

На правах рукописи



Гарушянц Софья Константиновна

**Структура и эволюционная динамика прокариотических сообществ
необычных местообитаний**

03.01.09 – математическая биология, биоинформатика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата биологических наук

Москва – 2019

Работа выполнена в Учебно-научном центре «Биоинформатика» Федерального государственного бюджетного учреждения науки Института проблем передачи информации им. А.А. Харкевича Российской академии наук.

Научный руководитель:

Гельфанд Михаил Сергеевич

кандидат физико-математических наук, доктор биологических наук, профессор
(Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации имени А.А. Харкевича Российской академии наук)

Официальные оппоненты:

Озолин Ольга Николаевна

доктор биологических наук, профессор, зав. лабораторией
(Институт биофизики клетки Российской академии наук – обособленное подразделение Федерального государственного бюджетного учреждения науки «Федеральный исследовательский центр «Пушкинский научный центр биологических исследований Российской академии наук»)

Равчеев Дмитрий Андреевич

кандидат биологических наук, научный сотрудник
(Ирландский национальный университет, Голуэй, Ирландия)

Ведущая организация:

Федеральное государственное учреждение «Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук»

Защита диссертации состоится « 15 » апреля 2019 года в 15:00 на заседании диссертационного совета Д 002.077.04 при Федеральном государственном бюджетном учреждении науки Институте проблем передачи информации имени А.А. Харкевича Российской академии наук (ИППИ РАН) по адресу: 127051, г. Москва, Большой Каретный переулок, д. 19, стр. 1.

С диссертацией можно ознакомиться в библиотеке ИППИ РАН, а также на сайте ИППИ РАН по адресу: http://iitp.ru/upload/content/1391/GSK_dissertation_final.pdf

Автореферат разослан «___» _____ 2019 г.

Ученый секретарь диссертационного совета
доктор биологических наук, профессор

Рожкова Г.И.

Общая характеристика работы

Актуальность темы исследования. Прокариоты – бактерии и археи – это самая широко распространённая группа организмов на Земле. Их изучение представляет значительный интерес, так как эти организмы обладают большим метаболическим разнообразием и имеют важное прикладное значение, в первую очередь в медицине (возбудители болезней, симбионты человека, продуценты антибиотиков), сельском хозяйстве (почвенные бактерии, патогены растений и животных) и биотехнологии (штаммы-продуценты, системы биоочистки). Изучение микроорганизмов из необычных местообитаний интересно не только с прикладной, но и с теоретической точки зрения, потому что позволяет понять, как эти организмы приспосабливаются к сложным условиям среды, например, к средам с малым количеством питательных веществ.

В связи с развитием технологий секвенирования появилась возможность изучать новые микроорганизмы не только в пробирке, но и методами биоинформатики, в частности, сравнительной геномики. Накопленные к настоящему времени данные о геномах и генах микроорганизмов позволяют реконструировать метаболизм бактерий и архей без дорогостоящих экспериментов. Большинство видов бактерий не удаётся культивировать в виде чистых культур, но современные технологии позволяют реконструировать геном даже для таких некультивируемых организмов. Тем самым, методами биоинформатики можно реконструировать метаболизм бактерий, которые не поддаются изучению при помощи других подходов. К настоящему времени методами сравнительной геномики были реконструированы метаболические карты для многих организмов. Не менее важной чем реконструкция метаболизма, является реконструкция регуляции транскрипции и трансляции, т.к. важно понимать при каких условиях включается работа конкретных метаболических путей и какие факторы на это влияют. Именно методы сравнительной геномики позволили описать разнообразие транскрипционной регуляции бактерий и частично архей.

Повсеместное распространение методов секвенирования способствовало возникновению новой области биоинформатики – метагеномики. Метагеномика занимается анализом метагеномов, т. е. суммарной ДНК или РНК из проб окружающей среды или полостей тела. Такой подход позволяет изучать как отдельные организмы, так и их взаимодействие в сообществе. Дополнительным преимуществом метагеномики является возможность исследовать не только культивируемые микроорганизмы, но и те, которые не удаётся вырастить в лаборатории в виде монокультуры.

Биоинформатический подход даёт возможность описывать не только современное состояние популяции, но и реконструировать предковые состояния. Примером успеха в этой области стало получение оценки уровня горизонтального переноса генов (ГПГ). Именно использование компьютерных методов позволило показать, что ГПГ

подвержены не только гены устойчивости к антибиотикам, но что в результате ГПГ любой ген может быть перенесён из одного организма в другой. Известно, что перенос ряда метаболических генов привел к образованию новых крупных таксонов, например, метаногенных архей. Другой пример реконструкции предковых состояний – это изучение изменений генома во времени. Так, применение методов биоинформатики позволило показать как изменяются геномы симбиотических бактерий и архей в ситуации слабого отбора, например, при симбиотическом образе жизни.

Цели и задачи исследования. Целью данной работы являлась характеристика особенностей бактериальных сообществ из необычных слабо изученных местообитаний и изучение механизмов адаптации микроорганизмов к условиям окружающей среды.

Для достижения этой цели были поставлены следующие задачи:

- 1) изучение динамики бактериальных сообществ микробных топливных элементов, характеризующихся стабильными и контролируемыми условиями;
- 2) исследование особенностей горизонтального переноса генов при приспособлении к условиям окружающей среды;
- 3) метаболическая реконструкция редуцированных симбиотических организмов и анализ механизмов их приспособления к узким нишам внутри организма-хозяина;
- 4) описание изменений функциональных систем клетки, сопряжённых с приспособлением к симбиотическому образу жизни, на примере рибосомы.

Научная новизна. В настоящей работе был исследован ряд экосистем необычных местообитаний, в том числе и одна искусственная контролируемая система, и изучены их свойства. Вопросы, поставленные в этой работе, для таких систем до этого практически не изучались. Оценка уровня ГПГ была получена как для переносов между близкими родственниками, так и для ГПГ между далёкими родственниками. Тем не менее, существующие оценки далёкого ГПГ были получены в условиях, когда было доступно значительно меньше данных о геномах бактерий и архей. Сейчас можно получить более аккуратные оценки потока горизонтально перенесённых генов между далёкими организмами. Если известны геномные последовательности для многих представителей одной таксономической группы, в которую произошёл ГПГ, то можно понять, как эволюционировал фрагмент генома после того, как был перенесён в новый таксон. Это было практически невозможно сделать ещё десять лет назад, когда полных геномных последовательностей было известно значительно меньше. Получение более аккуратной оценки для количества ГПГ позволяет точнее описать и функциональные категории генов, которые подвержены ГПГ. В настоящей работе на примере архей из семейства *Methanosarcinaceae* изучена роль ГПГ между бактериями и археями, а также описаны функциональные классы генов, чаще подвергающихся ГПГ.

В работе проведена метаболическая реконструкция бактерии из группы, которая до этого совсем не была изучена, – ядерного эндосимбионта инфузорий.

На различных системах было показано, что симбиотический образ жизни связан с ослаблением естественного отбора и приводит сначала к накоплению псевдогенов и мобильных элементов, а затем к редукции геномов, но не до конца понятно, какие именно гены и почему псевдогенизируются и исчезают из генома. Симбиотические бактерии с крайне редуцированными геномами могут иметь всего около двухсот генов. Эти гены должны кодировать абсолютно все необходимые для жизни функции, и тем самым 200 генов – это очень мало. Например, одна только бактериальная рибосома у свободноживущих бактерий кодируется приблизительно 60 генами. В данной работе на примере рибосомы исследуется редукция жизненно важной системы.

Теоретическая и практическая значимость. Полученные в данной работе результаты в первую очередь имеют фундаментальное значение, поскольку позволяют понять механизмы взаимодействий микроорганизмов внутри сообществ. Изучение бактерий, живущих в необычных местообитаниях, позволяет лучше описать весь спектр метаболических возможностей мира микробов. Описанный в работе метаболизм голоспоры – один из таких примеров. Описание на примере рибосомы пределов и механизма потери белков позволяет понять механизмы, по которым происходит редукция генома бактерий при слабом отборе. Понимание механизмов и распространённости ГПГ позволяет более аккуратно описывать процессы передачи генов между микроорганизмами. Проведённые нами исследования позволили исправить существовавшую ранее неверную оценку количества ГПГ в метаногенных археях.

Ряд полученных нами результатов имеет практическую значимость. Проанализированные метагеномные образцы микробных топливных элементов (МТЭ) дают возможность следить за динамикой и эффективностью прокариотических сообществ. Подробные знания об этой динамике позволяют предложить пути для оптимизации состава микробного сообщества, используемого в этом типе биореакторов, что может позволить повысить эффективность очистки сточных вод. Похожий подход может быть применён и к другим практически значимым сообществам, например, различного рода ферментерам.

Методология и методы исследования. Для решения поставленных задач использовались различные методы биоинформатики и сравнительной геномики. В задачах, которые предполагают секвенирование новых геномов и метагеномов, применялись стандартные подходы к анализу данных секвенирования нового поколения: полученные из прибора прочтения фильтровали по качеству и собирали для получения геномных последовательностей. Во всех задачах поиск гомологов и построение ортологических групп проводился на основе критерия двустороннего лучшего сходства. Помимо этого использовались методы построения множественных выравниваний и филогенетических деревьев (метод ближайших соседей и метод максимального правдоподобия); методы анализа аминокислотных и нуклеотидных последовательностей. Для проверки статистической значимости использовался

критерий χ^2 и корреляционный анализ.

Положения, выносимые на защиту. В микробных топливных элементах (МТЭ) присутствуют бактерии, способные как к прямому, так и к опосредованному переносу электронов на анод. Изученные сообщества из двух МТЭ способны достаточно эффективно очищать сточные воды. Более того, эффективная работа МТЭ может осуществляться разными микробными сообществами, где одни и те же функции выполняют микроорганизмы разных таксономических групп.

Вклад горизонтального переноса генов (ГПГ) в эволюцию *Methanosarcina* spp. ранее был переоценён, на самом деле только около 5% генов было получено в результате ГПГ. Гены, полученные в ходе ГПГ от бактерий, обычно переносятся без соответствующих транскрипционных регуляторов, и часто этот процесс приводит к переносу отдельных генов, а не целых оперонов. Чаще всего переносятся гены, связанные с метаболизмом, в особенности гены транспортёров.

Holospira spp. – ядерные эндосимбионты ресничных инфузорий – не способны синтезировать большинство необходимых для жизни низкомолекулярных соединений, таких как, например, аминокислоты, и у них полностью отсутствует весь центральный энергетический метаболизм. В качестве основного источника энергии *Holospira* spp. используют нуклеотиды, которые получают от хозяина. Ряд генов, специфичных для *Holospira* spp. и отсутствующих у близких к ним представителей отряда *Rickettsiales*, могут играть роль в образовании репродуктивной формы и развитии инфекции.

Состав рибосом в 214 бактериях с редуцированными геномами стабилен, однако отдельные рибосомные белки могут отсутствовать. Чаще отсутствуют белки большой субъединицы, а также белки, находящиеся на поверхности рибосомы и имеющие в среднем меньше контактов, чем универсальные белки. Потеря может происходить целыми функциональными блоками; в частности потеря триггер-фактора приводит к потере блока из трёх белков, связанных с регуляцией трансляции. Потеря последовательности анти-Шайна – Дальгарно не связана со степенью редукции генома и с потерями рибосомных белков.

Степень достоверности и апробация результатов. По материалам диссертации опубликовано три статьи в рецензируемых научных журналах. Результаты работы были представлены на международных конференциях: на международной конференции NETTAB & IB 2015 (15th International Workshop on Network Tools and Applications for Biology 2015 & 11th Integrative Bioinformatics International Symposium 2015, Бари, Италия), на международном семинаре RECESS'13 (Венеция, Италия), на Московских международных конференциях по вычислительной молекулярной биологии (Moscow Conference on Computational Molecular Biology – MCCMB'11, MCCMB'15); а также на конференциях «Информационные технологии и системы» (ИТиС'13 – Калининград, ИТиС'16 – Репино, ИТиС'17 – Уфа).

Структура и объем работы. Диссертация изложена на 137 страницах машинописного текста и содержит следующие разделы: введение, обзор литературы, результаты и

обсуждение в четырёх главах, заключение и выводы. В конце приведён список литературы. Материал включает 27 рисунков, 13 таблиц и список литературы, содержащий 326 ссылок.

Содержание работы

Глава 1 содержит обзор литературы по теме диссертации.

Прокариоты – это все одноклеточные живые организмы, которые не являются эукариотами, т. е. не имеют ядра, окружённого мембраной. Анализ последовательностей 16S рРНК показал, что прокариоты разбиваются на две большие группы организмов: археи и бактерии. Более того, позднее было показано, что эукариоты являются внутренней ветвью архей. Тем самым, термин «прокариоты» не описывает конкретную таксономическую группу организмов, а объединяет все организмы для которых характерна ко-транскрипционная трансляция.

Прокариоты приспособились к жизни в самых разнообразных местообитаниях на Земле: они могут жить везде, начиная от почв и океанов и заканчивая кишечными трактами животных. Каждое из таких сообществ уникально и населено своими видами бактерий и архей. Прокариоты обычно обитают в сложных сообществах, состоящих из многих видов. Образование таких сообществ основывается на возникновении сложных метаболических систем, где продукты жизнедеятельности одних организмов выступают как субстраты для других. Изучение состава и метаболических возможностей сообществ микроорганизмов стало возможно благодаря развитию технологий секвенирования и появлению метагеномного подхода.

Изучение бактерий и архей в контексте сообществ является критичным для понимания как функций конкретных генов, так и эволюции всего генома. Традиционно эксперименты проводятся на таких модельных организмах как *Bacillus subtilis* или *Escherichia coli*. Традиционный молекулярно-биологический и микробиологический подход предполагает изучение бактерий в монокультурах, и в таких условиях для многих генов было показано, что они не являются жизненно необходимыми. Оказывается, что многие из «неважных» генов на самом деле абсолютно критичны для выживания, когда бактерии сталкиваются с конкуренцией.

Таксономический и функциональный анализ микробных сообществ анодов микробных топливных элементов. На состав сообщества влияет целый ряд биотических и абиотических факторов. Поэтому особенной задачей является изучение динамики сообществ в ситуациях, когда все эти условия контролируются исследователем. Примером таких контролируемых сообществ могут служить биореакторы для очистки сточных вод. В промышленных условиях используются в основном анаэробные биореакторы (АБ), так как они предполагают меньший расход электроэнергии, чем аэробные решения. Идея технологии заключается в том, что прокариоты, помещённые в резервуар, разлагают органические отходы из протекающей через резервуар жидкости; при этом часто побочным продуктом такого разложения

является природный газ. При использовании реальных сточных вод в промышленности состав прокариотического сообщества может значительно меняться в зависимости от состава стока. Однако, если реактор помещён в лабораторные условия, состав и объёмы подаваемых питательных веществ можно контролировать. Дополнительно можно контролировать такие параметры, как освещённость, температуру, а при некоторых конструкциях можно ограничивать доступ других прокариот извне. В таких экспериментальных условиях можно изучать, например, как сообщества приспосабливаются к разным питательным средам. При этом, поскольку остальные параметры внешней среды зафиксированы, можно быть уверенным, что наблюдаемая динамика связана именно с конкретным изменением, а не с более сложным набором внешних факторов.

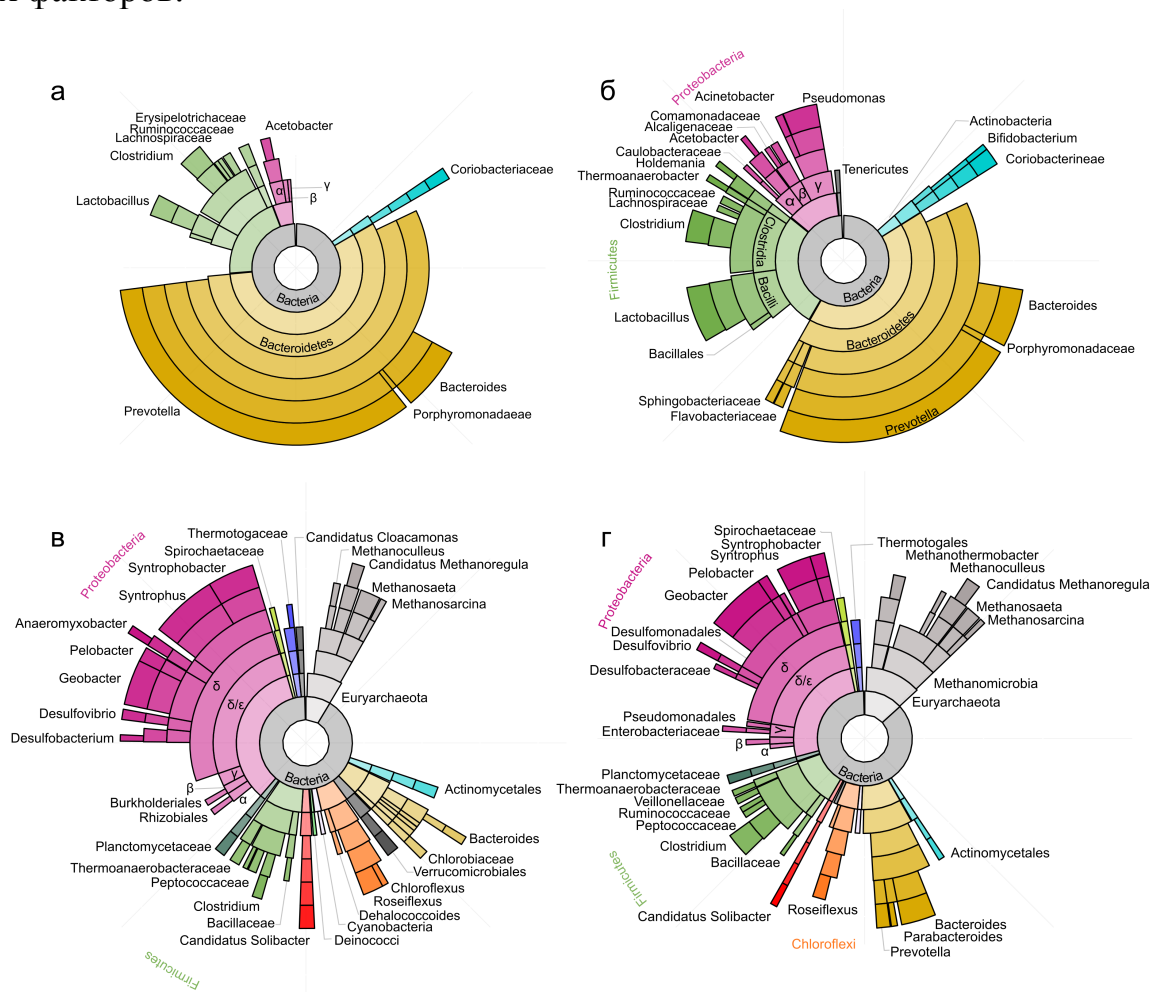


Рисунок 1. Таксономический состав метагеномных образцов. Диаграмма Крона для инокулята МТЭ UK (а), биоплёнки из МТЭ UK (б), инокулята МТЭ JP (в), биоплёнки МТЭ JP (г) показывает представленность разных таксонов в образце, где каждый круг соответствует определённому таксономическому уровню, а цвет круга соответствует типу или группе (уровень 3 в таксономии NCBI): жёлтый – *Bacteroidetes/Chlorobi* group, фиолетовый – *Proteobacteria*, светло-серый – *Archaea*, зеленый – *Firmicutes*, оранжевый – *Chloroflexi*, красный – *Acidobacteria*, салатовый – *Spirochaetes*, синий – *Thermotogae*, голубой – *Actinobacteria*, темно-зелёный – *Planctomycetes*, розовый – *Cyanobacteria*.

Одним из типов анаэробных биореакторов для очистки воды являются микробные топливные элементы (МТЭ). Этот тип биореакторов интересен тем, что является одной из самых многообещающих технологий экологически чистого производства энергии. Идея МТЭ заключается в том, что очистка сточных вод микроорганизмами приводит к одновременной выработке электроэнергии. Внутри МТЭ анаэробные микроорганизмы разрушают органические соединения и переносят полученные при этом электроны на электрод (анод), тем самым переводя энергию, полученную от очистки отходов, в электричество. Несмотря на то, что был описан целый ряд микроорганизмов, образующих биоплёнку на поверхности анода, комплексный метагеномный анализ сообществ электрогенных бактерий практически не проводился.

В **главе 2** представлены результаты полногеномного метагеномного анализа бактериальных сообществ МТЭ из Великобритании (UK) и Японии (JP), разработанных для очистки разных типов сточных вод в разных климатических зонах. Чтобы изучить, как изменяются со временем сообщества этих двух МТЭ, была выделена тотальная ДНК из активного ила, которым инокулировались МТЭ (эти образцы дальше называются UK-ref и JP-ref), и из биоплёнки с поверхности анода через 90 (UK) и 70 (JP) дней после запуска МТЭ (эти образцы дальше называются UK-mfc и JP-mfc). Образцы были секвенированы и собраны *de novo*.

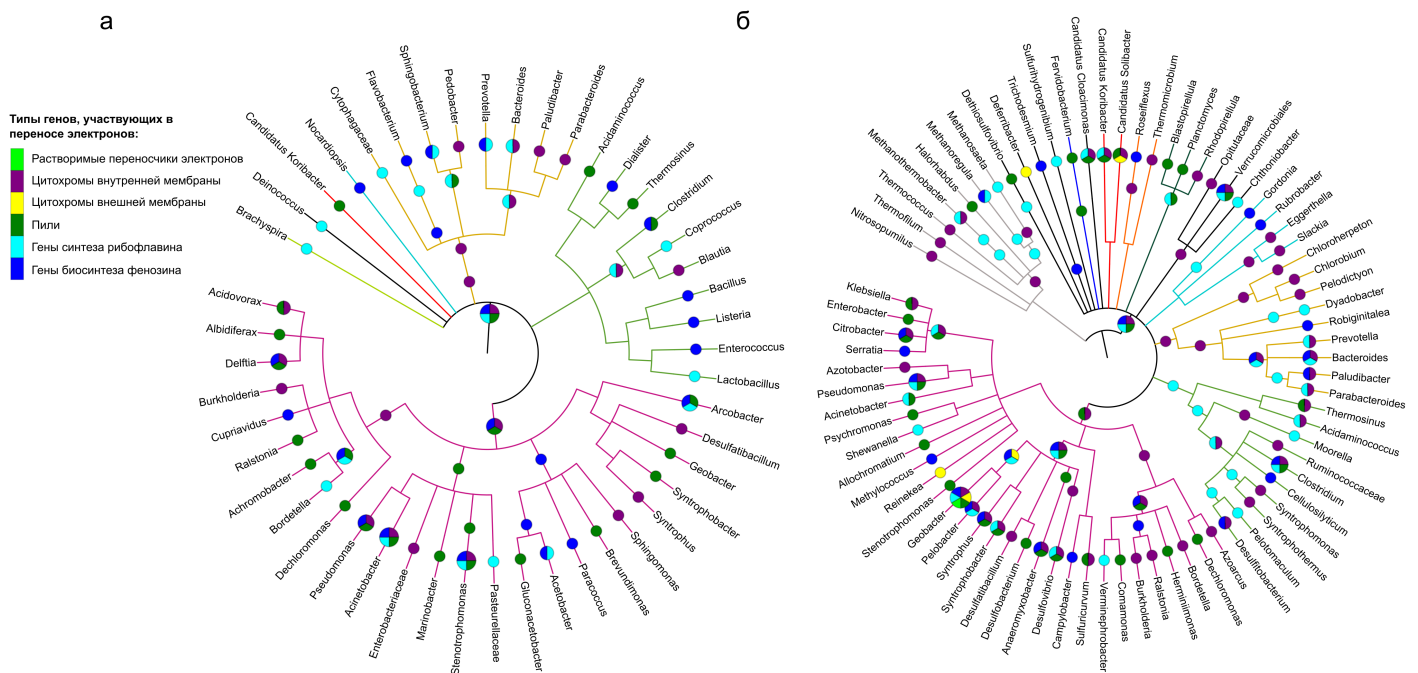


Рисунок 2. Распределение гомологов генов, вовлечённых в перенос электронов, по разным таксономическим группам для (а) МТЭ UK и (б) МТЭ JP. Ветви окрашены в соответствии с типом (Рисунок 1), круги на ветвях показывают наличие белков данного типа в данной таксономической группе.

Мы показали, что состав исходного активного ила значительно различался для двух МТЭ. В МТЭ UK (Рисунок 1а) встречались представители *Bacteroidetes*, *Firmicutes*,

Proteobacteria и *Actinobacteria*, причём *Bacteroidetes* составляли более половины сообщества. Самой распространенной группой в анаэробном активном иле из МТЭ JP были *Proteobacteria*, но в целом состав сообщества был значительно более разнообразен и содержал значительную долю архей, а также бактерий из таких таксонов, как *Chloroflexi*, *Thermotogae*, *Spirochaetes*, *Planctomycetes*, *Acidobacteria*, *Verrucomicrobia*, *Deinococcus-Thermus* и *Cyanobacteria* (Рисунок 1в). Оба исходных сообщества претерпели структурные изменения после 70-90 дней работы МТЭ, но преобладающие группы остались те же. Таксономический анализ показал, что *Bacteroidetes*, *Firmicutes* и *Proteobacteria* встречаются в разных образцах активного ила и преобладают в МТЭ, которые используются для очистки промышленных отходов разного происхождения.

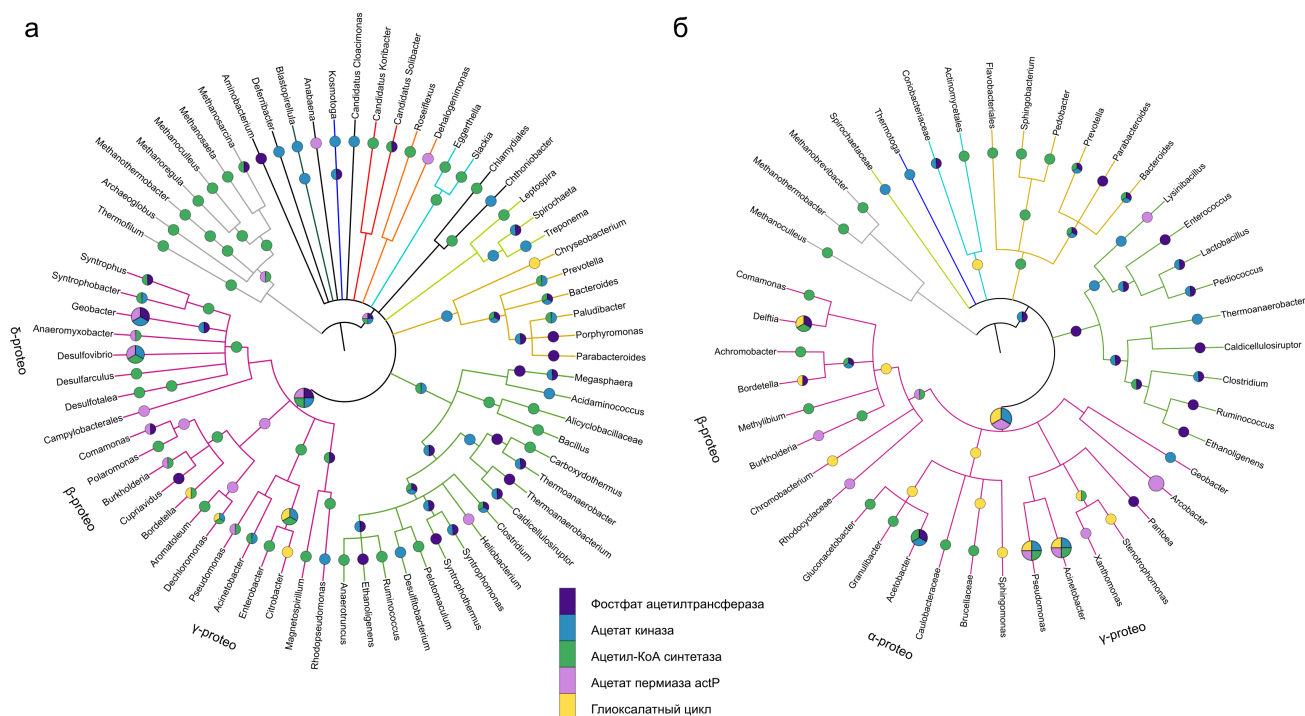


Рисунок 3. Метаболизм ацетата микробными сообществами с анода МТЭ. (а) МТЭ JP и (б) МТЭ UK. Ветви дерева окрашены в соответствии с типом, к которому относятся бактерии (см. Рисунок 2.2), а круги на ветках показывают присутствие определённых ферментов в данной группе бактерий.

Последовательности, кодирующие белки, похожие на образующие пили, встречались у *Proteobacteria*, особенно у *Deltaproteobacteria* и *Gammaproteobacteria*, в обоих биореакторах, причем в JP-mfc эти гены в основном принадлежали *Deltaproteobacteria*, а в UK-mfc – *Gammaproteobacteria*. Образец JP-mfc, но не UK-mfc, содержал гомологи *pilA* из *Geobacter* (Рисунок 2). Гены растворимых переносчиков электронов были найдены в обеих биоплёнках и были представлены у *Bacteroidetes*, *Firmicutes* и *Proteobacteria*. Эти гены имеют более широкую таксономическую представленность в МТЭ JP, где эти гены были найдены в *Deltaproteobacteria*, *Verrucomicrobia* и *Archaea*.

Переносчики электронов в *Alphaproteobacteria* были найдены только в МТЭ UK.

Для сообществ, характерных для анодов МТЭ, мы изучили распределение генов, важных для биodeградации преобладающих компонентов сточных вод, таких как ацетат, целлюлоза, крахмал и белки. В МТЭ JP эти пути были представлены у более разнообразного набора микроорганизмов, что, возможно, связано с большим общим разнообразием сообщества в этом МТЭ (Рисунок 3). Важнейшие ферменты катаболизма ацетата (ацетил-КоА синтаза, фосфат ацетилтрансфераза, ацетат киназа, ацетат пермеаза и ферменты глиоксалатного пути) встречаются в основном у *Proteobacteria*, *Firmicutes* и *Archaea* (Рисунок 3). Ферменты анаэробного разложения целлюлозы (эндоглюкоза, бета-глюкозидаза и экзоглюконаза) встречаются в основном у представителей *Firmicutes*, что позволяет предположить, что именно представители этой группы вносят основной вклад в разложение целлюлозы в обоих МТЭ. К разложению крахмала способны представители *Firmicutes*, *Bacteroidetes* и *Proteobacteria* в обоих МТЭ, тогда как в МТЭ JP крахмал могут разлагать еще и представители *Archaea*.

Горизонтальный перенос генов и эволюция геномов метаносарцин. Помимо обсуждавшегося выше постепенного изменения сообщества под воздействием внешних условий, существуют и другие механизмы, которые позволяют конкретным видам получить преимущество в меняющихся условиях. Это возможно благодаря активному обмену генетической информацией внутри микробных сообществ. Примером такой адаптации может служить, например, быстрое распространение генов устойчивости к антибиотикам.

Ключевую роль в распространении генетической информации и эволюции микроорганизмов играет горизонтальный перенос генов (ГПГ). ГПГ позволяет бактериям и археям быстро получать новые метаболические функции и адаптироваться к изменяющейся среде. Большая часть генов, вовлечённых в ГПГ, ассоциированы с патогенезом, симбиозом, метаболизмом и устойчивостью к антибиотикам. ГПГ происходит как благодаря способности некоторых бактерий к естественной трансформации одноцепочечной или двуцепочечной ДНК, так и во время конъюгации или за счёт активности мобильных генетических элементов, таких как плазмиды или фаги.

Интересной группой для изучения ГПГ являются метаногенные археи, в частности, археи из семейства *Methanosarcinaceae*. Геномы представителей рода *Methanosarcina* spp. являются одними из самых больших в надцарстве архей. Предполагается, что это связано с активным горизонтальным переносом генов (ГПГ) из бактерий. Считается, что гены бактериального происхождения в *Methanosarcinales* вовлечены в центральный метаболизм, особенно в пути синтеза сахаров, метаболизм серы и фосфатов и транспорт мелких молекул. Считается, что горизонтально перенесенные гены позволили *Methanosarcina* spp. приспособиться к очень широкому спектру экологических ниш. Оценки уровня горизонтального переноса в этой группе

организмов сильно расходятся, поэтому интерес представляет как аккуратная оценка количества перенесённых генов, так и изучение свойств таких генов.

ГПГ	Количество ГОб	Количество генов	Генов в МА	Генов в МВ	Генов в ММ
В общего предка <i>Methanosarcinaceae</i>	72	178	66	53	59
В общего предка <i>Methanosarcina</i>	143	351	141	128	82
В один из видов <i>Methanosarcina</i>	-	57	14	33	10
Всего	215	586	221	214	151

Таблица 1. Общее количество ГПГ в *Methanosarcina* spp. МА – *M. acetivorans*, МВ – *M. barkeri*, ММ – *M. mazei*.

В **главе 3** приводится подробный анализ ГПГ в семействе *Methanosarcinaceae* и рассматриваются особенности генов, которые были получены от бактерий. Нам удалось обнаружить 349 генов из 143 ГОб, которые вероятнее всего были перенесены из бактерий в общего предка всех *Methanosarcina*, гены еще из 72 ГОб были перенесены в общего предка *Methanosarcinaceae*. Отдельно мы проанализировали синглтоны, т.е. гены, не вошедшие в ГОб и характерные только для конкретного вида микроорганизмов. Такой анализ позволил описать недавние события ГПГ. Нам удалось показать, что 14 генов было недавно горизонтально перенесено в *M. acetivorans*, 33 гена в *M. barkeri* и 10 в *M. mazei* (Таблица 1). В случае *M. barkeri* нам удалось обнаружить случай ГПГ, включающий целый оперон из четырёх генов, связанных с системой CRISPR.

Количество случаев ГПГ, произошедших на уровне последнего общего предка всех *Methanosarcina* или *Methanosarcinaceae*, которое мы предсказали в этой работе, значительно отличается от оценок, полученных ранее другими авторами. Для того, чтобы понять с чем это связано, мы воспроизвели состав базы данных биологических последовательностей в разные годы, и повторили наш анализ на этих «исторических» базах. Значительное изменение числа предсказанных случаев ГПГ в *Methanosarcina* произошло в 2006 году, когда были отсекарованы две археи из семейства *Methanosarcinaceae*, и оказалось, что события, которые раньше считались ГПГ в общего предка *Methanosarcina*, на самом деле видимо произошли раньше в эволюции. Однако в любом случае доля ГПГ устойчиво падает, и, тем самым, влияние ГПГ на структуру генома *Methanosarcina* было переоценено в прошлых исследованиях, а полученная нами в данной работе оценка является достаточно точной.

Оказалось, что *Firmicutes* и *Proteobacteria* являются самыми частыми донорами генов: 104 ГОб содержат гены, перенесённые из *Firmicutes*, из которых для 12 удалось установить, что они были перенесены из *Clostridia*, а для 6, что из *Bacilli*; 66 ГОб получены в результате ГПГ из *Proteobacteria*; 12 из *Planctomycetes*; 6 из группы

Bacteroides/Chlorobi; 6 из *Actinobacteria*; 9 из *Cyanobacteria* (Рисунок 4).

Мы показали, что половина горизонтально перенесённых генов ассоциированы с метаболизмом. Если бы перенос происходил равномерно из всех функциональных категорий, то распределение горизонтально перенесённых генов было бы похоже на распределение количества КОГ для каждой функциональной категории, тем не менее мы видим, что эти распределения не совпадают (Рисунок 5). Среди перепредставленных категорий оказались защитные механизмы (группа D), все группы, ассоциированные с метаболизмом, за исключением липидного.

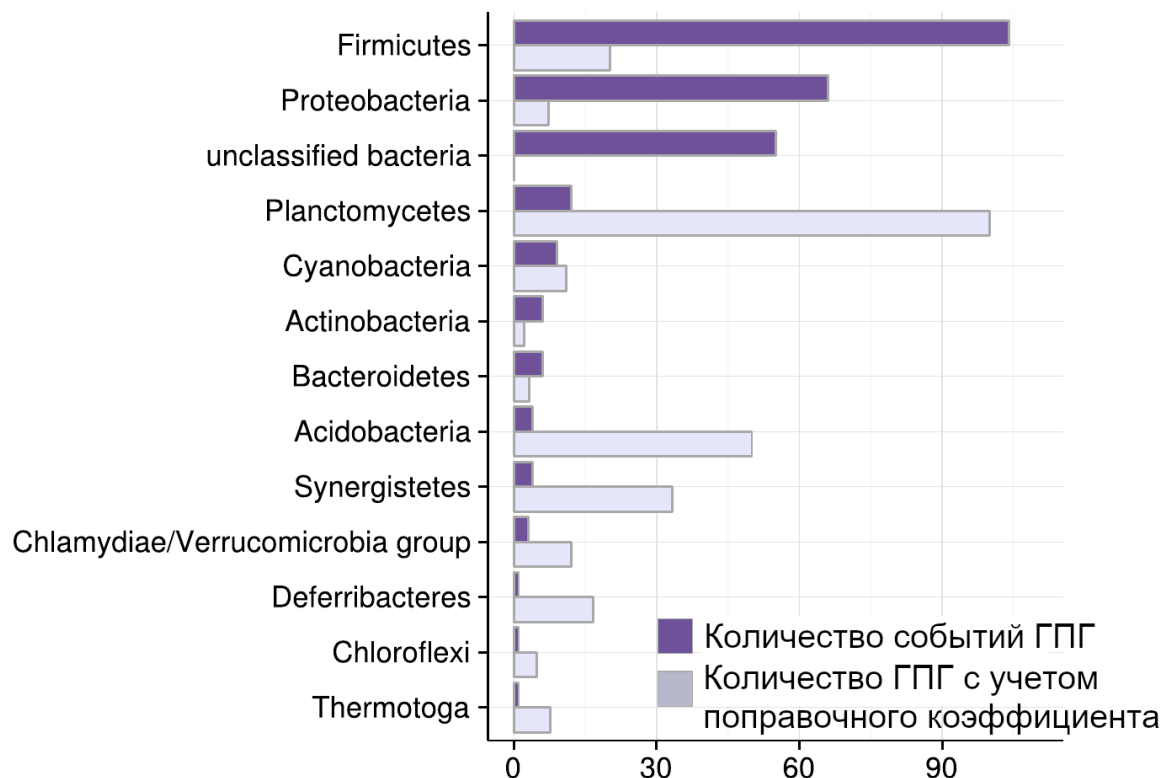


Рисунок 4. Таксономическое разнообразие перенесённых генов. Количество событий ГПГ для каждого донора показано фиолетовым, значения с учётом поправочного коэффициента представлены светло-фиолетовым цветом.

Более половины горизонтально перенесённых генов, которые удалось отнести к какому-либо функциональному классу, являются или ферментами, или транспортёрами (187 и 67 соответственно). Среди всех транспортёров и ферментов доля горизонтально перенесённых генов составляет 12-15% и 9-11%, соответственно (Рисунок 5б). Тем самым, ГПГ, ассоциированных с транспортом, – чуть более частое событие, чем перенос ферментов, эта разница является статистически значимой для всех геномов, но не значима для каждого генома в отдельности.

Из всех оперонов, содержащих горизонтально перенесенные гены, 10 оперонов являются общими для всех *Methanosarcina* spp. В целом гены, полученные в результате ГПГ, чуть реже включены в опероны, чем гены, характерные для архей (Рисунок 5в).

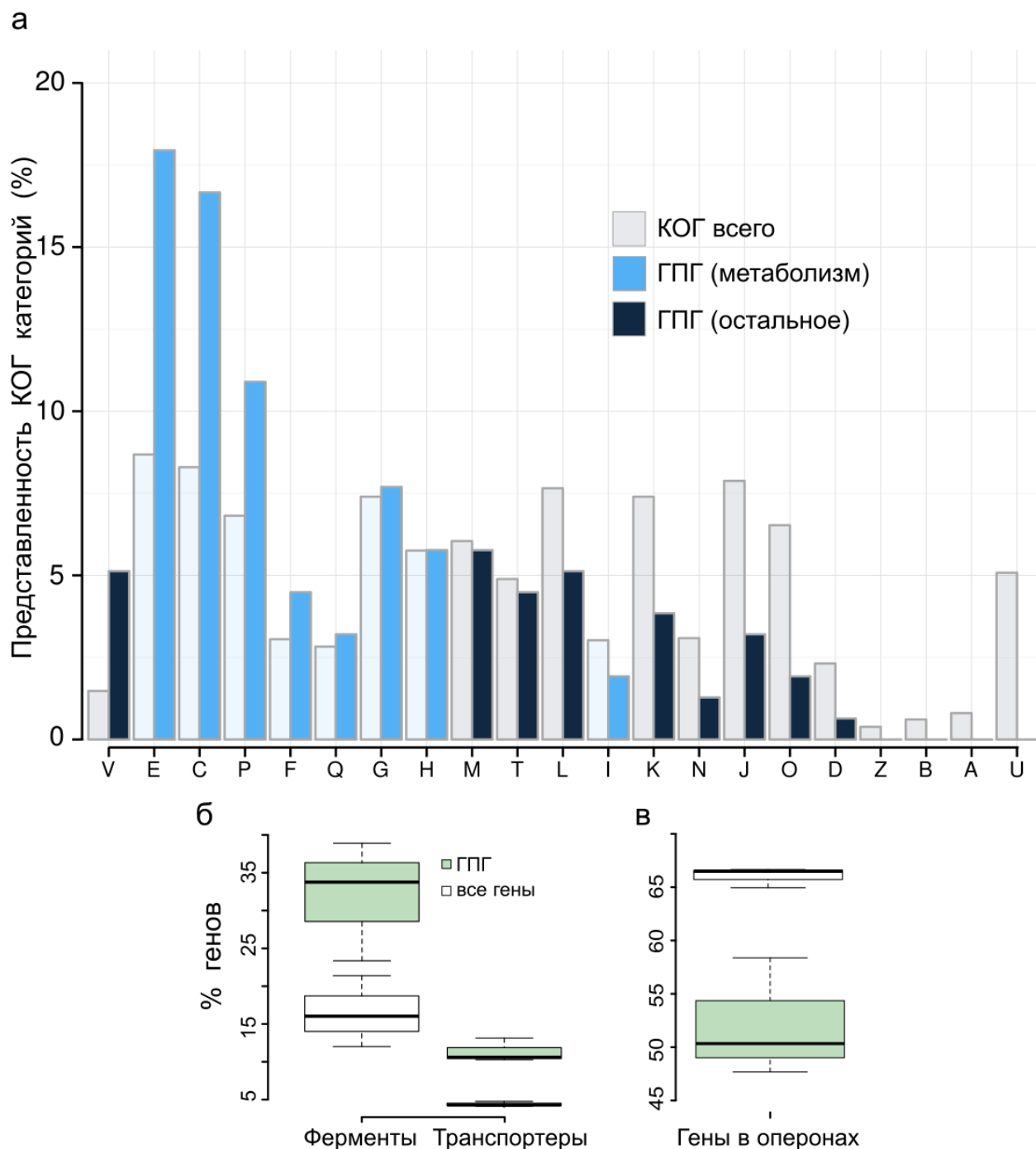


Рисунок 5. Функциональное распределение горизонтально перенесённых генов и их оперонная структура. (а) Представленность категорий КОГ среди горизонтально перенесённых генов в *Methanosarcinaceae* и в геномах бактерий и архей в целом. (б) Доля ферментов и транспортеров среди горизонтально перенесённых генов и в геномах в целом (см. таблицу 5). (в) Доля генов, включённых в опероны среди генов, полученных в результате ГПГ, и всех генов генома. (г) Диаграмма Венна, показывающая количество КОГ, содержащих транскрипционные факторы, характерные для бактерий (голубой) и для архей (жёлтый).

Сравнительный анализ геномов *Holospira* spp. – внутриядерных симбионтов инфузорий. Размеры бактериальных сообществ могут сильно варьировать, но обычно состоят из десятков видов. Крайний вырожденный случай сообществ – это сообщества состоящие всего из нескольких, а иногда и одного вида прокариот. Такие сообщества обычно представлены симбиотическими бактериями и иногда археями. В таких сообществах нельзя изучать динамику, но можно отслеживать изменения в геномах, способствующие приспособлению к определённому местообитанию. Размер популяции симбиотических сообществ сильно ограничен и сильно меняется во времени.

Симбиоз с бактериями характерен для многих эукариот. Бактерии-эндосимбионты хорошо изучены в растениях, насекомых и некоторых других организмах. В соответствии с первоначальным определением, симбиотические взаимоотношения не предполагают обязательной выгоды для организма-хозяина; симбионт использует хозяина как среду для жизни. Тем не менее, в ответ организм-хозяин может получать какие-либо необходимые соединения, которые он не может синтезировать сам.

Эндосимбиотические бактерии обычно имеют ряд характерных особенностей, таких как компактный геном, низкий GC-состав и короткие межгенные расстояния. Для таких бактерий также характерны минимальный приток новых генов, значительный генетический дрейф, возникающий из-за малого размера популяций, и повышенная скорость эволюции, возникающая из-за ослабленного отбора. Все это ведёт к образованию псевдогенов или полной потере генов и снижению метаболического потенциала. Часто наблюдаемая для симбиотических бактерий потеря генов, связанных с репарацией, ведёт к ещё большему увеличению скорости мутирования и к снижению GC-состава. Стабильные условия внутри организма хозяина не требуют сложных регуляторных систем, а тем самым не требуются длинные межгенные участки.

Большая часть эндосимбиотических бактерий могут передаваться только вертикально, тогда как представители рода *Holospira* способны передаваться от хозяина к хозяину. Все бактерии рода являются эндосимбионтами инфузорий рода *Paramecium*, причём каждый вид способен заражать только определённое ядро (макро- или микронуклеус) узкого набора видов.

В **главе 4** приводятся результаты сравнительно-геномного анализа представителей рода *Holospira*, включая геном отсеквенированного нами вида *Holospira curviuscula*, макронуклеарного эндосимбионта *Paramecium bursaria*. Полученная нами сборка генома *H. curviuscula* состоит из 152 скэффолдов (210 контигов) с N50 ~39 тысяч п.о. Общая длина сборки составляет 1.7×10^9 нуклеотидов, а GC-состав равен 37.6%. Сборка содержит 1594 гена, из которых 1555 белок-кодирующие. Нам удалось определить функции 683 генов. Анализ 16S рРНК показал, что собранный нами геном действительно принадлежит *H. curviuscula*. Более того, 16S рРНК *H. curviuscula* больше похожа на 16S рРНК *Holospira acuminata*, другого эндосимбионта *P. bursaria*, чем на 16S рРНК *Holospira* из другой инфузории – *P. caudatum* (Рисунок 6).

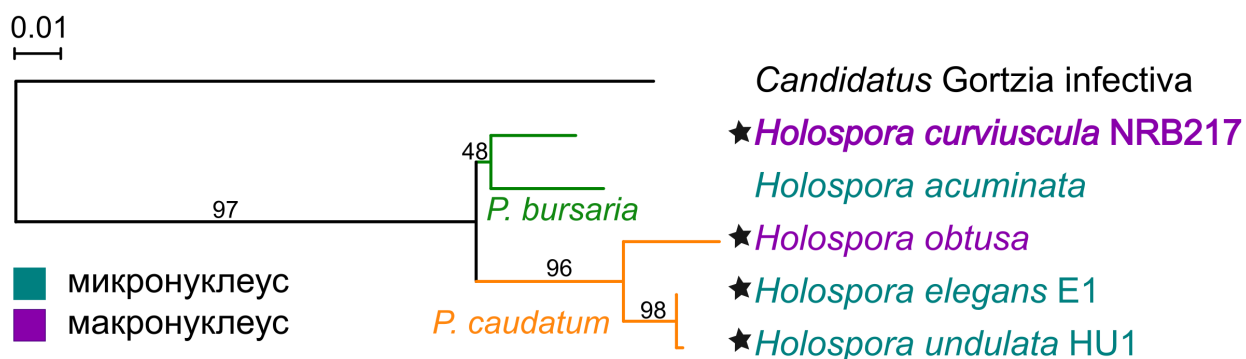


Рисунок 6. Филогенетическое древо *Holospora* spp. построено методом наибольшего правдоподобия при помощи программы PhyML. Фиолетовым цветом отмечены эндосимбионты, характерные для макронуклеуса инфузорий, голубым – для микронуклеуса. Для бактерий, отмеченных звёздочками, известны геномные последовательности.

Ни одна из бактерий из рода *Holospora* с известной последовательностью генома не способна синтезировать какие-либо аминокислоты. Частичные пути синтеза сохранились только для триптофана, лизина, глутамата и для превращения L-аланина в D-аланин, который является необходимым компонентом клеточной стенки (Рисунок 7а). Это означает, что все аминокислоты должны транспортироваться извне, т. е. голоспора должна получать их от хозяина. Голоспоры, как и риккетсии, не способны синтезировать хорионат и, видимо, вынуждены импортировать его или его производные извне.

Представители *Holospora*, как и риккетсии, не способны синтезировать пурины и пиримидины (Рисунок 7б), но способны превращать УТФ в ЦТФ. Тем самым, получается, что и все нуклеозид трифосфаты должны быть получены от хозяина. С другой стороны, *Holospora* имеет рибонуклеотид редуктазы, т. е. способна превращать рибонуклеотиды в дезоксирибонуклеотиды и наоборот.

Известные геномы представителей рода *Holospora* не содержат генов, связанных с получением энергии. В частности, у голоспор отсутствуют ферменты гликолиза, за исключением фосфоглицератмутаза (Рисунок 7в), все ферменты цикла Кребса за исключением малатдегидрогеназы, и нет F_1F_0 -АТФазы. Б. От пентозо-фосфатного пути у голоспор сохранилась только ферменты неокислительного этапа (Рисунок 7в), т. е. набор реакций, позволяющих превращать сахара друг в друга, а не получать НАДФН. При этом все голоспоры сохранили пируват-дегидрогеназу, т. е. способны превращать пируват в ацетил-КоА и далее в ацетоацетил-КоА и ацетоацетат с выработкой АТФ (Рисунок 7г). У всех голоспор есть набор рибонуклеотид-редуктаз, которые не только обеспечивают возможность преобразования рибонуклеотидов в дезоксирибонуклеотиды и наоборот, но и позволяют им использовать нуклеотиды или дезоксинуклеотиды в качестве источника энергии.

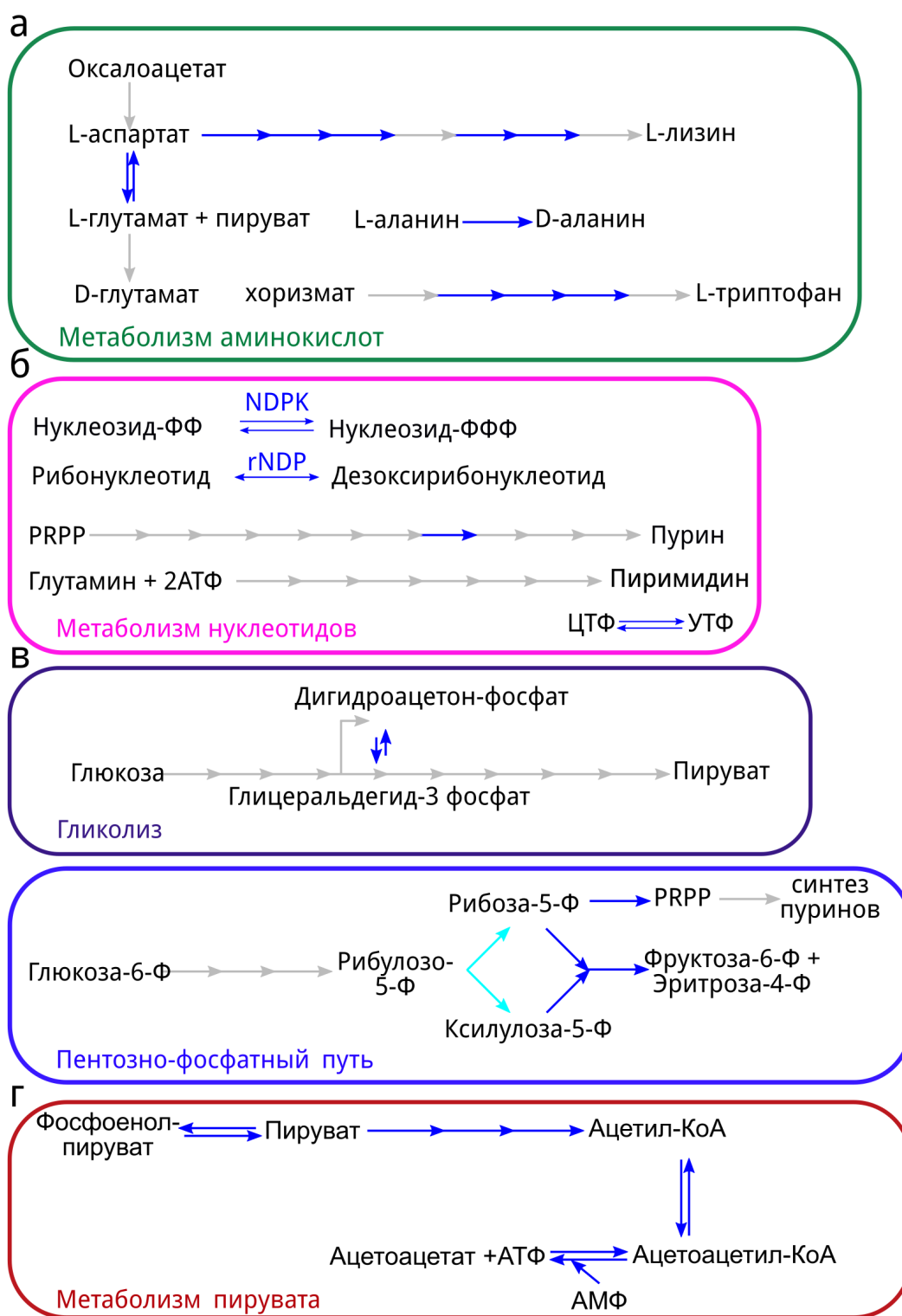


Рисунок 7. Некоторые особенности метаболизма *Holospora*: (а) метаболизм аминокислот; (б) метаболизм пуринов и пиримидинов; (в) энергетический метаболизм (сверху – гликолиз, снизу – пентозо-фосфатный путь); (д) метаболизм пирувата. Синими стрелками отмечены те реакции, для которых были найдены ферменты во всех геномах *Holospora* spp.; голубым – ферменты отсутствуют у *H. curviuscula*, но есть в геномах *H. undulata* и *H. obtusa*. PRPP – фосфориборибозил пиррофосфат; NDPK – нуклеозид-дифосфат киназа; rNDP – рибонуклеотид редуктаза.

живых организмов. Бактериальная рибосома состоит из малой (30S) и большой (50S) субъединиц, которые вместе образуют частицу размером 70S. Субъединицы рибосомы состоят из множества белков и трёх молекул рРНК – 16S, являющейся частью малой субъединицы, и 23S и 5S, образующих большую субъединицу. Основные каталитические функции рибосомы: образование пептидной связи, распознавание кодона и протягивание мРНК – выполняют молекулы рРНК. Более того, молекулы рРНК образуют остов, на который крепятся рибосомные белки (р-белки). Рибосома *Escherichia coli* состоит из 21 белка 30S субъединицы (S1 – S21) и 33 белков 50S субъединицы (L1 – L36). Роль р-белков заключается в стабилизации структуры органеллы и регуляции рибосомной активности. Несмотря на то, что ключевую роль в биосинтезе белков играет рРНК, состав р-белков очень консервативен во всех бактериях. Более того, 33 р-белка консервативны во всех живых организмах.



Рисунок 9. Тепловая карта, показывающая присутствие рибосомных белков в геномах изучаемых бактерий. Рибосомные белки расположены в строках и сгруппированы по субъединицам. Геномы расположены в столбцах, и сгруппированы по типам. Отсутствующие белки отмечены тёмно-синим цветом. Часто пропадающие белки (отсутствует как минимум в 19 штаммах из разных типов) отмечены красным цветом.

Однако, известно, что геномы некоторых бактерий содержат неполный набор генов рибосомных белков. Наименее полные рибосомы были обнаружены у эндосимбионтов насекомых с крайне редуцированными геномами.

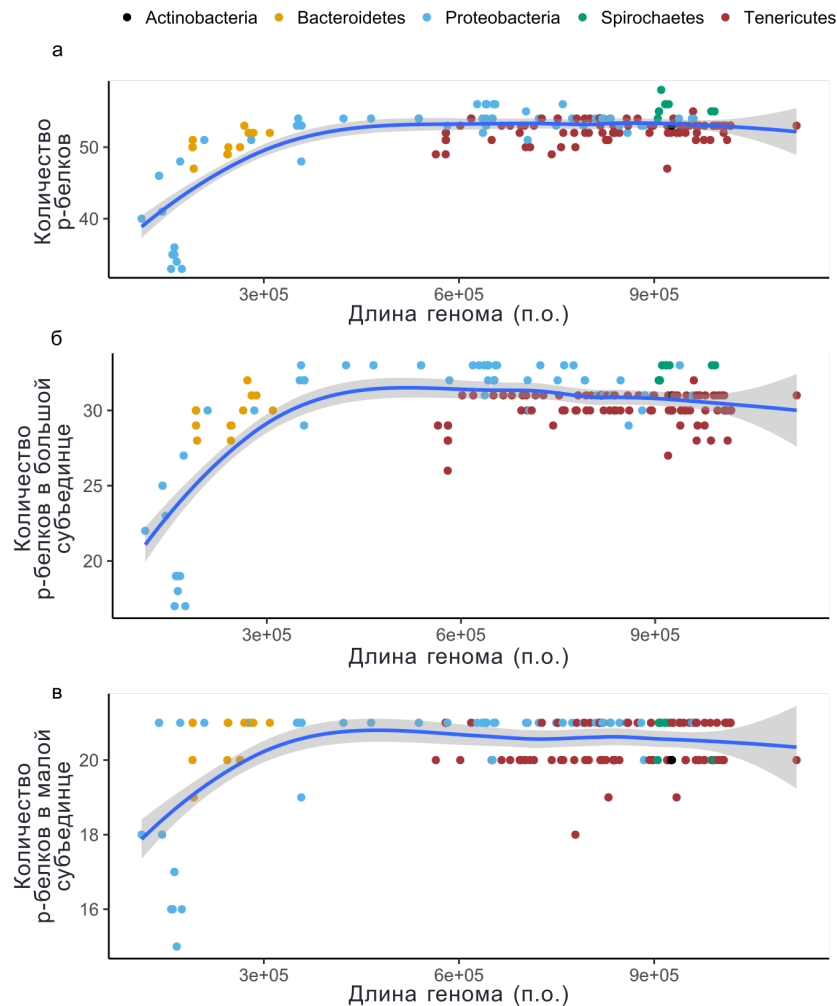


Рисунок 10. Зависимость количества рибосомных белков и длины генома (в нуклеотидах): (а) во всей рибосоме, (б) в 50S субъединице и (в) 30S субъединице. Каждая точка представляет собой один геном, точки раскрашены в соответствии с типом, цвета указаны вверху рисунка. Линия тренда построена при помощи полиномиальной регрессии по методу наименьших квадратов (синяя линия) с 95% доверительным интервалом (голубая полоса) при помощи стандартных средств пакета ggplot2 в R.

В **главе 5** рассматривается устройство рибосом в 214 бактериях с геномами размером менее 1 Мб. Для всех этих бактерий характерен симбиотический образ жизни. Мы проаннотировали заново все рибосомные белки в 214 бактериальных штаммах с короткими геномами (короче 1 Мб) из 38 родов из следующих типов: *Proteobacteria*, *Bacteroidetes*, *Spirochaetes*, *Tenericutes* и *Actinobacteria*. Все, за исключением 11 белков, отсутствуют хотя бы в одном штамме в нашей выборке. Рибосомные белки L9, L24,

L25, L29, L30, L32, L34, L36, S21 и TF пропадают чаще всего, каждый из них отсутствует как минимум в 19 штаммах из разных типов (Рисунок 9). Рибосомные белки малой субъединицы пропадают значительно реже, чем белки большой (Рисунок 9). Единственный часто исчезающий белок – это S21, он отсутствует в геномах 54 штаммов из семи родов, тогда как самый часто исчезающий белок большой субъединицы, L30, отсутствует в геномах 138 штаммов из 18 родов.

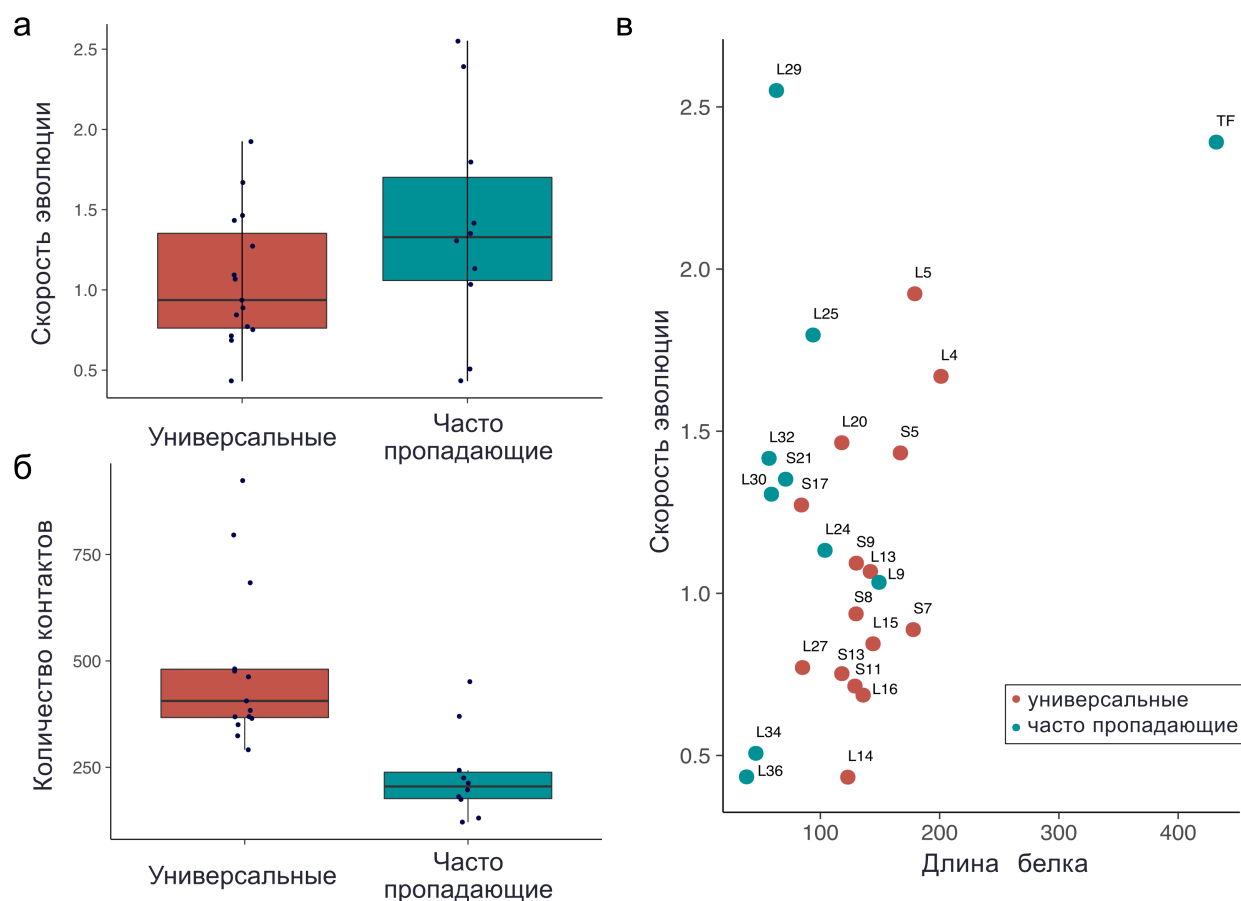


Рисунок 11. Сравнение характеристик универсальных и часто исчезающих рибосомных белков. (а) Разница в скоростях эволюции, (б) разница в количестве контактов между универсальными (отмечены красным) и часто исчезающими (отмечены синим) белками. (в) Зависимость скорости эволюции от длины белка.

Сравнение количества генов рибосомных белков в геноме и размера генома показало, что эти величины скоррелированы между собой ($r^2 = 0.6$, $p < 2.2 \times 10^{-16}$), это зависимость особенно ярко видна для крайне редуцированных геномов (Рисунок 10).

Для того, чтобы понять, есть ли какая-то связь между потерями р-белков, мы использовали коэффициент сходства Жаккара филогенетических профилей, т.е. векторов присутствия/отсутствия белков во всех исследуемых геномах. Нам удалось найти два кластера: L24 + L29 + L9 + TF и L25 + L30 + S21. Данные кластеризуются похожим

образом, если использовать коэффициент корреляции Пирсона.

Часто пропадающие белки по большей части расположены на поверхности рибосомы. Исключение из этого наблюдения составляют белки L34 и L36, которые глубоко погружены в структуру. Мы изучили, какие характеристики различаются у часто пропадающих и универсальных рибосомных белков. Для этого мы построили логистическую регрессию, пытаясь объяснить, какие из характеристик (скорость эволюции, длина белка и количество контактов с окружающими белками и рРНК) наилучшим образом объясняют имеющиеся паттерны исчезновения белков (Рисунок 11). Оказалось, что только количество контактов является значимым определяющим фактором ($p = 0.0211$).

Аннотация последовательностей анти-Шайна – Дальгарно в 16S рРНК показала, что они часто и независимо пропадают в разных таксонах, но исчезновение этих регуляторных последовательностей никак не связано с количеством потерянных р-белков ($r^2 = 0.28$, $p = 9.38 \times 10^{-5}$) или укорочением генома ($r^2 = 0.24$, $p = 0.00083$). Более того, потеря анти-ШД никак не связана с потерей какого-либо из часто пропадающих рибосомных белков.

Заключение

Исследование сообщества МТЭ полупромышленных масштабов, которые могут использоваться в реальных производствах, показало, что три бактериальные группы *Proteobacteria*, *Firmicutes* и *Bacteroidetes* составляют основу электрохимически активных биоплёнок в независимо инокулированных МТЭ, которые использовались для очистки промышленных сточных вод в разных климатических зонах. Сообщество микроорганизмов из МТЭ JP было более разнообразным и имело больший метаболический потенциал по сравнению с сообществом из МТЭ UK. МТЭ JP был остановлен почти сразу после сбора образцов с анода, тогда как МТЭ UK работал еще год, показывая стабильные результаты как очистки воды, так и производства электричества (неопубликованные данные). Это показывает, что изучаемые сообщества достаточно стабильны по производственным показателям, что может косвенно свидетельствовать о том, что и состав прокариот в них, вероятно, стабилен. То, что в случае МТЭ JP наблюдалось ещё и выделение природного газа, показывает, что возможно построить систему, которая не только бы очищала воду и генерировала электрический ток, но и производила какие-то количества природного газа усилиями одного сбалансированного сообщества микроорганизмов. На основании наших исследований можно сделать предположение, что в МТЭ UK можно было бы получить лучшие показатели производства электрического тока, если добавить в сообщество бактерии, способные к прямому переносу электронов, например, *Geobacter* spp. Большее количество метабеномных данных как для активного анаэробного ила, так и для электрогенерирующих биопленок позволит проводить более точный и направленный анализ компонентов этих сложных сообществ и разработать стабильные

и продуктивные микробные сообщества, приспособленные для очистки разных типов промышленных отходов.

Исследование потока горизонтально перенесённых генов в прокариотических сообществах на примере *Methanosarcina* spp. показало, что более ранние оценки интенсивности этого процесса были завышены. Мы считаем, что более правильная оценка такова: приблизительно 5% генов в геноме этих архей получены в результате ГПГ в последнего общего предка либо *Methanosarcina* spp., либо *Methanosarcinaceae*. Мы показали, что предпочтительно переносятся транспортные белки и ферменты. Есть ограничения на длину переносимых фрагментов, и длинные опероны переносятся редко. При этом гены переносятся без регулирующих их белков, а уже после переноса начинают экспрессироваться в составе уже имеющихся регулонов. Самые частые доноры генов – это *Firmicutes* и *Proteobacteria*.

Анализ геномов ядерных эндосимбионтов ресничных инфузорий из рода *Holospira* показывает, что приспособление к симбиотическому образу жизни связано с значительным сокращением генома и потерей целого ряда метаболических возможностей. *Holospira* spp. не способны синтезировать большинство необходимых для жизни низкомолекулярных соединений и у них полностью отсутствует весь центральный энергетический метаболизм. Даже эндосимбионты с меньшими геномами способны сами синтезировать хотя бы какие-то аминокислоты или сохранили какие-то фрагменты центрального метаболизма; представители же рода *Holospira* вынуждены получать все аминокислоты из внешней среды, т. е. от хозяина. *Holospira* не способны к гликолизу, у них отсутствуют путь Энтнера–Дудорова и пентозофосфатный путь. У голоспор нет также никаких ферментов цикла Кребса, что необычно для *Rickettsiales*, потому что даже у *Rickettsia* spp., бактерий с крайне редуцированным геномом, он сохраняется. Мы предполагаем, что *Holospira* не просто зависит от хозяина в получении энергии, но использует в качестве источника энергии нуклеотиды или рибонуклеотиды. Заражение голоспорами может либо увеличивать, либо понижать жизнеспособность инфузории хозяина в зависимости от условий, но молекулярные механизмы этого процесса не до конца понятны. Добавление к анализу отсеквенированной в нашей работе бактерии *Holospira curviuscula* позволило нам изучить не только особенности голоспор в одном определённом хозяине, но и найти ряд белков, потенциально важных для заражения макронуклеуса. Помимо этого, мы обнаружили ряд генов, специфичных для *Holospira* spp. и отсутствующих у близких к ним представителей отряда *Rickettsiales*, которые могут играть роль в образовании репродуктивной формы и развитии инфекции. Дополнительный экспериментальный анализ обнаруженных нами генов может помочь разобраться в процессе инфицирования голоспорой хозяина.

Для изучения функциональных изменений клетки под воздействием симбиотического образа жизни была составлена выборка из 214 бактерий с редуцированными геномами (короче 1 Мб) из разных таксономических групп. На этой выборке мы исследовали

изменения структуры важной клеточной органеллы – рибосомы. Нам удалось показать, что состав рибосом стабилен, хотя отдельные рибосомные белки могут отсутствовать. При этом сокращение генома приводит и к потере тех белков, которые обычно считаются консервативными. Рибосомные белки L34 и L36, классифицированные ранее как практически универсальные, часто отсутствуют в нашей выборке. Несмотря на то, что эти белки считаются необходимыми, они могут не влиять на жизнеспособность в условиях симбиоза. Отсутствие части белков, по всей видимости, снижает эффективность работы рибосом, но не полностью её нарушает, позволяя им функционировать в условиях стабильной окружающей среды, характерной для симбиотических бактерий. На нашей выборке видно, что чаще отсутствуют белки большой субъединицы, а также белки, находящиеся на поверхности рибосомы и имеющие в среднем меньше контактов, чем универсальные белки. Часто потери происходят целыми функциональными блоками; в частности, потеря триггер-фактора приводит к потере блока из трех белков, связанных с регуляцией трансляции. Потеря последовательности анти-Шайна – Дальгарно не связана со степенью редукции генома и с потерями рибосомных белков.

Выводы

1. Показано, что в микробных топливных элементах (МТЭ) присутствуют бактерии, способные как к прямому, так и к опосредованному переносу электронов на анод. Изученные сообщества из двух МТЭ способны достаточно эффективно очищать сточные воды. Обнаружено, что эффективная работа МТЭ может осуществляться разными микробными сообществами, где одни и те же функции выполняют микроорганизмы разных таксономических групп.
2. Вклад горизонтального переноса генов (ГПГ) в эволюцию *Methanosarcina* spp. ранее был переоценён. Показано, что на самом деле только около 5% генов было получено в результате ГПГ. Обнаружено, что гены, полученные в ходе ГПГ от бактерий, обычно переносятся без соответствующих транскрипционных регуляторов, и часто этот процесс приводит к переносу отдельных генов, а не целых оперонов. Чаще всего переносятся гены, связанные с метаболизмом, в особенности гены транспортёров.
3. Установлено, что *Holospira* spp. – ядерные эндосимбионты ресничных инфузорий – не способны синтезировать большинство необходимых для жизни низкомолекулярных соединений, таких как, например, аминокислоты, и у них полностью отсутствует весь центральный энергетический метаболизм. В качестве основного источника энергии *Holospira* spp. используют нуклеотиды, которые получают от хозяина. Ряд генов, специфичных для *Holospira* spp. и отсутствующих у близких к ним представителей отряда *Rickettsiales*, могут играть роль в образовании репродуктивной формы и развитии инфекции.
4. Показано, что состав рибосом в 214 бактериях с редуцированными геномами стабилен, при этом отдельные рибосомные белки могут отсутствовать. Обнаружено, что чаще отсутствуют белки большой субъединицы, а также белки, находящиеся на поверхности рибосомы и имеющие в среднем меньше контактов, чем универсальные белки. Потери происходят целыми функциональными блоками; в частности потеря триггер-фактора приводит к потере блока из трёх белков, связанных с регуляцией трансляции. Установлено, что потеря последовательности анти-Шайна – Дальгарно не связана со степенью редукции генома и с потерями рибосомных белков.

Список публикаций по теме диссертации

По теме диссертации опубликовано три статьи в рецензируемых международных научных журналах, входящих в основные библиометрические базы данных (PubMed, WoS и Scopus):

1. Kiseleva L. *, Garushyants S. K. *, Ma H., Simpson D. J. W., Fedorovich V., Cohen M. F., Goryanin, I. Taxonomic and functional metagenomic analysis of anodic communities in two pilot-scale microbial fuel cells treating different industrial wastewaters // *Journal of Integrative Bioinformatics*. 2015. Vol. 12, № 3. P. 1–15.
2. Garushyants S.K., Kazanov M.D., Gelfand M.S. Horizontal gene transfer and genome evolution in *Methanosarcina* // *BMC Evolutionary Biology*. 2015. Vol. 15, № 102. P. 1-14.
3. Garushyants S.K., Beliavskaia A. Y., Malko D. B., Logacheva M. D., Rautian M. S., Gelfand M. S. Comparative genomic analysis of *Holospira* spp., intranuclear symbionts of *Paramecia* // *Frontiers in Microbiology*. 2018. Vol. 9. P. 1-11.

Кроме того, результаты работы опубликованы в сборниках тезисов международных и российских конференций:

1. Garushyants S., Kazanov M. Horizontal gene transfer and genome evolution in *Methanosarcina* // *Proceedings of the International Moscow Conference on Computational Molecular Biology (MCCMB'11)*, Moscow, Russia, July 21-24, P.118. <http://mccmb.belozersky.msu.ru/2011/mccmb11.pdf>
2. Гарушянц С.К., Казанов М.Д., Гельфанд М.С. Исследование горизонтальных переносов генов и эволюции архей рода *Methanosarcina* // *Информационные технологии и системы 2013*, Калининград, Россия, 1-6 сентября 2013. ISBN 978-5-901158-23-4.
3. Garushyants S., Kiseleva L., Ma H., Simpson D.J., Fedorovich V., Cohen M.F., Goryanin I. Comparative metagenomic profiling of two pilot-scale microbial fuel cells treating industrial wastewaters // *Proceedings of the International Moscow Conference on Computational Molecular Biology 2015 (MCCMB'15)*, Moscow, July 16-19, 2015. ISBN 978-5-901158-27-2. <http://mccmb.belozersky.msu.ru/2015/proceedings/abstracts/120.pdf>
4. Garushyants S. Functional analysis of *Holospira* spp. - nuclear endosymbionts of *Paramecium* // *Информационные технологии и системы 2016*, Санкт-Петербург, Россия, 25-30 сентября 2016. <http://itas2016.iitp.ru/pdf/1570282334.pdf>
5. Николаева Д., Гарушянц С.К. Эволюция рибосомы у бактерий с короткими геномами. // *Информационные технологии и системы 2017*, Уфа, Россия, 14-18 сентября 2017. <http://itas2017.iitp.ru/papers/>