта ………………………………… | 72 |

* 1. ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящем отчете о НИР применяют следующие термины с соответствующими определениями

Dataflow programming – стиль визуального программирования, основанного на конструировании вычислительного процесса из функциональных блоков, соединенных друг с другом входными и выходными параметрами.

RESTful интерфейс – способ организации взаимодействия функциональных блоков распределенных вычислительных систем, реализуя доступ к объектам по протоколу HTTP специально формализованным образом.

Model-View-Controller (MVC) – архитектура разработки интерфейсов пользователя в объектно-ориентированных системах программирования, разбивая функции интерфейса и приложения между отдельными объектами, выполняющими роль Model, View и Controller.

Structure query language (SQL) – стандарт управления реляционными базами данных при помощи специального языка запросов.

New-generation sequencing (NGS) – направление исследований в биологии, основывающееся на анализе анализа ДНК образцов.

* 1. ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящем отчете о НИР применяют следующие сокращения и обозначения

|  |  |  |
| --- | --- | --- |
| CIM | – | Computationally independent model |
| HTTP | – | Hypertext transfer protocol |
| JSON | – | JavaScript Object Notation |
| MVC | – | Model-View-Controller |
| MDA | – | Model driven architecture |
| NGS | – | New generation sequencing |
| ОТЕ | – | Operational taxonomic element |
| OTU | – | Operational taxonomic unit |
| PIM | – | Platform independent model |
| SQL | – | Structured Query Language |
| RDF | – | Resource description framework |
| URL | – | Universial resource locator |
| ДНК | – | Дезоксирибонуклеиновая кислота |
| РНК | – | Рибонуклеиновая кислота |

* 1. ВВЕДЕНИЕ

Внедрение в исследования структурно-функциональной организации водных экосистем методов секвенирования нового поколения (NGS) и биоинформатических методов анализа больших объемов информации («больших данных») дает наиболее полное представления о таксономическом разнообразии сообществ микроэкориот и прокариот и позволяет проводить сравнительный и корреляционный анализ [1, 2]. Например, было показано, что корреляции между отдельными бактериальными таксонами намного сильнее, чем между бактериями и эукариотами или бактериями и факторами окружающей среды [3], выявлены экологические связи морских бактерий, архей и протист [4]. На основе таких ко-ассоциативных сетей стало возможным построение динамических моделей метапопуляций, которые предсказывают изменение численности различных организмов в зависимости от изменений факторов окружающей среды [5] и позволяет получить более глубокое понимание скрытых механизмов, обеспечивающих стабильность и динамику микробных сообществ [6]. Помимо фундаментальных знаний о природе моделирование сообществ имеет важный практический аспект в биотехнологии, биоэнергетике и биоремедиации, так как позволяет целенаправленно создавать многовидовые микробные сообщества с определенными функциями и предсказанным поведением [7].

Используемое в научных исследованиях ЛИН СО РАН программное обеспечение анализа ампликонов включает в себя различные библиотеки модулей обработки последовательностей, например, Mothur, USearch, статистические пакеты и среды разработки алгоритмов многомерного статистического анализа данных, например, R (https://www.r-project.org). Для проведения исследований с использованием обработки и анализа метагеномных данных специалисту требуются навыки программирования (составления сценариев) в командной оболочке операционной системы (Linux, Windows), запуска пакетов в распределенной вычислительной среде и кластерных вычислительных системах, а также программирования на языках R или Python. Одним из способов расширения круга пользователей сложных прикладных пакетов – это представление структуры модулей прикладного пакета в виде визуальной модели, а также разработки средств визуального проектирования вычислительного процесса. В результате пользователи-предметники, не являющиеся специалистами в области программирования, смогут самостоятельно формировать вычислительную процедуру из визуальных блоков.

Популярным подходом к представлению вычислительного процесса, который строится в виде комбинации исполняющих модулей, выступает программирование потоков данных (dataflow programming). Модули получают на вход некоторые данные, обрабатывают их и выдают на выход некоторый результат. Результаты вычислений передаются на вход других модулей. Подход развивается с 1970-х годов и имеет приложения в самых различных областях.

Подход удобен тем, что модули и сценарий их исполнения удобно задавать при помощи визуального редактора. Пользователь выбирает нужный модуль из существующего набора, добавляет его в рабочую область. Затем необходимо сцепить входы и выходы модулей (порты), обеспечив таким образом передачу информации. В результате должен получиться сетевой граф, не содержащий циклов и имеющий, как правило, один вход и один выход. На вход графа подаются данные, требующие обработки, на выход графа выдается результат, который сохраняется в файл или визуализируется при помощи соответствующих модулей.

Каждый модуль прикладного пакета Mothur получает на вход название файла и в результате работы создает новый файл. Функционирование модуля зависит от параметров, задаваемых пользователем индивидуального для каждого модуля. Разработанные программные средства позволяют представлять вычислительный процесс Mothur в виде блоков системы Rapidminer Studio, реализующих вышеупомянутый модель потоков данных. Система поддерживает возможность представления сценария в виде нового комплекного блока со своими портами ввода и вывода, а также возможность облачного хранения и исполнения сценариев, что позволяет создавать распределенные вычислительные среды. Богатый набор функций Rapidminer, а также различные сервисы, предоставляемые разработчиками, стали основной причиной выбора этой системы в качестве среды разработки информационно-вычислительных ресурсов проекта.

Для того чтобы установить и исследовать механизмы взаимодействия между видами в сообществах можно использовать корреляционный анализ. Как правило, одно и то же сообщество охарактеризовано либо в сезонной динамике, либо исследователями изучалось несколько повторных проб (образцов) из разных географических районов. Таким образом, результат работ представляет собой таблицу, характеризующую представленность определенного спектра видов в разных биологических образцах. Если иметься достаточно количество таких биологических образцов (10 и более) можно рассчитать корреляционную матрицу, отражающую взаимосвязи между видами в изучаемом сообществе. Положительные коэффициенты корреляции будут свидетельствовать о взаимном позитивном влиянии видов друг на друга. Отрицательные значение коэффициентов корреляции будет говорить о том, что в рассматриваемой паре виды конкурируют за какой либо общий ресурс, подавляя друг друга. Информацию из корреляционной матрицы можно представить в виде сети взаимодействий (ненаправленного графа) отражающего взаимосвязи между организмами. В свою очередь такой граф может служить шаблоном для разработки динамической модели в виде системы дифференциальных уравнений, описывающей и предсказывающей поведение численности (биомассы) видов с течением времени.

**1 Управление данными информационно-аналитической системы поддержки исследования микробиома озера Байкал**

Краткий результат главы.Выполнено проектирование и прототипирование информационно-аналитической системы поддержки исследования микробиома озера Байкал. Усовершенствована схема представления и хранения исследовательских метагеномных данных и OTU результатов. Реализован процесс автоматического разбора и загрузки OTU-данных, полученных в программе обработки данных высокопроизводительного секвенирования MOTHUR в базу метагеномных данных. Разработан веб-ориентированный RESTful-интерфейс прикладного программирования для доступа к данным исследования микробиома на основе архитектуры взаимодействия компонентов распределённого приложения в сети. Интерфейс обеспечивает чтение и запись данных для всех объектов текущей модели информационно-аналитической системы. Развёрнута инфраструктура проекта, включая систему совместной разработки программного обеспечения (доступна по адресу https://git.icc.ru), систему управления проектом (доступна адресу <https://redmine.icc.ru>).

**1.1 Материалы и методы**

Проектирование и реализация объектной модели информационно-аналитической системы поддержки исследования микробиома озера Байкал выполнена на платформе разработки программного обеспечения веб-приложений DJANGO с использованием шаблона проектирования MVC (Model-View-Controller) для разделения данных приложения, пользовательского интерфейса и управляющей логики. Проектирование усовершенствованной схемы данных осуществлялось в соответствии спостановкой задачи и предоставленных исследовательских данных ЛИН СО РАН. На основе разработанной схемы подготовлено метаописание таблиц базы данных. Сформированные классы объектной модели в терминах DJANGO позволили сгенерировать SQL-скрипты (Structured Query Language) создания структуры базы данных в формате СУБД (системы управления базами данных) PostgreSQL.

Для разработки программного интерфейса созданы управляющие скрипты: (1) сериализации (serializing) объектов модели разрабатываемой информационно-аналитической системы в формате JSON (JavaScript Object Notation); (2) создания представлений для каждого сериализатора класса этой модели; (3) подключения списка URL-ссылок доступных запросов интерфейса к платформе разработки Django; (4) настройки безопасности доступа к данным с поддержкой пользовательской авторизации.

**1.2 Результаты**

Усовершенствована схема данных исследования микробиома озера Байкал. На рисунке 1 представлена её упрощенная модель. Выполнено разделение на следующие модули:

* Модуль RPM обеспечивает общее описание проводимых исследовательских работ, включая данные по их исполнителям, публикациям, методикам, оборудованию и расходным материалам.
* Модуль WaterLIN служит для представления специфических данных исследования микробиома озера Байкал, включая описания отбираемых проб, измеримые параметры, результаты анализа проб, изучаемые ампликоны, полученные прочтения, результаты секвенирования.
* Модуль NGS организует данные, связанные с проведением вычислений по высокопроизводительному секвенированию ДНК/РНК последовательностей исследуемых проб воды и донных осадков, включая информацию по операционным таксономическим единицам, идентификаторам таксонов, используемым таксономическим базам данных, рабочим процессам (вычислительным цепочкам) обработки последовательностей и анализа результатов, списку программ, задействованных в этих рабочих процессах.



Рисунок 1 – Упрощенное представление схемы данных исследования  
микробиома озера Байкал

Разработан алгоритм генерации таблицы встречаемости OTU по выходным данным формата MOTHUR. На вход алгоритм получает MOTHUR файлы final.an.list и final.groups. Исследователь выбирает необходимую частоту встречаемости, для которой необходимо посчитать OTU. На выходе алгоритм генерирует файл с необходимой таблицей встречаемости. Ниже представлены блок-схемы основных функций алгоритма (см. рисунки. 2-6). На рисунке 2 представлен алгоритм подготовки финального списка, которая из файла с множеством результатов частот встречаемости выделяет нужную строку с заданной исследователем частотой встречаемости.

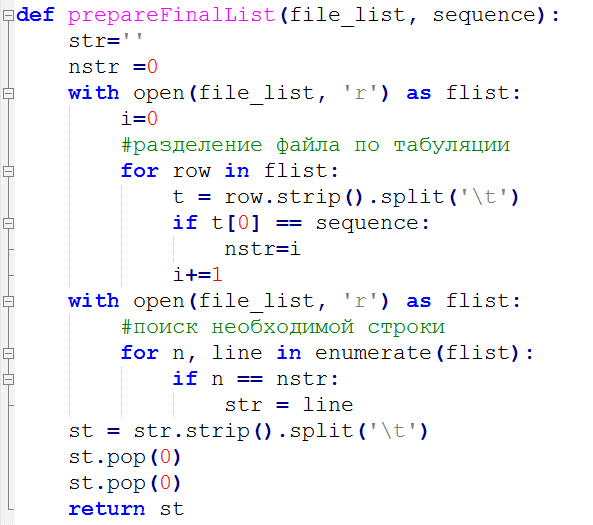


Рисунок 2 – Программа подготовки финального списка таблицы встречаемости OTU

На рисунке 3 показан алгоритм подготовки списка по файлу final.groups, который разделяет файл на строки и столбцы для последующего удобного поиска по файлу.

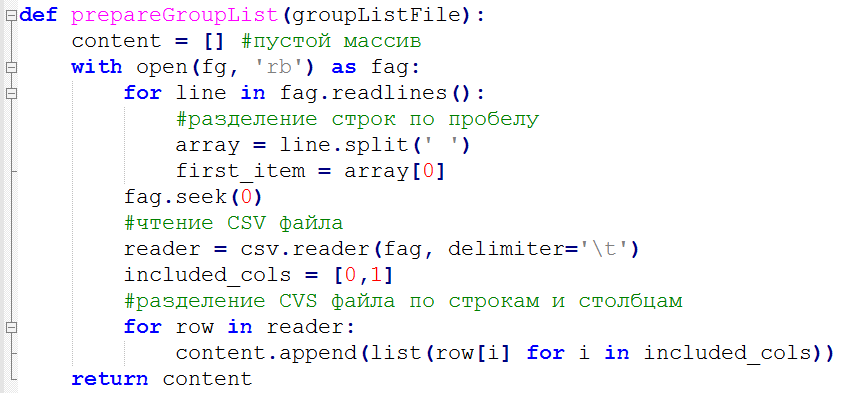


Рисунок 3 – Программа подготовки списка группы таблицы встречаемости OTU

Рисунок 4 демонстрирует функцию получения уникальной пробы, анализирует массив данных, полученных после подготовки списка по файлу final.groups и выдает на выходе список проб, состоящий из проб по содержимому файла final.groups.

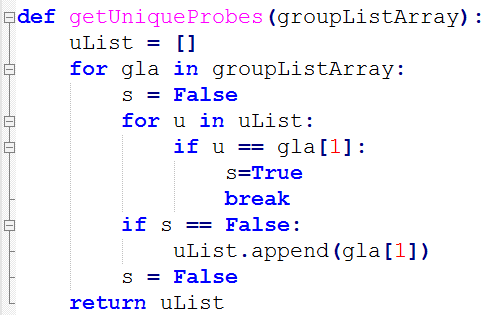


Рисунок 4 – Программа запроса уникальных проб из списка проб

На рисунке 5 представлен алгоритм поиска совпадений для заданного ампликона по данным, полученным из файла final.groups.

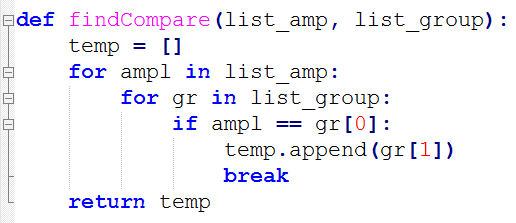


Рисунок 5 – Программа поиска совпадений заданного ампликона

На рисунке 6 представлен алгоритм генерации таблицы встречаемости OTU, задающий рабочий процесс вызова функций, представленных на рисунках 2-5. Алгоритм считает количество ампликонов для каждой пробы, происходит генерация двумерной матрицы и её наполнение статистическими подсчётами, а также сериализацию результатов в файл. Спроектированный алгоритм реализован на языке программирования PYTHON. Фрагмент таблицы встречаемости OTU, полученной в результате запуска этой реализованного, показан на рисунке 7.

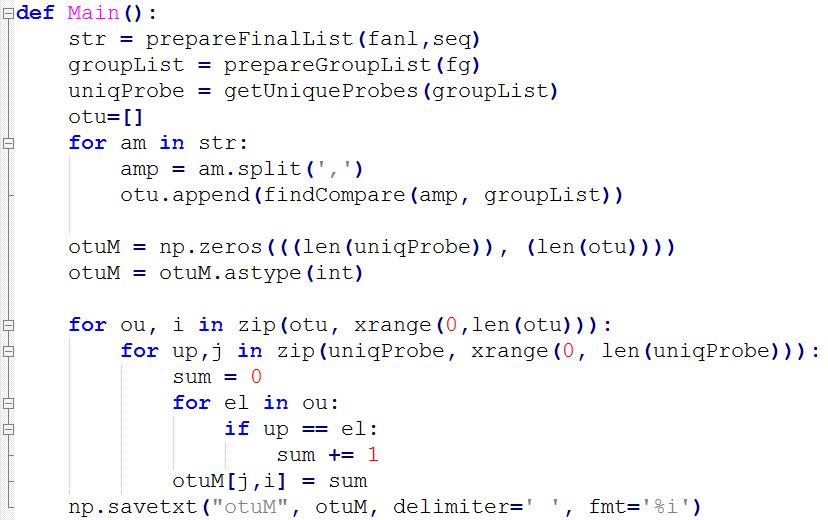


Рисунок 6 – Блок-схема алгоритма генерации таблицы встречаемости OTU

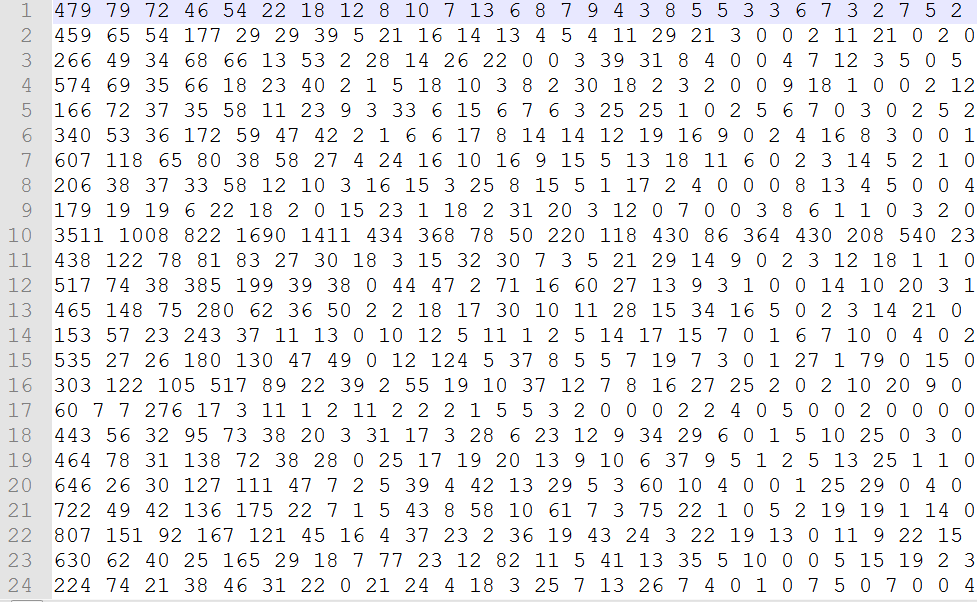


Рисунок 7 – Фрагмент таблицы встречаемости OTU, полученной в результате запуска реализованного алгоритма

Спроектирован и реализован программный веб-ориентированный интерфейс доступа к данным исследования микробиома озера Байкал на основе архитектуры взаимодействия компонентов распределенного приложения REST (Representational State Transfer) API. Этот интерфейс обеспечивает чтение и запись данных для всех объектов модели разрабатываемой информационно-аналитической системы. Тестовая версия интерфейса (см. рисунок 1.8) доступна по адресу https://ngs.icc.ru/api.

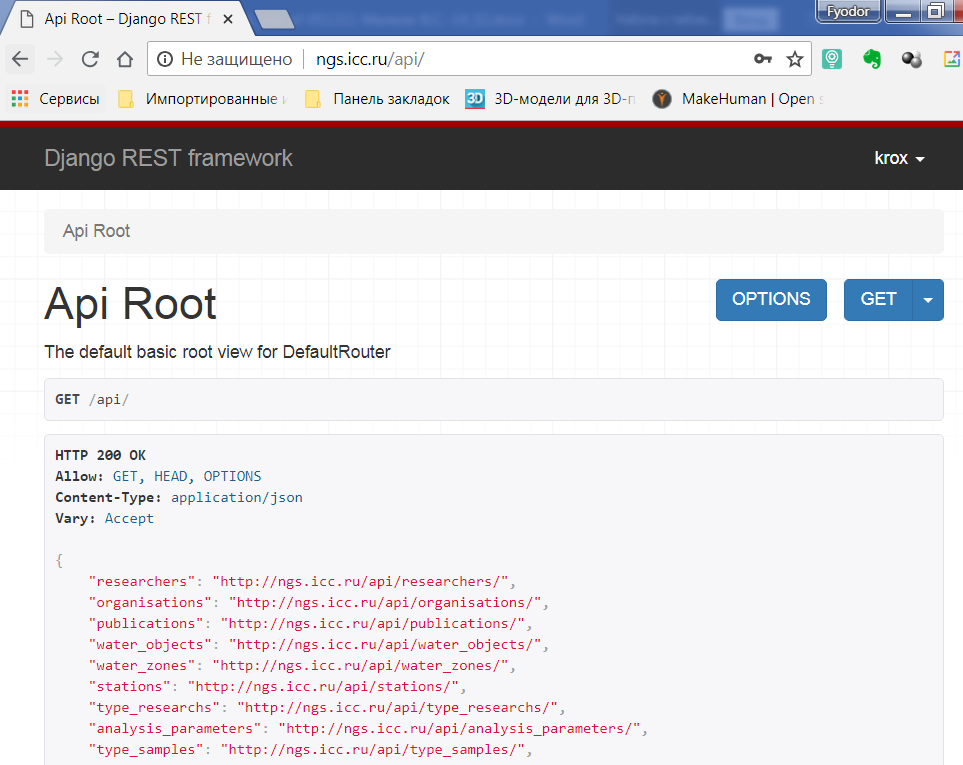


Рисунок 8 – Снимок экрана тестирования программного интерфейса (RESTful) доступа к данным исследования микробиома озера Байкал

Интерфейс определяет, что для отправки HTTP-запроса необходимо указать URL запроса, тип (POST или GET), а также данные для авторизации (имя пользователя и пароль). В ответе REST-интерфейс возвращает результат выполнения запроса в формате JSON. Пример отправки запроса для получения всех доступных URL для классов модели с использованием утилиты CURL выглядит следующим образом: curl -X GET --user user:password http://ngs.icc.ru/api. В результате будет получен следующий список всех доступных URL в формате JSON, как показано ниже на рисунке 9.

{"researchers":"http://127.0.0.1:8000/api/researchers/",  
"organisations":"http://127.0.0.1:8000/api/organisations/",  
"publications":"http://127.0.0.1:8000/api/publications/",  
"water\_objects":"http://127.0.0.1:8000/api/water\_objects/",  
"water\_zones":"http://127.0.0.1:8000/api/water\_zones/",  
"stations":"http://127.0.0.1:8000/api/stations/","type…

Рисунок 9 – Пример с фрагментом ответа в формате JSON с результатом выполнения запроса получения всех доступных URL ссылок интерфейса

Следующий пример показывает выполнение GET запроса для получения всех объектов класса Researcher, представляющего данные исполнителей исследовательских работ: curl -X GET --user user:password http://ngs.icc.ru/api/researchers. В результате выполнения данного запроса будет получен список всех исследователей, как показано на рисунке 10.

[{"user":1,"fio":"Ivanov","organisation":1,"lab":"laboratory 1","contacts":"890256660777","date\_added":"2018-08-02T01:26:39.586000+08:00"},{"user":null,"fio":"Petrov","organisation":1,"lab":null,

"contacts":null,"date\_added":"2018-12-07T10:43:31.039000+08:00"},{"user":null,"fio":"Ivanov","organisation":1,"lab":null,

"contacts":"892435556677","date\_added":"2018-12-07T11:14:29.374000+08:00"}]

Рисунок 10 – Пример с фрагментом ответа в формате JSON с результатом выполнения запроса получения всех объектов класса Researcher.

Дополнительно к запросам чтения данных, интерфейс также предоставляет возможность изменения данных. Например, для записи данных необходимо отправить запрос POST и указать необходимые параметры объекта. Пример записи нового объекта класса Researcher: curl –X POST –d “fio=Ivanov&organisation=1&contacts=892435556677” --user user:password http://ngs.icc.ru/api/researchers, получаемым ответом, как показано на рисунке 11.

{"user":null,"fio":"Ivanov","organisation":1,"lab":null,"contacts":"892435556677", "date\_added":"2018-12-07T11:14:29.374000+08:00"}

Рисунок 11 – Пример с фрагментом ответа в формате JSON с результатом выполнения записи нового объекта класса Researcher

**1.3** **Краткое обсуждение**

При исследовании микробиома озера Байкал методами высокопроизводительного секвенирования порождаются большие массивы метагеномных данных. Жизненный цикл этих данных включает этапы сбора, обработки, анализа, публикации и архивирования этих. Микробиологи не всегда способны самостоятельно организовать управление жизненным циклом своих исследовательских данных. Планирование и управление данными требует навыков работы со специализированным программным обеспечением. Предполагается, что разрабатываемая информационно-аналитической системы поддержки исследования микробиома озера Байкал обеспечит микробиологов возможностью упрощенного управления метагеномными данными на протяжении всего их жизненного цикла. В частности, она позволяет микробиологам хранить и публиковать в открытом и защищенном доступе свои первичные и обработанные данные, результаты исследования, а также самостоятельно создавать и хранить собственные рабочие процессы обработки и анализа метагеномных данных при исследовании микробиома озера Байкал.

Разработаны средства управления данными (в т. ч. схема данных и объектная модель информационно-аналитической системы поддержки исследования микробиома озера Байкал, веб-ориентированный пользовательский и программный интерфейс доступа к исследовательским данным) с целью обеспечения курирования больших объемов данных высокопроизводительного секвенирования.

# Визуальное моделирование проблемно-ориентированных пакетов

Краткий результат главы.Разработана программная система анализа исходного кода прикладного пакета Mothur, создающая абстрактную модель модулей в виде графа отношений элементов структуры пакета. Создана и формализована система знаний, при помощи которой по результатам вышеуказанного анализа осуществляется порождение исходного кода Java-модулей для системы визуального проектирования вычислительных процессов Rapidminer. В результате в Rapidminer представлены 143 процедуры обработки данных NGS. Разработанная технология позволит на следующих этапах проекта также сгенерировать визуальное представление модулей вычислительных сред NGS, таких систем, как QIIME, QIIME2, Usearch и др., что позволит унифицировать модель вычислительной процедуры анализа ампликонов и сделать ее независимой от используемого прикладного пакета.

На этапе 2018 года направлением развития автоматизации программирования сценариев метагеномного анализа являлась разработка системы интерпретации метаописаний операций программы Mothur в промежуточное представление, разработка процедуры трансформации данного представления в исходный код модулей системы Rapidminer и формирование библиотеки вычислительных модулей, которой могу пользоваться специалисты-предметники для решения биоинформатических задач.

Необходимость в автоматическом анализе исходного кода Mothur и синтезе соответствующих структур модулей Rapidminer определяется, во-первых, количеством модулей, которые необходимо представить в виде визуальных блоков, – 143, а, во-вторых, необходимостью поддерживать соответствие сгенерированных блоков новым версиям модулей Mothur, т.к. система находится в фаза активного развития. Кроме того, наличие такой модели позволяет на следующих этапах проекта разрабатывать процедуры автоматизации вычислительного процесса на кластерных вычислительных системах.

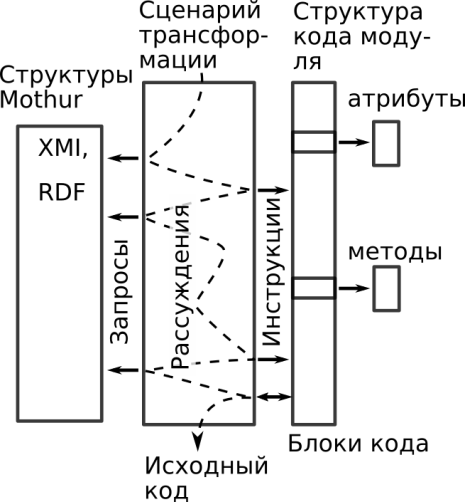
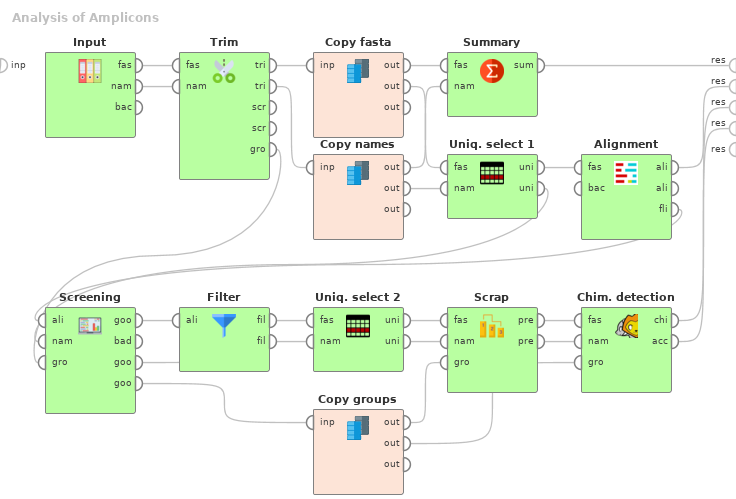
Процесс трансформации представляет собой анализ исходного кода Mothur, осуществляющийся при помощи алгоритмов, применяющих регулярные выражения для распознавания элементов структуры классов С++, задающих модули Mothur. Результаты анализа представляются в виде структур RDF (Resource description framework). Пример представления результата анализа модуля представлен на рисунке 12.

|  |  |
| --- | --- |
| **@prefix** **dc:** <http://purl.org/dc/elements/1.1/> .  *# ......*  **@prefix** **xsd:** <http://www.w3.org/2001/XMLSchema#> .  **ngsp:spec** a **ngsp:Specification** ;  **ngsp:module** **mothur:chimera-ccode**,  *# ......*  **mothur:venn** .  *# ......*  **mothur:chimera-ccode** a **ngsp:Module** ;  **ngsp:outputPattern** [ a **cnt:Chars** ;  **ngsp:parameterName** *"type"* ;  *# ......*  [ **ngsp:patternString** *"[filename],[tag],..."* ;  **dc:identifier** *"accnos"* ],  [ **ngsp:patternString** *"[filename],mapinfo"* ;  **dc:identifier** *"mapinfo"* ] ;  **cnt:chars** *"getOutputPattern( // ... "* ;  **ngsp:parameter** **mothur:chimera-ccode-fasta-parameter**,  *# ......*  **mothur:chimera-ccode-window-parameter** ;  **dc:title** *"chimera-ccode"* ;  **dcterms:description** *"detect chimeric sequences"* ;  **schema:citation** *"..."* ; **schema:softwareHelp** *"..."* ;  **nco:websiteURL** <http://www.mothur.org/wiki/Chimera-ccode> ;  **v:category** *"Sequence Processing"* .  **mothur:chimera-ccode-fasta-parameter** a **ngsp:Parameter** ;  **ngsp:important** true ;  **ngsp:multipleSelectionAllowed** false ;  **ngsp:optionsDefault** *""* ;  **ngsp:outputTypes** [ a **oslc:AllowedValues** ;  **dc:identifier** *"accnos"*,*"chimera"*,*"mapinfo"* ] ;  **ngsp:outputTypesOrig** *"chimera-mapinfo-accnos"* ;  **ngsp:required** true ; **ngsp:type** **mothur:InputTypes** ;  **dc:title** *"fasta"* ; **schema:sku** 1 . |  |

Рисунок 12 – Результат анализа класса в исходном коде прикладного пакета Mothur: представление в виде RDF (слева), диаграмма SysML (справа)

Трансформация RDF-представления производится алгоритмами, реализованными в логической и объектно-ориентированной парадигмах программирования. Для обеспечения трансформации развит предложенный ранее подход применения логического программирования в качестве языка реализации процедур трансформации моделей (знаний о трансформации) в рамках «Модельно-управляемой архитектуры» (MDA, Model driven architecture). Использованная реализация трансформации в виде программы на языке LogTalk расширена новыми объектами, представляющими собой знания о программной целевой платформе, системы программирования Java.

Трансформация представляется как сценарий синтеза платформозависимой модели (PSM) программного обеспечения (ПО) (рисунок 13). Сценарий представляет собой объект – инкапсуляцию процедур трансформации, посылающий запросы объектам, представляющим платформонезависимую модель (PIM), и управляющий объектами, синтезирующими структуру PSM. PIM представлена в виде RDF-графа (рисунок 13, а), а PIM – древовидная система блоков кода (объектов и литералов).

а) б)

Рисунок 13 – Общая структура процесса порождения исходного кода модуля (а) на основе данных анализа структуры прикладного пакета Mothur; представление части вычислительного процесса анализа ампликонов (б) в виде визуальной диаграммы Rapidminer

Используемая архитектора инфраструктуры трансформации представлена на рисунке 13. Инфраструктура состоит из сервисов (большие блоки). Сервисы «CIM» (Computationally independent model) и «PIM» являются источниками информации, они заполняют базу данных «Сервера модели Mothur» исходными данными для трансформации. Сервер модели и «Сервер онтологий» предоставляют модулю трансформации («Инструмент MDA») исходные данные о структуре модулей Mothur, а также дополнительную информацию, например, из Wikipedia (DBPedia.org), позволяющую дополнить семантическими данными синтезируемую программу.

В рамках проекта используемая архитектура подверглась изменениям в части представления исходной информации CIM в виде результата анализа исходного кода Mothur. Также произведена настройка (в виде классов-адаптеров) блока Сервера моделей на разработанную концептуальную модель «ngs/ngsp» структуры Mothur. Данная модель позволяет задавать различные отношения, описывающий модуль Mothur (ngsp) и термины их предметной области NGS.

Одной из сопутствующих задач являлось исследование новых возможностей LogTalk для представления процедур трансформации. При реализации процедур трансформации использованы большинство предоставляемых LogTalk синтаксических структур и вариантов объектных иерархий. Особенно оказались полезными параметризованные объекты, категории и их иерархии прототипов. Параметризованные объекты позволяют задавать контекст выполнения метода без создания экземпляра класса. Второй необычной но полезной структурой являются категории – инкапсуляция реализаций логически связанных друг с другом методов – и композиционное программирование. Все статические объекты компилятор LogTalk (например, параметризованные объекты) преобразует непосредственно в Prolog, при этом накладные расходы на производительность кода не больше 1% (рисунок 14).



Рисунок 14 – Архитектура инфраструктуры анализа и порождения исходного кода  
модулей Rapidminer

На рисунке 15 представлен отрывок исходного кода объекта-адаптера запросов сценария синтеза исходного кода к исходным данным RDF-модели. Адаптер содержит один параметр – граф RDF и реализует три метода-запроса «class», «attribute» и «method», предназначенные для распознавания класса, списков его атрибутов и методов, соответственно.

В приведенном отрывке также представлен (в комментарии) пример запроса к данному адаптеру. Результатом запроса, как и в любой реализации Prolog, является множество ответов, представленных в виде наборов значений переменных. В данном примере в качестве ответа выступают наборы имя и идентификатор метода заданного идентификатором «ClassID» класса. Процесс порождения исходного кода состоит из двух этапов. Первый – синтез модели генерируемого объекта, второй – собственно запись исходного кода объекта в определенное место в файле. На рисунке 16 представлен фрагмент генератора модуля Java. Известно, что согласно стандарту Java-1.8 в каждом модуле может находиться только один класс. В соответствии с данной спецификацией генератор модуля состоит из двух методов: генератора преамбулы и генератора класса. Генератор класса реализован в отдельном модуле. Кроме методов генерации модели исходного кода генератор модуля содержит вспомогательные методы, используемые на этапе порождения исходного кода конфигурационных файлов Rapidminer. Например, метод «module\_icon\_name» синтезирует имя файла, в котором храниться пиктограмма модуля, на основе имени класса.

|  |
| --- |
| :- **object**(**query**(\_**RDF**)).  :- protected(xmi/1).  :- public([**class**/2, **attribute**/3, **method**/3]).  xmi(XMI) :- parameter(1, XMI).  **class**(Name, ID):- *% Распознавание класса в RDF*  ::xmi(XMI),  XMI::rdf(ID,rdf:type,uml:*'Class'*),  XMI::rdf(ID,rdfs:label, literal(Name)).  **attribute**(Name, ClassID, ID):- *% ...атрибута...*  ::xmi(XMI),  XMI::graph(G),  XMI::rdf(ClassID, G:ownedAttribute, ID),  *% XMI::rdf(ID, rdf:type, uml:'Property'),*  XMI::rdf(ID, rdfs:label, literal(Name)).  **method**(Name, ClassID, ID):- *% ...метода...*  ::xmi(XMI),  XMI::graph(G),  XMI::rdf(ClassID, G:ownedOperation, ID),  XMI::rdf(ID, rdfs:label, literal(Name)).  :- **end\_object.**  *%* ***query****(****mothur****)::****method****(Name,ClassID,MethodID).* |

Рисунок 15 – Параметризованный объект LogTalk, адаптирующий запросы к структурам   
исходной RDF-модели

|  |
| --- |
| :- **object**(mothur\_module,  **specializes**(java\_module)).  :- public(**preamble**/0). *% генерирует начало модуля.*  **preamble**:-  ^^preamble,  ::set\_package(*'com.rapidminer.ngs.operator'*),  ::add\_skip(1),  **create\_object**(Imports, [instantiates(java\_import)],[],[]),  ::set\_block(imports(Imports)),  ::append(Imports),  ::add\_skip(1),  **create\_object**(ClassDef, [instantiates(mothur\_class)],[],[]),  ClassDef::preamble,  ::set\_block(class(ClassDef)),  ::append(ClassDef),  Imports::add(*'java.util.List'*),  Imports::add(*'com.rapidminer.operator.OperatorDescription'*),  Imports::add(*'com.rapidminer.operator.ports.InputPort'*),  *% . . . . . . .*  *% Указание родительского класса генерируемому классу.*  ClassDef::extends(*'MothurGeneratedOperator'*).  :- public(**module\_class**/2).  **module\_class**(Class, Name):- *% Генератор класса модуля*  ::current\_block(class(Class)),  Class::class\_name(Name).  :- public(**module\_name**/1).  **module\_name**(String):-  ::module\_class(\_,Name),  writef::swritef(String,*'%w.java'*, [Name]).  *% . . . . . . . . .*  :- public(**module\_icon\_name**/1).  **module\_icon\_name**(String):-  ::module\_class(Class,\_),  Class::item(name(Name)),  downcase\_atom(Name, DName),  writef::swritef(String, *'icon-operator-%w.png'*, [DName]).  :- **end\_object.** |

Рисунок 16 – Генератор модуля Rapidminer

Процесс записи синтезируемого кода реализуется предикатами метода «renderitem/2», который дополняется новыми знаниями при наследовании классов-генераторов. На рисунке 17 представлен пример такого дополнения. Метод «renderitem» дополнен правилом синтеза исходного кода элемента «imports». Основная процедура порождения исходного кода, как правило, последовательно обходит в глубину структуру модели кода «code\_block» и запускает методы «renderitem» для каждого найденного элемента.

|  |
| --- |
| :- **object**(java\_module,  **specializes**(code\_block)).  *% . . . . . . .*  **renderitem**(package(Name), String):-!,  root::iswritef(String, *'package %w;'*, [Name]).  **renderitem**(imports(Object), String):-!,  Object::render(String).  **renderitem**(Item, String):-  ^^renderitem(Item, String). % Возов в суперклассе  *% . . . . .*  :- public(**current\_imports**/1).  current\_imports(Object):-  ::item(imports(Object)).  :- public(**preamble**/0).  preamble. *% Do nothing*  :- **end\_object.** |

Рисунок 17 – Реализация экспорта исходного кода

Система Rapidmier Studio поддерживает блоки обработки данных в виде расширений системы (plug-ins). Каждый модуль расширения помимо генерируемых классов содержит метаописания, позволяющие классы интегрировать в визуальную среду, а также снабдить все информационные единицы текстами контекстно помощи на разных языках. Метаописания также генерируются автоматически, исходя из структуры класса и исходных RDF-данных. В результате окончательный вариант визуального представления процедур Mothur представляет собой набор блоков из которых строиться модель вычислительного процесса, пример которой представлен на рисунке 13, б.

Таким образом, в результате разработки данного направления создана и формализована система знаний, при помощи которой по результатам анализа исходного кода Mothur осуществляется порождение исходного кода Java-модулей для системы визуального проектирования вычислительных процессов Rapidminer. В виде блоков Rapidminer представлены 143 процедуры обработки данных NGS. На следующем этапе исследований по данному направлению необходимо разработать генераторы визуального представления модулей вычислительных сред NGS, таких как QIIME, QIIME2, Usearch и др., что позволит решить две задачи. Первая - это расширить методический комплекс алгоритмами других прикладных пакетов, вторая - это актуализировать процесс проектирования вычислений на текущее состояния отрасли NGS.

Таким образом, для поддержки исследований в метагеномном анализе, который позволяет описывать микробные сообщества с ранее недоступной точностью, на данном этапе разработана система формализованных знаний, позволяющих преобразовывать описания струкр модулей анализа NGS в визуальные модели. На основе этих моделей на следующих этапах проекта будут реализованы программы визуального программирования процедур анализа ампликонов, модули обеспечения хранения исходных и промежуточных данных в облачном хранилище, интегрировать процедуры и данных, организованные разными исследованиями, публиковать семантически-размеченные данные в Интернете, а также разрабатываться методики автоматизации проектирования распределенных вычислительных процессов на вычислительных кластерах.

**3 Алгоритм построения сетей взаимодействия видов в сообществах микроорганизмов на основе результатов мониторинговых экологических исследований с привлечением геномных и метагеномных данных**

Краткий результат главы. В рамках работы была реализована задача расчетов корреляционной матрицы на основе таблицы представленности видов, статистической коррекции значений коэффициентов корреляций и графической визуализации сети взаимодействий между организмами сообщества.

**3.1 Описание алгоритма**

В качестве программного средства для реализации алгоритма построения корреляционной сети была выбрана свободно распространяемая статистическая среда программирования R. Среда R содержит большое количество встроенных функций для статистической обработки данных кроме того для R доступен большой набор пакетов реализующих разные статистические методы. Общая блок схема алгоритма приведена на рисунке 18.

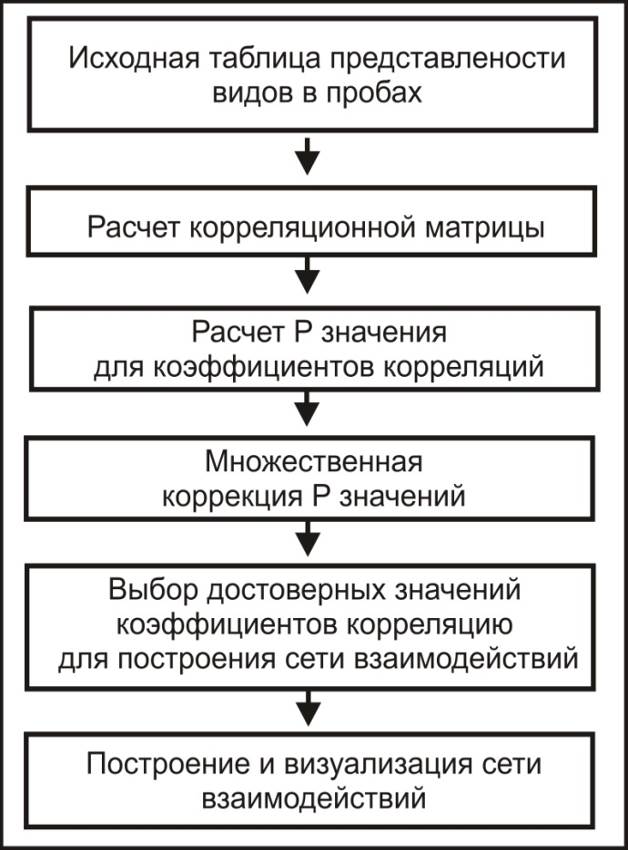


Рисунок 18 – Блок схема алгоритма построения корреляционной сети взаимодействия между организмами в сообществе

Алгоритм включает в себя несколько этапов обработки информации:

1) Считывание таблицы, характеризующей представленность таксонов в биологических пробах.

2) Тестирование выборок, представленности таксонов в пробах на соответствие закону нормального распределения с помощью теста Шапиро – Уилка.

3) Выбор способа расчета коэффициента корреляции. Коэффициент корреляции Пирсона при нормальных законах распределения таксонов в пробах и коэффициент корреляции Спирмена для не нормально законов распределений.

4) Расчет матрицы значений коэффициентов корреляции.

5) Расчет P значений, для определения достоверности значений коэффициентов корреляций (с помощью распределения Стьюдента для коэффициентов корреляции Пирсона или с помощью W статистики Спирмена).

6) Коррекция P значений на множественное тестирование с помощью метода Бенджамини — Хохберга.

7) Замена недостоверных значений (P>0.05) на ноль (недостоверные корреляции).

8) Использование матрицы коэффициентов корреляции для построения ненаправленного графа.

9) Визуализация ненаправленного графа в виде корреляционной сети взаимодействия организмов.

В качестве средств визуализации сети корреляционных взаимодействий использовались пакеты «igraph» [8] и «qgraph» [9] для языка программирования R.

Дополнительно функции пакета «igraph» в алгоритме позволяет рассчитать различные показатели, помощью которых можно определить важность таксонов в сообществах. К этим показателям относятся: центральность таксона - вида; количество связей, формируемых таксоном с другими видами.

**3.2 Апробация алгоритма**

Алгоритм был апробирован на данных опубликованных в работе [10]. Для анализа была взята таблица, характеризующая представленность различных видов бактерий в 27 образцах воды из фотического слоя пелагиали озера Байкал. Информация о представленности видов в пробах определялась с помощью метабаркодинга на основе региона V3-V4 16S рибосомной РНК. ПЦР продукты, полученные из смеси ДНК микроорганизмов расшифровывались с помощью секвенатора нового поколения GS 454 FLX (Roche, USA) в Лимнологического института СО РАН. Анализ данных метагеномного секвенирования был выполнен с помощью программы Mothur 1.19.0 [11].

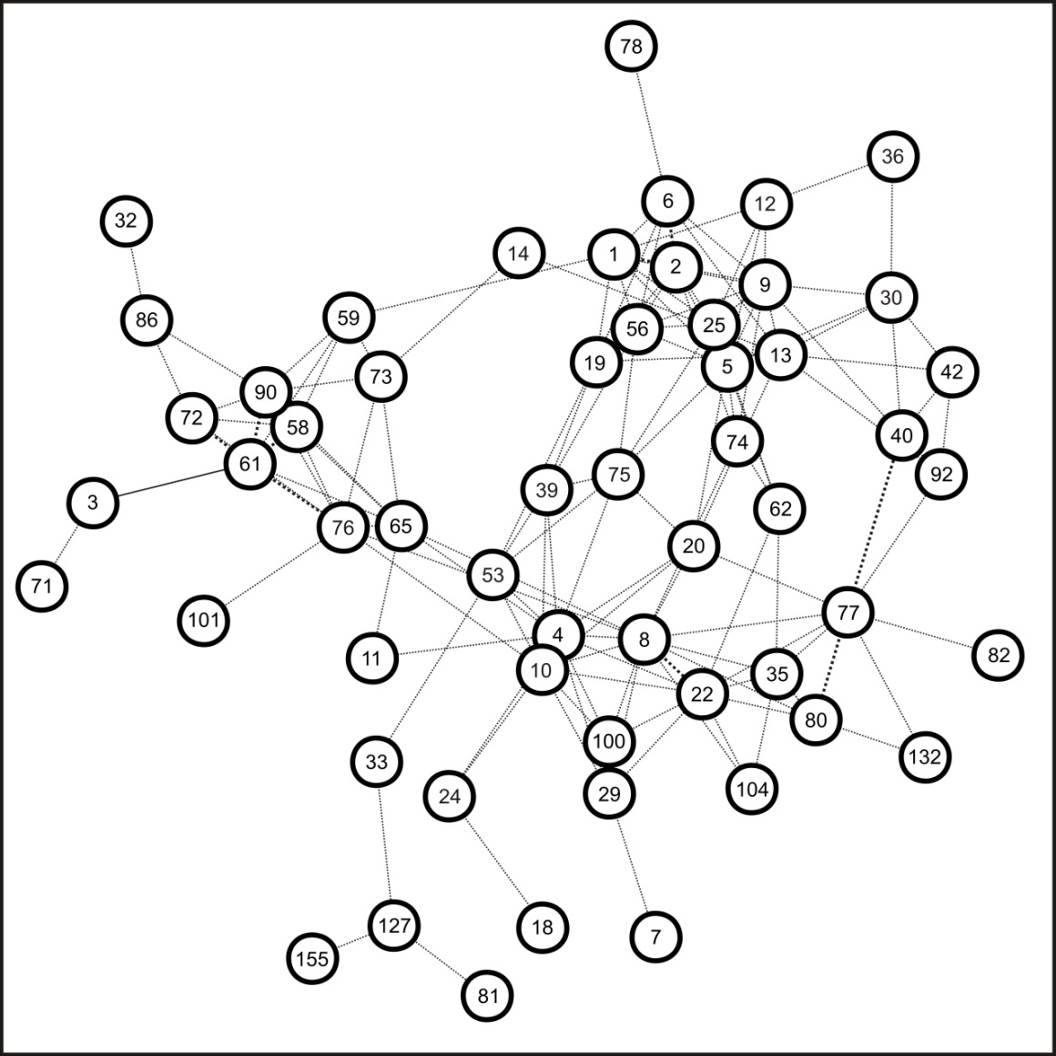


Рисунок 19 – Пример работы алгоритма построения корреляционной сети взаимосвязей между организмами в сообществе микроорганизмов. Рассматривается бактериальное сообщество экосистемы фатического слоя пелагиали озера Байкал. Цифрами обозначены виды бактерий, пунктирные линии отображают положительные корреляционные связи, сплошные линии отрицательные корреляционные связи, толщина линий соответствует силе взаимосвязи (чем толще линия, тем больше абсолютное значение коэффициента корреляции).

После первичного статистического анализ данных по разработанному алгоритму для расчетов коэффициентов корреляции был выбран ранговый критерия Спирмена. Минимально достоверные значения коэффициентов корреляции по абсолютному значению оказались больше либо равными 0.53 (α = 0.05). Реконструированная сеть межвидовых взаимодействий внутри сообщества бактерия приведена на рисунке 1.2. Сеть включила в себя 60 узлов соответствующих определенным видам бактерий и 149 связей между этими узлами. Анализ сети показал, что только одна связь из 149 отображает отрицательное значение коэффициента корреляции, все остальные связи были положительными. Это означает, что в исследованном сообществе озера Бакал большинство видов микроорганизмов находятся в мутуалистичских взаимоотношениях. Вероятно, в большинстве случаях продукты жизнедеятельности одного вида микроорганизмов являться субстратом для другого вида микроорганизмов. Конкуренция за общие ресурсы практически отсутствует.

Разработанный алгоритм построения сетей взаимодействия между организмами в сообществах позволяет выделить в исследуемой экосистеме ключевые звенья – виды, образующие каркас видов первичных продуцентов органического вещества и видов потребителей. Сеть взаимодействий позволяет выделить в основном сообществе с помощью кластеризации группы видов с более тесной связью друг с другом и менее связанные с остальными видами сообщества (структурные подгруппы видов). Анализ сети взаимодействий позволяет перейти к следующему этапу – разработки динамической системы для прогнозирования поведения сообщества и экосистемы оз. Бакал при изменении условий среды.

**4 Анализ проб фотического слоя озера Байкал 2018 г методами микроскопии и микробиологии и подготовка проб для секвенирования NGS**

Краткий результат главы.В ходе экспедиционных работ 2018 г в разные сезоны, начиная с ледового периода, произведен отбор 80 проб воды из разных экотопов оз. Байкал, квлючая зоны антропогенной нагрузки. В результате исследований определено количество (КОЕ/мл) и пространственное распределение органотрофных бактерий в верхнем слое (0-25 м) воды на 33 станциях в период с мая по сентябрь 2018 г. и общая численность бактерий (ОЧБ) в образцах из разных экотопов Байкала. Определена видовая структура, количественные показатели и пространственное распределение весеннего и летнего фитопланктона пелагической и литоральной зон Байкала в 2018 г. Оптимизирован метод выделения ДНК из природных образцов. Преимуществом метода является включение дополнительного этапа с непосредственной обработкой фильтра из аналитической трековой мембраны для концентрации биомассы из проб объемом 20 л. С использованием данного метода получены препараты ДНК из проб, отобранных из различных экотопов оз. Байкал в период весна-лето 2018 г.

**4.1 Отбор проб планктонных сообществ микроэукариот и прокариот из разных экотопов озера Байкал для метагеномного анализа**

В 2018 г. в разные сезоны, начиная с ледового периода, произведен отбор проб воды из разных экотопов оз. Байкал, включая районы повышенной антропогенной нагрузки (п. Листвянка, прол. Малое Море) и фоновые районы. Во время ледовых экспедиции в середине марта на центральной станции разреза Листвянка-Танхой (Южный Байкал) и станциях пролива Малое Море пробы отобраны с нижней поверхности льда водолазами с помощью шприцов; интегральные пробы отобраны батометром Нискина (НТФ Вольта, Россия) (V=20 л) (таблица 1); придонный слой воды – при помощи бентосной гравитационной трубки для отбора донных осадков. В период открытой воды пробы отбирали, начиная с конца мая до конца сентября, по всей акватории озера Байкал. Для сравнительного анализа взяты пробы воды из фотического (0–25м) слоя воды на 17 станциях пелагиали оз. Байкал. В ходе экспедиций на НИС «Верещагин» и НИС «Папанин» ЛИН СО РАН пробы воды отобраны с помощью системы батометров карусели SBE-32 (Сarousel Water ampler, фирмы Sea-Bird Electronics, Inc. США) и батометром Нискина для химического анализа, подсчета численности и биомассы фитопланктона, общей численности бактерий, численности культивируемых гетеротрофов на среде РПА:10 и определения таксономического состава сообществ бактерий и микроскопических эукариот методом NGS. Всего для метагеномного анализа отобрано 80 проб по 20 л (таблица 1).

Таблица 1 – Календарь отбора проб воды из озера Байкал в течение 2018 г.

| **№** | **Станция** | **Координаты** | **Глубина , м** | **март** | **май** | **июнь** | **июль** | **август** | **сентябрь** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Академический хребет | 53 55.6927  107, 99.2494 | 1, 5, 10, 15, 20, 25 | 15.03 |  | 1.06 | 30.07 |  | 24.09 |
| 2 | п. Листвянка (80 м от берега) | 51 51.8247 104 50.6089' | 0, 2, 3, 5, 7 | 23.03 |  |  | 11.07 |  |  |
| 3 | Варначка-Танхой (50 м от берега, р-н п. Б. Коты, глуб. 10 м) | 51 54.5′  105 06.21 | 0, 2, 3, 5, 7 |  |  |  | 11.07 |  |  |
| 4 | Бухта Песчаная (глубина 10 м.) | 52 15.514  105 42.270 | 0, 2, 3, 5, 7 |  |  |  | 12.07 |  |  |
| 5 | Бухта Ая (глубина 10 м.) | 52 47.332  106 36.445 | 0, 2, 3, 5, 7 |  |  |  | 15.07 |  |  |
| 6 | м. Арал - м. Хобой, глуб. 330 м | 53 26.563  107 40.447 | 1, 5, 10, 15, 20, 25 |  |  |  |  |  |  |
| 7 | М. Хобой (Сев. оконечность о. Ольхон напротив м. Хобой, 60 м от берега, глубина 10 м) | 53 24.466  107 47.289 | 0, 2, 3, 5, 7 | 12.03-13.03 |  |  | 15.07 |  |  |
| 8 | Мыс Зундук (м. Зундук – м. Будун, глубина 200 м) | 53 21.456  107 24.580 | 1, 5, 10, 15, 20, 25 | 12.03 |  |  | 14.07 |  |  |
| 9 | Прол. Ольхонские ворота, 100 м от берега, глуб. 10 м | 53 01.034  106 55.470 | 0, 2, 3, 5, 7 | 12.03 |  |  | 13.07 |  |  |
| 10 | Бухта Базарная (напротив п. Сахюрта, глубина 10 м) | 53 01.245  106 53.410 | 0, 2, 3, 5, 7 |  |  |  | 13.07 |  |  |
| 11 | Зал. Куркутский (300 м от берега, глубина 5 м) | 53 02.082  106 51.008 | 0, 2, 3 |  |  |  | 13.07 |  |  |
| 12 | Зал. Мухур (м. Шида, глуб. 6 м) | 53 03.274  106 48.160 | 0, 2, 3 |  |  |  | 13.07 |  |  |
| 13 | Зал. Хужирскиий (п. Хужир, глуб. 10 - 12 м) | 53 11.758  107 19.075 | 0, 2, 3, 5, 7 |  |  |  | 15.07 |  |  |
| 14 | Зал. Нюрганская губа (п. Песчанка, глуб. 10 - 12 м) |  | 0, 2, 3, 5, 7 |  |  |  | 14.07 |  |  |
| 15 | Мыс Уюга – мыс Цаган-Хушун (Курма напротив о. Ольтрек, глуб. 10 м) | 53 09.680  106 58.529 | 0, 2, 3, 5, 7 |  |  |  | 14.07 |  |  |
| 16 | Листвянка-Танхой | 51 48.021  104 59.066 | 1, 5, 10, 15, 20, 25 | 23.03 | 27.05 |  |  | 1.08 | 27.09 |
| 17 | 12 км от п. Култука | 51 40.578  103 52.309 | 1, 5, 10, 15, 20, 25 |  | 26.05 |  |  | 3.08 | 17.09 |
| 18 | Маритуй-Солзан | 51 38.710  104 13.715 | 1, 5, 10, 15, 20, 25 |  | 26.05 |  |  | 2.08 | 18.09 |
| 19 | Толстый-Снежная | 51 36.402  104 44.147 | 1, 5, 10, 15, 20, 25 |  | 27.05 |  |  | 3.08 | 18.09 |
| 20 | Кадильный-Мишиха | 51 46.731  105 22.528 | 1, 5, 10, 15, 20, 25 |  | 28.05 |  |  | 1.08 | 27.09 |
| 21 | Красный Яр- Харауз | 52 20.722  106 03.870 | 1, 5, 10, 15, 20, 25 |  | 29.05 |  | 31.07 |  | 26.09 |
| 22 | Анга-Сухая | 52 39.590  106 50.978 | 1, 5, 10, 15, 20, 25 |  | 29.05 |  | 25.07 |  | 26.09 |
| 23 | Ухан-Тонкий | 52 53.630  107 31.001 | 1, 5, 10, 15, 20, 25 |  | 30.05 |  | 26.07 |  | 25.09 |
| 24 | Хобой-Крестовый | 53 21.278  108 13.078 | 1, 5, 10, 15, 20, 25 |  | 31.05 |  | 26.07 |  | 24.09 |
| 25 | Заворотный-Сосновка | 54 16.860  108 44.473 | 1, 5, 10, 15, 20, 25 |  |  | 1.06 | 27.07 |  | 23.09 |
| 26 | Елохин-Давша | 54 27.052  109 04.164 | 1, 5, 10, 15, 20, 25 |  |  | 1.06 | 27.07 |  | 23.09 |
| 27 | Котельниковский-Амнундакан | 55 02.388  109 25.939 | 1, 5, 10, 15, 20, 25 |  |  | 3.06 | 28.07 |  | 22.09 |
| 28 | Байкальское-Турали | 55 19.487  109 28.707 | 1, 5, 10, 15, 20, 25 |  |  | 3.06 | 29.07 |  | 22.09 |
| 29 | Тыя-Немнянка | 55 33.968  109 35.597 | 1, 5, 10, 15, 20, 25 |  |  | 3.06 | 28.07 |  | 21.09 |
| 30 | 7 км от Нижнеангарска | 55 42.876  109 36.573 | 1, 5, 10, 15, 20, 25 |  |  | 2.06 | 28.07 |  | 21.09 |
| 31 | Малое Море центр | 53 14.500  107 15.416 | 1, 5, 10, 15, 20, 25 |  |  | 5.06 | 30.07 |  |  |

Продолжение таблицы 1

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **№** | **Станция** | **Координаты** | **Глубина , м** | **март** | **май** | **июнь** | **июль** | **август** | **сентябрь** |
| 32 | Малое Море выход (Хобой-Арал) | 53 26.563  107 40.447 | 1, 5, 10, 15, 20, 25 |  |  | 4.06 | 30.07 |  | 24.09 |
| 33 | Баргузинский залив | 53 27.245  108 44.387 | 1, 5, 10, 15, 20, 25 |  | 31.05 |  |  |  | 24.09 |
| 34 | Чивыркуйский залив | 53 42.564  109 06.384 | 1, 5, 10, 15, 20, 25 |  |  | 4.06 |  |  | 23.09 |

**4.2** **Микробиологический анализ**

Одноклеточные водоросли и бактерии важнейшие участники биогеохимических циклов в большинстве водных экосистем. Взаимодействия между первичными продуцентами и бактериями влияют на физиологию этих организмов, приводят к изменению условий окружающей их среды и участвуют в формировании разнообразия экосистемы [12]. Растворенные органические вещества, продуцируемые микроводорослями, усваиваются в основном гетеротрофными бактериями, поэтому изменение количественных и качественных характеристик фитопланктона влияет на численность и видовой состав бактерий [13].

Для определения общей численности бактерий (ОЧБ) использовали световую эпифлуоресцентную микроскопию. Для этого 5 мл интегральной пробы воды (с глубин 0 м, 5 м, 10 м, 15 м, 20 м, 25 м) фиксировали 1 % р-ром глутарового альдегида и окрашивали флуорохромным красителем ДАФИ (4,6-диамино-2-фенилиндол) (Sigma-Aldrich, США) в конечной концентрации 1 мкг/мл в течение 2-3 мин (Porter and Feig, 1980). Затем окрашенную пробу осаждали на поликарбонатный фильтр (Millipore, Ирландия) с диаметром пор 0,2 мкм с помощью фильтровальной установки (Sartorius, Германия). Фильтры промывали стерильной водой, помещали на предметное стекло в каплю иммерсионного нефлуоресцирующего масла (МиниМед, Росссия), выдерживали в темноте 20 мин. Полученные препараты анализировали с помощью инвертированного микроскопа Axiovert 200 (Zeiss, Германия) с ультрафиолетовой лампой HBO 50W/AC ASRAM при спектре возбуждения 365 нм. Для подсчета бактериальных клеток просматривали не менее 20 полей зрения на образец, вычисление ОЧБ проводили, как описано ранее [14]. Микрофотографии получали камерой PIXERA Penguin 600CL с программным обеспечением AXIOSET.

В результате исследований показано, что максимальная численность бактерий в период подледного цветения фитопланктона составляет 1.0 ×107 кл./мл в подледных обрастаниях на станции мыс Хобой (район Малое Море) (таблица 2). В подледном слое воды ОЧБ варьирует от 0.5 ×106 кл./мл до1.6×106 кл./мл. В период май-июнь максимальная численность 3.5×106 кл./мл зафиксирована в Чивыркуйском заливе, ОЧБ на разных станциях в фотическом слое воды варьировала от 0.5×106 кл./мл до 2.2×106 кл./мл (таблица 2).

Для культивирования органотрофных бактерий из природных образцов использовали рыбо-пептонный агар, разбавленный в 10 раз (РПА/10). При подборе среды учитывали, что микроорганизмы, развивающиеся в течение цветения фитопланктона, имеют гетеротрофный тип обмена веществ и являются типичными обитателями пресных водоемов с нейтральным рН. Посевы выполняли глубинным способом в 2-х повторностях и культивировали при температуре 25 °С.

Таблица 2 – Общая численность бактерий в подледных обрастаниях и фотическом слое воды из оз. Байкал.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Станция** | **ОЧБ, кл./мл** | | |
| **март** | **май** | **июнь** |
| 1 | Академический хребет, подледные обрастания | 0.7×106 |  |  |
| 2 | Академический хребет, фотический слой воды | 1.6×106 |  | 3.0×106 |
| 3 | Академический хребет, придон ная вода | 2.2×106 |  |  |
| 4 | Пролив Малое море Ольхонские ворота (100 м от уреза), подледные обрастания | 1.2 ×106 |  |  |
| 5 | Пролив Малое море Ольхонские ворота (100 м от уреза), фотический слой воды | 1.3 ×106 |  |  |
| 6 | п. Листвянка (80 м от берега напротив кафе Алания), фотический слой воды | 1.1×106 |  |  |
| 7 | Мыс Хобой (Северная оконечность о. Ольхон напротив м. Хобой, 60 м от берега, глубина 10 м), подледные обрастания | 1.0 ×107 |  |  |
| 8 | Мыс Хобой (Северная оконечность о. Ольхон напротив м. Хобой, 60 м от берега, глубина 10 м), фотический слой воды | 0.5 ×106 |  |  |
| 9 | Мыс Зундук (м. Зундук – м. Будун, глубина 200 м), фотический слой воды | 1.6×106 |  |  |
| 10 | Листвянка-Танхой, фотический слой воды | 1.2×106 | 0.9×106 |  |
| 11 | 12 км от Култука, фотический слой воды |  | 2.2×106 |  |
| 12 | Маритуй-Солзан, фотический слой воды |  | 1.7×106 |  |
| 13 | Толстый-Снежная, фотический слой воды |  | 0.5×106 |  |
| 14 | Кадильный-Мишиха, фотический слой воды |  | 1.4×106 |  |
| 15 | Красный Яр- Харауз, фотический слой воды |  | 0.9×106 |  |
| 16 | Анга-Сухая, фотический слой воды |  | 1.4×106 |  |
| 17 | Ухан-Тонкий, фотический слой воды |  | 1.0×106 |  |
| 18 | Хобой-Крестовый, фотический слой воды |  | 1.0×106 |  |
| 19 | Заворотный-Сосновка, фотический слой воды |  |  | 1.2×106 |
| 20 | Елохин-Давша, фотический слой воды |  |  | 1.4×106 |
| 21 | Тыя-Немнянка, фотический слой воды |  |  | 0.7×106 |
| 22 | Малое Море центр, фотический слой воды |  |  | 1.6×106 |
| 23 | Малое Море выход (Хобой-Арал), фотический слой воды |  |  | 1.4×106 |
| 24 | Баргузинский залив, фотический слой воды |  | 1.2×106 |  |
| 25 | Чивыркуйский залив, фотический слой воды |  |  | 3.5×106 |

В результате работы было определено количество органотрофных бактерий (КОЕ/мл), культивируемых на среде РПА/10 в верхнем слое 0-25 м воды на 33 станциях в период с мая по сентябрь (рисунок 20). В целом, выявлено, что численность органотрофных бактерий в период цветения весеннего фитопланктона в пелагической зоне озера выше чем в летний и осенний сезоны. Наименьшее количество органотрофов выявлено в литоральной зоне Южного и Среднего Байкала в июле (0-27 КОЕ/мл).

**4.3 Микроскопический анализ состава фитопланктона**

Для последующего сравнительного анализа таксономического разнообразия микроэукариот, полученных данными NGS, была проведена оценка численности и биомассы фитопланктона в отобранных в 2018 г. пробах (таблица 1). Качественные пробы, отбиранные большой сетью Джеди, фиксировали раствором Люголя с ацетатом натрия по общепринятой методике [15]. Для анализа видовой принадлежности микроводорослей собранный батометром материал сразу осаждали на фильтры (“Whatman”, США) с диаметром пор 1 мкм. После высушивания на воздухе фильтры приклеивали на столики для сканирующей электронной микроскопии (СЭМ), затем в лабораторных условиях напыляли золотом и анализировали с помощью сканирующего электронного микроскопа Philips SEM 525M (Голландия). Для количественного учета фитопланктона пробы концентрировали с помощью осадочного метода [16, 17], подсчет клеток микроводорослей проводили по методу Гензена [16] на разграфленном тонком предметном стекле в капле, взятой штемпель-пипеткой объемом 0,1 мл в двух повторностях с помощью светового микроскопа «Axiovert 200» ZEISS (Германия) c фотокамерой Pixera Penguin 600CL. Биомассу клеток определяли расчетным способом [18, 19].

Проведен микроскопический анализ пролива Малое Море и нескольких станций литоральной зоны. В весенний подледный период общая численность микроводорослей была ниже, чем в летний период (рис. 21), обнаружены диатомовые, хризофитовые водоросли и их стоматоцисты (рис. 22). Максимальные значения общей численности и биомассы отмечены на станции пролив Ольхонские Ворота, где доминирующим по численности видом диатомовая водоросль *Fragilaria radians* (Kützing) D.M.Williams & Round, по биомассе доминировал более крупноклеточный вид динофлагеллят *Gymnodinium baicalense* N.L.Antipova. На станциях мыс Зундук, мыс Хобой и Академический хребет наибольшей численности достигала зеленая водоросль *Monoraphidium griffithii* (Berkeley) Komárková-Legnerová. По биомассе на станциях мыс Зундук и Академический хребет доминировали диатомеи *Aulacoseira baicalensis* (Wislouch) Simonsen; на станции мыс Хобой – динофлагеляты *G. baicalense*. На пелагической станции Листвянка-Танхой (центр) по численности доминировали гаптофитовые *Chrysochromulina parva* Lackey, содоминирующими были зеленые водоросли *M. griffithii*, *Koliella longiseta* (Vischer) Hindak и диатомеи *Nitzschia graciliformis.* По биомассе доминировали динофлагелляты *Gymnodinium baicalense* и диатомеи *F. radians*. В литоральной зоне на станции п. Листвянка по численности доминировали диатомеи *Nitzschia graciliformis* Lange-Bertalot et Simonsen emend. Genkal et Popovskaya., по биомассе – динофлагелляты *Peridinium baicalense* Kiselev & V. Zvetkov (рисунок 23).

По данным микроскопического анализа выявлено, что весенний фитопланктон 2018 г., в основном, был представлен диатомовыми, хризофитовыми и зелеными водорослями (рисунок 24, 25). Основу биомассы и численности в Южном и Среднем Байкале стабильно составляла диатомовая водоросль *F. radians* (=*Synedra acus* subsp. *radians* (Kützing) Skabitchevsky), на долю которой приходилось 25-87% от общего количества фитопланктона.

В Северной части озера преобладающими, как по численности, так и по биомассе, были зеленые водоросли *K. longiseta* и представители рода *Monoraphidium*, на долю, которой приходилось до 50% от общего количества фитопланктона.На некоторых станциях были зафиксированы максимальные показатели диатомовой *N. graciliformis*,которая составляла от 50% до 65% от общего количества фитопланктона.

В июле количественные показатели микроводорослей варьировали в широком диапазоне. Общая численность микроводорослей изменялась в пределах от 45,21 до 1211,15 тыс. кл./л., а биомасса – от 0,038 до 0,72 г/м3 (рисунок 21). По данным СЭМ показано, что в пробах, отобранных в июле, были обнаружены диатомовые, динофитовые, сине-зеленые, хризофитовые водоросли и их стоматоцисты (рисунок 26). В пробах отмечено большое количество бентосных видов диатомовых водорослей, которые не составляли высокую численность, но вносили разнообразие в сообщества. Максимальное значение общей численности было отмечено, также как и в марте, на станции Ольхонские Ворота, где доминировал вид *M. Griffithii* (рисунок 27). Наибольшая биомасса была на станции мыс Зундук, где доминировали диатомеи *F. radians*. На станциях п. Листвянка и Варначка-Танхой по численности доминировал вид *M. griffithii*, по биомассе хризофитовые *Dinobryon cylindricum* O.E. Imhof. Наименее продуктивной была станция Варначка-Танхой (45,21 тыс. кл./л., 0,038 г/м3). Кроме того, зеленая водоросль *M. griffithii* по численности преобладала на станциях бухта Базарная, Куркутский залив и бухта Песчаная. В Мухорском заливе по численности доминировала сине-зеленая водоросль *Anabaena lemmermannii* P.G. Richter, по биомассе динофлагелляты *Cerathium hirundiella* (O.F. Müller) Dujardin.

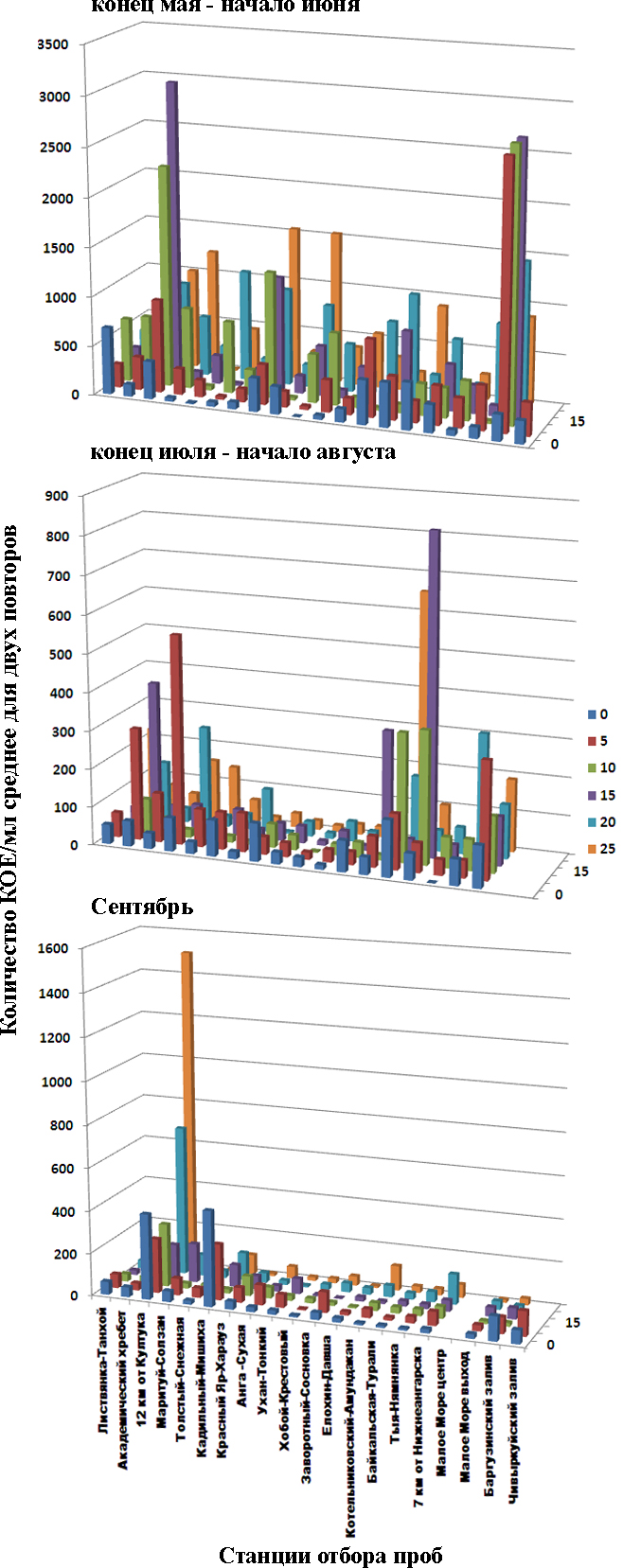


Рисунок 20 – Количество гетеротрофных бактерий (КОЕ/мл), культивируемых на среде РПА/10, 17 станций оз. Байкал в мае-сентябре 2018 г.

Рисунок 21 – Общая численность и биомасса микроводорослей пролива Малое Море, а также нескольких станций литоральной зоны южной и средней котловин озера Байкал в марте и июле 2018 г.

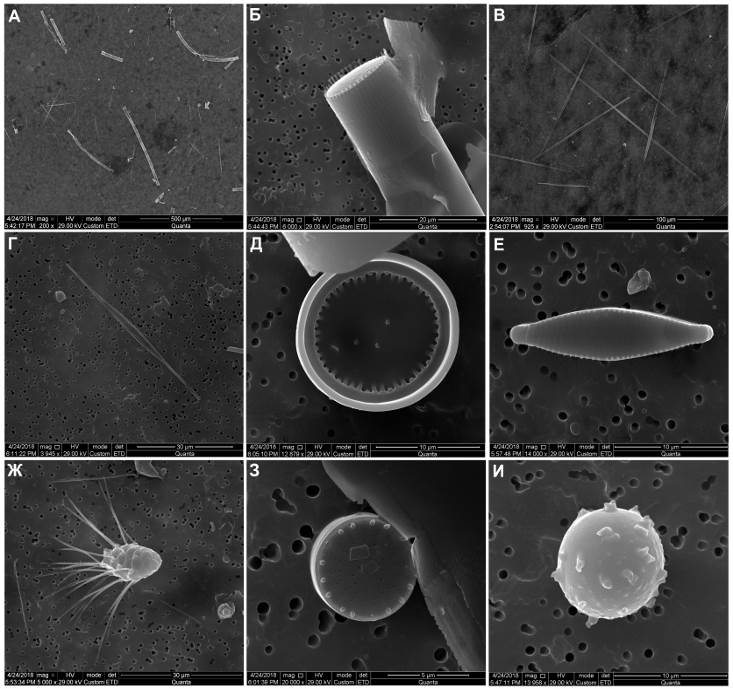


Рисунок 22 – Микроводоросли (СЭМ). Общий вид (А), диатомея *Aulacoseira islandica* (Б); *Fragilaria radians* (В); *Nitzschia graciliformis* (Г); *Cyclotella minuta* (Д); *Fragilaria* sp. (Е); хризофитовая *Mallomonas* sp. (Ж); центричская диатомея (З); стоматоциста (И)

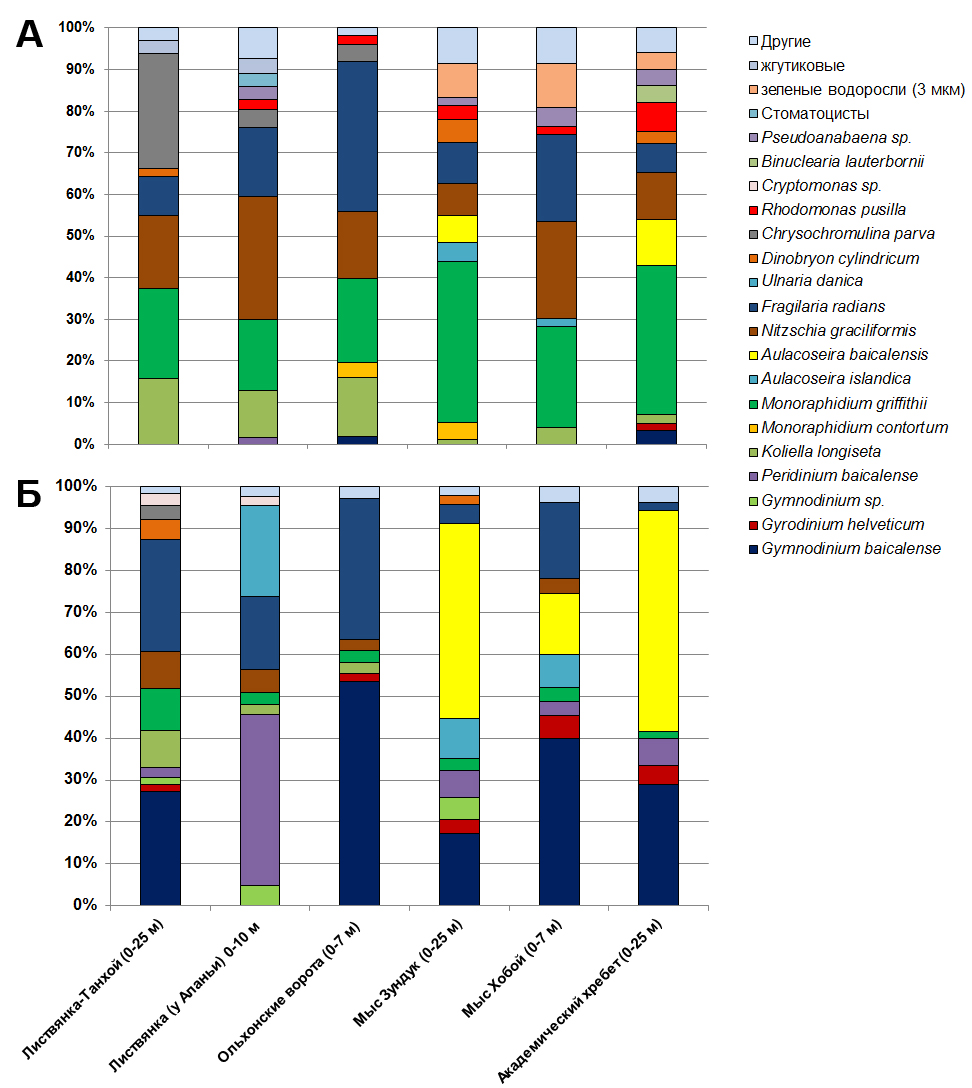


Рисунок 23 – Процентное соотношение видов микроводорослей пролива Малое Море и нескольких станций литоральной зоны южной и средней котловин озера Байкал в марте. По численности (А) и по биомассе (Б)

На станции мыс Зундук по численности – зеленая водоросль *Monoraphidium contortum* (Thuret) Komárková-Legnerová, а по биомассе – *F. radians*. На станции мыс Хобой по численности преобладали *Chrysochromulina parva* Lackey, по биомассе – *Gymnodinium* sp (рисунок 27).

Рисунок 24 – Относительное содержание видов весеннего фитопланктона пелагиали Байкала по биомассе в 2018 г.

Рисунок 25 – Относительное содержание численности видов весеннего фитопланктона пелагиали Байкала в 2018 г.

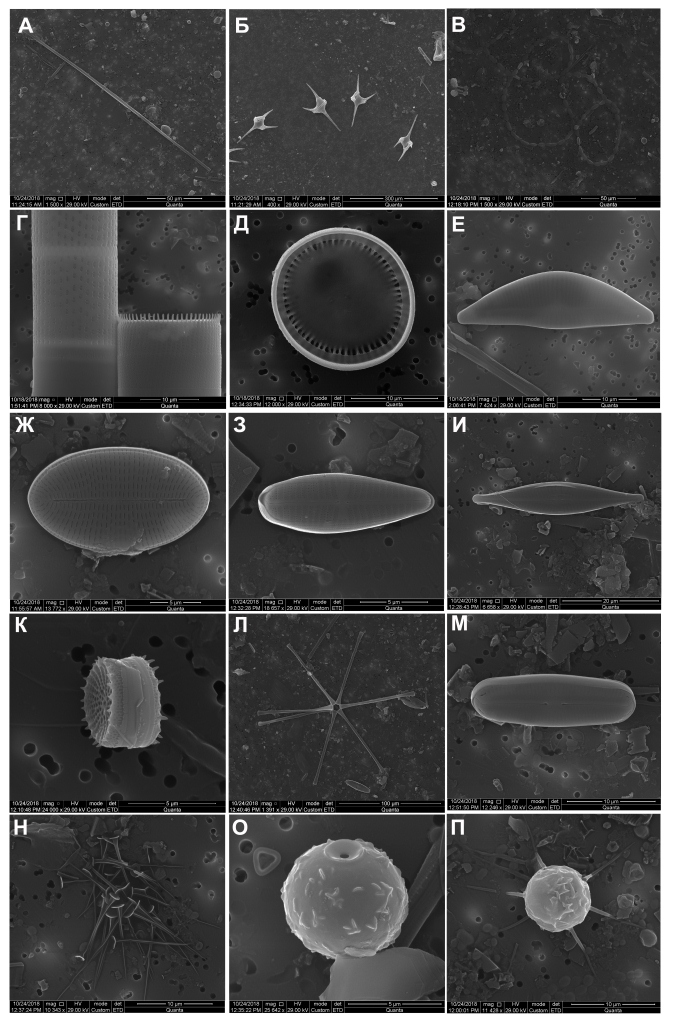


Рисунок 26 – Микроводоросли пролива Малое Море и нескольких станций литоральной зоны южной и средней котловин озера Байкал в июле 2018 года (СЭМ). *Fragilaria radians* (А); *Cerathium hirundiella* (Б); *Anabaena* sp. (В); диатомеи *Aulacoseira baicalensis* и *A. islandica* (Г); *Cyclotella minuta* (Д); *Encyonema* sp. (Е); *Cocconeise* sp. (Ж); *Gomphonema* sp. (З); *Navicula* sp. (И); *Stephanodiscus* sp. (К); *Asterionella formosa* (Л); *Achnanthidium* sp. (М); хризофитовые (Н, П); стоматоциста (О)

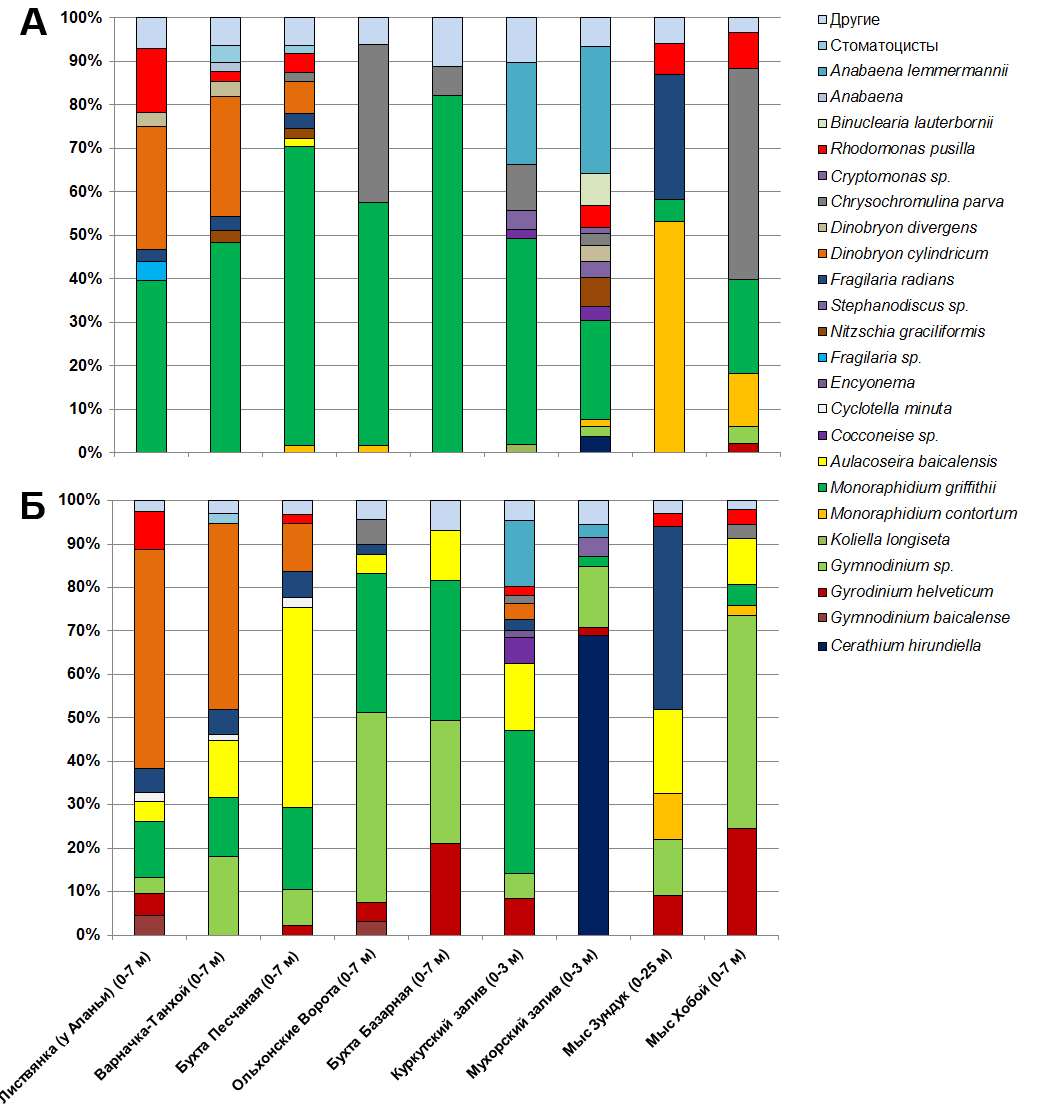


Рисунок 27 – Процентное соотношение видов микроводорослей пролива Малое Море и нескольких станций литоральной зоны южной и средней котловин озера Байкал в июле. По численности (А) и по биомассе (Б)

**4.4** **Оптимизация метода выделения суммарной ДНК из проб водной толщи для секвенирования библиотек ампликонов фрагментов маркерных генов 16S (прокариоты) и 18S (эукариоты) рРНК**

Оптимизирован метод выделения ДНК из природных образцов. Преимуществом метода является включение дополнительного этапа с непосредственной обработкой фильтра из аналитической трековой мембраны, который использовался для концентрации биомассы из проб объемом 20 л. При использовании данного метода получены препараты ДНК из проб, отобранных из различных экотопов оз. Байкал в период весна-лето 2018 г.

При анализе данных, полученных по результатам секвенирования библиотек ампликонов фрагментов генов 16S и 18S рРНК из различных образцов воды из озера Байкал 2017 г. с использованием высокопроизводительной платформы Illumina MiSeq было показано, что имеются различия в уровне вариабельности результатов при использовании различных фильтров при фильтрации проб воды. При получении ДНК из водных образцов фильтрование бактерий осуществляли с помощью поликарбонатного и нитроцеллюлозного фильтров с диаметром пор 0,22 мкм. Показано, что при использовании поликарбонатных фильтров для фильтрации бактерий из водных природных проб результаты имеют более высокую воспроизводимость. Нами выдвинуто предположение, что причиной расхождение результатов может быть плотное связывание биомассы с фильтрами и, как следствие, исключение ее из дальней процедуры выделения ДНК. Поэтому в рамках выполнения работ по проекту в этом году была поставлена задача доработать метод выделения ДНК с тем, чтобы максимально избежать предполагаемых потерь.

В ходе выполнения работ был оптимизирован метод экстракции ДНК, разработанный нами ранее и основанный на ферментативной обработке клеток (отчет по проекты «Фундаментальные исследования и прорывные технологии как основа опережающего развития Байкальского региона», направление 4, часть 4.1). Преимуществом предложенного метода является включение дополнительного этапа с непосредственной обработкой фильтра из аналитической трековой мембраны (аналог поликарбонатных фильтров), который использовался для концентрации биомассы из проб объемом 20 л. Применение такого подхода значительно увеличивает количество выделяемой ДНК и позволяет избежать потери биомассы, которая остается необратимо связана с фильтром. С использованием данного метода получены 37 препарата ДНК из проб, отобранных из различных экотопов Байкала в период весна-осень 2018 г.

Биомассу из проб (V=20 л) осаждали на фильтры из аналитической трековой мембраны (Реатрек, Россия) с диаметром пор 0,2 мкм с помощью аппарата фильтрационного АФ-142 с компрессором и напорной емкостью из нержавеющей стали (Владисарт, Россия), затем фильтры погружали в контейнер со стерильным ТЕ-буфер (30 мл) и замораживали. Образцы хранили при -80 °С до дальнейшего использовании. Перед выделением ДНК образцы плавно размораживали при 4 °С. Фильтры измельчали в стерильных условиях, добавляли к ним биомассу, осажденную из ТЕ-буфера центрифугированием в течении 30 мин при 16100 об./мин, 4 °С. К измельченным фильтра доливали стерильного ТЕ-буфер, добавляли лизоцим до концентрации 2,5 мг/мл и инкубировали при 37 °С, 60 мин. Затем к образцам добавляли раствор SDS до конечной концентрации 1 %, дважды замораживали при 20 °С и оттаивали при 55 °С. После добавления раствора протиназы К до конечной концентрации 182 мкг/мл инкубировали образцы с периодическим перемешиванием в течение 60 мин при 55 °С. Процедуру инкубации и добавление протеиназы К повторяли дважды. Лизат отбирали и последовательно экстрагировали фенолом, насыщенным ТЕ-буфером рН 8.0, смесью фенол : хлороформ : изоамиловый спирт в соотношении 25:24:1 и смесью хлороформ : изоамиловый спирт в соотношении 24:1. ДНК из раствора осаждали добавлением 1/10 от объема раствора 3 М ацетата натрия рН 5.2 и двух объемов этилового спирта, выдерживали ночь при -20 °С и цинтрифугировали 30 мин при 16100 об./мин, 10 °С. Осадок ДНК дважды промывали 70 % этанолом, подсушивали на воздухе и растворяли в 200 мкл стерильного ТЕ-буфера (таблица 3).

Таблица 3 – Характеристика препаратов ДНК.

| № | Шифр | Описание образца | Тип образца | Объем пробы | Концентра-ция ДНК (нг/мкл) | Растворитель |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | BD25 | Пр. Ольхонские Ворота п/лед 12.03.18 | Геномная ДНК | 200 | 76,5 | ТЕ |
| 2 | BD26 | Пр. Ольхонские Ворота инт. 0-7 м 12.03.18 | Геномная ДНК | 200 | 190,0 | ТЕ |
| 3 | BD27 | О. Ольхон м. Хобой п/лед 2 л 13.03.18 | Геномная ДНК | 200 | 640,0 | ТЕ |
| 4 | BD28 | Малое Море мыс Хобой инт. 0-7 м, 20 л, 13.03.18 | Геномная ДНК | 200 | 141,5 | ТЕ |
| 5 | BD29 | Малое Море мыс Зундук инт. 0-25 м | Геномная ДНК | 200 | 181,5 | ТЕ |
| 6 | BD30 | Академический хребет п/лед, 3 л, 15.03.18 | Геномная ДНК | 200 | 77,0 | ТЕ |
| 7 | BD31 | Академический хребет инт. 0-25б 20 л, 15.03.18 | Геномная ДНК | 200 | 178,5 | ТЕ |
| 8 | BD32 | Академический хребет придон 1л 15.03.18 | Геномная ДНК | 200 | 665,0 | ТЕ |
| 9 | BD33 | Листвянка-Танхой инт. 0-25, 20 л, 23.03.18 | Геномная ДНК | 200 | 290,0 | ТЕ |
| 10 | BD34 | п. Листвянка, станция Алания инт. 0-5 м, 23.03.18 | Геномная ДНК | 200 | 82,5 | ТЕ |
| 11 | BD35 | Кругобайкальская эксп. май-июнь 2018 Листвянка-Танхой, 20 л инт. 0-25 м | Геномная ДНК | 200 | 25,5 | ТЕ |
| 12 | BD36 | Кругобайкальская эксп. май-июнь 2018 пр Малое Море центр, 20 л инт. 0-25 м. | Геномная ДНК | 100 | 515,0 | ТЕ |
| 13 | BD37 | Кругобайкальская эксп. май-июнь 2018 пролив Малое Море вход, 20 л. инт. 0-25 м. | Геномная ДНК | 200 | 475,0 | ТЕ |

Продолжение таблицы 3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № | Шифр | Описание образца | Тип образца | Объем пробы | Концентра-ция ДНК (нг/мкл) | Растворитель |
| 14 | BD38 | Пос. Листвянка ст. Алания 0-7 м, 20 л 11.07.18 | Геномная ДНК | 200 | 71,5 | ТЕ |
| 15 | BD39 | Варначка-Танхой инт. 0-7 м, 20 л, 11.07.18 | Геномная ДНК | 200 | 45,0 | ТЕ |
| 16 | BD40 | Бух. Песчанная инт. 0-7м, 20 л, 12.07.17 | Геномная ДНК | 200 | 40,5 | ТЕ |
| 17 | BD41 | бухта Ая, инт 0-7, 20 л, 15.07.18 | Геномная ДНК | 200 | 480,0 | ТЕ |
| 18 | BD42 | Малое Море мыс Хобой инт. 0-7, 20 л, 15.07.18 | Геномная ДНК | 200 | 570,0 | ТЕ |
| 19 | BD43 | Пр. Ольхонские Ворота инт. 0-7м, 20 л., 13.07.18 | Геномная ДНК | 200 | 490,0 | ТЕ |
| 20 | BD44 | Бухта Базарная (прол. Сахюрта, глуб. 10 м) 13.07.18 | Геномная ДНК | 200 | 400,0 | ТЕ |
| 21 | BD45 | Зал. Мухор (м. Шида, глубина 6 м), инт. 0-2 м, 20 л, 13.07.18 | Геномная ДНК | 200 | 285,0 | ТЕ |
| 22 | BD46 | Зал. Куркутский (напротив п. Куркут, г. 5 м) инт 0-2м, 20л., 13.07.18 | Геномная ДНК | 200 | 350,0 | ТЕ |
| 23 | BD47 | Зал. Хужирский (п. Хужир, глубина 10-12м), инт 0-7 м, 20 л, 14.07.18 | Геномная ДНК | 200 | 485,0 | ТЕ |
| 24 | BD48 | Мыс Уюга - мыс Цаган-Хужун (Курма, пролив о. Ольтрек, глуб. 10 м) инт. 0-7 м, 20 л, 14.07.18 | Геномная ДНК | 200 | 470,0 | ТЕ |
| 25 | BD49 | Зал. Нюрганская Губа (п. Песчанка, глуб. 10-12 м) 14.07.18 | Геномная ДНК | 200 | 385,0 | ТЕ |
| 26 | BD50 | Листвянка-Танхой инт. 0-25 м, 20 л, летняя КБ 2018 | Геномная ДНК | 200 | 725,0 | ТЕ |
| 27 | BD51 | Прол. Малое Море центр инт. 0-25 м, 20 л, летняя КБ 2018 | Геномная ДНК | 200 | 366,0 | ТЕ |
| 28 | BD52 | Прол. Малое Море вход инт. 0-25 м, 20 л, летняя КБ 2018 | Геномная ДНК | 200 | 430,0 | ТЕ |
| 29 | BD53 | Прол. Малое Море вход инт. 0-25 м, 20 л, осень КБ 2018 | Геномная ДНК | 200 | 403,0 | ТЕ |
| 30 | BD54 | Листвянка-Танхой инт. 0-25 м, 20 л, осенняя КБ 2018 | Геномная ДНК | 200 | 390,0 | ТЕ |

**4.5 Секвенирование**

Для секвенирования библиотек ампликонов фрагментов генов 16S и 18S рРНК на платформе Illumina MiSeq с целью анализа состава микробного сообщества оз. Байкал 28.11.2018 г. в ООО “Генотек” (Москва) в рамках исполнения Контракта № 0334100021718000037-0009343-01 от 26.10.2018 г. было направлено 76 образцов геномной ДНК, выделенной из водной толщи и донных осадков.

В ходе выполнения работ, будет получено 105 библиотек ампликонов, из них 36 библиотек – фрагментов генов V3-V4 16S рРНК эубактерий; 17 библиотеки - фрагментов генов V2-V3 16S рРНК архей; 52 библиотеки - фрагментов генов V8-V9 18S рРНК эукариот. В настоящее время в рамках исполнения Контракта ООО “Генотек” ведется подготовительный этап работ.

**5 Реконструкция полных геномов выделенных в культуры перспективных для практического применения штаммов бактерий**

Выделены и охарактеризованы анаэробные термофильные бактерии рода *Geobacillus.* Штаммы выделены из донных осадков разных районов, включая метановый сип Посольская банка, грязевые вулканы Маленький и Кукуйского каньона во время экспедиции в июле 2013-2015 гг. Пробы осадка отбирали c соблюдением стерильности. Посев производился на питательные среды и помещался в термостат при Т +65 ºС. Принадлежность штаммов к роду *Geobacillus* установлена на основе фенотипических, физиолого-биохимических характеристик и структуры гена 16S рРНК. В таблице 4 даны координаты отобранных станций, их глубины, характеристика осадка, а также типы пробоотборников.

Таблица 4 – Характеристика станций пробоотбора

|  |  |  |  |
| --- | --- | --- | --- |
| станция/ глубина водной толщи, м | Пробоотборник | Координаты | Характеристика осадка в слое, где были выделены штаммы |
| 2013 грязевой вулкан «К-2», Кукуйский каньон, 940 м, Центральная котловина | \*GC-3 | 52,589683°N 106,770733°E | 40 см восстановленный диатомовый алеврито-пелитовый ил, серого цвета, газонасыщенный |
| 2015 влк «Маленький», 1371 м, Южная котловина | GC-1 | 51,919990N 105,634760E | 15-16 см алеврито-пелитовый диатомовый ил, серого цвета, насыщен гидротроилитом |
| 2015 Посольская банка, 478 м., метановый сип | \*\*BC-7 | 52,03550°N  105,84317°E | 0-5 см алеврито-пелитовый диатомовый ил серого цвета с многочисленными вклю-ченииями гидротроилита. В отложениях присутствуют следы дегазации. |

\* GC- гравитационная трубка

\*\* BC – коробчатый пробоотборник (грейфер)

Условия культивирования:Для получения накопительных культур брали 0,5 г осадка, помещали в 50 мл жидкой среды и тщательно перемешивали, затем из анаэробной болтушки отбирали 3 мл и добавляли в 30 мл среды. Использовали модифицированную среду Видделя для пресноводных форм бактерий [20], которая содержит полный набор витаминов и дрожжевой экстракт (0,5 г/л). В качестве восстановителя добавляли сульфид натрия (2,5 мМ). Этот же состав среды использовали для поддержания накопительных культур бактерий. Культивирование осуществляли в течение 30 суток, при температуре 60 °С. В качестве контроля служили флаконы с незасеянной средой. Изолирование культур *Geobacillus* из осадка вели путем последовательных пересевов с последующим выделением колоний на среде Видделя с добавлением агара – 12 г/ л. Выращивали посевы в термостате при температуре 60 °С. В дальнейшем культуры поддерживали на модифицированной среде Луриа-Бертани (г/ л): KNO3 - 1,0; дрожжевой экстракт - 0,5; триптон или пептон - 1,0; агар - 10,0; гилрит - 4,0; рН = 7,2-7,8)

Морфологиябактерий исследована с помощью эпифлуоресцентного микроскопа “AxioImager M1”(Сarl Zeiss) и сканирующего электронного микроскопа Quanta 200 (FEI Сompany, США) с ускоряющим напряжением в 20 киловольт на базе ЦКП “Ультрамикроанализ”. В первом случае бактерии окрашивали ДАФИ (4,6-диамидино-2-фенилиндол) в рабочей концентрации 10 мкг/мл в течение 5–8 мин при комнатной температуре. Во втором, суспензию клеток фиксировали 4 % раствором параформальдегида в 0,066 М буфере Соренсена при комнатной температуре в течение ночи. Далее клетки отфильтровывали на поликарбонатные фильтры (диаметр пор 0,2 мкм) и проводили обезвоживание в серии растворов этанола (30 %, 50 %, 70 %, 96 %, 100 %) и ацетоне (100 %). Затем проводили сушку в критической точке на приборе CPD 030 Critical Point Dryer (BalzersUnion, Лихтенштейн). Фильтры были помещены на алюминиевые столики и напылены золотом на установке SCD 004 (BalzersUnion, Лихтенштейн).

Анализ последовательностей гена 16S рРНК.Выделение ДНК проводили по модифицированной методике ферментативного лизиса с последующей фенол-хлороформной экстракцией [21]. Амплификацию ДНКпроводили в следующем режиме: 94 °С - 2 мин (1 цикл); денатурация – 94 °С - 45 сек, отжиг – 52 °С - 45 сек, элонгация - 72°С - 60 сек (30 циклов); финальная элонгация – 72 °С - 10 мин (1 цикл). В работе были использованы праймеры, комплементарные наиболее консервативным участкам гена 16S рРНК: 27L-AGAGTTTGATCMTGGCTCAG, 1350R- GACGGGCGGTGTGTACAAG [22, 23]. Секвенирование проводили на геномном анализаторе ABI 3130XL Genetic Analyser («Applied Biosystems») с использованием реактива BigDye Terminator Kit v.3.1 (Applied Biosystems) в ЦКП «Геномика», Новосибирск. Последовательности, полученные в результате секвенирования, сравнивали с последовательностями из международного банка данных NCBI c помощью программы BLASTN (http://www.ncbi.nlm.nih.gov/blast). Структуры были проанализированы с использованием программы ClustalW V 1.4 (http://www.ebi.ac.uk/clustalw). Сравнение последовательностей и построение филогенетических древ осуществляли с помощью пакета программ МЕGA (http:// www.megafile.co) с использованием алгоритмов группирования «Neighbor-Joining», модели «Kimura-2-parameters». Статистическая достоверность ветвления оценивалась с помощью "bootstrap-анализа", с использованием соответствующей функции той же программы.

Морфология клеток исследованных штаммов представлена одиночными длинными прямыми или слегка изогнутыми палочками (рисунок 28). Деление с образованием перетяжек (бинарное деление). Клетки при окраске по Граму – положительные. Организмы образуют споры, имеющие овальную форму. В клетке споры локализованы терминально. Культуры терморезистентные. Споры выдерживают прогрев при 100°С в течение 5 мин. При росте на жидкой среде наблюдается однородное помутнение среды с образованием рыхлого осадка в конце роста. Штамм **PB15** подвижен за счётперитрихиально расположенных жгутиков, а штаммы **М15, G3, G4** неподвижны. На плотной среде (модифицированная среда LB) штаммы образуют колонии округлой формы с ровным краем, лишенные пигмента (**PB15, М15),** либо имеют светло-бежевую пигментацию (**G3, G4**), слизистой консистенции, однородной структуры.

|  |  |  |  |
| --- | --- | --- | --- |
|  | а |  | б |
|  | в |  | г |

Рисунок 28 – Морфология клеток штаммов: а- PB15; b- М15; c- G3; d- G4

Физиологические свойства. Штаммы термофильные, оптимум роста – 55–60 °С. Верхняя граница роста – 80 °С, нижняя – 46 °С. Штаммы пресноводные поэтому максимум роста наблюдали в среде при отсутствии хлорида натрия. Штаммы **PB15, М15** росли в интервале 0-1.5 % NaCl, а **G3, G4-**до 2,5 %. Все штаммы нейтрофильны.Область рН для роста 6.8–8.2, с оптимумом 7.0–7.2. При росте штаммы нуждаются в нитрате (нитрат калия). Каталазоположительные. Оксидазоотрицательные. В аэробных условиях в качестве источника энергии и углерода использует глюкозу, фруктозу, рамнозу, маннозу, мальтозу, сахарозу. Растут на агаре с эскулином, гидролизуют крахмал. Для роста необходимы дрожжевой экстракт и витамины. Не используют ксилозу, арабинозу, лактозу, рафинозу, салицин, сорбит, инозит, дульцит, адонит, цитрат, ацетат, малонат натрия, пируват, сукцинат лактат, мочевину, желатину, метанол, этанол, бутанол, пропанол.

Филогенетический анализ.Для того чтобы уточнить таксономическое положение исследованных штаммов, проведено сравнение последовательностей с известными представителями фенотипических групп термофильных *Geobacillus.* В работе использовали последовательности типовых штаммов (валидно описанных), представителей этого рода: *G. thermoleovorans, G. kaustophilus, G. thermocatenulatus, G. thermodenitrificans, G. stearothermophilus* и *G. thermoglucosidasius,* которые являются одними из 15 официально зарегистрированных видов *Geobacillus* [24], а также последовательности штаммов *Parageobacillus toebii* B-1024, *Geobacillus gargensis* Gaт, выделенных в районах горячих источников Байкальской рифтовой зоны (рис. 29). Родовые названия штаммов приведены с учетом ревизии этого рода [25]. На филогенетическом древе все последовательности ложатся в один кластер, “*Geobacillus thermoleovorans* group”, который объединяет в себе истинных геобацилл. Уровень сходства последовательностей гена 16S рРНК штаммов **PB15, М15, G3** – 99 %, а идентичность штамма **G4** и других последовательностей этого кластера не превышает 97 %. Штаммы из района Посольской банки и Кукуйского каньона ближе к *G. thermoleovorans* а штамм **М15** к *G. kaustophilus*. Из двух штаммов выделена ДНК (штаммы G4 и Pb15) и передана для полногеномного секвенирования.

**6 Структура бактериальных сообществ фотического слоя озера Байкал в весенний, летний и осенний периоды 2017 г. по данным NGS**

Краткий результат главы. C помощью NGS секвенирования фрагментов генов 16S рРНК выявлено богатство (в общем 1166 ОТЕ с 97% сходства) и разнообразие бактериальных сообществ фотического слоя (0-25 м) трех котловин озера Байкал, заливов и пролива весной, летом и осенью.

**6.1 Материалы и методы**

Для исследования бактериальных сообществ в пелагиали озера Байкал были отобраны пробы в южной (Маритуй-Солзан – SMS; Листвянка-Танхой – SLT), средней (Красный Яр – Харауз – MKKh; Ухан-Тонкий – MUT), северной (Елохин-Давша – NED; Байкальское-Турали – NBT) котловинах, Чивыркуйском (ChB) и Баргузинском (BB) заливах и проливе Малое море (MM) в течение трех экспедиций в весенний (27/05-5/06), летний (18/07-26/07) и осенний (2/09-8/09) периоды. Из отобранных проб была выделена суммарная ДНК и затем проведена амплификация региона V3-V4 гена 16S рРНК и последующее высокопроизводительное секвенирование (NGS) этих фрагметнов генов на Illumina. Анализ прочтений, полученных в результате секвенирования, проводили в программе Usearch v. 10 [26]. Парные прочтения региона V3-V4 16S рРНК были объединены и проведены фильтрация прочтений и дальнейший анализ согласно tutorial для Usearch. Количество прочтений в образцах было нормализовано. Таксономию определяли с помощью Sintax в Usearch [27]. Референсной базой данных для Sintax была rdp\_16s\_v16.fa.

**6.2 Результаты**

В результате было получено в среднем 70100 прочтений на образец со средней длиной прочтений 446 нуклеотидов (таблица 5). Общее количество прочтений в 27 образцах составляло 1892698. Количество операционных таксономических единиц (ОТЕ) в разных котловинах и по сезонам варьировало от 318 до 570 (таблица 5). Для всех образцов графики зависимости количества ОТЕ от количества прочтений выходят на плато (рисунок 29). Это показывает, что количество прочтений, полученных в результате секвенирования достаточно для охвата всего богатства бактериальных сообществ. Разнообразие по индексу Shannon\_e среди всех образцов было выше в заливах (BB, ChB). Наименьшее разнообразие бактериальных сообществ по индексу Shannon\_e было в осенний период (2/09-8/09), в сравнении с весной (27/05-5/06) и летом (18/07-26/07) (таблица 5). Индекс Simpson выше у осенних сообществ, что подтверждает их наименьшее разнообразие в сравнении с сообществами других сезонов (таблица 5). Индекс Chao1 был похож на количество ОТЕ (рисунок 29), следовательно, мы охватили все богатство ОТЕ в пробах (таблица 5).

Состав бактериальных сообществ трех котловин и трех сезонов (весна, лето, осень), в основном был представлен Actinobacteria, Bacteroidetes, Cyanobacteria, Betaproteobacteria, Alphaproteobacteria, Gammaproteobacteria, Verrucomicrobia, Planctomycetes, Firmicutes, Acidobacteria (рисунок 30). В весенний период было наибольшее соотношение Actinobacteria, Bacteroidetes, Betaproteobacteria в сообществах, в сравнении с другими сезонами (рисунок 30). От весны к осени увеличивается соотношение Cyanobacteria в сообществах (рисунок 30).

Таким образом, было показано, что в фотическом слое (0-25 м) различных котловин озера Байкал в ходе трех сезонов от весны к осени в бактериальных сообществах снижается соотношение Actinobacteria, Bacteroidetes и Betaproteobacteria и увеличивается соотношение Cyanobacteria.

Кластерный анализ для ОТЕ (97% сходства) бактериальных сообществ, построенный с использованием метрики UniFrac показал, что большинство сообществ формируют один общий кластер (рисунок 31). В пределах большого кластера были подкластеры бактериальных сообществ весны, осени и весны-лета. С сообществами весны кластеризовались некоторые из сообщества лета, и с другими летними сообществами кластеризовались сообщества осени (рисунок 31).

Таблица 5 – Количество прочтений, богатство и разнообразие для ОТЕ (97% сходства) бактерий в пробах фотического слоя (0-25 м) озера Байкал весной, летом и осенью.

| Код образцов | Образцы | Количество прочтений | Количество ОТЕ (97% сходства) | Chao1 | Simpson | Shannon\_e |
| --- | --- | --- | --- | --- | --- | --- |
| BD1 | SMS\_27/05 | 70088 | 414 | 414 | 0.0473 | 3.91 |
| BD27 | SLT\_27/05 | 70120 | 484 | 484 | 0.0539 | 3.91 |
| BD7 | MKKh\_29/05 | 70060 | 402 | 402 | 0.066 | 3.55 |
| BD2 | MUT\_30/05 | 70134 | 409 | 409 | 0.0394 | 4.02 |
| BD3 | NED\_1/06 | 70096 | 386 | 386 | 0.0615 | 3.58 |
| BD4 | NBT\_2/06 | 70120 | 321 | 321 | 0.091 | 3.34 |
| BD8 | BB\_31/05 | 70068 | 431 | 431 | 0.0335 | 4.09 |
| BD6 | ChB\_4/06 | 70077 | 404 | 404 | 0.0238 | 4.27 |

Продолжение таблицы 5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Код образцов | Образцы | Количество прочтений | Количество ОТЕ (97% сходства) | Chao1 | Simpson | Shannon\_e |
| BD5 | MM\_5/06 | 70138 | 402 | 402 | 0.0445 | 3.89 |
| BD9 | SMS\_18/07 | 70081 | 348 | 348 | 0.0695 | 3.42 |
| BD30 | SLT\_19/07 | 70092 | 371 | 371 | 0.0581 | 3.64 |
| BD15 | MKKh\_21/07 | 70116 | 500 | 503.1 | 0.0909 | 3.34 |
| BD10 | MUT\_26/07 | 70091 | 332 | 333.8 | 0.0665 | 3.35 |
| BD11 | NED\_22/07 | 70059 | 394 | 395.8 | 0.0661 | 3.69 |
| BD12 | NBT\_23/09 | 70068 | 367 | 368.3 | 0.113 | 3.32 |
| BD16 | BB\_25/07 | 70147 | 505 | 506 | 0.0651 | 3.65 |
| BD13 | MM\_22/07 | 70137 | 416 | 420 | 0.158 | 2.71 |
| BD14 | ChB\_25/07 | 70119 | 412 | 412.4 | 0.0986 | 3.03 |
| BD17 | SMS\_2/09 | 70127 | 402 | 404.7 | 0.289 | 1.97 |
| BD32 | SLT\_3/09 | 70127 | 318 | 321.5 | 0.316 | 1.81 |
| BD23 | MKKh\_4/09 | 70151 | 433 | 434.1 | 0.166 | 2.64 |
| BD18 | MUT\_11/09 | 70115 | 557 | 560 | 0.179 | 2.79 |
| BD19 | NED\_7/09 | 70126 | 414 | 416 | 0.23 | 2.44 |
| BD20 | NBT\_7/09 | 69978 | 390 | 392.9 | 0.252 | 2.33 |
| BD21 | MM\_5/09 | 70121 | 453 | 453.6 | 0.179 | 2.78 |
| BD22 | ChB\_8/09 | 70064 | 570 | 570.4 | 0.0673 | 3.78 |
| BD24 | BB\_8/09 | 70078 | 462 | 462.9 | 0.054 | 3.69 |

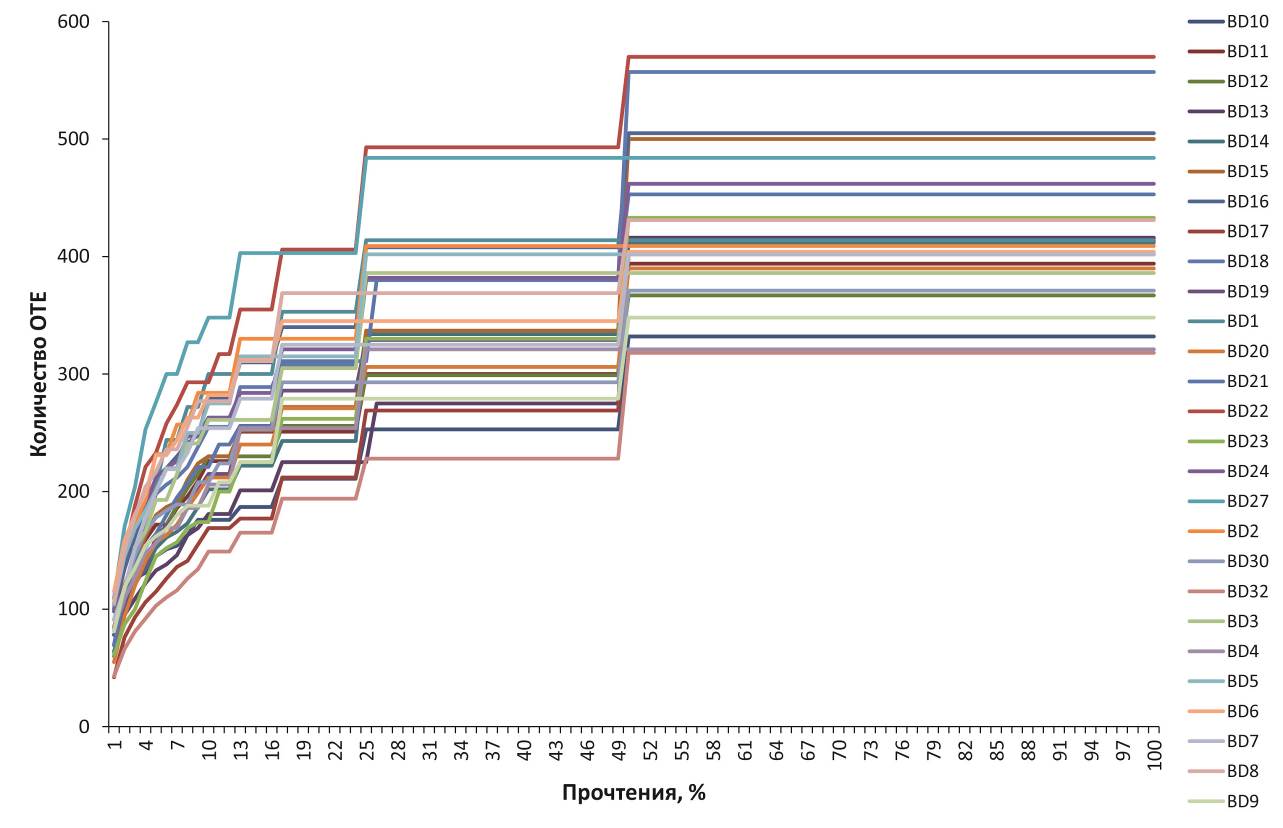


Рисунок 29 – Графики разряжения (rarefaction) прочтений 16S рРНК для 27 образцов  
из фотического слоя озера Байкал 2017 г.

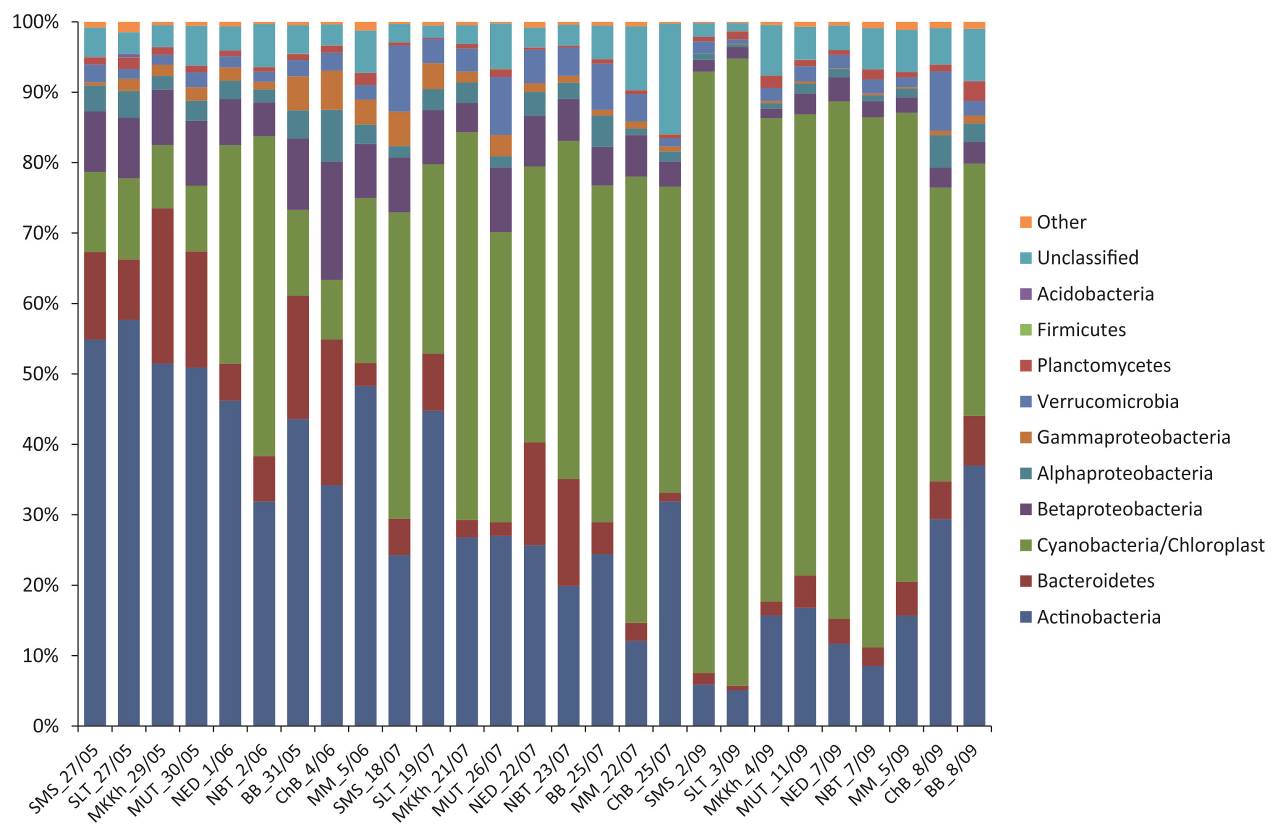


Рисунок 30 –Соотношение прочтений Bacteria в образцах фотического слоя озера Байкал весной (27/05-5/06), летом (18/07-26/07) и осенью (2/09-8/09)

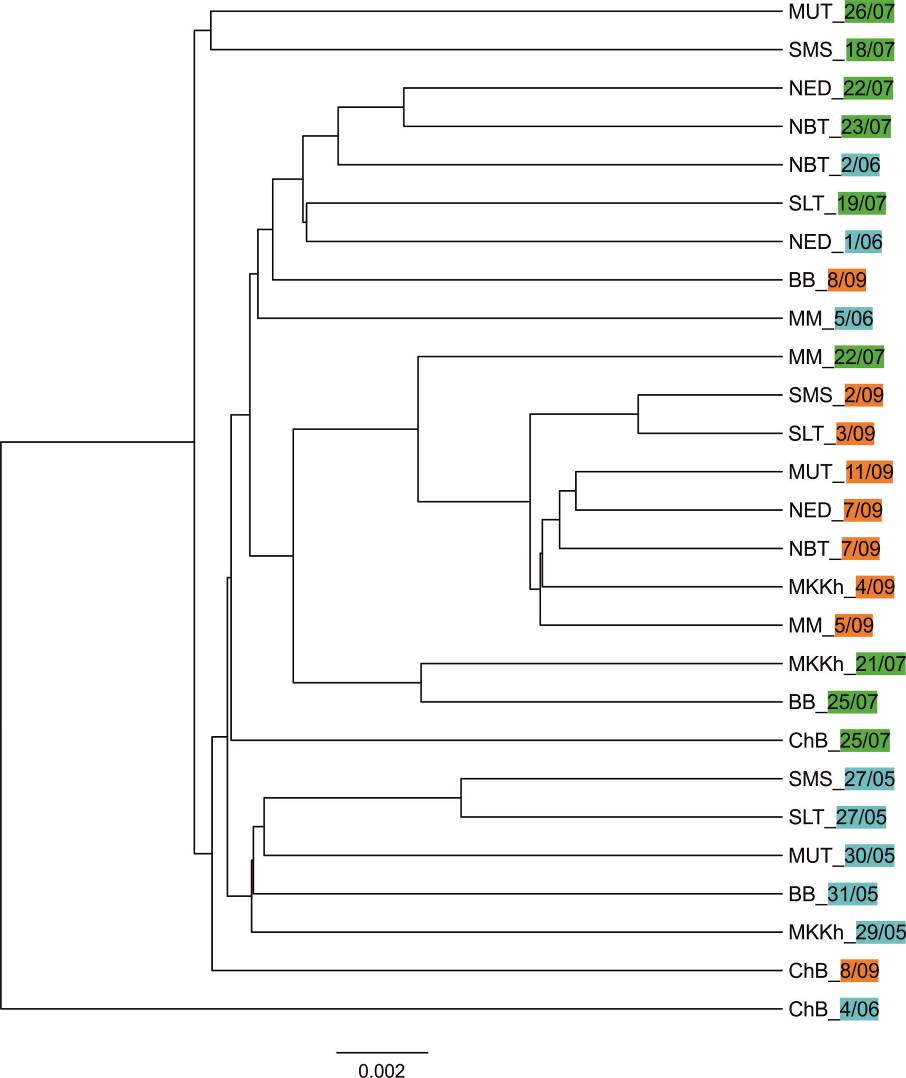


Рисунок 31 – Кластерный анализ для ОТЕ (97% сходства) бактериальных сообществ, построенный с использованием метрики UniFrac. Синим отмечены сообщества весны;  
зеленым – сообщества лета

**7 Сукцессия сообществ бактерий и микроэукариот в южной котловине озера Байкал в 2017 г. по данным NGS**

Краткий результат главы. От весны к осени в бактериальных сообществах снижалась доля представителей Actinobacteria, Bacteroidetes и Betaproteobacteria и увеличивалась доля Cyanobacteria. С марта по сентябрь фотическом слое пелагической станции южной котловины Байкала происходила сезонная смена сообществ микроэукариот. В общем выявлено 424 ОТЕ (97 % сходства) микроэукариот. Весной выявлены наибольшая доля Diatomeae (диатомеи) и паразитов эукариот Cryptomycota. У гетеротрофных Ciliophora не было резкого изменения доли по сезонам. Chrysophyceae имели наибольшую долю весной после вскрытия озера ото льда и осенью.

**7.1 Материалы и методы**

На пелагической станции южной котловины озера Байкал была исследована сукцессия (смена сообществ во времени) сообществ бактерий и микроэукариот. Были отобраны пробы в 8 временных точках с марта по сентябрь. В марте и апреле пробы отбирали со льда, в другие месяцы с НИС «Г.Ю. Верещагин» и НИС «Академик Коптюг». На станции Листвянка-Танхой в 2017 температура воды увеличивалась от весны (0.4 °С) к лету (9.78 °С), затем снижалась осенью (7.09 °С) (таблица 6). Данные температуры в ледовый период были измерены самостоятельно и в дальнейшие сезоны получены сотрудниками лаборатории гидрологии и гидрофизики ЛИН СО РАН. В марте толщина льда составляла 63 см, заснеженность 100% и толщина снега 4 см. В апреле толщина льда составляла 57 см и снег отсутствовал. Концентрации Si, PO43-, P общ н/фильтр, NO3- были выше в период весенней гомотермии 27.05 и 6.06 (таблица 6). Концентрации Si снижались летом (10/07, 19/07) и затем повышались осенью (3/09, 14/09) (таблица 6).

Таблица 6 – Средние значения параметров среды в верхнем 0-25 м слое на станции Листвянка-Танхой с марта по сентябрь 2017 г.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Дата | Время | Т, °С | pH | Si, мг/л | PO43-, мг/л | P общ н/ф-р, mg/L | NO2-, мг/л | NO3-, мгл | NH4+, мг/л |
| 29.03.2017 | 13.30 | 0.4 | 8.09 | 0.37 | 0.013 | 0.009 | 0.005 | 0.23 | 0.013 |
| 12.04.2017 | 13.00 | 0.8 | 8.00 | 0.31 | 0.014 | 0.012 | 0.005 | 0.34 | 0.004 |
| 27.05.2017 | 14.00 | 3.6 | 8.01 | 0.54 | 0.024 | 0.013 | 0.001 | 0.46 | 0.001 |
| 06.06.2017 | 10.30 | 3.7 | 8.00 | 0.49 | 0.026 | 0.013 | 0.001 | 0.54 | 0.000 |
| 10.07.2017 | 14.00 | - | - | 0.14 | 0.015 | - | 0.003 | 0.26 | 0.006 |
| 19.07.2017 | 4.20 | 9.78 | - | 0.19 | 0.014 | 0.007 | 0.002 | 0.19 | 0.002 |
| 3.09.2017 | 7.10 | 7.3 | 8.06 | 0.37 | 0.010 | 0.011 | 0.007 | 0.30 | 0.004 |
| 14.09.2017 | 2.10 | 7.09 | 8.05 | 0.32 | 0.019 | 0.011 | 0.003 | 0.19 | 0.004 |

**7.2 Результаты**

В весенний подледный период (29/03, 12/04) были наибольшие численность и биомасса фитопланктона (529×103 кл./л, 441×103 кл./л, и 568 мг/м3, 466 мг/м3, соответственно) в сравнении с другими сезонами и доминирующими по биомассе были хризофитовая *Dinobryon cylindricum* O.E.Imhof (29/03) и диатомея *Fragilaria radians* (Kützing) D.M.Williams & Round (=*Synedra acus* subsp. *radians* (Kützing) Skabichevskii) (12/04) (рисунок 31). *F. radians* доминировала по биомассе с весны до лета (12/04-19/07) (рисунок 31 B). Центрическая диатомея *Cyclotella baicalensis* Skvortzov and Meyer доминировала по биомассе осенью (14/09). Динофлагелляты *Gymnodinium baicalense* Antipova и *Peridinium baicalense* Kiselev and Zvetkov выявлены только в подледный период (рисунок 31 B). Зеленая мелкоклеточная микроводоросль *Monoraphidium griffithii* (Berkeley) Komárková-Legnerová(3/09) доминировала по численности в подледный период (29/03, 12/04), летом и осенью (10/07-14/09) (рисунок 31 A).

Параллельно в тех же пробах исследована сукцессия бактериальных сообществ. Парные прочтения V3-V4 16S рРНК, полученные в результате секвенирования, были объединены и проанализированы в программе Usearch. В результате было получено в среднем 50000 прочтений на образец со средней длиной прочтений 446 нуклеотидов (таблица 7). Общее количество прочтений в 9 образцах составляло 450032. Для всех образцов графики зависимости количества ОТЕ от количества прочтений выходят на плато (рисунок 32) и индекс Chao1 сходен с количеством ОТЕ, это подтверждает, что было охвачено все богатство ОТЕ в пробах (таблица 7). Количество ОТЕ в образцах варьировало от 276 до 444 (таблица 7).

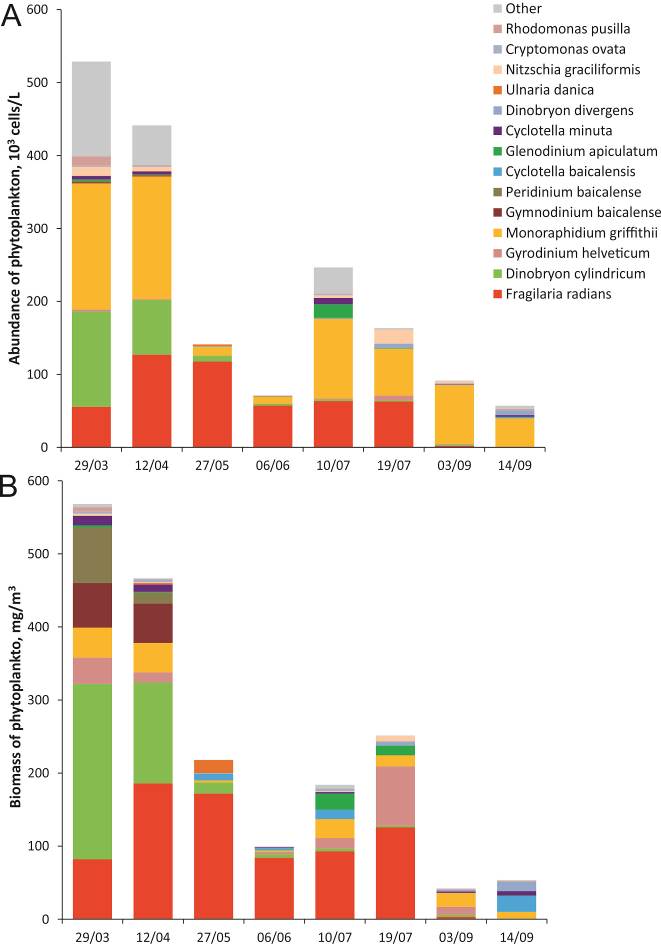


Рисунок 31 –Динамика численности и биомассы фитопланктона в верхнем 0-25 м слое на станции Листвянка-Танхой с марта по сентябрь 2017

Таблица 7 – Количество прочтений, богатство и разнообразие для ОТЕ (97% сходства) бактерий в образцах фотического слоя (0-25 м) на станции Листвянка-Танхой в озере Байкал с марта по сентябрь.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Код образцов | Образцы | Количество прочтений | Количество ОТЕ (97 % сходства) | Chao1 | Simpson | Shannon\_e |
| BD25 | 29/03 | 50005 | 278 | 278 | 0.0639 | 3.59 |
| BD26 | 12/04 | 50000 | 295 | 295 | 0.0576 | 3.64 |
| BD27 | 27/05 | 50016 | 440 | 440 | 0.0563 | 3.85 |
| BD28 | 6/06 | 49961 | 444 | 444.5 | 0.0581 | 3.83 |
| BD29 | 10/07 | 49982 | 315 | 316.2 | 0.0496 | 3.7 |
| BD30 | 19/07 | 50024 | 317 | 317.8 | 0.0644 | 3.52 |
| BD31 | 3/09 (пол.) | 50020 | 339 | 339.2 | 0.201 | 2.64 |
| BD32 | 3/09 (цел.) | 50007 | 276 | 283.4 | 0.334 | 1.75 |
| BD33 | 14/09 | 50017 | 368 | 379.7 | 0.147 | 3.05 |

Разнообразие по индексу Shannon\_e было ниже и индекс Simpson выше в осенних образцах (3/09, 14/09), что подтверждает их наименьшее разнообразие в сравнении с сообществами других сезонов (таблица 7). Было отобрано два образца 3/09, которые фильтровали на поликарбонатные (пол.) 0.2 мкм фильтры и целлюлозные (цел.) 0.2 мкм фильтры (таблица 7). В образце, профильтрованном через поликарбонатный фильтр, было выше богатство и разнообразие (таблица 7), поэтому лучше использовать этот фильтр для исследования бактериальных сообществ.

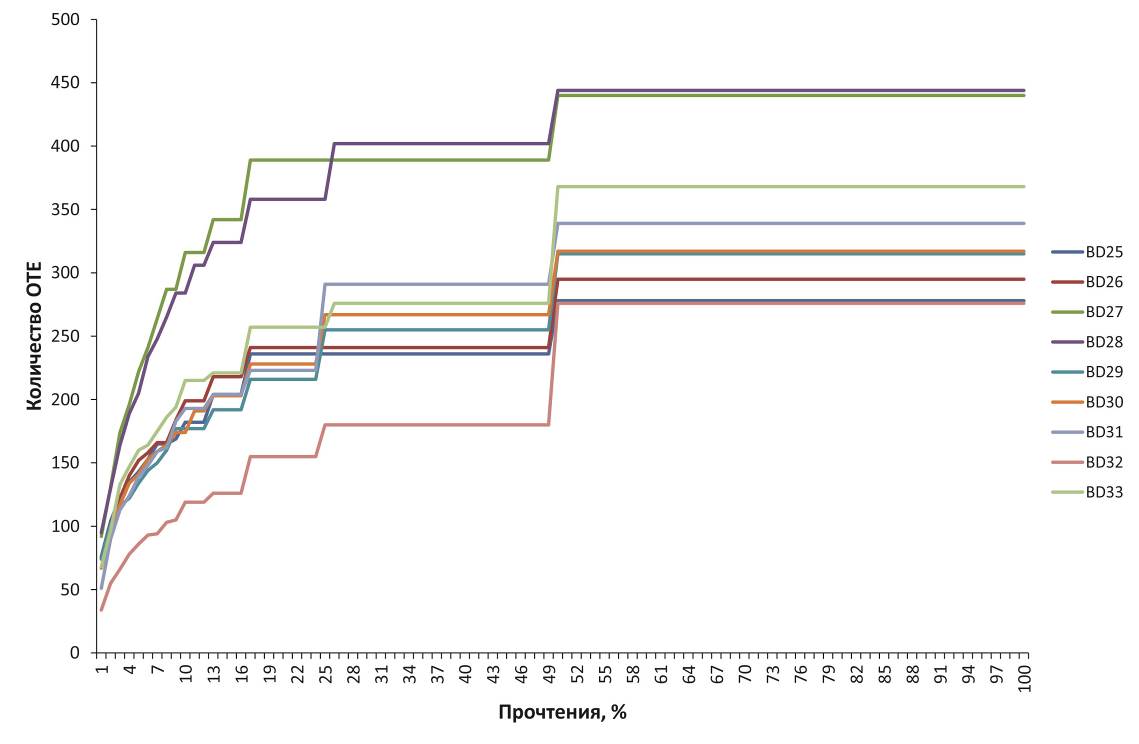


Рисунок 32 – Графики разряжения прочтений V3-V4 16S рРНК для 9 образцов  
из фотического слоя станции Листвянка-Танхой в озере Байкал в 2017 г.

Состав бактериальных сообществ на станции Листвянка-Танхой был представлен Actinobacteria, Bacteroidetes, Cyanobacteria, Betaproteobacteria, Alphaproteobacteria, Gammaproteobacteria, Verrucomicrobia, Planctomycetes, Acidobacteria (рисунок 33). Весной и летом был сходный состав сообществ с доминированием Actinobacteria (45-59% от общего количества прочтений в образце). Осенью доминирующим филумом становился Cyanobacteria (58-89%) (рисунок 33).

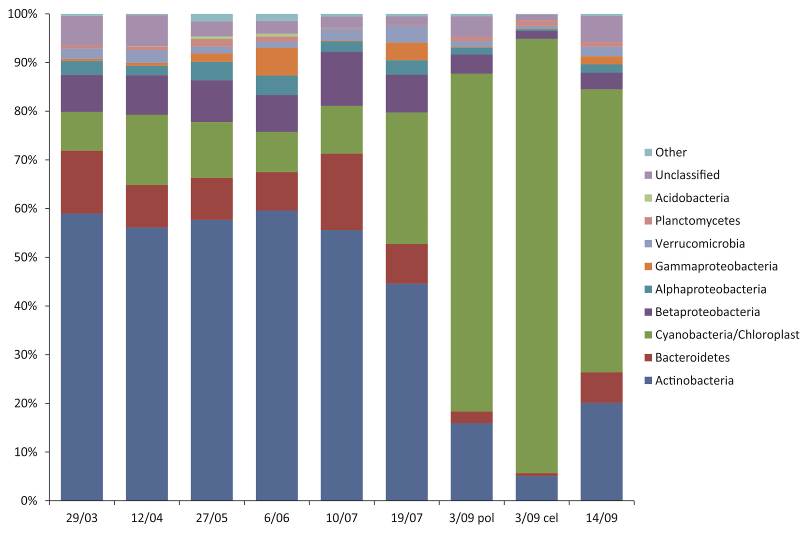


Рисунок 33 – Соотношение прочтений Bacteria в образцах фотического слоя станции Листвянка-Танхой в озере Байкал с марта по сентябрь

Кластерный анализ с использованием метрики UniFrac показал, что бактериальные сообщества формируют два кластера. В первый входят сообщества весны и лета, во второй – осени (рисунок 34).

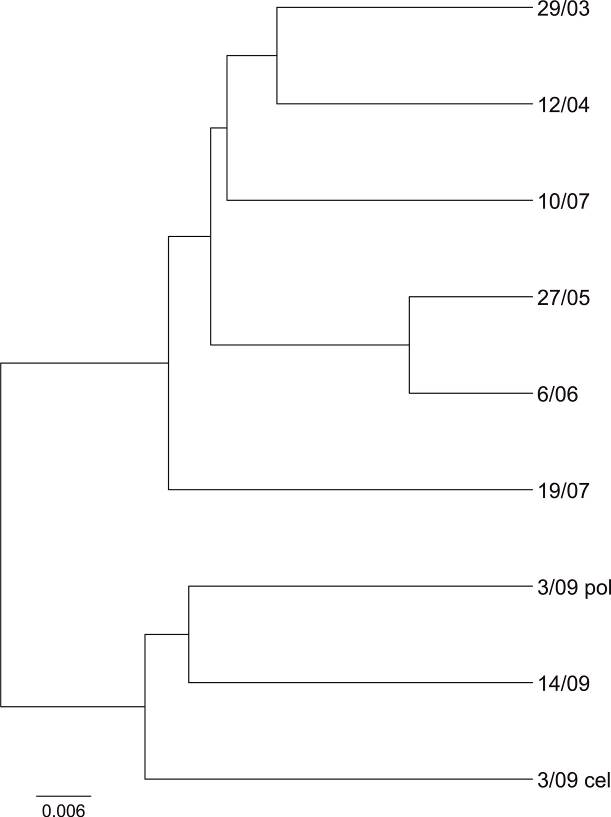


Рисунок 34 – Кластерный анализ для ОТЕ (97% сходства) бактериальных сообществ,  
построенный с использованием метрики UniFrac

Анализ данных NGS выявил сукцессию сообществ микроэукариот. Парные прочтения региона V4 гена 18S рРНК эукариот объединены и проанализированы в программе Usearch. В результате было получено в среднем 37000 прочтений на образец (таблица 8). Общее количество прочтений для 9 образцов составляло 333003. Для всех образцов графики зависимости количества ОТЕ от количества прочтений выходят на плато (рис. 8) и индекс Chao1 сходен с количеством ОТЕ, таким образом, мы охватили все богатство ОТЕ в пробах (таблица 8). Количество ОТЕ в образцах варьировало от 179 до 239 (таблица 8). Разнообразие по индексу Shannon\_e и индекс Simpson были сходны в различные сезоны. В образце, профильтрованном через поликарбонатный фильтр, было выше богатство и разнообразие, по сравнению с использованием целлюлозного фильтра (таблица 8).

Таблица 8 – Количество прочтений, богатство и разнообразие для ОТЕ (97% сходства) микроэукариот (V4 18S рРНК) в образцах фотического слоя (0-25 м) на станции Листвянка-Танхой в озере Байкал с марта по сентябрь 2017.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Код образцов | Образцы | Количество прочтений | Количество ОТЕ (97% сходства) | Chao1 | Simpson | Shannon\_e |
| BD25 | 29/03 | 37007 | 184 | 185.2 | 0.0616 | 3.49 |
| BD26 | 12/04 | 37011 | 192 | 192.3 | 0.0874 | 3.21 |
| BD27 | 27/05 | 37019 | 196 | 196.5 | 0.121 | 2.75 |
| BD28 | 6/06 | 36986 | 183 | 183 | 0.075 | 3.16 |
| BD29 | 10/07 | 36986 | 183 | 183.4 | 0.0502 | 3.46 |
| BD30 | 19/07 | 36991 | 194 | 194.2 | 0.0448 | 3.62 |
| BD31 | 3/09 (пол.) | 37010 | 209 | 210.9 | 0.051 | 3.58 |
| BD32 | 3/09 (цел.) | 36992 | 179 | 179.9 | 0.0581 | 3.43 |
| BD33 | 14/09 | 37001 | 239 | 240.5 | 0.0346 | 3.88 |

Парные прочтения V8-V9 18S рРНК эукариот объединены и проанализированы в программе Usearch. Таксономию определяли с помощью Sintax в Usearch [27]. Референсной базой данных для Sintax была silva\_18s\_v123.fa. В результате было получено в среднем 62000 прочтений на образец (таблица 9) со средней длиной прочтений 367 нуклеотидов. Общее количество прочтений для 9 образцов составляло 558039. Количество ОТЕ (богатство) при анализе сообществ микроэукариот по региону V8-V9 18S рРНК было выше (таблица 9), чем при анализе сообществ по региону V4 18S рРНК (таблица 9).

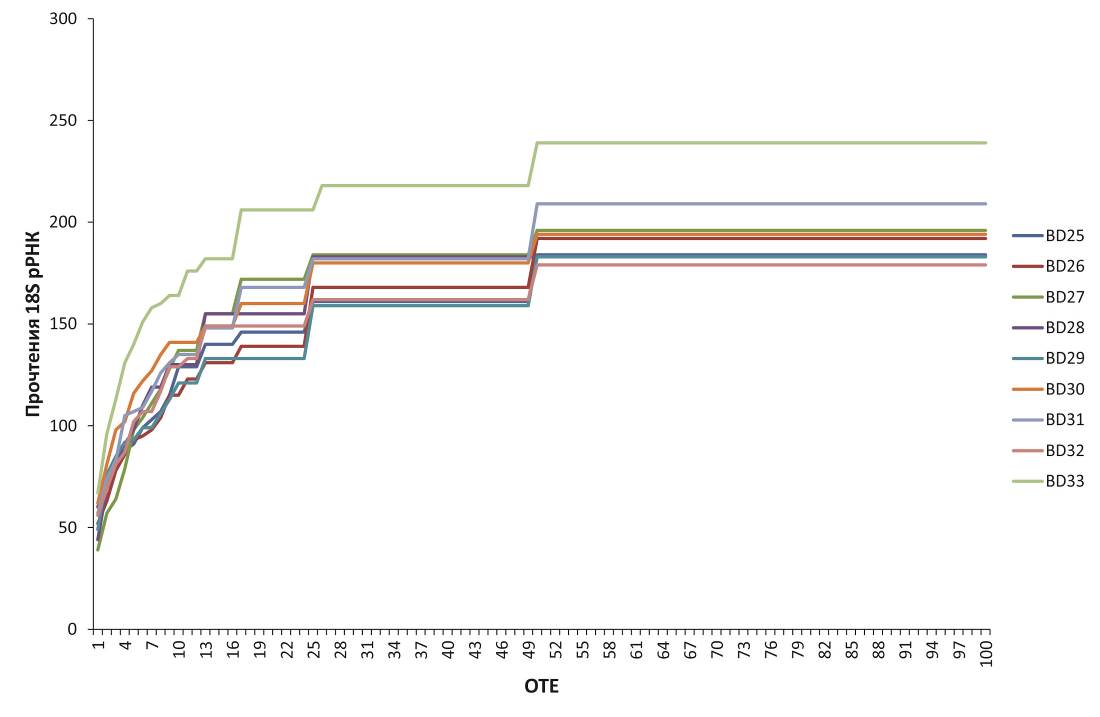


Рисунок 33 – Графики разряжения прочтений региона V4 гена 18S рРНК для 9 образцов  
из фотического слоя станции Листвянка-Танхой в озере Байкал в 2017 г.

Поэтому в дальнейшем анализ сообществ микроэукариот следует проводить по V8-V9 18S рРНК. Разнообразие сообществ микроэукариот в различных образцах было сходным (таблица 9).

Таблица 9 – Количество прочтений, богатство и разнообразие для ОТЕ (97% сходства) микроэукариот (V8-V9 18S рРНК) в образцах фотического слоя (0-25 м) на станции Листвянка-Танхой в озере Байкал с марта по сентябрь.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Код образцов | Образцы | Количество прочтений | Количество ОТЕ (97% сходства) | Chao1 | Simpson | Shannon\_e |
| BD25 | 29/03 | 62015 | 243 | 243.3 | 0.0662 | 3.46 |
| BD26 | 12/04 | 62017 | 225 | 226.1 | 0.0618 | 3.43 |
| BD27 | 27/05 | 61985 | 231 | 231.3 | 0.111 | 2.86 |
| BD28 | 6/06 | 62009 | 226 | 226 | 0.0715 | 3.25 |
| BD29 | 10/07 | 62012 | 224 | 226.4 | 0.0556 | 3.42 |
| BD30 | 19/07 | 61990 | 240 | 241.3 | 0.0524 | 3.54 |
| BD31 | 3/09 (пол.) | 62013 | 241 | 242.2 | 0.0488 | 3.61 |
| BD32 | 3/09 (цел.) | 61988 | 179 | 179 | 0.0538 | 3.44 |
| BD33 | 14/09 | 62010 | 235 | 235.6 | 0.0387 | 3.79 |

Сообщества эукариот были в основном представлены Dinoflagellata, Ciliophora, Chrysophyceae, Diatomeae, Dictyochophyceae, Cryptomonadales, Bicosoecida, Kathablepharidae, Prymnesiophyceae, Cryptomycota, Basidiomycota, Chytridiomycota, Cercozoa, Choanoflagellida, Arthropoda, Unassigned Eukaryota (рисунок 34). Соотношение Diatomea было выше в весенний период (12/04-6/06). Chrysophyceae вносили наибольший вклад в состав сообществ весной (27/05) и осенью (3/09, 14/09). Соотношение гетеротрофных Ciliophora в различные сезоны варьировало в пределах от 9% до 24%. Паразитические Cryptomycota имели наибольшее соотношение весной после вскрытия озера ото льда 27/05 и 6/06 и составляли 12% и 8% соответственно (рисунок 34).

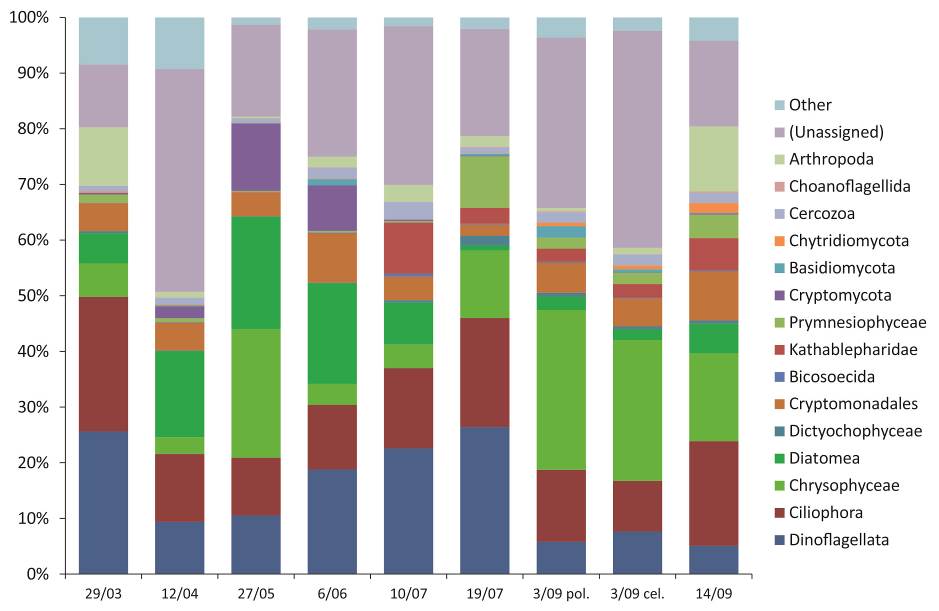


Рисунок 34 – Соотношение прочтений Eukaryota (по V8-V9 18S рРНК) в образцах  
фотического слоя станции Листвянка-Танхой в озере Байкал с марта по сентябрь

Кластерный анализ с использованием метрики UniFrac показал, что сообщество подледного периода (29/03) было отдельно от большого кластера с сообществами других периодов (рисунок. 35). Большой кластер подразделялся на два подкластера весна с началом лета (10/07) и осень со второй половиной лета (19/07) (рисунок 35).

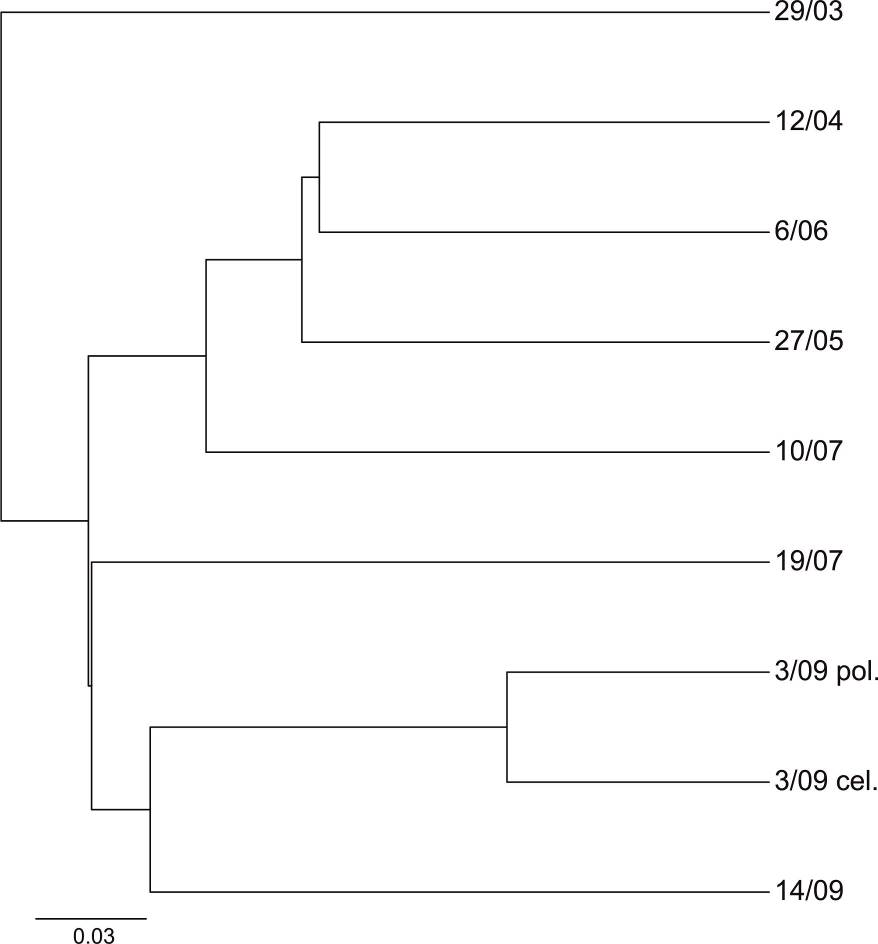


Рисунок 35 – Кластерный анализ для ОТЕ (97% сходства) сообществ микроэукариот (по V8-V9 18S рРНК), построенный с использованием метрики UniFrac

Таким образом, в данной работе была выявлена сукцессия сообществ бактерий и микроэукариот в фотическом слое (0-25 м) пелагической станции южной котловины озера Байкал с марта по сентябрь. В бактериальных сообществах от весны к осени увеличивалось соотношение Cyanobacteria и снижалось соотношение других таксонов. В сообществах микроэукариот весной выявлены наибольшие соотношения Diatomeae (диатомеи) и паразитов эукариот Cryptomycota. У гетеротрофных Ciliophora не было резкого изменения соотношения по сезонам. Chrysophyceae имели наибольшее соотношение весной после вскрытия озера ото льда и осенью.

Для исследования сообществ бактерий и микроэукариот проводилось NGS секвенирование фрагментов генов 16S рРНК и 18S рРНК на платформе Illumina. Полученные прочтения анализировались с помощью программы Usearch, которая работает быстрее и выполняет более строгую фильтрацию данных, чем используемая ранее программа Mothur. Количество прочтений на образец варьировало от 37000 до 70000. При анализе сообществ бактерий и микроэукариот в пробах фотического слоя озера Байкал, отобранных в 2017 г., в общем было получено 3233772 прочтения.

При анализе бактериальных сообществ 2017 г. было получено 50000-70000 прочтений на образец, что позволило выявить наибольшее богатство (278-557 операционных таксономических единиц – ОТЕ с 97 % сходства) сообществ, чем в нашей работе, выполненной ранее [10], в которой для большинства бактериальных сообществ фотического слоя Байкала среднее количество прочтений составляло 1500 и количество ОТЕ было от 51 до 265, что недостаточно для определения всего богатства бактериальных сообществ. В результате в бактериальных сообществах фотического слоя (0-25 м) трех котловин озера Байкал (6 станций), Баргузинского и Чивыркуйского заливов, и пролива Малое Море за три сезона (весна, лето, осень) выявлено в общем 1166 ОТЕ. От весны к осени в бактериальных сообществах снижалась доля представителей Actinobacteria, Bacteroidetes и Betaproteobacteria и увеличивалась доля Cyanobacteria.

**ЗАКЛЮЧЕНИЕ**

В 2018 г. для поддержки исследований в метагеномном анализе, который позволяет описывать микробные сообщества с ранее недоступной точностью, на данном этапе разработана система формализованных знаний, позволяющих преобразовывать описания структур модулей анализа NGS в визуальные модели. На основе этих моделей на следующих этапах проекта будут реализованы программы визуального программирования процедур анализа ампликонов, модули обеспечения хранения исходных и промежуточных данных в облачном хранилище, интегрировать процедуры и данных, организованные разными исследованиями, публиковать семантически-размеченные данные в Интернете, а также разрабатываться методики автоматизации проектирования распределенных вычислительных процессов на вычислительных кластерах.

На основе ранее полученных данных секвенирования показано, что наибольшей разрешающей способностью на уровне таксонов низшего ранга (роды, виды) обладает фрагмент, включающий районы V2–V3 16S rRNA. Этот фрагмент позволяет провести более подробное разделение пула нуклеотидных последовательностей на основе генетических дистанций на операционные таксономические единицы (ОТЕ, или OTU) видового ранга. Статистически сходящиеся оценки видового разнообразия для доминирующей части сообщества (пул видов, представляющих 95 % последовательностей) могут быть получены при размерах выборок от 10 000 до 15 000 последовательностей. Относительная ошибка оценок индекса видового разнообразия Шеннона при этом будет менее 4 %.

Для анализа структурно-функциональной организации микробиома разработан алгоритм построения сетей взаимодействия между организмами в сообществах, который позволяет выделить в исследуемой экосистеме ключевые звенья – виды, образующие каркас первичных продуцентов органического вещества и видов его потребителей. Сеть взаимодействий позволяет выделить в основном сообществе с помощью кластеризации группы видов с более тесной связью друг с другом и менее связанные с остальными видами сообщества (структурные подгруппы видов). Анализ сети взаимодействий позволяет перейти к следующему этапу – разработки динамической системы для прогнозирования поведения сообщества и экосистемы оз. Байкал при изменении условий среды.

Проведен анализ данных секвенирования (NGS) проб фотического слоя воды (0-25 м) пелагической станции в южной котловине оз. Байкал (Листвянка-Танхой) с марта по сентябрь 2017 г. (восемь временных точек). При анализе сообществ микроэукариот средняя длина прочтений V8-V9 18S рРНК составила 360 нуклетотидов, среднее количество прочтений на образец – 62 000. Выявленное количество ОТЕ составляет 225-241 на образец, что превышает недавно опубликованные зарубежными авторами данные (от 15 до 224) для проб, отобранных в южной котловине на различных глубинах в июле 2009 г. [28]. Для всех образцов в бактериальных сообществах выявлено в общей сложности 603 ОТЕ, представляющих отдельные таксоны, и для микроэукариот – 424. Нами впервые методом NGS выявлена сезонная динамика сообществ бактерий и микроэукариот в фотическом слое южной котловины оз. Байкал. Показано, что от весны к осени увеличивается доля Cyanobacteria и снижается доля других таксонов. В сообществах микроэукариот в весенний период наибольшая доля от всех полученных для пробы ОТЕ принадлежит диатомовым водорослям, что согласуется с данными микроскопических исследований. Высокая доля ОТЕ паразитов эукариот Cryptomycota, которые другими методами в каком разнообразии ранее не выявлялись. Доля гетеротрофных Ciliophora по сезонам существенно не изменяется. Хризофитовые вносят наибольший вклад в состав сообществ весной после вскрытия озера ото льда и осенью, что согласуется с данными, полученными нами ранее с помощью микроскопического анализа [29, 30]. Полученные данные свидетельствуют о перспективности NGS для исследований экосистемы оз. Байкал.

**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Karsenti E., Acinas S.G., Bork P., Bowler C., De Vargas C., Raes J., Sullivan M., Arendt D., Benzoni F., Claverie J.M., Follows M., Gorsky G., Hingamp P., Iudicone D., Jaillon O., Kandels-Lewis S., Krzic U., Not F., Ogata H., Pesant S., Reynaud E.G., Sardet C., Sieracki M.E., Speich S., Velayoudon D., Weissenbach J., Wincker P. A holistic approach to marine eco-systems biology // PLoS Biol. – 2011. – Vol. 9. – № 10. – P. e1001177.

2. Karsenti E. Towards an ‘Oceans Systems Biology’ // Mol. Syst. Biol. – 2012. – Vol. 8. – P. 575.

**3. Gilber J.A.** Steele J.A., Caporaso J.G., Steinbrück L., Reeder J., Temperton B., Huse S., McHardy A.C., Knight R., Joint I., Somerfield P., Fuhrman J.A., Field D. Defining seasonal marine microbial community dynamics // **ISME J.** – 2012. – Vol. 6. – P. 298–308.

4. Steele J.A., Countway P.D., Xia L., Vigil P.D., Beman J.M., Kim D.Y., Chow C.E., Sachdeva R., Jones A.C., Schwalbach M.S., Rose J.M., Hewson I., Patel A., Sun F., Caron D.A., Fuhrman J.A. Marine bacterial, archaeal and protistan association networks reveal ecological linkages // ISME J. – 2011. – Vol. 5. – P. 1414–1425.

5. Dam P., Fonseca L.L., Konstantinidis K.T., Voit E.O. Dynamic models of the complex microbial metapopulation of Lake Mendota // Syst. Biol. Appl. – 2016. – Vol. 2. – P.16007.

6. Cardona C., Weisenhorn P., Henry C., Gilbert J.A. Network-based metabolic analysis and microbial community modeling // Curr. Opinion Microbiol. – 2016. –Vol. 31. – P. 124–131.

7. Johns N.I. Blazejewski T., Gomes A.L., Wang H.H. Principles for designing synthetic microbial communities // Curr. Opinion Microbiol. – 2016. – Vol. 31. – P. 146–153.

8. Csardi G., Nepusz T. The igraph software package for complex network research // InterJournal, Complex Systems. – 2006. – № 1695. – P. 1–9.

9. Epskamp S., Cramer A. O., Waldorp L.J., Schmittmann V.D., Borsboom D. qgraph: Network visualizations of relationships in psychometric data // Journal of Statistical Software. – 2012. – № 48. – P. 1–18.

10. Mikhailov I.S., Zakharova Y.R., Bukin Y. S., Galachyants Y.P., Petrova D.P., Sakirko, M.V., Likhoshway, Y.V. Co-occurrence networks among bacteria and microbial eukaryotes of Lake Baikal during a spring phytoplankton bloom // Microbial Ecology. – 2018. – P. 1-14.

11. Schloss P.D., Westcott S.L., Ryabin T., Hall J.R., Hartmann M., Hollister E.B., Sahl J.W. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities // Applied and environmental microbiology. – 2009. – №75. – P. 7537–7541.

12. [Amin S.A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Amin%20SA%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Hmelo L.R](https://www.ncbi.nlm.nih.gov/pubmed/?term=Hmelo%20LR%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [van Tol H.M](https://www.ncbi.nlm.nih.gov/pubmed/?term=van%20Tol%20HM%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Durham B.P](https://www.ncbi.nlm.nih.gov/pubmed/?term=Durham%20BP%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Carlson L.T](https://www.ncbi.nlm.nih.gov/pubmed/?term=Carlson%20LT%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Heal K.R](https://www.ncbi.nlm.nih.gov/pubmed/?term=Heal%20KR%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Morales R.L](https://www.ncbi.nlm.nih.gov/pubmed/?term=Morales%20RL%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Berthiaume C.T](https://www.ncbi.nlm.nih.gov/pubmed/?term=Berthiaume%20CT%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Parker M.S](https://www.ncbi.nlm.nih.gov/pubmed/?term=Parker%20MS%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Djunaedi B](https://www.ncbi.nlm.nih.gov/pubmed/?term=Djunaedi%20B%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Ingalls A.E](https://www.ncbi.nlm.nih.gov/pubmed/?term=Ingalls%20AE%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Parsek M.R](https://www.ncbi.nlm.nih.gov/pubmed/?term=Parsek%20MR%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)., [Moran M.](https://www.ncbi.nlm.nih.gov/pubmed/?term=Moran%20MA%5BAuthor%5D&cauthor=true&cauthor_uid=26017307)A, [Armbrust E.V](https://www.ncbi.nlm.nih.gov/pubmed/?term=Armbrust%20EV%5BAuthor%5D&cauthor=true&cauthor_uid=26017307). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. // [Nature.](https://www.ncbi.nlm.nih.gov/pubmed/26017307) – 2015. – Vol. 522(7554). P. 98–101.

13. [Teeling H](https://www.ncbi.nlm.nih.gov/pubmed/?term=Teeling%20H%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Fuchs B.M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Fuchs%20BM%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Becher D](https://www.ncbi.nlm.nih.gov/pubmed/?term=Becher%20D%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Klockow C](https://www.ncbi.nlm.nih.gov/pubmed/?term=Klockow%20C%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Gardebrecht A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gardebrecht%20A%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Bennke C.M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bennke%20CM%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Kassabgy M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Kassabgy%20M%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Huang S](https://www.ncbi.nlm.nih.gov/pubmed/?term=Huang%20S%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Mann A.J](https://www.ncbi.nlm.nih.gov/pubmed/?term=Mann%20AJ%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Waldmann J](https://www.ncbi.nlm.nih.gov/pubmed/?term=Waldmann%20J%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Weber M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Weber%20M%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Klindworth A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Klindworth%20A%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Otto A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Otto%20A%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Lange J](https://www.ncbi.nlm.nih.gov/pubmed/?term=Lange%20J%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Bernhardt J](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bernhardt%20J%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Reinsch C](https://www.ncbi.nlm.nih.gov/pubmed/?term=Reinsch%20C%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Hecker M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Hecker%20M%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Peplies J](https://www.ncbi.nlm.nih.gov/pubmed/?term=Peplies%20J%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Bockelmann F.D](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bockelmann%20FD%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Callies U](https://www.ncbi.nlm.nih.gov/pubmed/?term=Callies%20U%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Gerdts G](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gerdts%20G%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Wichels A](https://www.ncbi.nlm.nih.gov/pubmed/?term=Wichels%20A%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Wiltshire K.H](https://www.ncbi.nlm.nih.gov/pubmed/?term=Wiltshire%20KH%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Glöckner F.O](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gl%C3%B6ckner%20FO%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Schweder T](https://www.ncbi.nlm.nih.gov/pubmed/?term=Schweder%20T%5BAuthor%5D&cauthor=true&cauthor_uid=22556258)., [Amann R](https://www.ncbi.nlm.nih.gov/pubmed/?term=Amann%20R%5BAuthor%5D&cauthor=true&cauthor_uid=22556258). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom // Science. – 2012. – Vol. 366 (6081). – P. 608-611.

14. Gerhardt P., Costilow R., Krieg N.R., Murrey R.G.E., Nester E.W., Phillips G.B., Wood W.A. Manual of methods for general bacteriology. – Washington: Amer. Soc. Microbiol., 1981.

15. Садчиков А.П. Методы изучения пресноводного фитопланктона. – М.: Университет и школа, 2003. – 157 с.

16. Киселев И.А. Планктон морей и континентальных водоемов. – Л.: Наука, 1969. – 657 с.

17. Абакумов В.А. Руководство по гидробиологическому мониторингу пресноводных экосистем. – СПб.: Гидрометеоиздат, 1992. – 318с.

18. Макарова И.В., Пичкилы Л.О. К некоторым вопросам методики вычисления биомассы фитопланктона // Ботан. журн. – 1970. – Т. 55. – № 10. – С. 1488-1494.

19. Белых О.И., Бессудова А.Ю., Гладких А.С., Кузьмина А.Е., Помазкина Г.В., Поповская К.И., Сороковикова Е.Г., Тихонова И.В., Усольцева М.В. Руководство по определению биомассы видов планктона пелагиали оз. Байкал. Методическое пособие. –Иркутск: Изд-во ИГУ, 2011. – 51с.

20. Widdel F., Back F. The genus *Desulfotomaculum* // Dworkin M., Falkow S., Rosenberg E., Schleifer K.H., Stackebrandt E. (eds.). The Prokaryotes. Ecophysiology and Biogeochemistry. – Springer, New York, 1992. – P. 787-794.

21. Sambrook J., First E.F., Maniatis T. Molecular cloning. A laboratory manual. – New York: Cold Spring Harbor Laboratory Press, 1989. – 350 p.

22. Денисова Л.Я., Белькова Н.Л., Тулохонов И.И., Зайчиков Е.Ф. Биоразнообразие бактерий на различных глубинах южной котловины озера Байкал, выявленное по последовательностям 16S рРНК // Микробиология. – 1999. – Т. 68. – № 4. – С. 547–556.

23. Brosius J., Ullrich A., Paker M.A., Gray A., Dull T.J., Gutell R.R., Noller H.F. Construction and fine mapping for recombinant plasmids containing the rrnB ribosomal RNA operon of E. coli // Plasmid. – 1981. – Vol. 6. – P.112–118.

24. Coorevits A., Dinsdale A.E., Halket G., Lebbe L., De Vos P., Van Landschoot A., Logan N.A. Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus*, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. thermoglucosidans* (nom. corrig., formerly 'thermoglucosidasius'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov //Int. J. Syst. Evol. Microbiol. – 2012. – Vol. 62. – P. 1470–1485.

25. Brumm P., Land M.L., Hauser L.J., Jeffries C.D., Chang Y.-J., and Mead D.A. Complete genome sequences of *Geobacillus* sp. Y412MC52, a xylan-degrading strain isolated from obsidian hot springin Yellowstone National Park // Stand. Genomic Sci. – 2015. – Vol. 10. – P. 81.

26. Edgar R.C. Search and clustering orders of magnitude faster than BLAST // Bioinformatics. – 2010. – V. 26. – № 19. – P. 2460–2461.

27. Edgar R.C. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. bioRxiv. 2016.

28. Yi Z., Berney C., Hartikainen H., Mahamdallie S., Gardner M., Boenigk J., Cavalier-Smith T., Bass D. High-throughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations // FEMS Microbiol. Ecol. – 2017. – Vol.93. – № 8. – P. fix073.

29. Фирсова А.Д., Кузьмина А.Е., Томберг И.В., Потемкина Т Г., Лихошвай Е.В. Cезонная динамика формирования стоматоцист хризофитовых водорослей в планктоне Южного Байкала // Известия РАН. Серия биологическая. – 2008. – № 5. – С. 589–596.

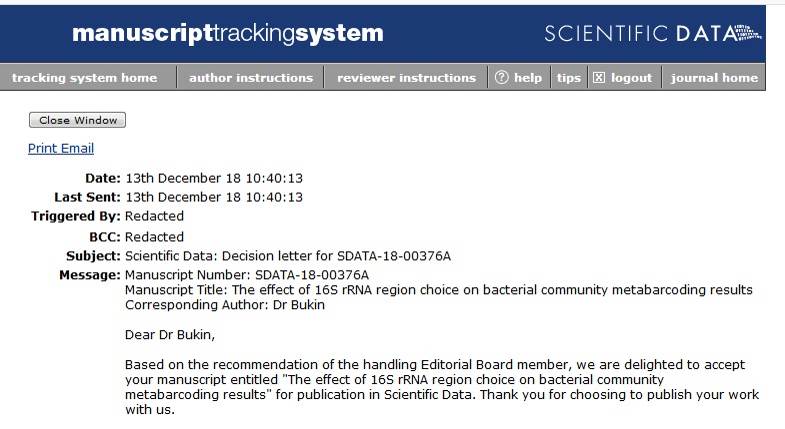
30. Bessudova A., Firsova A.D.,Likhoshway Y.V. Silica-scaled chrysophytes of Lake Baikal // Acta Biologica Sibirica. – 2017. – Vol. 3. – № 3. – P. 47-56.

* 1. Приложение А  
     Публикации по теме проекта за 2018 год

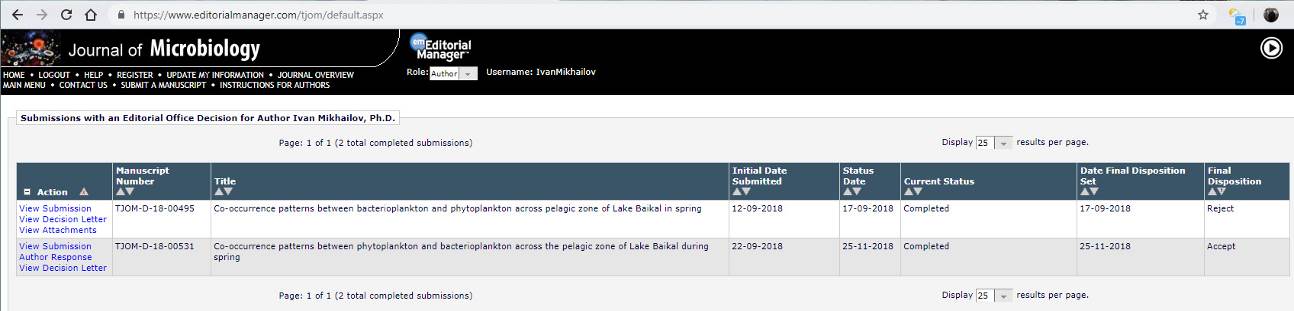
1. Cherkashin E., Kopaygorodsky A., Kazi L., Shigarov A., Paramonov V. Model driven architecture implementation using linked data // Procs. of 24th International Conference, ICIST 2018, Vilnius, Lithuania, October 4-6, 2018. P. 412-423. **WoS, SCOPUS**. https://doi.org/10.1007/978-3-319-99972-2\_34

2. Букин Б.С., Михайлов И.С., Черкашин Е.А., Алгоритм построения сетей взаимодействия видов в сообществах на основе результатов мониторинговых экологических исследований [Текст] // Доклады Международной конференции «Математическая биология и биоинформатика». Под ред. В.Д. Лахно. Том 7. Пущино: ИМПБ РАН, 2018. Статья No e93. doi: 10.17537/icmbb18.24. **РИНЦ**

3. Bukin Yu.S, Galachyants Yu.P., Morozov I.V., Bukin S.V., Zakharenko A.S., Zemskaya T.I. The effect of 16S rRNA region choice on bacterial community metabarcoding results // Scientific Data – accepted, **WoS** **, SCOPUS.** *Принята к опубликованию, возможен выход в 2018.*



1. Ivan S. Mikhailov, Yuri S. Bukin, Yulia R. Zakharova, Marina V. Usoltseva, Yuri P. Galachyants, Maria V. Sakirko, Vadim V. Blinov, Yelena V. Likhoshway.Co-occurrence patterns between bacterioplankton and phytoplankton across pelagic zone of Lake Baikal in spring //J. Microbiol. (Korea) -accepted, publication in April 2019. **WoS, SCOPUS**.



5. Cherkashin E., Shigarov A., Malkov F., and Morozov A. An Instrumental Environment for Metagenomic Analysis. Proc. Int. Conf. on Information Technologies in the Research of Biodiversity. 2018. *(Принято в печать в издании, индексируемом Web of Science Core Collection, выходит в 2019).*

* 1. Приложение Б  
     Информация о количестве статей в виде таблицы

|  |  |  |
| --- | --- | --- |
| Индикатор | Ед. измерения | 2018 |
| Количество публикаций в ведущих российских и международных журналах по результатам исследований, полученных в процессе реализации проекта | единиц | 2 (3) |
| Количество публикаций в мировых научных журналах, индексируемых в базе данных «Сеть науки» (WEB of Science) и/или Scopus | единиц | 1 (2) |
| РИНЦ (исключая WEB of Science и Scopus) | единиц | 2 (3) |

* 1. ПРИЛОЖЕНИЕ В   
     Этапы выполнения проекта

*Этапы* 2018 г.:

1. Проектирование и прототипирование средств управления данными и рабочими процессами веб-ориентированной информационно-аналитической инфраструктуры для поддержки исследования микробиома оз. Байкал и курирования больших объемов данных высокопроизводительного секвенирования.

2. Отбор проб из разных экотопов по всему Байкалу, выделение ДНК, секвенирование библиотек, таксономический анализ сообществ.

3. Разработка алгоритма, позволяющего на основе результатов корреляционного анализа построить сеть взаимодействий организмов в изучаемом сообществе.

4. Реконструкция полных геномов выделенных в культуры перспективных для практического применения штаммов бактерий.