

Федеральное государственное бюджетное учреждение науки  
Институт проблем передачи информации им. А.А. Харкевича  
Российской академии наук

На правах рукописи



Гарушнянц Софья Константиновна

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

03.01.09 – математическая биология, биоинформатика

Диссертация на соискание ученой степени кандидата биологических наук

Научный руководитель:  
доктор биологических наук, профессор  
Михаил Сергеевич Гельфанд

Москва, 2019

## Оглавление

Введение.....	5
Глава 1. Обзор литературы.....	12
1.1 Разнообразие прокариот.....	12
1.2 Сообщества микроорганизмов.....	14
1.3 Микробные топливные элементы.....	19
1.4 Распространение генетической информации внутри микробных сообществ.....	22
1.4.1 <i>Methanosarcina</i> .....	23
1.5 Примеры крайне редуцированных сообществ – эндосимбионты.....	24
1.5.1 <i>Holospira</i> .....	27
1.5.2. Редукция геномов у эндосимбионтов.....	28
Глава 2. Таксономический и функциональный анализ микробных сообществ анодов микробных топливных элементов.....	31
2.1 Материалы и методы.....	31
2.1.1 Режим работы МТЭ.....	31
2.1.2 Сбор образцов, выделение ДНК и секвенирование.....	32
2.1.3 Анализ метагеномных контигов.....	33
2.1.4 Филогенетический анализ.....	33
2.1.5 Функциональная аннотация.....	34
2.1.6 Анализ функций, необходимых для успешной работы МТЭ.....	35
2.2 Результаты и обсуждение.....	35
2.2.1 Работа МТЭ и сбор образцов.....	35
2.2.2 Таксономический анализ МТЭ.....	37
2.2.3 Функциональный анализ и идентификация генов, необходимых для очистки сточных вод.....	40
2.2.4 Гены, необходимые для переноса электронов на анод.....	44
Глава 3. Горизонтальный перенос генов и эволюция геномов метаносарцин.....	46
3.1 Материалы и методы.....	46
3.1.1 Геномные последовательности.....	46
3.1.2 Построение ортологических рядов.....	47
3.1.3 Идентификация ГПП.....	47
3.1.4 Филогенетические деревья для <i>Methanosarcinaceae</i> .....	49
3.1.5 Определение КОГ категорий.....	49

3.1.6	Определение функциональных классов.....	49
3.1.7	Определение оперонной структуры.....	50
3.1.8	Зависимость результатов от состава белковой базы данных.....	50
3.2	Результаты.....	50
3.2.1	Идентификация событий ГПГ.....	50
3.2.2	Реконструкция новейшей истории: воспроизведение состава базы данных последовательностей и предсказанного уровня ГПГ.....	54
3.2.3	Первый контроль: ГПГ в <i>Thermotogaceae</i> .....	55
3.2.4	Второй контроль: ГПГ в <i>Thermococcaceae</i> .....	56
3.2.5	Таксономическое разнообразие перенесенных генов.....	56
3.2.6	Функциональное разнообразие перенесенных генов.....	58
3.2.7	Оперонная структура.....	62
3.2.8	Экспрессия горизонтально перенесенных генов.....	64
3.3	Обсуждение результатов.....	64
Глава 4.	Сравнительный анализ геномов <i>Holospira</i> spp. – внутриядерных симбионтов инфузорий.....	68
4.1	Материалы и методы.....	68
4.1.1	Выращивание бактерий и выделение геномной ДНК.....	68
4.1.2	Секвенирование геномной последовательности, сборка и аннотация.....	69
4.1.3	Построение филогенетического древа <i>Holospira</i> spp.....	71
4.1.4	Анализ повторов.....	71
4.1.5	Ортологические группы и полногеномные сравнения.....	71
4.2	Результаты.....	74
4.2.1	Геном <i>Holospira curviuscula</i> .....	74
4.2.2	Сравнение геномов <i>H. undulata</i> и <i>H. elegans</i> .....	76
4.2.3	Геномы <i>Holospira</i> spp. содержат множественные повторы и последовательности профагов.....	78
4.2.4	<i>Holospira</i> spp. не способны синтезировать аминокислоты и некоторые другие важные соединения.....	80
4.2.5	<i>Holospira</i> spp. использует нуклеотиды в качестве основного источника энергии.....	82
4.2.6	Секреторные системы и потенциальные инвазины.....	83
4.2.7	Гены, специфичные для <i>Holospira</i> .....	84
4.3	Обсуждение результатов.....	85

Глава 5. Эволюция рибосом у симбиотических бактерий.....	88
5.1 Материалы и методы.....	88
5.1.1 Выборка бактериальных геномов.....	88
5.1.2 Аннотация рибосомных белков.....	96
5.1.3 Определение событий потери рибосомных белков.....	98
5.1.4 Измерение скорости эволюции.....	98
5.1.5 Подсчет количества контактов.....	98
5.1.6 Анализ рРНК.....	99
5.1.7 Статистический анализ и визуализация данных.....	99
5.2 Результаты.....	99
5.2.1 У бактерий с короткими геномами отсутствуют некоторые рибосомные белки.....	99
5.2.2 Потеря рибосомных белков зависит от степени редукции генома.....	102
5.2.3 Паттерны потери рибосомных белков.....	103
5.2.4 Часто теряющиеся белки меньше контактируют с остальными частями рибосомы по сравнению с универсальными.....	104
5.2.5 Укорочение рРНК и потери рибосомных белков.....	106
5.2.6 Потеря анти-ШД происходила много раз в разных таксонах.....	106
5.3 Обсуждение результатов.....	107
Заключение.....	110
Выводы.....	113
Благодарности.....	114
Список литературы.....	115

## Введение

### 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 привёл 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.

### **Степень разработанности темы**

Изучению отдельных геномов и метагеномов методами биоинформатики посвящено множество работ (см. обзоры [1–6]). Количество доступной информации такого рода постоянно растёт, так, в базе данных Genbank содержатся десятки тысяч полных последовательностей бактериальных геномов разной степени родства [7]. Такое огромное количество информации позволяет достаточно точно реконструировать функции отдельных генов в еще не изученных геномах или сообществах, имея только последовательности ДНК. Тем самым, можно заниматься изучением эволюции и метаболизма в немодельных организмах. К настоящему моменту многие из этих немодельных организмов остаются неизученными. Например, для интересного природного объекта – внутриядерных эндосимбионтов инфузорий – была достаточно подробно изучена морфология и некоторые особенности процесса инфицирования хозяина [8–10], тогда как метаболизм подробно не исследовался [11,12]. Другой интересный пример – эволюция базового аппарата трансляции – рибосом. Структура рибосомы была очень подробно изучена экспериментально у ряда модельных объектов [13–16]. Было установлено, что у бактерий с крайне короткими геномами рибосома может значительно редуцироваться [17], помимо этого в ряде экспериментов было показано, что в лабораторных условиях бактерии могут терять гены некоторых рибосомных белков [18–21]. Изменения структуры рибосом немодельных бактерий с редуцированными геномами не изучались, хотя именно эта группа представляет наибольший интерес. Бактерии с

редуцированными геномами – это симбиотические или паразитические организмы, на геномы которых действует более слабый естественный отбор. Изучение геномов таких организмов, принадлежащих к разным таксономическим группам, позволяет наблюдать большое количество независимых изменений в структуре рибосомы, а тем самым помогает лучше объяснить, по какой причине пропадают конкретные гены. Изучение геномов немодельных архей показало, что, по всей видимости, возникновение метаногенных архей связано с массовым горизонтальным переносом метаболических генов из бактерий [22–24]. При этом оценки уровня ГПГ в метаногенах значительно отличаются в разных исследованиях.

### **Цель 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 контролируруемыми 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.

### **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  $\chi^2$  и корреляционный анализ.

## 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 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.

## Положения, выносимые на защиту

1. 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.

2. 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.

3. *Holospora* 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 *Holospora* spp. используют нуклеотиды, которые получают от хозяина. Ряд генов, специфичных для *Holospora* spp. и отсутствующих у близких к ним представителей отряда *Rickettsiales*, могут играть роль в образовании репродуктивной формы и развитии инфекции.

4. 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.

### 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/>

# Глава 1. Обзор литературы

## 1.1 Разнообразие прокариот

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 [25,26]. Более того, позднее было показано, что эукариоты являются внутренней ветвью архей [27]. Тем самым, термин «прокариоты» не описывает конкретную таксономическую группу организмов, а объединяет все организмы для которых характерна ко-транскрипционная трансляция [28].

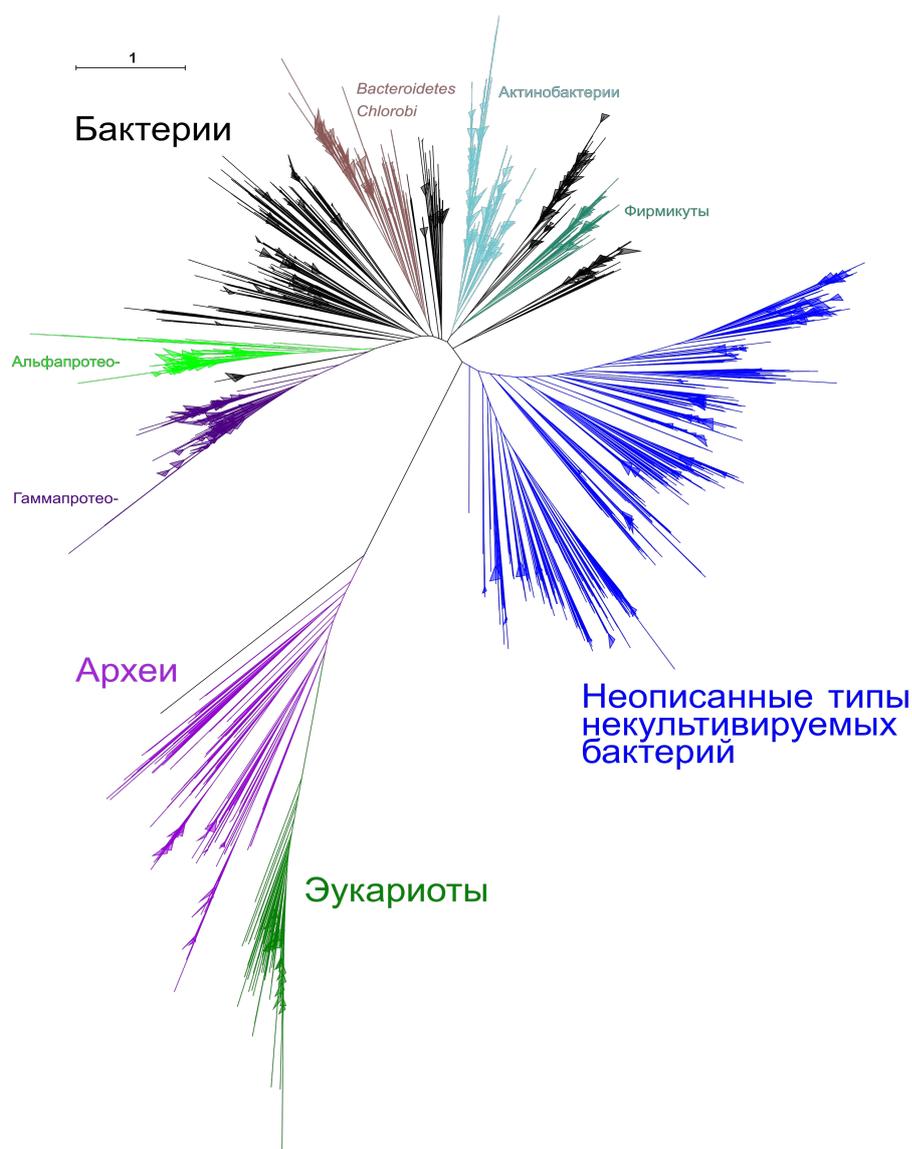
Прокариоты – самые распространённые живые существа на нашей планете (если не считать живыми существами вирусов и фагов). Считается, что количество прокариот на Земле составляет от  $4 \times 10^{30}$  до  $6 \times 10^{30}$ , что значительно превосходит количество всех других живых существ на нашей планете [29]. Эта оценка была получена в предположении, что наибольшее количество микроорганизмов обитает в мировом океане или пресных водоёмах, в почвах или донных осадках. Вклад других экологических ниш на этом фоне крайне мал. Тем самым, указанное выше число является оценкой снизу количества прокариот на нашей планете. При этом вклад бактерий в синтез органических соединений сравним с таковым растений [29].

В отличие от эукариот, многие виды которых хорошо характеризованы, наши представления о разнообразии прокариот кардинально изменилось за последние несколько лет [30]. Стало понятно, что разнообразие прокариотических форм значительно превосходит разнообразие всех других организмов (Рисунок 1.1). Эти новые классы и порядки прокариот не удавалось охарактеризовать ранее, так как имевшиеся технические решения позволяли описывать только микроорганизмы, поддающиеся культивированию [31].

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. Представители таких сообществ обладают уникальными метаболическими особенностями, которые позволили им приспособиться к конкретным условиям [32]. Многие из этих специфичных реакций не способны осуществлять никакие другие живые организмы. Так, например, азотфиксация полностью происходит за счет

деятельности различных групп архей и бактерий, тогда как эукариоты к фиксации азота не способны [33], другой важный процесс, метаногенез, происходит только благодаря дыханию нескольких групп архей [34]. Более того, метаболизм прокариот обычно значительно быстрее, чем эукариотический, тем самым, круговорот веществ за счет деятельности архей и бактерий происходит значительно быстрее [35].

Прокариоты обычно обитают в сложных сообществах, состоящих из многих видов (за редкими исключениями, которые будут рассмотрены ниже). Образование таких сообществ основывается на возникновении сложных метаболических систем, где продукты жизнедеятельности одних организмов выступают как субстраты для других.



**Рисунок 1.1.** Разнообразие живых организмов. Археи отмечены фиолетовым цветом, эукариоты – зеленым; бактерии, которые раньше не удавалось обнаружить, – синим; культивируемые бактерии – черным. Отдельные группы культивируемых бактерий дополнительно раскрашены в следующие цвета: темно-фиолетовый – *Gamma*proteobacteria (гаммапротеобактерии); салатный – *Alphaproteobacteria* (альфапротеобактерии); коричневый – *Bacteroidetes* / *Chlorobi*; голубой – *Actinobacteria* (актинобактерии); сине-зеленый – *Firmicutes* (фирмикуты). Древо адаптировано из [30].

## 1.2 Сообщества микроорганизмов

Изучение состава и метаболических возможностей сообществ микроорганизмов стало возможно благодаря развитию технологий секвенирования и появлению метагеномного подхода. Идея этого метода заключается в том, что из пробы (например, воды или почвы), выделяется тотальная ДНК, которая затем секвенируется. Возможно как секвенирование отдельных фрагментов, например, генов 16S рРНК, так и всей ДНК пробы. Из такого эксперимента в первом случае можно получить всю информацию о видовом разнообразии пробы, а во втором ещё и информацию о метаболических возможностях микроорганизмов из данного местообитания [36]. Впервые термин «метагеном» был использован при поиске новых потенциальных антибиотиков из некультивируемых почвенных бактерий [37]. Секвенирование полного метагенома впервые было сделано для оценки разнообразия микроорганизмов, обитающих в Саргассовом море [38]. На основании анализа фрагментов генов 16S рРНК удалось обнаружить как минимум 1800 разных видов бактерий, из которых 148 не удалось отнести к известным родам или видам. Более того, удалось не только описать бактериальный состав сообщества, но и оценить его метаболические возможности; было показано, что все эти бактерии вместе содержат порядка миллиона генов, из которых семьдесят тысяч не имели известных близких гомологов [38].

Важно понимать, что определение видов прокариот в метагеномных образцах является сложной задачей, равно поэтому выше обсуждаются именно филогруппы, а не виды. Стандартный подход предполагает определение филоотипов, или операционных таксономических единиц (ОТЕ), т. е. таких наборов последовательностей генов 16S рРНК, которые идентичны как минимум на 97% или 98.7%, устанавливаемый порог варьирует в разных исследованиях и второй считается более оптимальным [39]. Как ОТЕ соотносятся с признанными микробиологическими видами не до конца понятно и сильно зависит от дизайна эксперимента [40]. Ключевым выводом из исследования разнообразия сообществ

стало то, что удалось понять, что есть лишь небольшое количество видов прокариот, которые встречаются в большом количестве разных ниш (например, в разных типах почв, в кишечнике и в океанах), в то время как большое количество прокариот встречается лишь в нескольких местообитаниях [41].

Для лучшего описания метаболического потенциала конкретного сообщества можно использовать не только тотальную ДНК, но и тотальную РНК, т. е. секвенировать метатранскриптом сообщества. Обычно такой подход применяется совместно с метагеномным анализом. Так, например, удалось определить какие бактерии вносят наибольший вклад в разложение нефти, разлившейся в Мексиканском заливе [42].

Благодаря развитию метагеномики, удалось показать, что сообщества микроорганизмов из разных местообитаний различаются по составу и количеству входящих в них видов. На состав сообществ значительно влияет концентрация соли, содержание фосфатов, температура и концентрации азота [43,44]. Состав микроорганизмов в экологической нише зависит ещё и от других микроорганизмов, населяющих это местообитание. Каждый вид внутри сообщества имеет свою собственную роль и ресурсы [45]. Реконструкция метаболических сетей 155 видов бактерий с полными геномами и сравнение представленности этих видов в 124 образцах метагенома кишечника человека показало, что бактерии, которым для жизни требуются похожие наборы метаболитов, значительно реже встречаются вместе, чем бактерии с разными метаболическими возможностями [46].

Сообщества могут состоять как из нескольких десятков, так и тысяч видов [41]. Было показано, что свободноживущие сообщества обычно более разнообразны, чем симбиотические [41]. Считается, что одними из самых богатых и разнообразных являются сообщества почвы [47].

Бактерии активно взаимодействуют внутри сообществ и реагируют на внешние стимулы. Было показано, что в зависимости от среды отличаются и расстояния, на которых микроорганизмы способны взаимодействовать. Например, абсолютно неподвижные бактерии в почве могут не взаимодействовать, даже находясь на маленьких расстояниях друг от друга [48], тогда как в кишечнике животных взаимодействия иногда происходят на расстоянии нескольких метров [49]. Тем самым, определение сообщества является достаточно расплывчатым, высокое разнообразие микроорганизмов почвы, по-видимому, во многом связано с наличием большого количества малых ниш, каждая из которых населена своим набором микроорганизмов.

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 [45].

В этом разделе рассматриваются только сообщества прокариот, с основным упором на метаболические особенности и разнообразие бактерий. Понятно, что роль вирусов и эукариотических организмов может быть достаточно велика, но вклад этих организмов в общую структуру сообществ изучен хуже, и для его описания имеется меньше данных. В последние несколько лет виромы (т. е. совокупность всех вирусов сообщества) стали изучать активнее; как и в случае микробиомов, лучше всего исследованы вирусные сообщества кишечника [50–52]. Было показано, что вирусы могут значительно влиять на физиологическое состояние хозяина, даже в тех случаях, когда не приводят к болезням. Так, например, было показано, что вирусные инсулин-подобные пептиды понижают уровень глюкозы в крови у мышей [53].

В ряде исследований метагеномов из разных поверхностей тела человека, и в особенности кишечника, было показано, что индивидуальный состав сообществ может значительно отличаться [54]. Более того, было сделано предположение, что в кишечнике человека может встречаться три класса сообществ (так называемые «энтеротипы») с разными доминирующими видами [55]. Результаты последующих исследований оказались противоречивы: количество выделяемых энтеротипов составляло два [56–60], три [61–63] или четыре [64]. Было показано существование промежуточных вариантов [65]. Более того, оказалось, что состав метагенома кишечника может значительно изменяться в течение жизни, например, в раннем детстве или старости [66]. Тем не менее, исследование микробиомов приматов показало, что как минимум некоторые виды, в частности, представители семейств *Bacteroidaceae* и *Bifidobacteriaceae*, коэволюционировали с приматами на протяжении как минимум 15 миллионов лет [67]. Тем самым, состав таких сообществ в какой-то степени стабилен в эволюции. В итоге, наиболее сбалансированная точка зрения, видимо, заключается в том, что действительно существуют три основных набора микроорганизмов, которые описывают большую часть метагеномов кишечника человека, но в заметном количестве встречаются и промежуточные переходные формы [68]. Существование энтеротипов может быть частично объяснено за счет немного разных

функций, выполняемых этими сообществами, например, они отличаются по способности разлагать соединения углерода. В частности, было показано, что *Prevotella* (род, характерный для одного из энтеротипов) чаще всего встречается в микробиомах людей, употребляющих богатую клетчаткой пищу [68].

Ещё одна группа хорошо изученных сообществ – это сообщества морских микроорганизмов. Исследование морских образцов из 68 географических точек по всему миру, находящихся в эпи- и мезопелагических слоях, показало, что морские местообитания крайне разнообразны и сильно стратифицированы по глубине в первую очередь за счёт изменения температуры воды [69]. Верхние слои океана в тропическом и субтропических поясах крайне бедны с точки зрения разнообразия прокариот, а расстояния между соседними бактериями могут превышать сотни радиусов клетки. Эти местообитания почти полностью заселены представителями одного рода – *Prochlorococcus*, который является наиболее изученным морским микроорганизмом и самым распространённым фотосинтезирующим организмом на Земле. Такое разнообразие местообитаний объясняется тем, что существует большое количество групп видов, адаптированных к разной степени освещённости, и потому толща воды оказывается разбита на множество маленьких ниш, где преобладают разные штаммы [70]. Удалось показать, что перемешивание микробных сообществ океана на разных глубинах происходит за счёт населённых бактериями органических частиц, которые падают на дно [71].

Ещё одна группа относительно хорошо изученных сообществ – это бактерии и археи почвы. Они представляют особый интерес потому, что в этой среде плотность и разнообразие бактерий являются наибольшими (плотность сравнима только с концентрацией бактерий в кишечнике). Считается, что в одном грамме почвы содержится около  $10^8$  клеток прокариот [72]. Более того, микробиом почвы оказывает огромное воздействие на здоровье и разнообразие наземных животных и растений. Помимо этого, многие открытия в области микробиологии, такие как открытие некоторых классов антибиотиков, например, стрептомицина [73], фиксации азота или окисления аммония, связаны именно с работой почвенных микробиологов. Интерес представляет ещё и тот факт, что почва населена большим количеством микроорганизмов, которые невозможно культивировать. Так, в образцах почвы из Центрального парка в Нью-Йорке только 16,2% ОТЕ удалось классифицировать [74]. Только 1% наблюдаемых в микроскоп микроорганизмов в почве удаётся культивировать [75]. Почва представляет собой набор отдельных плотно упакованных экологических ниш, которые могут находиться друг от

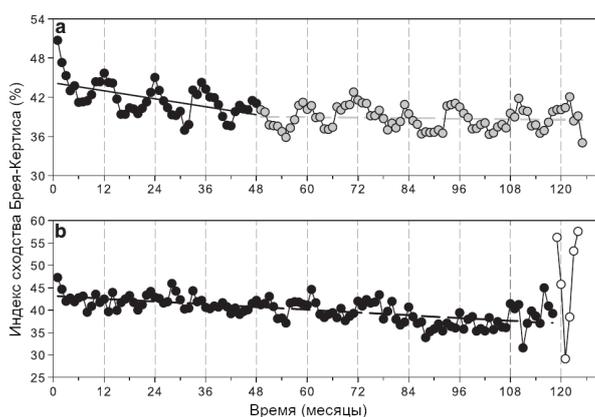
друга на расстоянии всего нескольких миллиметров, но значительно отличаться по своим абиотическим свойствам (структуре, pH, содержанию кислорода, концентрации органических веществ, солёности и концентрации азота) и по микробному составу. Даже в почвах похожего состава свойства могут значительно отличаться за счёт наличия ризосферы, обводненности или звериных нор [76]. Совсем не все прокариотические таксоны представлены в почве равномерно, большая часть бактерий и архей являются достаточно редкими, и только некоторые таксоны являются частыми, например, *Verrucomicrobia* в почвах прерий [77]. Анализ 66 уникальных образцов почв показал, что есть бактерии и археи, характерные почти для всех типов почв: самыми распространёнными являются *Acidobacteria* среди бактерий и *Thaumarchaeota* среди архей [78].

Отдельной интересной задачей является изучение динамики микробных сообществ. Такие исследования также стали возможны благодаря развитию метагеномики. Поскольку часто взаимодействие между организмами в сложных системах нельзя выявить напрямую, можно изучать изменение состава сообщества во времени или в зависимости от меняющихся абиотических факторов. Скоординированное изменение представленности разных видов говорит о том, что они, вероятно, взаимодействуют.

Изменение состава сообществ во времени активно изучалось для морских сообществ. Скоординированные изменения могут происходить очень быстро, иногда даже в течение нескольких часов [79]. Такие изменения характерны для быстро растущих бактерий, например, таких как многие *Gamma*proteobacteria и *Flavobacteriia*, а также некоторые *Alphaproteobacteria* [80]. Особый интерес представляют сезонные изменения, связанные с изменением угла падения солнечных лучей, длины светового дня и температуры. Такие изменения по-разному влияют на сообщества на разных глубинах (Рисунок 1.2).

Как уже говорилось выше, 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 [82]. При использовании

реальных сточных вод в промышленности состав прокариотического сообщества может значительно меняться в зависимости от состава стока [82,83]. 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 1.2.** Изменение разнообразия микробного сообщества во времени: (а) на глубине 0-5 м, (б) на глубине 30 метров [81].

### 1.3 Микробные топливные элементы

Одним из типов 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, тем самым переводя энергию, полученную от очистки отходов, в электричество [84].

Эти системы показали свою эффективность на уровне малых установок

лабораторного размера. В настоящий момент ведётся активная разработка более крупных систем, которые могли бы применяться в реальном производстве [85–87]. МТЭ по сравнению с классическими системами аэробной очистки вод потребляют меньше электроэнергии. Это важно, так как на очистку сточных вод приходится 1–4% всей электроэнергии расходуемой в странах Европы и США [88]. Считается, что интеграция МТЭ в систему очистки сточных вод приведёт к значительному росту эффективности биоремедиации. Дополнительным плюсом МТЭ является и то, что, в отличие от большинства биореакторов, производящих биогаз, МТЭ способны работать в широком спектре температур и не требуют специальной обработки полученных на входе органических соединений [84,89].

Основной упор в разработке более крупных МТЭ делается на более эффективный дизайн реакторов и поиск недорогих материалов, что должно сократить стоимость производства и увеличить продуктивность, необходимую для очистки больших объемов сточных вод [86,87,90–93]. Однако нельзя забывать о том, что ключевую роль в работе таких систем играет правильный состав используемых микроорганизмов [93]. Улучшение качества очистки сточных вод возможно только за счёт подбора микроорганизмов, которые будут эффективно разлагать широкий спектр органических соединений.

Для решения этой задачи можно использовать несколько подходов. Самым распространённым является тестирование отдельных культур бактерий в экспериментальных условиях [94–101]. Такой подход позволяет определить не только эффективность передачи электронов на анод, но и механизмы, по которым такой перенос происходит. Очевидным минусом является то, что тестировать в таких условиях можно только микроорганизмы, которые удаётся культивировать.

Альтернативный подход заключается в использовании методов метагеномики. Например, можно определять состав и возможности бактериальных сообществ, которые показали эффективную работу в условиях МТЭ разных размеров [102–104]. Более того, в такой постановке можно следить и за динамикой сообщества биореактора, т. е. фиксировать изменения микробиома МТЭ по сравнению с сообществом, которое было использовано для инокуляции. Отслеживание динамики позволяет решить сразу две задачи: во-первых, понять насколько микробный инокулят устойчив во времени, и, во-вторых, установить, какие микроорганизмы ответственны за производство электричества. Наиболее эффективное сообщество должно иметь много генов, участвующих в утилизации органики и во внеклеточном переносе электронов.

Бактерии, находящиеся в МТЭ, способны переносить электроны на анод как

напрямую, так и при помощи растворимых переносчиков электронов [105,106]. В качестве переносчиков электронов могут выступать самые разные органические соединения, более того, добавление искусственных соединений, таких как метилвиологен или тионин, позволяет увеличить эффективность переноса электронов [106].

Лучше всего механизмы переноса электронов описаны для представителей родов *Geobacter* (*Deltaproteobacteria*) и *Shewanella* (*Gammaproteobacteria*). У представителей рода *Geobacter* способности к переносу электронов значительно варьируют [107]. Транскрипционный анализ биоплёнок, образованных *Geobacter*, показал, что эффективность переноса отличается даже между разными слоями сообщества [108]. Пленки, состоящие только из *Geobacter anodireducens* или из смеси этих бактерий с другими, активно переносили электроны только с внешней поверхности плёнки, тогда как слой мертвых клеток, находящихся в непосредственном контакте с анодом, играл роль проводника [109]. Бактерии родов *Geobacter* и *Shewanella* для переноса электронов, полученных в результате катаболизма, сквозь клеточную оболочку на внешнюю проводящую поверхность используют цитохромы типа с, ассоциированные с мембраной [110]. Помимо этого, для представителей рода *Geobacter* был показан перенос электронов при помощи структур, похожих на пили: для переноса электронов сквозь многослойные биоплёнки использовались пили IV типа, состоящие из белка PilA [111].

У *Shewanella* перенос электронов может осуществляться через белки внешней мембраны [105] или при помощи производимых клетками этого рода растворимых переносчиков электронов [112,113]. Предполагается, что синтез переносчиков может усиливаться под воздействием электрического тока в МТЭ [114]. Секретируемые флавины, в первую очередь рибофлавин, производимые представителями рода *Shewanella*, – лучше всего описанные переносчики электронов в МТЭ [115–117]. Известно также, что феназин-подобные метаболиты, продуцируемые родственными шеванеллам бактериями из рода *Pseudomonas*, могут выступать в роли переносчиков электронов как для самих *Pseudomonas*, так и для других Грам-положительных бактерий [118].

Высокомолекулярные органические соединения не могут быть напрямую использованы бактериями для построения собственных клеток, а должны быть гидролизованы экзоферментами [119]. Только низкомолекулярные соединения могут быть непосредственно использованы бактериальными клетками, и тем самым быть убранными из жидкой среды, т. е. приводить к очистке воды. Одним из таких ключевых соединений является ацетат [54]. Например, *Geobacter* использует его в качестве донора электронов. Поэтому в многих экспериментальных работах, в которых изучались свойства

производящих электричество бактерий, в качестве питательной среды использовали ацетат [120–123]. Более того, показано, что *Geobacter sulfurreducens* способен полностью его окислять [124]. Ацетат также является одной из ключевых промежуточных молекул при разложении крупных органических соединений [125–127]. Накопление ацетата в среде приводит к переходу бактерий из экспоненциальной к замедленной фазе роста [126,128]. Поэтому ацетат является одной из ключевых молекул для оценки эффективности не только производства электричества, но и очистки воды. Ключевым путём производства ацетата является путь, включающий фосфат-ацетилтрансферазу (Pta) и ацетаткиназу (AckA) [129]. Расщепление ацетата возможно либо с участием тех же ферментов, либо по более эффективному пути, включающему ацетил-КоА-синтетазу (Acs) [130]. Было показано, что *E. coli* для расщепления ацетата в основном использует первый путь [131].

Все проводившиеся до нашей работы исследования метагеномов МТЭ производились только для реакторов малого размера (<0.5 л), и в них изучался исключительно таксономический состав прокариот, тогда как функциональный анализ не проводился вообще [132,133]. В главе 2 рассматривается динамика двух микробных сообществ МТЭ и проводится таксономический и функциональный анализ.

## **1.4 Распространение генетической информации внутри микробных сообществ**

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 [134].

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 [135–137]. Большая часть генов, вовлечённых в ГПГ, ассоциированы с патогенезом, симбиозом, метаболизмом и устойчивостью к антибиотикам [138]. Горизонтальный перенос генов происходит как благодаря способности некоторых бактерий к естественной трансформации одноцепочечной или двуцепочечной ДНК [139], так и во время конъюгации или за счёт активности мобильных генетических элементов,

таких как плазмиды или фаги [140].

ГПГ в бактериях достаточно хорошо изучен, и были опубликованы некоторые оценки его скорости [141–147]. ГПГ в археях до нашей работы был охарактеризован значительно хуже, хотя существовали предположения, что он играет важную роль в возникновении крупных клад архей [148].

Было показано, что ГПГ может происходить как между близкими, так и между далёкими организмами [24,149]. В соответствии с гипотезой сложности (the complexity hypothesis), горизонтальный перенос генов, которые кодируют белки, взаимодействующие с большим количеством белков, происходит реже, чем генов, чьи продукты имеют меньше связей [150,151]. Кроме того, гены предпочтительнее переносятся между геномами с похожими частотами использования кодонов [152].

Экспериментально было показано, что ГПГ между бактериями и археями возможен почти для всех групп генов, за исключением малого количества генов, токсичных для реципиента [153].

Интересной группой для изучения ГПГ в археях являются метаногенные археи, в частности, археи из семейства *Methanosarcinaceae*. Предполагалось, что большой размер их геномов связан с массовым ГПГ из бактерий [23]. Оценки уровня горизонтального переноса в этой группе организмов сильно расходятся [22,23,154], поэтому интерес представляет как аккуратная оценка количества перенесённых генов, так и изучение свойств таких генов.

### 1.4.1 *Methanosarcina*

Наиболее интересными для изучения ГПГ являются археи из рода *Methanosarcina* (семейство *Methanosarcinaceae*), которые обладают одними из самых крупных геномов среди архей. Другие представители семейства *Methanosarcinaceae*, такие как психрофил *Methanococcoides burtonii* [155], галофил *Methanohalobium evestigatum* (NCBI Assembly ID: ASM19665v1), *Methanohalophilus mahii* [21], *Methanlobus psychrophilus* [156], *Methanosalsum zhilinae* (NCBI Assembly ID: ASM21799v1), имеют более короткие геномы, чем *Methanosarcina* spp.; исключение составляет *Methanomethylovorans hollandica* [157].

Представители семейства *Methanosarcinaceae*, как и все *Methanosarcinales*, встречаются в самых различных местообитаниях и обладают самым большим и разнообразным набором метаболических путей среди архей. *Methanosarcinales* объединяют такие особенности, как способность к ацетокластическому метаногенезу;

наличие цитохромов; генов, кодирующих субъединицы А, К и N дегидрогеназы редуцированного кофермента F420; фосфоглицерат мутаза и аденилат киназа бактериального типа; негистонового хромосомного белка MC1, вовлечённого в процесс конденсации хромосом; длинного варианта белка ScpB [158].

Было показано, что горизонтальный перенос одного короткого оперона из *Clostrida* в общего предка либо *Methanosarcinaceae*, либо *Methanosarcinales* кардинально изменил метаболические возможности этого таксона, сделав его представителей способными к использованию органических соединений с метильной группой в качестве субстрата для метаногенеза [24,159]. Первые работы, в которых был проанализирован ГПГ из бактерий в архей рода *Methanosarcina*, показали, что перенос затрагивал гены всех функциональных классов [23,160,161]. По этим оценкам, около 30% генов *M. acetivorans* имеют бактериальное происхождение [23], а половина генома *M. burtonii* имеет нетипичный состав олигонуклеотидов и большое количество транспозонов, что косвенно свидетельствует о большом количестве ГПГ [155].

Более детальные результаты скорости и направления ГПГ в археях были получены при анализе базы arCOG [154] и реконструкции филогенетических деревьев по отдельным генам [22]. Оба эти исследования пришли к похожим выводам, что наибольшее количество ГПГ произошло на уровне общего предка *Haloarchaea* (более 1000 событий) и отдельно на ветке *Methanomicrobia* на разных филогенетических уровнях. В этих же работах было показано, что *Methanosarcinales* скорее всего получили сотни метаболических генов в результате ГПГ из эубактерий [22,154]. При этом вопрос о том, как происходил перенос в общего предка *Methanosarcina* или *Methanosarcinaceae*, из каких именно бактерий происходил ГПГ в этой группе, и где именно перенесённые гены расположены в геноме, подробно не изучался.

В главе 3 мы приводим более аккуратную оценку уровня горизонтального переноса в *Methanosarcina* spp. и *Methanosarcinaceae*, описываем потенциальных доноров горизонтально перенесённых генов, а также изучаем их функции, оперонную структуру и экспрессию.

## **1.5 Примеры крайне редуцированных сообществ – эндосимбионты**

Как уже говорилось выше, 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 [162,163]. В таких сообществах нельзя изучать динамику, но можно отслеживать изменения в геномах, способствующие приспособлению к определённому местообитанию. Помимо того, что симбиотические сообщества обычно содержат только несколько, а иногда и только один вид бактерий или архей, размер популяции таких сообществ сильно ограничен и сильно меняется во времени [17].

Симбиоз с бактериями характерен для многих эукариот. Бактерии-эндосимбионты хорошо изучены в растениях, насекомых и некоторых других организмах [164–171]. В соответствии с первоначальным определением [172], 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, такие как аммоний, производимый азотфиксирующими бактериями, незаменимые аминокислоты, производимые эндосимбионтами насекомых [168,173,174] или витамины в случае мухи цеце [175], либо приобретать новые свойства. Некоторые симбионты, например, некоторые штаммы *Wolbachia* spp. защищают хозяина от патогенных вирусов [176]; симбионты могут быть нужны для созревания иммунной системы, как *Wigglesworthia* для мухи цеце [177]. Другие эндосимбионты, такие как *Caedibacter*, создают условия, способствующие лучшему выживанию заражённого хозяина. *Caedibacter* – это симбионт инфузорий из рода *Paramecium*. Нерепродуктивная форма этой бактерии содержит Р-тельца, содержимое которых токсично для инфузорий, не заражённых симбионтом. В репродуктивной форме эндосимбионт может заражать новых инфузорий. Поскольку заражённые инфузории постоянно выделяют наружу своих симбионтов, то незараженные инфузории не только того же, но и близких видов достаточно быстро либо погибают, либо оказываются тоже заражены [178]. Эндосимбионт инфузорий из рода *Euplotes* необходим для деления клеток хозяина [179,180]. Однако большинство симбиотических взаимоотношений не до конца понятны, и неизвестно, приносит ли такой союз пользу хозяину, особенно в случае факультативных эндосимбионтов [181,182].

Эндосимбиотические бактерии обычно имеют ряд характерных особенностей, таких как компактный геном, низкий GC-состав и короткие межгенные расстояния [17,183,184]. Для таких бактерий также характерны минимальный приток новых генов, значительный генетический дрейф, возникающий из-за малого размера популяций, и повышенная скорость эволюции (Рисунок 1.3), возникающая из-за ослабленного отбора

[174,185,186]. Все это ведёт к образованию псевдогенов или полной потере генов и снижению метаболического потенциала [17,185]. Часто наблюдаемая для симбиотических бактерий потеря генов, связанных с репарацией, ведёт к ещё большему увеличению скорости мутирования [17] и к снижению GC-состава [168,187,188]. Стабильные условия внутри организма хозяина не требуют сложных регуляторных систем, а тем самым не требуются длинные межгенные участки [17].



**Рисунок 1.3.** Филогенетическое древо бактерий с короткими геномами (менее 1Мб), построенное на основании набора универсальных рибосомных белков (см. главу 5). Листья раскрашены в соответствии с типом, к которому относится данная бактерия: *Actinobacteria* – чёрный, *Bacteroidetes* – жёлтым, *Proteobacteria* – голубым, *Spirochaetes* – зелёным, *Tenericutes* – красным, и неклассифицированные бактерии – серым.

### 1.5.1 *Holospora*

Одними из интересных симбиотических бактерий являются представители рода *Holospora*. Интересны они потому, что занимают уникальную экологическую нишу – они населяют либо макро-, либо микронуклеус инфузорий из рода *Paramecium*. *Holospora* spp. – это облигатные внутриядерные симбионты, относящиеся к подтипу *Alphaproteobacteria* и порядку *Rickettsiales*. Предполагается, что *Holospora* spp., так же, как и другие *Rickettsiales*, являются паразитами. На основании недавнего анализа последовательностей 16S рРНК некоторые исследователи предложили базальную ветку рикеттсий, куда входят все бактерии, похожие на *Holospora*, выделить в отдельный порядок *Holosporales* [189,190], но мы рассматриваем обе группы вместе, используя традиционное таксономическое определение порядка *Rickettsiales*, поскольку они имеют похожий образом жизни и обладают сопоставимым метаболическим потенциалом.

Род *Holospora* на настоящий момент состоит из девяти видов, каждый из которых способен населять только определённый вид инфузорий и только определённое ядро [191]. *Holospora* не могут делиться вне клеток хозяина, но способны выходить из заражённых клеток, чтобы заразить новые. В отличие от других симбиотических бактерий, у *Holospora* сложный жизненный цикл, который состоит из двух морфологически разных жизненных форм. Короткая репродуктивная форма встречается только в ядре хозяина, эта форма способна делиться, тогда как длинная инфекционная форма не делится и способна заражать новые клетки [10,192]. Инфекционная форма имеет вытянутые очертания, обширная периплазма на её дистальном конце необходима для успешного заражения. Инфекционная форма проникает в клетки *Paramecium* из пищеварительных вакуолей, куда она попадает вместе с пищевыми бактериями, но не переваривается, как остальные бактерии, а выходит в цитоплазму, откуда дальше проникает в ядро. Было предложено два возможных механизма проникновения в ядро. *Holospora* или способна разрушать пищевые вакуоли, и каким-то образом попадать в транспортные везикулы [193]; или же *Holospora* достигает ядра, двигаясь по актиновым филаментам хозяина [10,194]. Было показано, что *Holospora obtusa* содержит периплазматический белок размером 89 кДа, который необходим для проникновения в клетки [8]. Для проникновения в макронуклеус *H. obtusa* секретирует белок размером 63 кДа [195]. Помимо этого, было показано, что в процесс проникновения в макронуклеус вовлечены белки размером 15 кДа и 39 кДа [196]. Как происходит переход от репродуктивной к инфекционной форме, остаётся не до конца понятным, но известно, что секретируемый белок размером 5,4 кДа экспрессируется только в переходном

состоянии и в инфекционной форме [9].

В присутствии *Holospora*, *Paramecium* способны расти, делиться и размножаться [193]. Было показано, что *Holospora obtusa* способствует тому, что *Paramecium caudatum* лучше переносит повышенные температуры, считается, что это может быть связано с повышенной экспрессией *hsp70* клетками *H. obtusa*, но прямого объяснения этого феномена не существует [197]. Также известно, что голоспоры помогают хозяину лучше переносить повышенную солёность среды и осмотический шок [198]. Более того, было показано, что *Paramecium biaurelia*, инфицированные *H. caryophila*, лучше растут в экспоненциальной фазе роста [199]. Несмотря на описанные выше эффекты, было показано, что голоспоры могут снижать приспособленность хозяина. Так, присутствие *Holospora elegans* в ядрах *Paramecium* может приводить к развитию нефункционального макронуклеуса после конъюгации [200]. *Holospora undulata* увеличивает смертность инфузории-хозяина, особенно в условиях нехватки пищи [201], а высокие концентрации инфекционной формы в ядре препятствуют клеточному делению [198]. Таким образом, взаимоотношения *Holospora* spp. и инфузории-хозяина – это сложная система, в которой выгода для хозяина возникает только при определённом наборе условий.

Метаболизм *Rickettsiales*, и, в частности *Rickettsia* spp., был подробно изучен [202–205]. *Rickettsia* для своего успешного существования должна импортировать из хозяина как минимум 51 метаболит [206]. Представители рода *Rickettsia* не способны синтезировать аминокислоты и нуклеотиды, не способны к гликолизу и должны импортировать такие соединения, как кофермент А, пируват, ФАД, биотин и другие.

К моменту начала исследования было секвенировано три генома *Holospora*, все они эндосимбионты *Paramecium caudatum*: живущая в макронуклеусе *H. obtusa* и специфичные для микронуклеуса *H. elegans* и *H. undulata* [11,12]. Анализ универсальных для рода ортологов показал, что 572 однокопийных генов являются общими для всех трёх геномов, и что *Holospora*, по всей видимости, должна сильно зависеть от метаболитов, предоставляемых хозяином [12]. Однако детальная реконструкция метаболизма не проводилась. В главе 4 описан геном нового представителя рода *Holospora* – *H. curviuscula* и проведена реконструкция метаболизма и сравнительный анализ всех представителей *Holospora* spp., дополнительно приводится сравнение *Holospora* spp. с бактериями из порядка *Rickettsiales*.

### 1.5.2. Редукция геномов у эндосимбионтов

Достаточно долго считалось, что самые короткие среди прокариот геномы характерны для бактерий рода *Mycoplasma* (класс *Mollicutes*). Геномы у представителей этого рода не бывают короче 500 кб и содержат как минимум 470 генов [207]. Однако в 2006 году были секвенированы эндосимбиотические бактерии листоблошек – *Candidatus Carsonella rudii* [208] – и оказалось, что их геном содержит всего 182 гена, а его длина составляет всего 160 кб. Позднее был открыт ещё ряд эндосимбионтов насекомых с очень короткими геномами [165]. Важно отметить, что геномы микоплазм, по-видимому, являются близкими к минимально возможным для бактерий, способных к самостоятельной репликации *in vitro*. В экспериментах по искусственному синтезу минимального бактериального генома на базе *M. mycoides* было показано, что для независимого от хозяина образа жизни минимально необходимо 473 гена, и, что интересно, функции большей доли этих генов неизвестны [209].

Известно, что в разных таксонах бактерий редукция геномов происходила независимо, причём в условиях эндосимбиоза возможна значительно большая редукция метаболизма, чем для свободно живущих организмов [17]. В любом случае приспособление к симбиотическому образу жизни приводит к постепенному сокращению длины генома и потере целого ряда важных генов [17].

Особый интерес представляет редукция универсальных биологических систем, связанная с приспособлением к симбиотическому образу жизни. Одной из таких систем является рибосома – универсальная органелла, характерная для всех живых организмов. Бактериальная рибосома состоит из малой (30S) и большой (50S) субъединиц, которые вместе образуют частицу размером 70S [13,210]. Субъединицы рибосомы состоят из множества белков и трёх молекул рРНК – 16S, являющейся частью малой субъединицы, и 23S и 5S, образующих большую субъединицу [210]. Основные каталитические функции рибосомы: образование пептидной связи, распознавание кодона и протягивание мРНК – выполняют молекулы рРНК [14,15,211]. Более того, молекулы рРНК образуют остов, на который крепятся рибосомные белки (р-белки) [212]. Рибосома *Escherichia coli* состоит из 21 белка 30S субъединицы (S1 – S21) и 33 белков 50S субъединицы (L1 – L36) [16,213]. Роль р-белков заключается в стабилизации структуры органеллы и регуляции рибосомной активности [214]. Несмотря на то, что ключевую роль в биосинтезе белков играет рРНК, состав р-белков очень консервативен в бактериях [215]. Более того, 33 р-белка консервативны во всех живых организмах [215–217].

Состав бактериальных рибосом изучался очень детально. Анализ большого количества линий *E. coli*, нокаутных по одному из рибосомных генов, показал, что эти

бактерии способны к росту в отсутствие S6, S15, S20, S21, L1, L9, L11 и L25. При этом в большинстве случаев нокаут каждого из генов, кодирующих перечисленные выше белки, приводил к снижению скорости роста [19]. В другом исследовании, где также анализировались нокаутные линии *E. coli*, было показано, что девять р-белков (L15, L21, L24, L27, L29, L30, L34 и S17) не являются необходимыми для жизни как минимум в экспериментальных условиях [218]. Похожее исследование, проведённое с бактерией *Bacillus subtilis*, позволило показать, что эти бактерии способны к росту без 20 р-белков [219]. На основании этих результатов был сделан вывод, что р-белки меньшего размера являются достаточно молодыми и менее важными [219,220]. Филогенетический анализ 995 полных бактериальных геномов показал, что гены, кодирующие 44 р-белка являются необходимыми, гены, кодирующие белки S16, L9, L19, L31, L34 и L36, очень редко отсутствуют, тогда как гены, кодирующие белки S21, S22, Thx, L25 и L30, отсутствуют у многих бактерий [18]. Анализ трансляционного аппарата *Mollicutes* показал, что пять генов, кодирующих белки S14, S22, L7, L25 и L31, отсутствуют во всех геномах этой группы; ген, кодирующий белок L30, отсутствует почти во всех геномах; а ген, кодирующий белок S1, был независимо потерян семь раз [20]. Наконец, анализ эндосимбиотических бактерий с крайне редуцированными геномами из разных филогенетических групп, показал, что только 17 из 21 белка малой субъединицы и 16 из 32 белков большой субъединицы присутствуют у всех бактерий [17]. Например, геном *Candidatus Tremblaya princeps* – эндосимбионта червецов из семейства *Pseudococcidae*, содержит только 139 генов и тем самым просто не может кодировать все необходимые рибосомные гены [17]. Хотя списки необходимых р-белков немного отличаются в разных исследованиях, во всех описанных исследованиях белки L25, L30, L31, S21, S22 и Thx не являлись необходимыми. В главе 5 описываются особенности редукции рибосомы у бактерий с редуцированными геномами.

## **Глава 2. Таксономический и функциональный анализ микробных сообществ анодов микробных топливных элементов**

Совмещение процесса биоремедиации и переноса электронов из клеток на электрод делает микробные топливные элементы (МТЭ) многообещающей технологией для очистки сточных вод. Несмотря на то, что был описан целый ряд микроорганизмов, образующих биоплёнку на поверхности анода, комплексный метагеномный анализ сообществ электрогенных бактерий практически не проводился. В этой главе представлены результаты полногеномного метагеномного анализа бактериальных сообществ МТЭ, разработанных для очистки разных типов сточных вод в разных климатических зонах. Состав и метаболические возможности сообщества с анода сравниваются в обоих случаях с сообществом из анаэробного биореактора (АБ), которое использовалось для инокуляции МТЭ. Таксономический анализ показал, что *Proteobacteria*, *Bacteroidetes* и *Firmicutes* преобладали как в АБ, так и на аноде в МТЭ. Функциональный анализ позволил выявить группы микроорганизмов, необходимых для биоразложения органических соединений, часто встречающихся в жидких пищевых отходах. Помимо этого, наблюдалось накопление метаногенных архей в биоплёнках на поверхности анодов, что могло бы в дальнейшем позволить получать при очистке сточных вод не только электричество, но и природный газ. Наконец, в составе анодного сообщества нам удалось обнаружить ряд микроорганизмов, потенциально способных как к прямому, так и к опосредованному переносу электронов на анод. В этой главе показано, как изменяется микробное сообщество при переходе от АБ к МТЭ.

### **2.1 Материалы и методы**

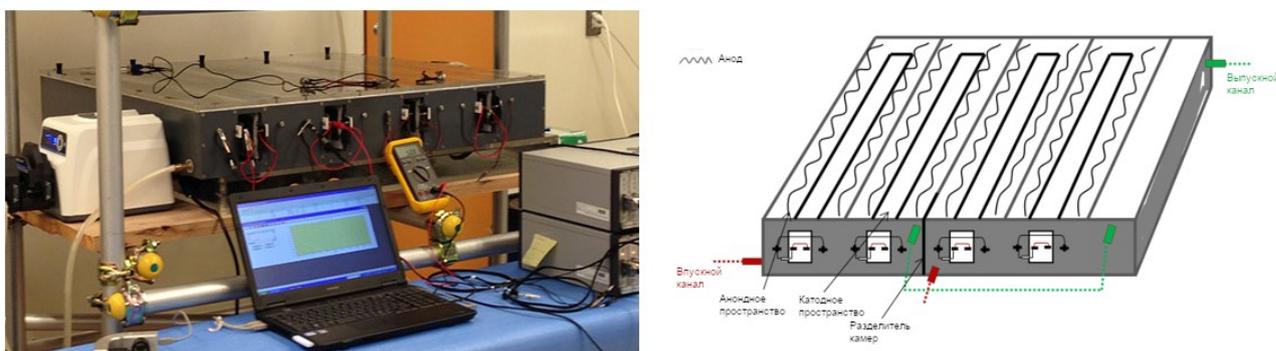
#### **2.1.1 Режим работы МТЭ<sup>1</sup>**

Два МТЭ полупромышленного масштаба были произведены компанией MPower World Ltd (Шотландия, Великобритания) и тестировались на настоящих сточных водах в разнообразных условиях. Проверялась способность МТЭ к генерации электричества и уровень очистки по ХПК

---

<sup>1</sup> Работа по разделу 2.1.1 и большей части раздела 2.1.2 до работы с ридами не выполнялась диссертантом, но эти описания приведены здесь для полноты.

(химическому потреблению кислорода). Используемые МТЭ по конструкции являются модификацией горизонтального проточного биореактора (Рисунок 2.1) [18]. Первый биореактор (UK) располагался в Эльвингстонском Научном парке (Elvingston Science Park) в Шотландии, его объем составлял 64 л, и он использовался для очистки сточных вод от отходов производства виски. Второй биореактор (JP) располагался в Окинавском институте науки и техники (OIST Graduate University) в Японии; его объем составлял 60 л. Этот биореактор использовался для очистки сточных вод от отходов производства авамори (бренд «Мизухо», Наха, Япония). Исходные сточные воды разбавлялись до среднего ХПК 10 г/л. Оба реактора имели 4 пары электродов с воздушным катодом. МТЭ инокулировались активным илом из АБ из местных водоочистных сооружений: UK – из станции очистки отходов от производства виски (Elgin, UK); JP – из станции очистки завода Ogiou (Окинава, Япония). Значение pH в МТЭ UK поддерживалось на уровне 4 – 5,5, а в JP – 6,2.



**Рисунок 2.1.** МТЭ JP. Внешний вид МТЭ (слева) и схематическое описание устройства (справа).

### 2.1.2 Сбор образцов, выделение ДНК и секвенирование

Образцы биоплёнки с анодов в случае МТЭ UK были собраны через 90 дней после запуска, а в случае JP через 70 дней после запуска. Из-за разной конструкции использованных анодов образцы анодных сообществ собирались по-разному для разных МТЭ: UK – образец биофильма был собран с поверхности анода; JP – фрагмент карбоновой ткани был срезан с анода. Все образцы сразу после сбора были помещены в LifeGuard Soil Preservation Solution в соответствии с инструкциями производителя (MO BIO Laboratories, Carlsbad, CA, USA). ДНК была выделена из каждого образца при помощи PowerMax soil DNA isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA, USA). Качество выделения ДНК было оценено при помощи спектрофотометрии и электрофореза в агарозном геле. Около 5 мкг ДНК было использовано для приготовления библиотеки. Полученные библиотеки были отсеквенированы на Roche 454 GS FLX Titanium (454 Life

Science, Branford, CT, USA). Все действия производились в соответствии с рекомендациями производителя [221]. Полученные риды были обрезаны по качеству, и праймеры были удалены. Для дальнейшего анализа были отобраны только риды длиннее 400 п.о. Риды были собраны *de novo* при помощи GS DeNovo Assembler version 2.8. Для дальнейшего анализа были использованы только контиги длиннее 500 п.о.

### 2.1.3 Анализ метагеномных контигов

Полученные контиги и несобранные одиночные риды были загружены в MG-RAST (MetaGenome Rapid Annotation with Subsystem Technology) [222] под следующими названиями: Japan МТЭ AD sludge (4540100.3, 4540109.3); Japan MFC anode samples (4539687.3, 4540099.3); Japan MFC anode cloth and granules (4540129.3, 4540130.3, 4540190.3, 4540191.3); UK MFC AD sludge (4540110.3, 4543057.3); UK MFC anode sample (4539688.3, 4540098.3).

Отдельно в MG-RAST были загружены образцы, содержащие несобранные риды, полученные с образцов анода МТЭ JP. Это было сделано, чтобы сравнить результаты, получаемые при анализе контигов и несобранных ридов. При помощи стандартных функций MG-RAST были удалены все фрагменты, соответствующие геному человека. В случае работы с несобранными ридами дополнительно при помощи стандартной опции MG-RAST были удалены дублицированные риды; такой фильтр позволяет избавиться от дубликатов, полученных в ходе ПЦР при пробоподготовке.

Несмотря на то, что MG-RAST разработан для анализа ридов любой длины, он не дает возможности посчитать представленность разных таксономических групп для образцов с последовательностями разной длины и разным покрытием. Поэтому результаты поиска при помощи BLAT, полученные на сервере MG-RAST, были сгружены и проанализированы при помощи скриптов, специально написанных на языке программирования perl.

### 2.1.4 Филогенетический анализ

Первоначальная филогенетическая аннотация каждого метагеномного образца проводилась на основании фрагментов гена 16S рРНК. Результаты кластерного анализа (уровень сходства 97%) последовательностей фрагментов генов 16S рРНК были загружены с сервера MG-RAST, и таксономическая принадлежность для всех последовательностей

была определена при помощи SILVA Aligner (минимальный уровень сходства с последовательностью из базы 0.85) и RDP classifier (значение степени уверенности больше 0.8) [223]. SILVA Aligner использовался в качестве основного источника аннотации, тогда как классификация по RDP использовалась только в случаях, когда последовательность не удавалось классифицировать по SILVA.

Отдельно была проведена таксономическая аннотация всех последовательностей для того, чтобы определить представленность разных таксонов в пробе. Результаты поиска при помощи BLAT против базы данных RefSeq на сервере MG-RAST были скачены с сервера. Среди лучших находок выбирались только такие, уровень сходства которых с аннотируемой последовательностью был не менее 60%. В случае, если в базе было много последовательностей, похожих на аннотируемую, выбирались только такие находки, уровень сходства которых не более чем на 10% отличался от лучшей находки. Длина всех отбираемых находок должна была превышать 50 нуклеотидов, а E-value должно было быть не более  $10^{-5}$ . Находки, выбранные таким образом для каждой аннотируемой последовательности, использовались для определения таксона, к которому относится эта последовательность, по принципу последнего общего предка (LCA): исследуемая последовательность приписывалась к таксону самого низкого ранга, к которому относились все отобранные находки. Представленность каждого таксона вычислялась как количество ридов, приписанных к данному таксономическому рангу; при этом риды, включённые в состав контига, учитывались индивидуально. В работе использовалась таксономия микроорганизмов из NCBI Taxonomy [224].

### 2.1.5 Функциональная аннотация

Для функциональной аннотации генов использовалась приписанная MG-RAST классификация базы данных кластеров ортологичных генов (Clusters of Orthologous Genes, COG, КОГ) [225] и базы ортологов Киотской энциклопедии генов и геномов (Kyoto Encyclopedia of Genes and Genomes, KEGG) [226] со следующими параметрами: длина аминокислотного выравнивания – 80 аминокислот; E-value –  $10^{-5}$ ; уровень сходства как минимум 60%. Таксономическое распределение молекулярной функции определялось как таксономическое распределение последовательностей, в которых она была найдена.

Представленность отдельных КОГ или групп ортологов из KEGG (КО) вычислялась как количество ридов, в которых встречается данный КОГ или КО, умноженное на среднее покрытие последовательности. Представленность каждой

категории – это сумма представленностей каждого КОГ или КО, относящихся к этой категории. Для описания наиболее распространенных КОГ в метагеномах анодных сообществ использовались все КОГ с представленностью больше 0.25% в анодных образцах. Эти значения сравнивались с представленностью соответствующих КОГ в исходных сообществах.

### **2.1.6 Анализ функций, необходимых для успешной работы МТЭ**

Список ферментов, необходимых для биосинтеза фенозин-1-карбоксилата, был получен из базы данных MetaCyc [227]. Список белков (всего 35), вовлечённых в передачу электронов на электрод, был получен из статей [106,228–232]. Аминокислотные последовательности этих белков использовались для поиска при помощи BLAST похожих белков в метагеномных образцах (45% сходства и минимальная длина выравнивания 60 аминокислот, E-value < 10<sup>-5</sup>).

Последовательности генов, связанные с утилизацией ацетата, были взяты из баз данных KEGG, COG и CDD: K00625 – фосфат ацетилтрансфераза; K01895 – ацетил-КоА синтаза; K00925 – ацетат киназа; K14393 и PRK09395 (Идентификатор в базе данных CDD) – ацетат пермеаза actP; K01637 и COG2224 – изоцитрат лиаза; K01638 – малат синтаза. В случае, если использовалась матрица из CDD, поиск производился при помощи HMMer (порог на вес совпадения (bitscore) 120).

Для поиска транспортёров аминокислот использовалась база COG. Были найдены все КОГ, содержащие «Amino acid transport» в описании. Находки были обнаружены только для трех КОГ: COG0531, COG0697, COG0833.

Найденные гены были размечены на таксономическом дереве NCBI. Для визуализации деревьев использовался веб-сервер iTOL [233].

## **2.2 Результаты и обсуждение**

### **2.2.1 Работа МТЭ и сбор образцов<sup>2</sup>**

За два месяца бесперебойной работы оба МТЭ (JP и UK) показали хорошие результаты как по очистке сточных вод от органики, так и по производству электричества.

---

<sup>2</sup> Работа по сбору образцов из МТЭ и подготовке библиотек для секвенирования не является частью диссертационной работы.

Однако МТЭ JP имел чуть лучшие показатели [234]. Помимо этого, в МТЭ JP наблюдалось образование пузырьков в анодной камере, которые, как оказалось, содержали достаточно высокую концентрацию метана. 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* (см. Материалы и методы). Суммарная статистика для каждого образца приведена в Таблице 2.1.

	МТЭ JP инокулум (JP-ref)	МТЭ JP биоплёнка с анода (JP-mfc)	МТЭ UK инокулум (UK-ref)	МТЭ UK биоплёнка с анода (UK-mfc)
Средняя длина ридов (нуклеотиды)	680.3	676.2	636.9	793.6
Общее количество ридов	543979	1443239	545744	706890
Средняя длина после фльтрации (нуклеотиды)	764.9	760.8	706.1	841.4
Количество ридов после фильтрации	460545	1219296	465036	652877
Суммарная длина после фльтрации(млн нуклеотидов)	352.3	927.6	328.4	549.3
Сборка метагеномных образцов				
% собранных ридов	34.75	55.45	87.46	72.31
Количество контигов	11689	25481	8445	15800
Количество контигов после фильтрации	11685	25471	8439	15791
N50	2321	3848	3450	2687
Длина самого большого контига	41621	220385	71492	62941
Суммарная длина контигов (млн нуклеотидов)	24.4	71.9	21.2	36
Количество несобранных ридов				
Количество синглтонов	279453	543974	85257	163085
Общая длина синглтонов (млн нуклеотидов)	208.1	403.4	59.7	133.9

	МТЭ JP инокулум (JP-ref)	МТЭ JP биоплёнка анода (JP-mfc)	МТЭ UK инокулум (UK-ref)	МТЭ UK биоплёнка анода (UK-fc)
Общая длина метагенома (млн нуклеотидов)	232.3	475.4	80	169.9
Функциональная аннотация				
Количество (%) ОРС с приписанной классификацией по COG	62162 (18.2%)	134641(19.7%)	22764(19.6%)	63037(27.5%)
% ОРС с предсказанной функцией по KEGG	16.11	17.68	21.28	25.49

**Таблица 2.1.** Общая статистика для секвенированных образцов метагеномов.

### 2.2.2 Таксономический анализ МТЭ

Чтобы оценить таксономическое разнообразие, было решено на первом этапе использовать фрагменты генов 16S рРНК, которые удалось найти в ридсах или в полученных контигах. Понятно, что такой подход не даёт приписать таксономию к каждой последовательности, но позволяет оценить разнообразие имеющихся образцов. Для того, чтобы приписать таксономический ранг к каждой последовательности в библиотеке, мы использовали метод наименьшего общего предка (LCA). Качественно результаты, полученные при анализе 16S рРНК, похожи на то, что получается при анализе всех последовательностей методом LCA, поэтому дальше мы решили работать только с результатами последнего подхода, так как он позволяет точнее количественно оценить таксономический состав проб. Результаты этого анализа приведены на Рисунок 2.2.

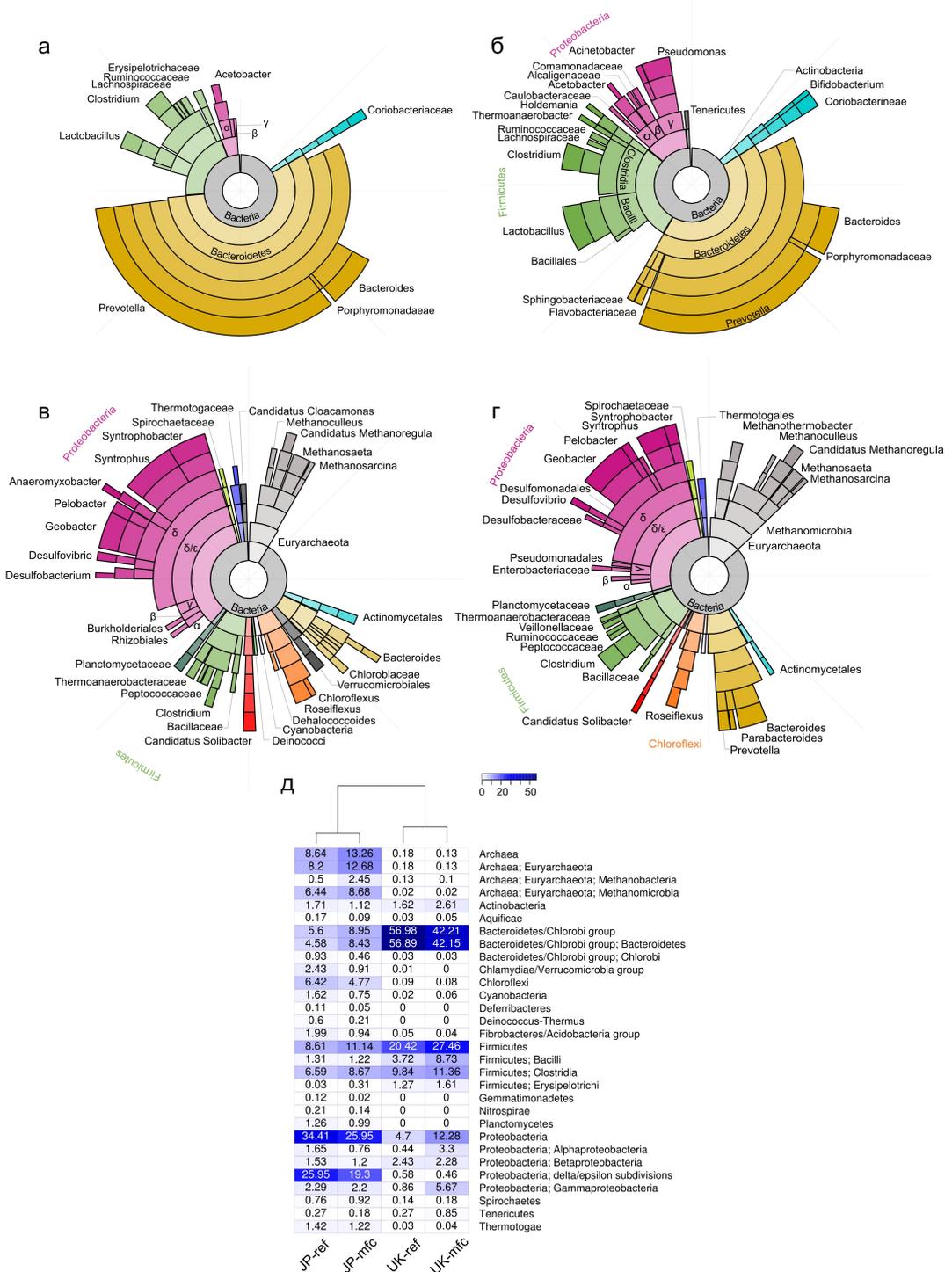
Состав исходного активного ила значительно различался для двух МТЭ. В МТЭ UK (Рисунок 2.2а) встречались представители *Bacteroidetes*, *Firmicutes*, *Proteobacteria* и *Actinobacteria*, причём *Bacteroidetes* составляли более половины сообщества (Рисунок 2.2д). Самой распространенной группой в анаэробном 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 2.2в).

Оба исходных сообщества претерпели значительные структурные изменения после 70-90 дней работы МТЭ. В МТЭ UK в биоплёнке на аноде увеличилось количество *Proteobacteria*, *Firmicutes* и *Actinobacteria* по сравнению с инокулумом. Несмотря на то, что количество *Bacteroidetes* уменьшилось, они все равно преобладали в биоплёнке с анода. На уровне родов значительно выросло количество *Pseudomonas* и *Lactobacillus* (Рисунок 2.2б).

В биопленке из МТЭ JP (JP-mfc) выросло количество *Bacteroidetes*, *Firmicutes* и *Archaea*. *Proteobacteria* остались самой широко распространённой группой в МТЭ JP, но их количество уменьшилось по сравнению с исходным активным илом (Рисунок 2.2г,д). Преобладающими родами в JP-mfc стали *Clostridium*, *Bacteroides*, *Prevotella*, *Pelobacter*, *Geobacter*, *Paludibacter* и *Methanothermobacter*.

Таким образом, 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. Стабильное преобладание этих групп было ранее показано для реакторов, вырабатывающих природный газ [235,236], и на анодах малых МТЭ, которые использовались для очистки муниципальных сточных вод [132], тогда как в средах, богатых ацетатом, наблюдалось преобладание только *Bacteroidetes* и *Proteobacteria* [133]. Метаногенные *Archaea* были также обнаружены в JP-mfc (Рисунок 2.2г), что хорошо соотносится с наблюдавшимся выделением природного газа в этом МТЭ.

Найденные на поверхности анода микроорганизмы, в том числе хорошо описанные представители рода *Geobacter*, способные к переносу электронов, скорее всего вовлечены в процесс прямой передачи электронов на поверхность анода. Кроме того, среди идентифицированных организмов была найдена архея *Methanothermobacter*, для которой ранее было показано, что ее присутствие повышает уровень производимого электричества в смешанных биоплёнках [237]. *Paludibacter* наблюдали в биореакторах, производящих водород [238], однако не было показано, что эта бактерия как-то влияет на производство электричества.



**Рисунок 2.2. Таксономический состав метагеномных образцов.** Диаграмма Крона для инокулума МТЭ UK (а), биоплёнки из МТЭ UK (б), инокулума 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*, салатный – *Spirochaetes*, синий – *Thermotogae*, голубой – *Actinobacteria*, темно-зелёный – *Planctomycetes*, розовый – *Cyanobacteria*. (д) Тепловая карта, показывающая представленность наиболее распространённых типов и классов; числа показывают процент представленности таксона в образце.

### 2.2.3 Функциональных анализ и идентификация генов, необходимых для очистки сточных вод

В микробных сообществах таксономическое разнообразие не всегда связано с функциональным разнообразием. В настоящем исследовании была предпринята попытка исследования полных метагеномов именно для того, чтобы оценить метаболические возможности изучаемых сообществ. Помимо крупных функциональных категорий, мы изучили особенности путей, связанных с утилизацией органических соединений, и представленность генов, необходимых для передачи электронов на электрод, т.е. генов, связанных с двумя важнейшими функциями МТЭ.

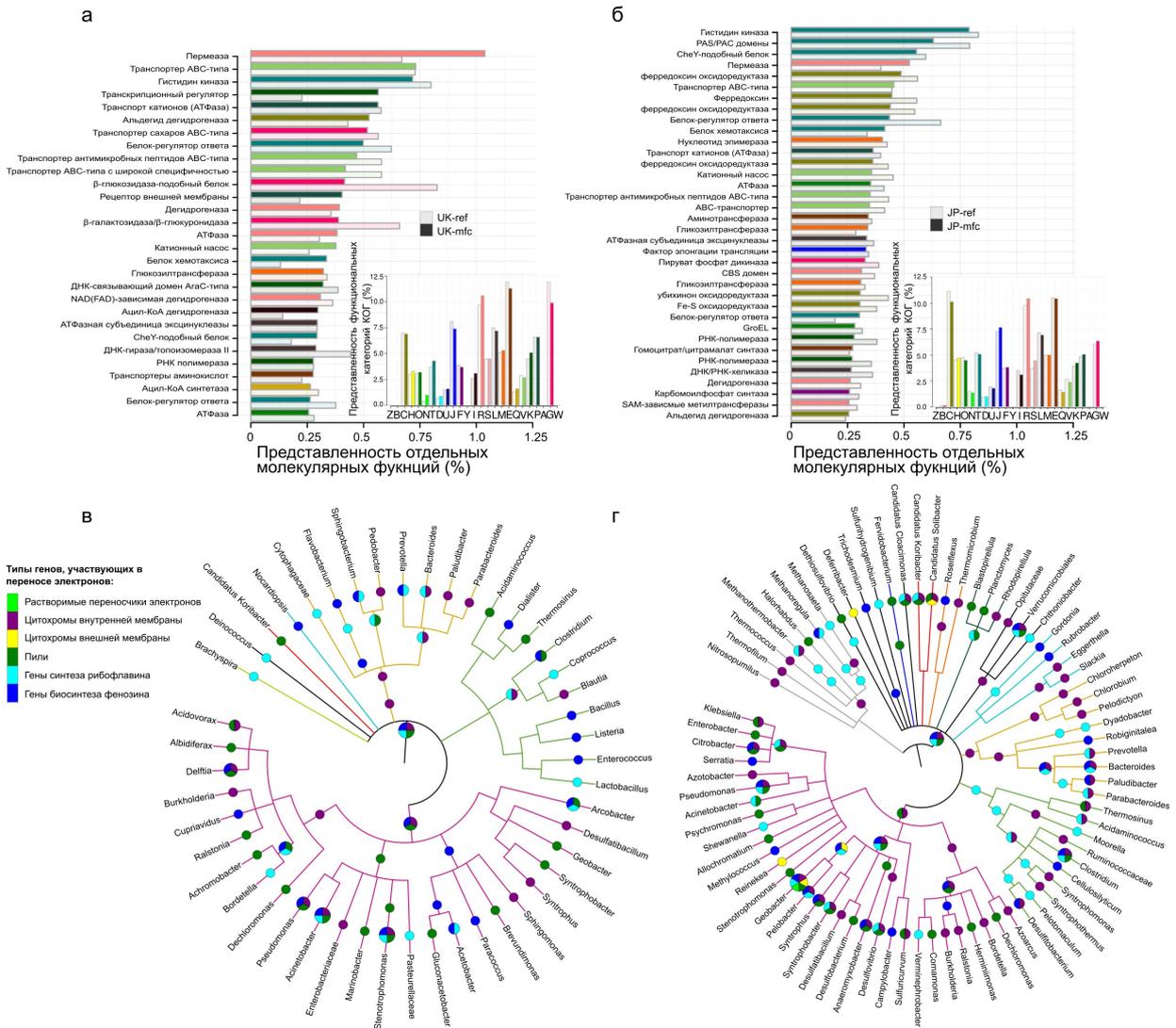
В МТЭ УК преобладали гены, относящиеся к категориям «Транспорт и метаболизм углеводов» (G) и «Транспорт и метаболизм аминокислот» (E), тогда как в МТЭ JP преобладали гены из категории «Производство и преобразование энергии» (C) по классификации функциональных категорий COG (Рисунок 2.3а,б). Такая разница в представленности самых распространённых функциональных категорий связана с различиями в таксономическом составе сообществ разных МТЭ, т.к. она вызвана генами из самых распространённых в этих образцах таксонов. В МТЭ УК преобладающая группа *Bacteroidetes* состояла из сахаролитических бактерий, которые получают энергию расщеплением углеводов и белков, и, тем самым, содержат много генов, необходимых для разложения органики. Преобладание категории C в МТЭ JP связано с наличием в сообществе архей, *Proteobacteria* и *Firmicutes*, т.е. тоже тех групп, которые преобладают в этом сообществе.

По сравнению с исходными сообществами, в биоплёнке с анода наблюдалось увеличение числа генов, связанных с хемотаксисом, передачей сигнала и мембранным транспортом. Гены, представленность которых значительно изменилась в биоплёнке, важны для ее формирования и для выработки энергии (Таблица 2.2).

Сравнение функциональных категорий, преобладающих в сообществе с анода и в инокууме, показало, что на аноде перепредставлена категория «Транспорт и метаболизм кофакторов» (H), а также гены с только предсказанной общей функцией (R) и гены с неизвестной функцией (S), что позволяет предположить, что существует много генов, необходимых для формирования биоплёнки, которые ещё не описаны.

Идентификатор КОГ	Кат	Описание КОГ	Изменение представленности в биоплёнке по сравнению с инокулумом(%)		Представленность функции в образце биоплёнки (%)	
			МТЭ JP	МТЭ UK	МТЭ JP	МТЭ UK
COG2199	T	домен GGDEF	50.28	75.57	0.14	0.18
COG0526	O	Тиол-дисульфид изомеразы и тиоредоксины	36.99	50.16	0.1	0.1
COG0488	R	АТФный компонент транспортной системы ABC с дублированным АТФазным доменом	36.39	20.49	0.21	0.38
COG1373	R	АТФаза (суперсемейство AAA+)	33.07	49.75	0.1	0.12
COG0621	J	2-метилтиоаденин синтетаза	31.49	31.65	0.11	0.13
COG0258	L	5'-3' экзонуклеаза (включая N-концевой домен PolI)	29.15	30.25	0.14	0.1
COG0749	L	ДНК полимеразы I - 3'-5' экзонуклеазный и полимеразный домены	27.17	30.25	0.12	0.1
COG0477	R	Пермеазы из суперсемейства major facilitator	23.99	35.45	0.52	1.04
COG1674	D	АТФазы FtsK/SpoIIE и похожие белки	19.85	29.2	0.11	0.16
COG0840	T	Метил-зависимый белок хемотаксиса	18.82	60.78	0.41	0.34
COG0018	J	Аргинил-тРНК синтетаза	17.77	13.21	0.1	0.2
COG0513	L	ДНК и РНК хеликазы из суперсемейства II	15.87	28.37	0.14	0.18
COG0438	M	Гликозилтрансферазы	15.76	19.79	0.34	0.24
COG0531	E	Транспортеры аминокислот	15.32	18.28	0.12	0.28
COG0370	P	Белок В системы транспорта Fe <sup>2+</sup>	13.89	10.68	0.13	0.14
COG0012	J	Предсказанные ГТФазы, предположительный регулятор трансляции	11.77	17.28	0.1	0.15
COG0468	L	RecA/RadA рекомбиназа	11.7	5.14	0.15	0.14
COG1109	G	Фосфоманномутаза	11.43	31.6	0.19	0.17

**Таблица 2.2.** Молекулярные функции (по классификации COG), наиболее сильно отличающиеся в инокулуме и биоплёнке. Кат – функциональная категория в базе данных COG.

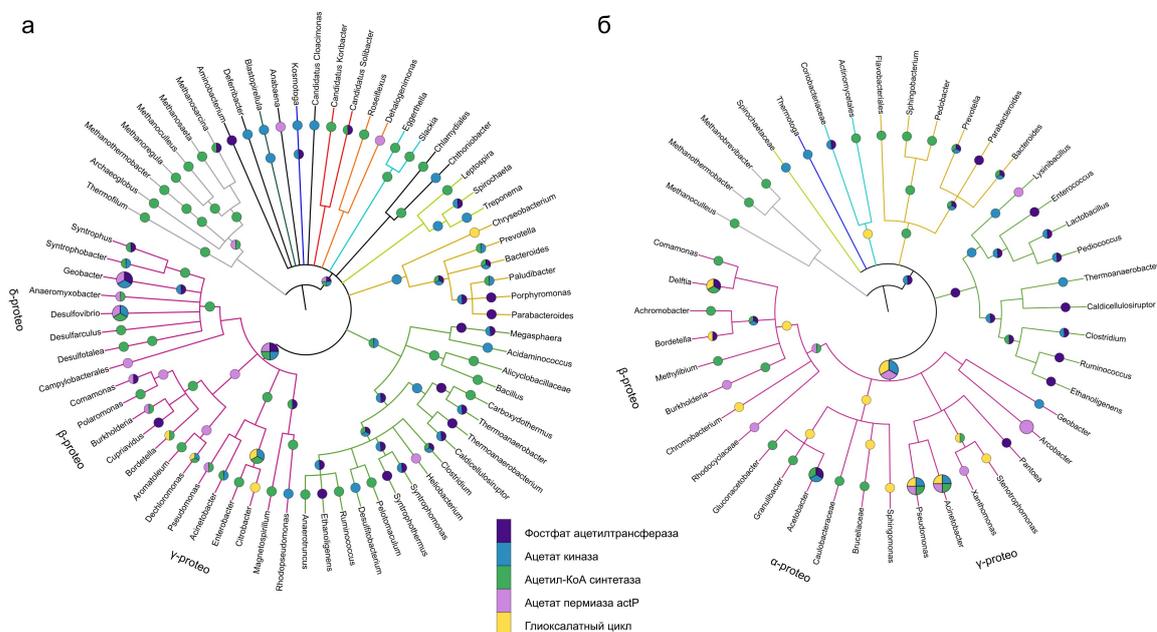


**Рисунок 2.3. Представленность молекулярных функции и категорий в метагеномных образцах МТЭ.** Наиболее представленные (не менее 0.25%) КОГ в биоплёнке с анодов для (а) МТЭ УК и (б) МТЭ JP. Группы раскрашены в соответствии с категориями; прозрачные полоски показывают представленность соответствующей группы в исходном сообществе. (в,г) Распределение гомологов генов, вовлечённых в перенос электронов, по разным таксономическим группам для (в) МТЭ УК и (г) МТЭ JP. Ветви окрашены в соответствии с типом (Рисунок 2.2), круги на ветвях показывают наличие белков данного типа в данной таксономической группе. Описание отдельных функций в представлено в таблице 3.

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 2.4, 2.5).

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 2.4). Ферменты анаэробного разложения целлюлозы (эндоглиукоза, бета-глюкозидаза и экзоглиуконаза) встречаются в основном у представителей *Firmicutes*, что позволяет предположить, что именно представители этой группы вносят основной вклад в разложение целлюлозы в обоих МТЭ (Рисунок 2.5). К разложению крахмала способны представители *Firmicutes*, *Bacteroidetes* и *Proteobacteria* в обоих МТЭ, тогда как в МТЭ JP крахмал могут разлагать еще и представители *Archaea*.

Внеклеточное расщепление белков может осуществляться за счёт целого ряда внеклеточных протеаз, которые встречаются в разных группах микроорганизмов. Понять, какие именно протеазы осуществляют внеклеточное расщепление белков, – нетривиальная задача, так как такие протеазы могут быть очень разными и не все из них хорошо охарактеризованы. Непрямой способ оценить вклад разных микроорганизмов в разложение белков в среде – это найти группы с наибольшим количеством транспортёров аминокислот. В обоих МТЭ наибольшее количество таких транспортёров найдено у представителей *Clostridia*. Дополнительно в МТЭ UK эта функция частично осуществляется за счет *Lactobacillus*, а в МТЭ JP – за счет *Delta*- и *Gammaproteobacteria* и частично архей.



**Рисунок 2.4.** 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.

### 2.2.4 Гены, необходимые для переноса электронов на анод

В образцах биоплёнок из обоих МТЭ нам удалось найти гены, кодирующие белки, необходимые как для прямого, так и для опосредованного переноса электронов. Однако в разных биореакторах к переносу электронов оказались способны разные бактерии (Рисунок 2.3в,г).

Последовательности, похожие на цитохромы типа с, были найдены в биоплёнках с анодов обоих МТЭ и принадлежали бактериям из различных семейств *Deltaproteobacteria* или *Gamma proteobacteria*. Все гены, необходимые для прямого переноса электронов с внешней мембраны (*omcABESZ*, *mtrF*), и периплазматические цитохромы (*ppcA*, *mtrA/D*), были найдены только в МТЭ JP и принадлежали *Geobacter*, *Deferribacter*, *Reinekea* и неопределённым бактериям. Последовательности, похожие на цитохромы внутренней мембраны (*macA*, *cymA*, *sirC*), были найдены в сообществах обоих МТЭ, но были значительно шире представлены в МТЭ JP. Большая часть найденных цитохромов в МТЭ JP найдена в разных *Deltaproteobacteria*.

Последовательности, кодирующие белки, похожие на образующие пили, встречались у *Proteobacteria*, особенно у *Deltaproteobacteria* и *Gamma proteobacteria*, в обоих биореакторах, причем в JP-mfc эти гены в основном принадлежали *Deltaproteobacteria*, а в UK-mfc – *Gamma proteobacteria*. Образец JP-mfc, но не UK-mfc, содержал гомологи *pilA* из *Geobacter*.

Было показано, что пили, образованные чувствительным к маннозе гемагглютинином (Msh-пили), способствуют генерации тока большой силы представителями рода *Shewanella* [239]. Гены, кодирующие белки, похожие на образующие Msh-пили, представлены у широкого спектра организмов, особенно в МТЭ JP, где *msh*-гены были найдены в *Firmicutes*, *Planctomycetes*, *Verrucomicrobia*, *Thermotogae*, *Acidobacteria*, разных классах *Proteobacteria* и *Archaea*. В МТЭ UK гены, похожие на *msh*, были найдены в *Firmicutes* и в разных классах *Proteobacteria*. Интересно, что большая часть фрагментов *msh*-генов, отнесённых к *Epsilonproteobacteria*, были больше всего похожи на соответствующие гены из *Arcobacter butzleri* – электрогенной бактерии, выделенной ранее из МТЭ похожей конструкции [240].

Для описания опосредованного переноса электронов при помощи растворимых переносчиков мы изучили представленность гена *ribB*, кодирующего ключевой фермент синтеза рибофлавина, и генов, вовлечённых в путь синтеза феназин-1-карбоксиамида



## Глава 3. Горизонтальный перенос генов и ЭВОЛЮЦИЯ ГЕНОМОВ МЕТАНОСАРЦИН

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. 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. В этой главе приводится подробный анализ ГПГ в роде *Methanosarcina* и рассматриваются особенности генов, которые были получены от бактерий.

### 3.1 Материалы и методы

#### 3.1.1 Геномные последовательности

Все последовательности геномов были скачаны из базы данных Genbank ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

В работе исследовались четыре группы видов: (1) *Methanosarcina* spp. (*M. acetivorans* C2A (NC\_003552; NC\_002097) [160], *M. barkeri* str. Fusaro (NC\_007349) [161], *M. mazei* Go1 (NC\_007355; NC\_003901) [23]); (2) все отсеквенированные на конец марта 2012 года представители семейства *Methanosarcinaceae* (*Methanosarcina* spp., *Methanococcoides burtonii* DSM 6242 (NC\_007955) [155], *Methanohalobium evestigatum* Z-7303 (NC\_014253), *Methanohalophilus mahii* DSM 5219 (NC\_014002) [241]); (3) *Pyrococcus* spp. (*P. horikoshii* OT3 (NC\_000961) [242], *P. abyssi* GE5 (NC\_000868; NC\_001773) [243], *P. furiosus* DSM 3638 (NC\_003413)[244]); и (4) все представители семейства *Thermotogaceae* (*Thermotoga thermarum* DSM 5069 (NC\_015707), *Thermotoga* sp. RQ2(NC\_010483.1) [245], *Thermotoga petrophila* RKU-1 (NC\_009486.1) [246], *Thermotoga neapolitana* DSM 4359 (NC\_011978.1), *Thermotoga naphthophila* RKU-10 (NC\_013642), *Thermotoga maritima* MSB8 (NC\_000853) [31], *Thermotoga lettingae* TMO (NC\_009828), *Thermosipho melanesiensis* BI429 (NC\_009616), *Thermosipho africanus* TCF52B (NC\_011653), *Petrotoga mobilis* SJ95 (NC\_010003), *Kosmotoga olearia* TBF 19.5.1

(NC\_012785) [247], *Fervidobacterium nodosum* Rt17-B1 (NC\_009718)).

### 3.1.2 Построение ортологических рядов

На первом шаге группы ортологичных белков (ГОб) были построены для каждой пары видов в группе. В результате попарных сравнений при помощи BLASTP [248] был получен список лучших двунаправленных находок (ЛДН). Находки игнорировались, если сходство между ними было меньше 50% (40% для *Methanosarcinaceae*) или если длина выравнивания была меньше 2/3 более короткого белка в паре. Если два паралога в одном геноме были больше похожи друг на друга, чем на ЛДН из другого генома, оба паралога добавлялись в ГОб. На основании попарного графа были построены максимальные компоненты связанности. Весь процесс осуществлялся при помощи набора специализированных скриптов, написанных на Oracle RDBMS Express Edition. Были построены отдельно ГОб для всех представителей рода *Methanosarcina* (*Methanosarcina* ГОб) и для всех представителей семейства *Methanosarcinaceae* (*Methanosarcinaceae* ГОб). Дополнительно были построены ГОб для *Thermotogaceae* и ГОб для *Pyrococcus* spp.

### 3.1.3 Идентификация ГПГ

Для идентификации событий ГПГ был написан ряд скриптов на языке программирования perl. Для каждого белка в ГОб производился поиск находок в неизбыточной базе данных белковых последовательностей (nr), скаченной с NCBI (дата выхода 2011-07-16), при помощи программы BLASTP с параметрами по умолчанию. Находкам приписывалась таксономия на основании базы данных таксономии NCBI (<http://www.ncbi.nlm.nih.gov/taxonomy>).

Если для каждого белка в анализируемой ГОб три лучших находки были в домене *Bacteria* или максимум три из 20 лучших находок принадлежали археям (порог на сходство 30%, длина блока выравнивания не менее 50, и находка должна покрывать как минимум 75% длины исследуемого белка), ГОб отбиралась для дальнейшего анализа.

На следующем шаге отбирались 100 лучших белковых находок для каждого члена ГОб. Для исследуемых и отобранных белков строилось множественное выравнивание при помощи программы MUSCLE версии 3.6 [249] с параметрами по умолчанию. Качество выравниваний оценивалось при помощи программы GUIDANCE [250] и вручную. Если менее половины колонок в выравнивании для данного ГОб имели вес меньше, чем 0.93

(стандартный порог, используемый GUIDANCE), ГОБ исключалась из дальнейшего анализа. Для каждой отобранной ГОБ на основании полученного множественного выравнивания строилось два дерева: методом ближайших соседей со 100 бутстрэп испытаниями (NJ) при помощи ClustalW, и методом наибольшего правдоподобия (ML) при помощи программы PhyML v3.0 с параметрами по умолчанию [251]. Полученные деревья визуализировались при помощи сервера iTOL [233].

NJ и ML деревья анализировались независимо. Никакой специальной процедуры для определения положения корня дерева не использовалось, но если археи, не относящиеся к *Methanosarcinaceae*, образовывали плотную кладу на дереве, корень вручную помещался между ветвью, на которой находились эти археи, и ветвью, на которой оказывались бактерии. Если такой группы не было или на дереве присутствовали только представители *Methanosarcinaceae*, корень помещался в место расхождения крупных бактериальных таксонов. Если все члены ГОБ образовывали единую кладу внутри бактериальной ветви (где не было белков из других архей), и на дереве NJ (бутстрэп поддержка > 70), и на дереве ML, такая ГОБ считалась полученной в результате ГПГ. Необходимо отметить, что использование двух независимых процедур построения дерева позволяет делать достаточно точные выводы о наличии ГОБ без использования внешней группы, т.е. даже если точное положение корня неизвестно; исключением может служить только случай, когда настоящий корень на дереве находится между белками бактерий и белками анализируемой ГОБ.

Мы считали, что ГПГ произошёл в последнего общего предка *Methanosarcina* spp., если все члены *Methanosarcina* ГОБ, содержащей белки как минимум из двух разных видов *Methanosarcina*, оказывались на одной ветви с бактериями, не содержащей других архей. Считалось, что ГПГ произошёл в общего предка *Methanosarcinaceae*, если все члены *Methanosarcinaceae* ГОБ (как минимум два белка из разных *Methanosarcina* и как минимум один белок из других *Methanosarcinaceae*) оказывались на одной ветви с бактериями, где не было белков из других архей. Донором ГПГ считался общий таксон для всех бактерий на изучаемой ветви дерева.

При анализе белков, не вошедших в группы (синглтонов), использовалась схожая процедура, только вместо ГОБ выступал индивидуальный белок.

Для воспроизведения результатов, полученных в других работах, использовался только первый шаг анализа (включающий поиск при помощи BLAST) без дальнейшего построения деревьев. В таком случае результатом ГПГ считались ГОБ, для которых лучшие находки были среди бактерий.

### 3.1.4 Филогенетические деревья для *Methanosarcinaceae*

Конкатенированное выравнивание рибосомных белков было построено при помощи программы MUSCLE [249]. Для этого выравнивания использовался ранее предложенный набор самых консервативных белков [17]: S2, S5, L2, L3, L4, L6, двух копий S4, и фактора элонгации EF-TuA. Выравнивания 16S rRNA были скачаны из базы данных SILVA [223]. Эти выравнивания использовались для построения NJ-дерева, как описано выше.

### 3.1.5 Определение КОГ категорий

Белкам приписывалась ортологическая группа КОГ на основании записей в базе данных COG IMG (<https://img.jgi.doe.gov/cgi-bin/w/main.cgi?section=FindFunctions&page=ffoAllCogCategories>). Для каждого изучаемого организма перепредставленность (O) для данной КОГ категории считалась как:

$$O = \frac{COG_{obs} - COG_{exp}}{COG_{exp}},$$

где  $COG_{exp}$  – нормализованное количество КОГ в данной категории либо среди всех генов архей или бактерий, а  $COG_{obs}$  – количество наблюдаемых КОГ в данной категории среди горизонтально перенесённых генов.

### 3.1.6 Определение функциональных классов

Количество транскрипционных факторов оценивалось отдельно для каждого вида *Methanosarcina* по количеству белков в COG категории «Регуляция транскрипции» (“Transcription regulation”).

Транспортные белки определялись на основании поиска соответствующих доменов (поиск проводился по ключевому слову “transport” в названии семейства) в базе данных PFAM (<http://pfam.xfam.org/>) [252]. Общее количество белков, для которых можно найти домены в базе PFAM, было посчитано как только для белков, вошедших в ГОБ, так и для всех генов *Methanosarcina*.

Для определения количества ферментов использовалась EFICAz EC classification [253]. Из этой базы данных были получены записи о всех белках *Methanosarcina* spp.,

которые содержали ЕС классификацию, т.е. являлись ферментами.

### 3.1.7 Определение оперонной структуры

Координаты и направления генов были получены из геномных файлов (Genbank формат). Межгенные расстояния были посчитаны с помощью скрипта на языке программирования Perl. Мы определяли оперон как набор последовательных однонаправленных генов с небольшими межгенными расстояниями. Подробно процедура выбора межгенных расстояний описана в разделе 3.2.7. Длина горизонтально перенесённого фрагмента вычислялась как суммарная длина последовательных горизонтально перенесённых генов и межгенных расстояний между ними. Для тривиальных оперонов эта длина равнялась длине гена. В этом анализе синглтоны не учитывались.

Для оценки уровня экспрессии использовались все опубликованные ранее данные из экспериментов по определению экспрессии и белкового состава *Methanosarcina* [43-50].

### 3.1.8 Зависимость результатов от состава белковой базы данных

Для оценки изменений в предсказанном уровне ГПГ у *Methanosarcina* в зависимости от года публикации результатов, была создана серия баз данных, имитирующих состояние Genbank в каждый изучаемый год (с 2001 по 2011). В основе такой базы данных лежала современная избыточная белковая база данных Genbank, из которой были удалены все археи, отсеквенированные позже изучаемого года. Для каждой такой базы данных производился анализ, основанный только на результатах BLAST, описанный выше.

## 3.2 Результаты

### 3.2.1 Идентификация событий ГПГ

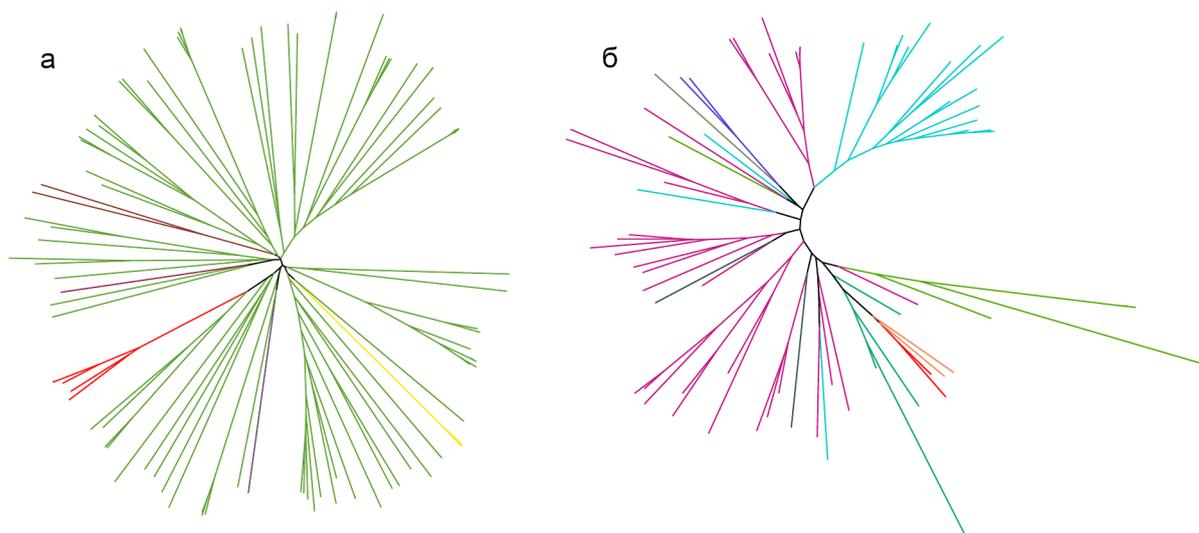
Для *Methanosarcina* spp. и *Methanosarcinaceae* были построены группы ортологичных белков (ГОб). Белки попадали в одну ГОб, если имели сходство 50% и 40% соответственно (Табл. 3.1).

	Количество генов в геноме	<i>Methanosarcinaceae</i> ГОБ (3122)		<i>Methanosarcina</i> ГОБ (2808)	
		Количество белков, включённых в ГОБ	Количество синглтонов	Количество белков, включённых в ГОБ	Количество синглтонов
<i>Methanococcoides burtonii</i> DSM 6242	2273	1811	462	–	–
<i>Methanohalobium evestigatum</i> Z-7303	2254	1699	555	–	–
<i>Methanohalophilus mahii</i> DSM 5219	1987	1713	274	–	–
<i>Methanosarcina acetivorans</i> C2A	4540	3325	1215	2852	1688
<i>Methanosarcina barkeri</i> str. Fusaro	3624	3062	562	2548	1076
<i>Methanosarcina mazei</i> Go1	3370	2888	482	2537	833

**Таблица 3.1.** Статистика исходных ГОБ до фильтрации (см. ниже) в *Methanosarcina* и *Methanosarcinaceae*. Общее количество ГОБ для каждой категории представлено в заголовке в скобках.

Все ГОБ, где как минимум половина белков были аннотированы как транспозазы, были исключены из дальнейшего анализа (30 групп *Methanosarcina* spp. и 38 групп *Methanosarcinaceae*). В финальную выборку ГОБ для *Methanosarcina* spp. попали 65% всех белков (2778 ГОБ), из них 94% (2624) содержали только по одному белку для каждого вида, тогда как оставшиеся содержали паралоги как минимум в одном из видов. Из выборки ГОБ для *Methanosarcinaceae* были удалены все ГОБ, которые уже учтены при составлении ГОБ для *Methanosarcina*, и в дальнейший анализ были включены только такие группы, которые содержали хотя бы один белок из *Methanosarcinaceae* spp. В результате этих процедур 55% от исходных ГОБ для *Methanosarcinaceae* (1702; 9993 белка из 6 видов) были включены в финальную выборку, из них 1375 ГОБ не содержали паралогов.

Для поиска случаев ГПГ сначала производился поиск при помощи BLASTP (см. Материалы и методы), и если среди находок были бактериальные белки, для максимум 100 лучших находок строились филогенетические деревья (методом ближайших соседей с бутстрэпами и методом наибольшего правдоподобия), всего 736 наборов деревьев (Рисунок 3.1). Если белки *Methanosarcina* spp. оказывались в одной кладе с бактериальными белками, но не с другими белками архей, соответствующие гены мы считали горизонтально перенесёнными из бактерий. В результате описанного выше анализа 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*.



**Рисунок 3.1.** Примеры филогенетических деревьев для некоторых ГОБ. Филогенетические деревья, построенные методом наибольшего правдоподобия, для (а) ацетат киназы AskA, вовлечённой в метаногенез и (б) для неохарактеризованной альфа/бета гидролазы (PF12695). *Methanosarcina* spp. показаны красным, другие представители *Methanosarcinaceae* – оранжевым, *Firmicutes* – светло-зелёным, *Actinobacteria* – голубым, *Synergistetes* – коричневым, *Thermotoga* – тёмно-фиолетовым, *Proteobacteria* – фиолетовым, *Dictyoglomi* – жёлтым, *Cyanobacteria* – изумрудным, *Aquificae* – серым, археи кроме *Methanosarcinaceae* – тёмно-синим.

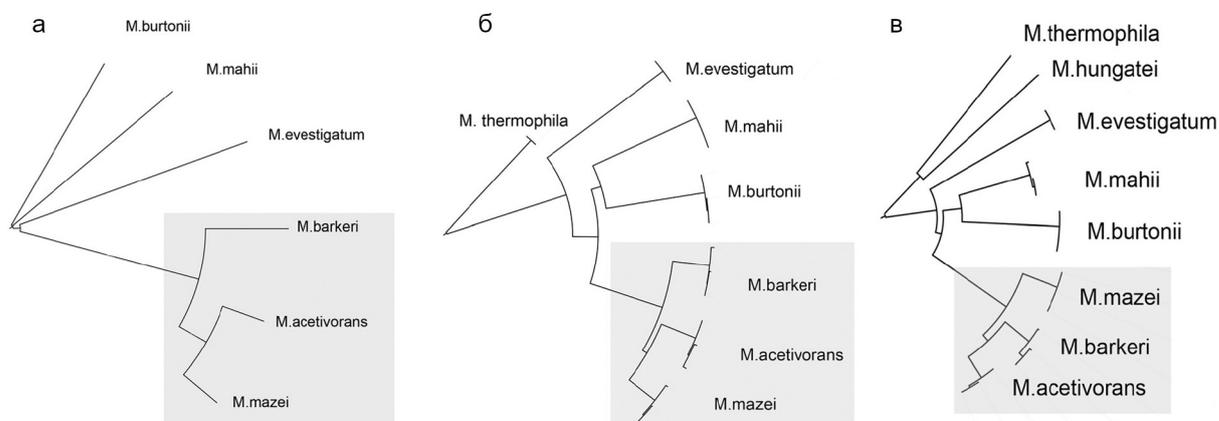
Отдельно мы проанализировали синглтоны, т.е. гены, не вошедшие в ГОБ и характерные только для конкретного вида микроорганизмов. Такой анализ позволил описать недавние события ГПГ. Нам удалось показать, что 14 генов было недавно горизонтально перенесено в *M. acetivorans*, 33 гена в *M. barkeri* и 10 в *M. mazei*. В случае *M. barkeri* нам удалось обнаружить случай ГПГ, включающий целый оперон из четырёх генов, связанных с системой CRISPR. Если просуммировать приведённые выше числа, то всего в геноме *M. acetivorans* был обнаружен 221 ген, полученный в результате ГПГ из бактерий; в геноме *M. barkeri* 214 генов и 151 ген в геноме in *M. mazei* (Табл. 3.2). Тем самым, по нашим оценкам, около 5% генов *Methanosarcina* были горизонтально перенесены из бактерий в геном общего предка *Methanosarcinaceae* либо *Methanosarcina*, или же непосредственно в геном конкретного вида.

Геном *M. mazei* значительно меньше, чем геномы других представителей данного рода, и в ее геноме осталось только 82 гена горизонтально перенесённых из бактерий в последнего общего предка *Methanosarcina* spp. Это почти в два раза меньше, чем в двух других видах.

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

Филогении семейства *Methanosarcinaceae*, реконструированные по 16S рРНК [254,255] и по консервативным белкам [256], не очень хорошо совпадают. На основании анализа белков трансляционного аппарата и 23S рРНК *M. mazei* оказывается ближе к *M. acetivorans* (Рисунок 3.2а, б), тогда как на основании анализа последовательности 16S рРНК *M. acetivorans* оказывается ближе к *M. barkeri* (Рисунок 3.2в). Мы обнаружили 12 ГОБ, полученных в результате ГПГ и содержащих гены только из *M. mazei* и *M. acetivorans*, 67 ГОБ, содержащих только гены *M. acetivorans* и *M. barkeri*, и ни одного ГОБ с горизонтально перенесёнными генами только из *M. mazei* и *M. barkeri*. Если верна филогения на основе 16S рРНК, то множество горизонтальных переносов произошло в общего предка *M. acetivorans* и *M. barkeri*. В противном случае, *M. mazei* потеряла большую долю генов, полученных в результате ГПГ в последнего общего предка всех *Methanosarcina*.

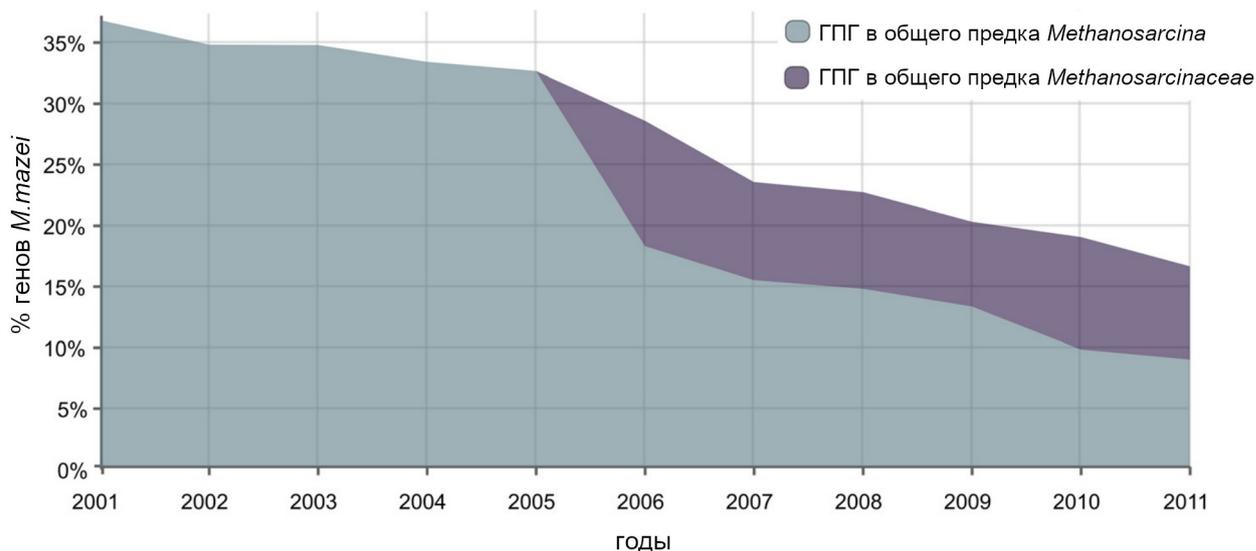


**Рисунок 3.2.** Филогенетическое древо *Methanosarcinaceae*. Филогенетические деревья построены методом ближайших соседей на основании белков трансляционного аппарата (а), 23S рРНК (б) и 16S рРНК(в).

### 3.2.2 Реконструкция новейшей истории: воспроизведение состава базы данных последовательностей и предсказанного уровня ГПГ

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 [23,161]. Можно предположить два возможных объяснения этому: либо наша процедура могла быть недостаточно чувствительна, либо количество предсказаний значительно зависит от состава базы данных биологических последовательностей. Для того, чтобы различить эти гипотезы, мы воспроизвели состав базы данных биологических последовательностей в разные годы, и повторили наш анализ на этих «исторических» базах. В каждой такой базе данных состав бактериальных генов был зафиксирован на уровне 2011 года, тогда как состав генов архей менялся: в базе для 2001 содержались только гены архей, отсекуенных строго до 2002 и т.п.. Мы не отслеживали отдельные гены архей, которые могли быть отсекуены до публикации полного генома. Для того, чтобы соотнести нашу оценку и опубликованные ранее, была воспроизведена и процедура анализа: ГПГ считались все случаи, когда лучшие находки в BLAST имели бактериальное происхождение без последующей реконструкции филогенетических деревьев. Оцененная таким образом доля бактериальных белков в геноме *M. mazei* составила 37% (928 генов) в 2001 году, а после этого только падала до 8% в 2011 году (222 гена). Дополнительно около 7,5% генов в базе данных 2011 года (194) составляли гены, предположительно перенесенные в общего предка *Methanosarcinaceae* (Рисунок 3.3). Оценка доли ГПГ производилась только для *M. mazei*, так как только для этого генома ранее были сделаны точные оценки. Поскольку в нашем анализе мы использовали только гены, включенные в ГОБ, и не учитывали синглтоны, мы получили 928 генов против 1043, описанные в предыдущей работе [23].

Как показывает этот анализ, значительное изменение числа предсказанных случаев ГПГ в *Methanosarcina* произошло в 2006 году, когда были отсекуены 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.



**Рисунок 3.3.** Зависимость доли генов предсказанных, как горизонтально перенесенных, в зависимости от года выпуска белковой базы данных. Голубая область – ГПГ в *Methanosarcina* spp., фиолетовая – ГПГ в *Methanosarcinaceae*.

### 3.2.3 Первый контроль: ГПГ в *Thermotogaceae*

Бактерия *Thermotoga maritima* долгое время считалась подверженной ГПГ, и по ранним оценкам предполагалось, что около 24% генов этой термофильной бактерии были горизонтально перенесены из архей [257]. Позднее было показано, что доля горизонтально перенесенных генов значительно меньше, только 8-11%, даже с использованием очень мягких порогов для BLAST [258]. Для того, чтобы разобраться, какова настоящая доля ГПГ в геноме *T. maritima* мы применили подход, описанных выше, только в отличие от прошлого случая в имитированных базах данных менялось как количество белков архей, так и количество белков бактерий. Для этого нами были построены ГОБ для всех отсеквенированных *Thermotogaceae*: *Thermotoga* spp., *Thermosipho* spp., *Petrotoga mobilis*, *Kosmotoga olearia* и *Fervidobacterium nodosum*. Синглтоны и ГОБ, содержащие транспозазы, не рассматривались. В анализе использовались только ГОБ, содержащие белки *T. maritima*. Мы проанализировали 1761 ГОБ и нашли только 28, содержащих гены, возможно горизонтально перенесённые из архей либо в последнего общего предка данного рода или семейства (меньше 1% всех генов в геноме). Это значительно меньше, чем оценки, полученные ранее [246,257–259].

Мы повторили процедуру не для ГОБ, а для всех белков *T. maritima*. Нам не

удалось воспроизвести результаты 1999 года с симулированной базой данных: по нашим оценкам с самыми низкими порогами на сходство только 10% генов (188) могли быть возможно перенесены из архей, тогда как в исходной статье [257] таких генов было найдено 451. Нам не удалось повторить и результаты 2008 года, т. к., по нашей оценке, только 4.6% генов (86) могло быть горизонтально перенесено, что меньше, чем в ранее предсказанные 204 открытых рамки считывания [258].

Скорее всего, такое различие в предсказаниях связано с тем, что в исследовании 1999 года использовались все записи, присутствовавшие на тот момент в базе данных, в том числе отдельные отсеквенированные гены, тогда как мы использовали только полные геномы.

### 3.2.4 Второй контроль: ГПГ в *Thermococcaceae*

Для того, чтобы оценить количество ложноположительных предсказаний, мы повторили нашу процедуру предсказания ГПГ на геномах архей из рода *Pyrococcus*. Археи *Pyrococcus* spp. из семейства *Thermococcaceae* достаточно хорошо изучены, и считается, что их геном практически не подвергался ГПГ. Анализ базы данных arCOGs показал, что общий предок всех *Thermococcales* получил достаточно много новых генов, но их происхождение не совсем ясно [154], тогда как более поздних событий ГПГ не наблюдалось. Для того, чтобы найти потенциальные случаи ГПГ из бактерий в *Pyrococcus* spp., мы использовали тот же подход, что и в случае *Methanosarcina* spp. Все гены *Pyrococcus* spp. образуют ГОБ без синглтонов. Около 76% всех ГОБ являются универсальными и не содержат паралогов. Мы нашли только одну ГОБ, содержащую гипотетические белки, предположительно полученную в результате ГПГ в последнего общего предка *Thermococcaceae*. Это показывает, что, действительно, геномы представителей *Methanosarcinaceae* очень динамичны по сравнению с другими археями.

### 3.2.5 Таксономическое разнообразие перенесенных генов

Далее мы попытались выяснить, из каких именно групп бактерий в основном происходит ГПГ в общего предка *Methanosarcina* spp. или *Methanosarcinaceae*. 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

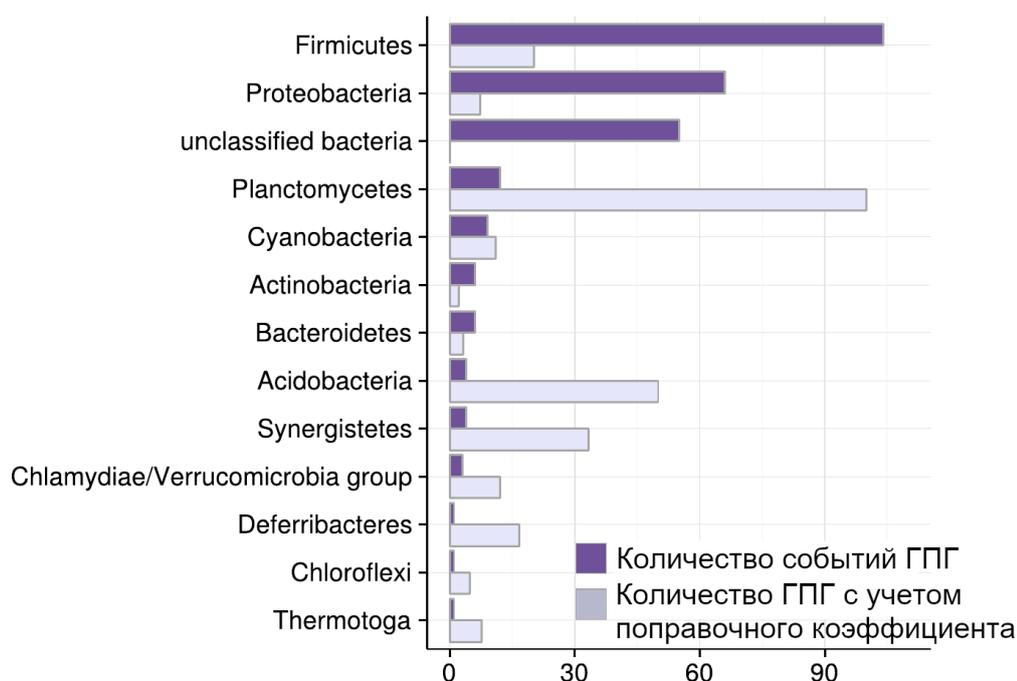
из *Proteobacteria*; 12 из *Planctomycetes*; 6 из группы *Bacteroides/Chlorobi*; 6 из *Actinobacteria*; 9 из *Cyanobacteria* (Табл. 3.3). Источник ГПГ не удалось установить для 46 ГОБ и 9 синглтонов. Несмотря на то, что *Actinobacteria* – это хорошо изученный таксон с большим количеством отсеквенированных геномов, нам удалось найти мало случаев ГПГ, где бы они выступали в качестве донора, и все обнаруженные события произошли у последнего общего предка *Methanosarcina*. Более того, все перенесённые из *Actinobacteria* гены потеряны у *M. mazei*.

Таксономическая группа	ГПГ в общего предка <i>Methanosarcina</i> (количество ГОБ)	ГПГ в общего предка <i>Methanosarcinaceae</i> (количество ГОБ)	Всего
<i>Firmicutes</i>	79	25	104
<i>Proteobacteria</i>	50	16	66
unclassified bacteria	27	28	55
<i>Planctomycetes; Planctomycetacia</i>	10	2	12
<i>Cyanobacteria</i>	7	2	9
<i>Bacteroidetes/Chlorobi group; Bacteroidetes</i>	3	3	6
<i>Actinobacteria</i>	6	0	6
<i>Synergistetes</i>	4	0	4
<i>Acidobacteria</i>	4	0	4
<i>Chlamydiae/Verrucomicrobia group</i>	3	0	3
<i>Thermotoga</i>	1	0	1
<i>Chloroflexi</i>	1	0	1
<i>Deferribacteres</i>	1	0	1
<i>Deinococcus-Thermus; Deinococci</i>	0	0	0

**Таблица 3.3.** Таксономия доноров ГПГ. В данном случае, *Methanosarcina* означает ГПГ либо в последнего общего предка *Methanosarcina*, либо более поздние события.

Поскольку *Firmicutes* и *Proteobacteria* – это самые распространённые группы бактерий с наибольшим количеством секвенированных геномов, ГПГ из этих групп может быть переоценен. Действительно, было показано, что количество уникальных ГОБ растёт с увеличением пан-генома [260,261], тем самым лучше отсеквенированные таксоны содержат больше разных генов в своих пан-геномах. Для того, чтобы учесть этот эффект, мы ввели поправочный коэффициент на изученность таксона, который определяется как количество ГОБ, содержащих гены, перенесённые из данного таксона, делённое на количество отсеквенированных для данного таксона геномов (Рисунок 3.4). Такой

коэффициент предполагает линейный рост размера пан-генома с ростом количества геномов, что хорошо соотносится с литературными данными [260]. После такой нормализации оказалось, что наиболее перепредставленные таксоны-доноры – это *Planctomycetes*, *Synergistetes* и *Firmicutes*. Для бактерий из этих таксономических групп, так же, как и для *Proteobacteria* и *Bacteroidetes* было показано, что населяют одни и те же экологические ниши, что и *Methanosarcina* spp. [221,262,263].



**Рисунок 3.4. Таксономическое разнообразие перенесенных генов.** Количество событий ГПГ для каждого донора показано фиолетовым, значения с учетом поправочного коэффициента представлены светло-фиолетовым цветом.

### 3.2.6 Функциональное разнообразие перенесенных генов

Другой интересный вопрос – это какие функциональные категории генов чаще всего подвержены ГПГ. Для того, чтобы исследовать этот вопрос, ко всем найденным нами горизонтально перенесённым генам мы приписали соответствующие им категории COG, 86% (506 белков) имело аннотацию в базе данных COG. Мы показали, что половина горизонтально перенесённых генов ассоциированы с метаболизмом (Табл. 3.4). 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 3.5а). Среди перепредставленных категорий оказались защитные механизмы (группа D), все группы, ассоциированные с метаболизмом, за исключением липидного (I): транспорт и метаболизм аминокислот (E); энергетический обмен (C); транспорт и метаболизм нуклеотидов (F); транспорт и метаболизм неорганических ионов (P); транспорт и метаболизм углеводов (G); биосинтез, транспорт и катаболизм вторичных метаболитов (Q); транспорт и катаболизм кофакторов (H).

Категория из базы данных COG		<i>Methanosarcina</i> ГОБ						<i>Methanosarcinaceae</i> ГОБ							
		Кол-во	Enz	Tr	1	2	3	Кол-во	Enz	Tr	2	3	4	5	6
Хранение и обработка генетической информации		18 (9,7%)						7 (11,7%)							
J	Трансляция, биогенез рибосомы	6	3	0	2	2	2	2	1	0	0	0	1	1	0
K	Транскрипция	7	1	0	5	1	1	3	0	0	0	0	0	2	1
L	Репликация, рекомбинация и репарация	5	2	0	2	1	2	2	1	0	0	0	0	1	1
Клеточные процессы и сигнальные пути		26 (14%)						9 (15%)							
V	Защитные механизмы	11	7	3	4	5	2	1	1	0	0	1	0	0	0
T	Механизмы передачи сигнала	5	2	0	3	2	0	1	0	0	0	0	0	0	1
M	Биогенез клеточной стенки и мембраны	8	1	0	4	1	3	5	2	2	0	1	2	2	0
N	Клеточная подвижность	1	1	0	0	1	0	0	0	0	0	0	0	0	0
O	Посттранскрипционная модификация и шапероны	1	0	0	0	1	0	2	0	0	0	0	1	1	0
Метаболизм		95 (51,3%)						37 (54%)							
C	Энергетический обмен	21	10	0	3	14	4	10	2	0	0	1	6	3	0
E	Метаболизм и катаболизм аминокислот	25	10	8	1	13	11	9	5	1	0	0	2	1	6
F	Транспорт и метаболизм нуклеотидов	7	5	0	1	4	2	3	2	0	0	0	0	0	3
G	Транспорт и метаболизм углеводов	10	2	2	1	4	5	3	1	0	0	1	1	0	1
H	Транспорт и метаболизм коферментов	10	3	0	1	7	2	4	1	0	1	0	1	0	2
I	Транспорт и метаболизм липидов	3	0	0	0	1	2	0	0	0	0	0	0	0	0
P	Транспорт и метаболизм неорганических ионов	16	2	9	2	9	5	6	0	5	0	0	0	1	5

Категория из базы данных COG		<i>Methanosarcina</i> ГОБ						<i>Methanosarcinaceae</i> ГОБ							
		Кол-во	Enz	Tr	1	2	3	Кол-во	Enz	Tr	2	3	4	5	6
Q	Биосинтез, транспорт и катаболизм вторичных метаболитов	3	0	0	0	2	1	2	1	0	0	0	1	0	1
Функция плохо описана		46 (24,8%)						16 (23,5%)							
R	Известна только общая функция	31	10	2	10	13	8	11	2	0	0	0	3	2	6
S	Функция неизвестна	15	0	0	1	4	10	5	0	0	0	0	1	3	1

**Таблица 3.4.** Функциональное разнообразие горизонтально перенесённых генов. Tr. – количество ГОБ, содержащих транспортёры, Enz. – количество ГОБ, содержащих ферменты. Цифры в заголовке означают количество геномов, в которых встречаются представители данного ГОБ. Синглтоны также включены в анализ.

Помимо этого, мы разделили все горизонтально перенесённые гены на четыре группы: гены домашнего хозяйства, ферменты, транспортёры и транскрипционные факторы. В первую категорию были включены гены, ассоциированные с репликацией, транскрипцией и трансляцией, а также гены синтеза клеточной стенки.

Количество транскрипционных факторов и генов домашнего хозяйства в нашей выборке было оценено по КОГ, тогда как количество транспортёров и ферментов было оценено по базам данных PFAM и EFICAZ, соответственно.

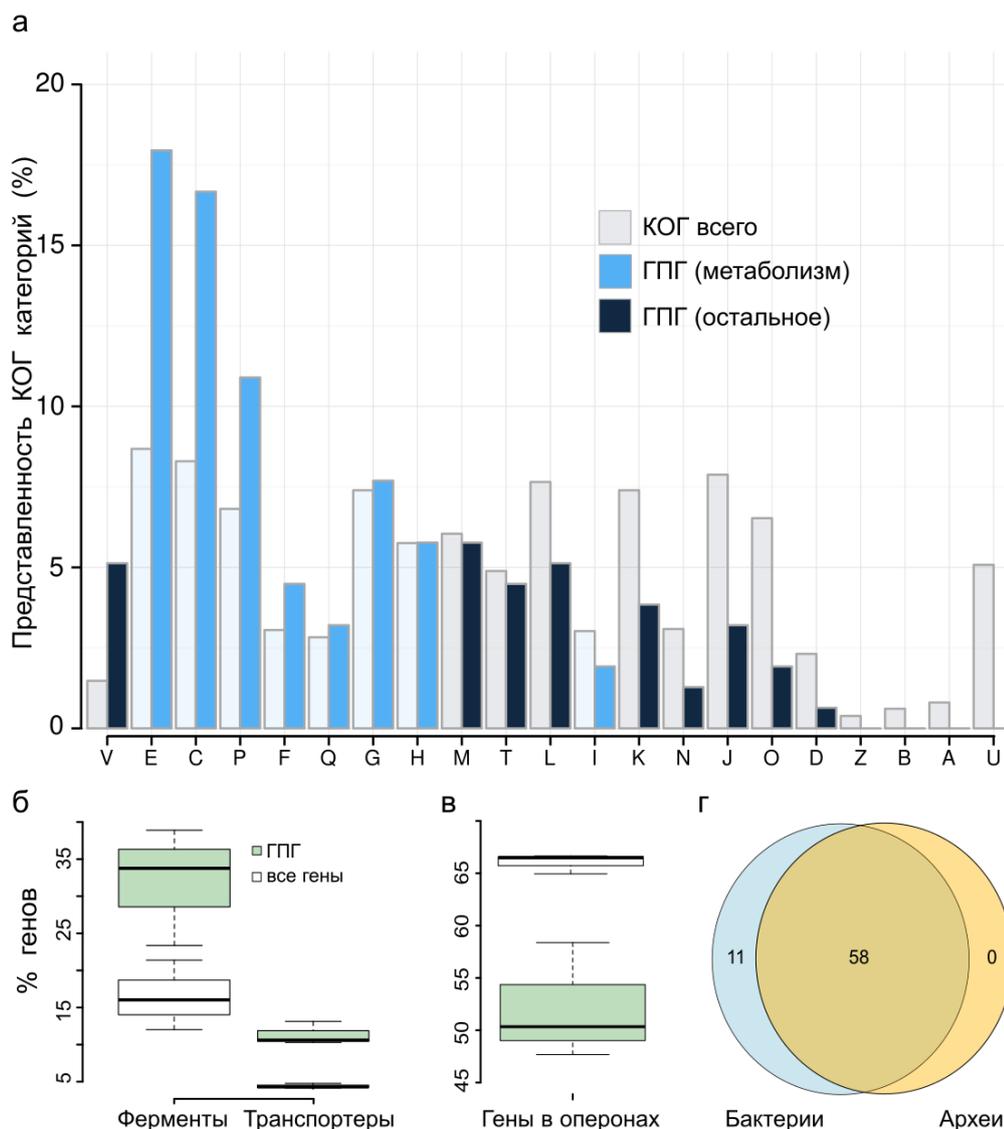
Среди горизонтально перенесённых генов есть только 7 ГОБ, содержащих гены домашнего хозяйства: дигидроуридин синтаза, цистеинил-тРНК-синтетаза; две ацетилтрансферазы или метилтрансферазы, более точные функции которых не удалось установить; и три гена, ассоциированных с системой репарации ДНК.

Из 226 транскрипционных факторов, описанных для *Methanosarcina* spp., только семь (относящихся к двум ГОБ) получены в результате ГПГ. Приблизительно 4.5% всех генов *Methanosarcina* spp. были охарактеризованы нами как транспортные (505 белков), что хорошо соотносится с ранее описанными наблюдениями [264]. 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 (Табл. 3.5). 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 (Табл. 3.5).

	Ферменты			Транспортные белки			Тест $\chi^2$ <i>p</i>
	Кол-во среди ГПГ	Кол-во во всем геноме	% ГПГ	Кол-во среди ГПГ	Кол-во во всем геноме	% ГПГ	
<i>M. acetivorans</i>	86 (38.9%)	971 (21.4%)	8.9	29 (13.1%)	194 (4.3%)	14.9	0.0094
<i>M. barkeri</i>	50 (23.4%)	435 (12%)	11.5	22 (10.3%)	172 (4.7%)	12.8	0.6562
<i>M. mazei</i>	51 (33.8%)	540 (16%)	9.4	16 (10.6%)	139 (4.1%)	11.5	0.4663
All	187 (31.9%)	1946 (16.9%)	9.6	67 (11.4%)	505 (4.4%)	13.3	0.0163

**Таблица 3.5.** Статистика встречаемости транспортных белков и ферментов среди как генов, полученных в результате ГПГ в *Methanosarcinaceae*, так и по геному в целом. Проценты в скобках обозначают долю генов либо среди всех горизонтально перенесённых генов, либо от всех генов в данном геноме, в столбце “% ГПГ” показан процент горизонтально перенесённых генов в данном геноме. Статистическая значимость посчитана для нулевой гипотезы, что доля ферментов и транспортёров среди горизонтально перенесённых генов такая же, как и по геному в целом.

Эти результаты показывают, что гены, вовлечённые в метаболизм, как и гены, ассоциированные с механизмами защиты (например, система рестрикции-модификации и хлорамфеникол О-ацетилтрансфераза), часто переносятся из бактерий в археи. Гены, ассоциированные с катаболизмом и метаболизмом липидов, наоборот, переносятся очень редко, что может быть объяснено радикальными отличиями в структуре мембран архей и бактерий [41]. Наблюдаемый нами частый ГПГ ферментов и транспортёров соответствует описанным ранее литературным данным для разных групп микроорганизмов [145,149,260] (Рисунок 3.5б). Несмотря на то, что случаи ГПГ и ферментов [24], и транспортёров [22] были описаны для архей и раньше, скорости этих переносов никогда не сравнивались напрямую.



**Рисунок 3.5.** 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).

### 3.2.7 Оперонная структура

Для того, чтобы охарактеризовать опероны *Methanosarcina* (см. определение в Материалах и методах), были посчитаны все межгенные расстояния для каждого генома *Methanosarcina* spp. Для того, чтобы наилучшим образом описать опероны, мы

использовали первоначально шесть разных порогов на максимальное расстояние между генами: 300, 250, 200, 150, 100 и 50 пн. В результате первичного анализа мы остановились на пороге в 150 пн, так как при более длинных порогах гены, полученные в результате ГПГ, сливались в длинные опероны с генами, характерными для архей, а при более коротких порогах менее половины генов оказывались в оперонах.

По нашей оценке примерно две трети генов *Methanosarcina* spp. входят в нетривиальные опероны. Из всех оперонов, содержащих горизонтально перенесенные гены (Табл. 3.6), 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 (Табл. 3.6, Рисунок 3.5в). Средний размер оперона, содержащего горизонтально перенесённые гены, составляет 3.5 гена, тогда как средняя длина оперонов в геномах *Methanosarcina* немного больше и составляет 4.1 гена. 50 оперонов *M. acetivorans*, 32 – *M. barkeri* и 33 – *M. mazei* содержат как горизонтально перенесенные, так и характерные только для архей гены. Средняя длина горизонтально перенесенного фрагмента составляет 1300 нуклеотидов (см. раздел 3.1.7), тогда как самый длинный перенесенный фрагмент составляет более 7000 нуклеотидов и содержит девять генов натриевой АТФазы. Эти гены формируют один длинный оперон, характерный для *M. acetivorans* и *M. barkeri*. Для этого оперона было ранее показано, что он был перенесен в результате одного события горизонтального переноса [265].

	Во всём геноме				ГПГ				Тест $\chi^2$
	Одиночные гены	Гены в оперонах	Кол-во оперонов	%	Одиночные гены	Гены в оперонах	Кол-во оперонов	%	<i>p</i>
<i>M. acetivorans</i>	1543	3063	768	66,5%	92	129	80	58,4%	0.0087
<i>M. barkeri</i>	1296	2402	570	65,0%	112	102	61	47,7%	5E-008
<i>M. mazei</i>	1145	2290	543	66,7%	75	76	56	50,3%	1E-005

**Таблица 3.6.** Включение генов в опероны среди горизонтально перенесённых генов, так и в геноме в целом. Гены в оперонах – гены, включённые в нетривиальные опероны. Статистическая значимость посчитана для нулевой гипотезы, что доля генов, включённых в опероны, одинаковая среди горизонтально перенесённых генов и по геному в целом.

### 3.2.8 Экспрессия горизонтально перенесенных генов

Для того, чтобы понять, экспрессируются ли гены, полученные в результате ГПГ, мы изучили доступные протеомные и экспрессионные данные для трёх видов *Methanosarcina* (см. Материалы и методы). По данным четырёх экспериментов, проведённых на *M. mazei* [266–269], 82 из 151 гена, предположительно полученные *M. mazei* в результате ГПГ, экспрессируются в разных условиях окружающей среды [270]. Для 30 из этих генов было показано изменение экспрессии в отсутствие гистона НМm; 66 генов по-разному экспрессировались в средах с разными источниками азота; 6 показывали высокий уровень экспрессии на средах с метанолом, а 10 – на средах с ацетатом. В этих экспериментах суммарно для 55% генов была показана дифференциальная экспрессия в разных условиях, т. е. фракция дифференциально экспрессирующихся горизонтально перенесённых генов не ниже, чем в среднем по геному.

На основании анализа протеомных данных, доступных для *M. acetivorans*, удалось показать, что в клетках этих архей присутствуют 13 белков (6.7%) из горизонтально перенесённых (для сравнения в этих экспериментах удалось обнаружить продукты 9.1% всех генов, присутствующих в геноме) [271,272]. На основании анализа экспрессионных данных [273], была показана дифференциальная экспрессия ещё 17 генов (12.6%) в средах с метанолом или ацетатом (всего в этом эксперименте была показана дифференциальная экспрессия для 27.5% всех генов *M. acetivorans*). В случае *M. barkeri* наблюдалась дифференциальная экспрессия семи горизонтально перенесённых генов (3.6%) в условиях разной аэрации, тогда как всего в геноме 40 (1.6%) генов, меняющих свою экспрессию в этих условиях [274]. В целом, в 46% от всех ГОБ, полученных в результате ГПГ, как минимум для одного гена из группы показана дифференциальная экспрессия, что говорит о том, что как минимум половина генов, полученных *Methanosarcina* spp. в результате ГПГ, дифференциально экспрессируются, и, следовательно, находятся под регуляцией. Следует отметить, что это оценка снизу, поскольку экспериментальные данные доступны лишь для ограниченного числа условий.

## 3.3 Обсуждение результатов

ГПГ играет важную роль в разнообразии микроорганизмов и их адаптации к окружающей среде. Считается, что захват генов при помощи горизонтального переноса позволяет быстро адаптироваться к новым экологическим нишам [135,144,275] и играет важную роль в видообразовании [24,148]. Было показано, что ГПГ в бактериях играет

важную роль в быстром распространении генов, ответственных за устойчивость к антибиотикам [276]. У бактерий *Rhodobacter capsulatus* были обнаружены специальные агенты переноса генов (GTA), по происхождению являющиеся вирусными частицами [277]. Было показано, что в морских популяциях бактерий уровень ГПГ генов устойчивости к антибиотикам, находящихся в GTA, очень высок [278].

Известны оценки скорости ГПГ для разных групп эу- и прокариот [279]. В частности, было показано, что все представители рода *Methanosarcina* обладают очень динамичным геномом и подвержены ГПГ из бактерий [23,160,161]. В более поздних работах была изучена скорость ГПГ в *Methanosarcinales* [22,154], но подробно горизонтальный перенос в семейство *Methanosarcinaceae* не изучался, более того не были изучены функции генов, перенесённых в последнего общего предка *Methanosarcina* spp. или *Methanosarcinaceae*. По нашей оценке, около 5% всех генов *Methanosarcina* были получены в результате ГПГ из бактерий, из них 3.5% от всех генов были перенесены в последнего общего предка *Methanosarcina* spp. или позже, а 1.5% – результат ГПГ в последнего общего предка *Methanosarcinaceae*. Полученная нами оценка хорошо соотносится с результатами анализа arCOG, где было идентифицировано 321 событие ГПГ [154]. Однако, в отличие от нашей работы, для этих генов не был определён донор, из которого был произведён перенос.

Для того, чтобы понять, почему в ранних работах вклад ГПГ в геном *Methanosarcina* spp. был завышен [23,160], мы провели поиск по имитированным базам данных, воспроизводящим состав белковых баз данных в разные годы. Оказалось, что количество предсказанных событий действительно зависит от состава базы данных. Помимо этого, мы оценили количество ГПГ в *Thermotoga maritima*, и не смогли воспроизвести исходные результаты. Мы нашли только 28 событий ГПГ из архей в *Thermotoga*, что значительно меньше, чем полученные ранее результаты [246,257,259,280]. В обоих случаях эту разницу в наблюдениях удаётся объяснить за счёт различий в составе базы данных в разные годы. Поскольку здесь мы отслеживали только достаточно новые события, произошедшее не раньше, чем на уровне последнего общего предка семейства, мы не рассматривали часть событий ГПГ, которые могли произойти сильно раньше. С другой стороны, в ранних работах из-за неравномерного покрытия секвенированными геномами таксономического пространства такие ГПГ могли определяться как ГПГ в конкретный вид или род. Так, например, известно много событий ГПГ в *Methanosarcinales* [22].

Считается, что ГПГ может произойти между любыми группами прокариот с

похожей частотой использования кодонов [152,153], но для того, чтобы ГПГ действительно произошёл, организмы должны иметь общие места обитания [281], а перенесённые гены не должны быть токсичны в новом контексте [153,282]. Анализ arCOG не выявил заметных таксономических предпочтений ГПГ [283]. Однако в нашем анализе получается, что хотя бы для *Methanosarcina* spp. такие предпочтения могут существовать, и ГПГ из *Clostridia* и *Proteobacteria* случается чаще, чем из остальных групп. Так как эти группы одновременно являются самыми хорошо изученными, мы попытались учесть этот эффект, и провести нормировку. Наша оценка очевидно имеет ряд ограничений просто потому, что группы, для которых отсеквенировано большое количество геномов, совсем не обязательно равномерно покрыты, и к тому же многократные события ГПГ могут дополнительно осложнять ситуацию. Поэтому мы приводим и нормированную, и ненормированную оценки, замечая при этом, что обе показывают заметные таксономические предпочтения при выборе доноров ГПГ, одно из которых, вероятно, – предпочтение *Firmicutes*.

Функциональный анализ горизонтально перенесённых генов показал, что и ферменты, и транспортные белки часто появляются в геноме в результате ГПГ. Транспортные белки по нашим оценкам переносятся чуть чаще, что кажется логичным, потому что перенос новых ферментов предполагает встраивание их в имеющуюся систему метаболических реакций, тогда как транспортёры менее зависят от других систем клетки. Ранее было показано, что ГПГ генов, вовлечённых в большое количество белок-белковых взаимодействий, происходит реже [150,151]. Возможно, однако, что предпочтение транспортёров объясняется ограниченностью наших данных, так как транспортёры специфичные для архей плохо изучены, и мы могли пропустить некоторые группы транспортёров, что влияет на нормировку [264].

Теоретически, быстрое получение в результате ГПГ новых транспортных белков должно являться хорошим механизмом адаптации к новым условиям, позволяя получать новые ионы и соединения, которые были недоступны до этого. Известно, что ABC-транспортёры подвержены ГПГ и часто переносятся между разными доменами, и потому в ситуациях, когда такие транспортёры попали в *Methanosarcina* spp. не напрямую, а из другой археи, которая пока ещё неотсеквенирована, мы будем делать неправильные предположения о доноре ГПГ. И наоборот, если ген, полученный *Methanosarcina* spp. от бактерий, был перенесен дальше в других отсеквенированных архей, мы не сможем детектировать такое событие как ГПГ. Поэтому мы концентрируемся в этом исследовании в основном на недавних событиях, где вклад таких сложных переносов должен быть

минимален. Важно отметить, что несмотря на то, что мембраны архей и бактерий значительно отличаются, наши результаты показывают, что транспортёры могут работать на обоих типах мембран.

Интересно, что более половины генов, полученных в результате ГПГ в *M. mazei*, функционируют, т. е. для них была описана дифференциальная экспрессия, что значит, что их экспрессия в клетке каким-то образом регулируется. Общее количество транскрипционных регуляторов в *Methanosarcinales* соответствует предсказаниям, полученным для всех прокариотических организмов [284]. Понятно, что такая оценка может быть не совсем точной, потому что специфичные для архей КОГ не описаны, т. е. возможно, что целые классы регуляторов нам просто не известны (Рисунок 3.5г). Мы обнаружили только два случая переноса регуляторов транскрипции, тогда как если бы регуляторы переносились вместе с регулируемыми генами, мы бы наблюдали значительно больше таких событий. Тем самым перенесённые гены, по всей видимости, встраиваются уже в имеющиеся регулоны.

Более того, поскольку события ГПГ, которые мы изучали, произошли недавно, нам удалось оценить предполагаемую длину переносимых фрагментов, считая, что они не подвергались дальнейшим геномным перестройкам. Средняя длина переносимого фрагмента – это 1200 нт, что приблизительно соответствует длине одного белок-кодирующего гена. Более того многие гены, полученные в результате ГПГ, включены в опероны, содержащие также гены, специфичные только для архей. Возможно, что такое ограничение на длину фрагментов связано с предпочтительным механизмом переноса.

## Глава 4. Сравнительный анализ геномов *Holospora* spp. – внутриядерных симбионтов инфузорий

Большая часть эндосимбиотических бактерий могут передаваться только вертикально, тогда как представители рода *Holospora* 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 этой главе приводятся результаты сравнительно-геномного анализа представителей рода *Holospora*, включая геном отсеквенированного нами вида *Holospora curviuscula*, макронуклеарного эндосимбионта *Paramecium bursaria*.

### 4.1 Материалы и методы

#### 4.1.1 Выращивание бактерий и выделение геномной ДНК<sup>3</sup>

*Holospora curviuscula* является облигатным симбионтом *Paramecium bursaria*, и не культивируется, поэтому бактерия выращивалась внутри клеток хозяина. *Holospora curviuscula* штамма NRB217, выбранная для этого эксперимента, получена из *Paramecium bursaria*, собранной в Новосибирском Академгородке. Эта заражённая линия *P. bursaria* хранится в коллекции культур симбионт содержащих инфузорий лаборатории Кариологии одноклеточных организмов в Научном парке СПбГУ. Клетки инфузории-хозяина культивировались при комнатной температуре в среде, содержащей экстракт латука и в присутствии *Enterobacter aerogenes* в качестве источника пищи для инфузории. Культура *P. bursaria*, зараженной *H. curviuscula*, была сконцентрирована при помощи центрифугирования (10 минут при 4500g) и затем гомогенизирована с использованием 10%-го раствора детергента Nonidet P-40 (Sigma-Aldrich Cat No. 21-3277 SAJ). Инфекционная форма *H. curviuscula* была выделена из гомогената путём центрифугирования в градиенте перколла (Sigma-Aldrich Cat No. P1644) как описано ранее [285]. Геномная ДНК была выделена при помощи набора DNeasy Blood and Tissue kit

---

<sup>3</sup> Выращивание *H. curviuscula*, выделение ДНК и секвенирование, описанные в разделах 4.1.1 и 4.1.2, не являются частью настоящей диссертационной работы, а приведены здесь для полноты и лучшего понимания.

(QIAGEN Cat No. 69504) с использованием немного изменённого протокола – время инкубирования клеток в ATL буфере, содержащем протеиназу К, было увеличено до 16 часов.

#### 4.1.2 Секвенирование геномной последовательности, сборка и аннотация

Было подготовлено две библиотеки для секвенирования: одна библиотека для парных ридов при помощи MiSeq (2x250) (PE) со стандартной длиной вставки, а другая – со вставкой 4 кб (MP). Количество ридов полученных для MP библиотеки составило 2.5 млн ридов, а для PE – 4.5 млн. Полученные MP риды были отфильтрованы при помощи NextClip v1.3 [286] так, что в рассмотрении были оставлены только риды категории A (т. е. такие риды, где оба рида в паре содержали адаптеры). Затем и MP, и PE риды фильтровались по качеству при помощи trimmomatic версии 0.33 (Leading 15, sliding -15, slidingwindow -4:25). После фильтрации осталось 1.3 млн ридов MP библиотеке и 1.7 млн ридов в PE библиотеке, эти риды использовались для дальнейшего анализа. Сборка проводилась при помощи программы platanus версии 1.2.1. Полученная сборка была выложена в DDBJ/ENA/GenBank с идентификационным номером PNHС00000000. Версия, описываемая здесь, это версия PNHС01000000.

Для оценки качества сборки *Holospira* spp. мы использовали список универсальных PFAM доменов [287]. Мы считали, что домен присутствует в данном геноме, если он был найден при помощи программы hmmsearch из пакета HMMer (with default parameters) [288], и значение bias было меньше, чем вес данной находки. Для всех доменов из списка, для которых находок не было найдено, мы провели дополнительный поиск среди 60 геномов *Rickettsiales* и 23 других эндосимбиотических бактерий, относящихся к *Proteobacteria* и имеющих геномы менее 0.8 Mb (полный лист приведен в Табл. 4.1).

Название генома	Тип	Класс	Порядок	Семейство	Размер
<i>Buchnera aphidicola</i> APS	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	655725
<i>Buchnera aphidicola</i> Bp	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	618379
<i>Buchnera aphidicola</i> Cc	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	422434
<i>Buchnera aphidicola</i> Ct	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	444925
<i>Buchnera aphidicola</i> G002	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	651316
<i>Buchnera aphidicola</i> LL01	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	641799
<i>Buchnera aphidicola</i> Sg	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	641454
<i>Buchnera aphidicola</i> Ua	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	627953

<i>Candidatus</i> Blochmannia chromaioides 640	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	791219
<i>Candidatus</i> Blochmannia floridanus	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	705557
<i>Candidatus</i> Blochmannia pennsylvanicus BPEN	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	791654
<i>Candidatus</i> Blochmannia vafer BVAf	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	722593
<i>Candidatus</i> Carsonella ruddii CE, Thao2000	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	unclassified	unclassified	162589
<i>Candidatus</i> Carsonella ruddii CS, Thao2000	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	unclassified	unclassified	162504
<i>Candidatus</i> Carsonella ruddii HC, Thao2000	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	unclassified	unclassified	166163
<i>Candidatus</i> Carsonella ruddii HT, Thao2000	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	unclassified	unclassified	157543
<i>Candidatus</i> Carsonella ruddii PC, NHV	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	unclassified	unclassified	159923
<i>Candidatus</i> Carsonella ruddii PV	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	unclassified	unclassified	159662
<i>Candidatus</i> Moranella endobia PCIT	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	538294
<i>Candidatus</i> Moranella endobia PCVAL	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	538203
<i>Candidatus</i> Riesia pediculicola USDA	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	582127
<i>Wigglesworthia glossinidia</i> endosymbiont of <i>Glossina brevipalpis</i>	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	703004
<i>Wigglesworthia glossinidia</i> endosymbiont of <i>Glossina morsitans</i>	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Enterobacteriales</i>	<i>Enterobacteriaceae</i>	719535

**Таблица 4.1.** Штаммы эндосимбиотических бактерий, использовавшихся для поиска минимально необходимых генов

Полученные контиги были проаннотированы при помощи PROKKA [289]. Метаболические пути были реконструированы при помощи SEED [290] и BioCyc [291]. Для поиска генов известных бактериоцинов в геноме использовались базы данных APD3 [292] и BAGEL3 [293], а также PFAM. Для поиска генов синтеза вторичных метаболитов был использован сервер AntiSMASH [294]. Трансмембранные спирали в белках были предсказаны при помощи сервера TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>). Специфичность транспортёров была предсказана с помощью TCDB [295]. Программы SignalP и SecretomeP были использованы для предсказания сигнальных пептидов и секретируемых белков, соответственно [296,297]. Для предсказания мотива twin-arginine использовался сервер TatP [298]. Белки-автотранспортёры были предсказаны поиском PFAM доменов: PF03797, PF12951, PF05662, PF05658, PF11924, PF11557, PF16168,

PF15403, PF03895 – при помощи пакета HMMer. Для поиска профаговых участков использовался сервер PHAST [299]. Помимо этого, мы искали такие участки поиском по всем PFAM доменам с ключевыми “transposase” и “phage” (всего 328 PFAM доменов).

### 4.1.3 Построение филогенетического древа *Holospora* spp.

Для того, чтобы реконструировать филогению *Holospora* spp., использовался пакет PhyML [251]. Последовательности генов 16S рРНК *H. curviuscula* NRB217, *H. elegans* E1, *H. obtusa* и *H. undulata* HU1 были взяты из полных геномных последовательностей, тогда как гены 16S рРНК *H. acuminata* и *Candidatus Gortzia infectiva* были скачены из базы данных SILVA (идентификаторы: KC164379 и HE797907, соответственно). Для определения положения корня на дереве использовалась последовательность 16S рРНК из *Rickettsiales* (идентификационный номер в базе данных SILVA: CP009217). Филогенетическое древо было реконструировано при помощи программы PhyML версии 3.1, использовалось 100 бутстрэп испытаний, модель замен GTR и параметр -o tlr.

### 4.1.4 Анализ повторов

Для того, чтобы оценить количество повторов в геномах *Holospora*, риды были выравнены на геномную сборку при помощи программы Bowtie 2 (--end-to-end mode) [300] и было посчитано покрытие для каждой позиции генома. В случае *H. curviuscula* были использованы только PE риды, полученные для данного исследования; для *H. obtusa* и *H. undulata* риды были скачены из базы данных DDBJ (идентификационные номера DRP001203 и DRA001008). Фрагмент считался дублированным, если он был длиной более 20 нп, и его покрытие было как минимум в 1.6 раза больше, чем медианное по контигу.

### 4.1.5 Ортологические группы и полногеномные сравнения

Для того, чтобы изучить филогенетические взаимоотношения между *H. undulata* и *H. elegans*, были построены парные полногеномные выравнивания для всех отсеквенированных видов *Holospora* при помощи программы MAUVE (snapshot 2015-02-13) [301]. Из этих выравниваний были получены попарные выравнивания универсальных генов (см. выше) и посчитано количество однонуклеотидных замен. Все колонки с

пробелами для данной пары видов не рассматривались.

Для того, чтобы изучить особенности геномного репертуара *Holospora* spp. по сравнению с другими представителями *Rickettsiales*, были построены ГОБ при помощи программы OrthoMCL v. 2.0.9 [302] с использованием MCL версии 14-137. Все геномы, кроме секвенированного нами, были скачаны из NCBI. Полный лист использованных геномов *Rickettsiales* приведен в Табл. 4.2.

<b>Название бактерии</b>	<b>Количество генов</b>
<i>Anaplasma centrale</i> str. Israel	925
<i>Anaplasma marginale</i> str. Florida	942
<i>Anaplasma phagocytophilum</i> str. Dog2	1258
<i>Anaplasma phagocytophilum</i> str. HZ	1058
<i>Anaplasma phagocytophilum</i> str. HZ2	1055
<i>Anaplasma phagocytophilum</i> str. JM	1256
<i>Caedibacter varicaedens</i>	1622
<i>Candidatus</i> Caedibacter acanthamoebae	1683
<i>Candidatus</i> Hepatobacter penaei	1057
<i>Candidatus</i> Hodgkinia cicadicola Dsem	169
<i>Candidatus</i> Midichloria mitochondrii IricVA	1161
<i>Candidatus</i> Neoehrlichia lotoris str. RAC413	903
<i>Candidatus</i> Odysella thessalonicensis L13	2391
<i>Candidatus</i> Paracaedibacter acanthamoebae	2151
<i>Candidatus</i> Paracaedibacter symbiosus	2267
<i>Candidatus</i> Rickettsia amblyommii	1162
<i>Candidatus</i> Rickettsia asemboensis	2322
<i>Candidatus</i> Rickettsia gravesii BWI-1	1152
<i>Candidatus</i> Xenolissoclinium pacificensis L6 (sea squirt metagenome)	956
<i>Ehrlichia canis</i> str. Jake	925
<i>Ehrlichia chaffeensis</i> str. Arkansas	883
<i>Ehrlichia chaffeensis</i> str. Sapulpa	805
<i>Ehrlichia minasensis</i>	942
<i>Ehrlichia muris</i> AS145	863
<i>Ehrlichia ruminantium</i> str. Gardel	950
<i>Ehrlichia ruminantium</i> str. Welgevonden	916
<i>Ehrlichia</i> sp. HF	855
endosymbiont of <i>Acanthamoeba</i> sp. UWC36	4095
<i>Holospora elegans</i> E1	1212
<i>Holospora obtusa</i> F1	1117
<i>Holospora undulata</i> HU1	1224
<i>Neorickettsia helminthoeca</i> str. Oregon	772
<i>Neorickettsia risticii</i> str. Illinois	759
<i>Neorickettsia sennetsu</i> str. Miyayama	753
<i>Orientia chuto</i> str. Fuller	770
<i>Orientia tsutsugamushi</i> str. Boryong	1565
<i>Orientia tsutsugamushi</i> str. Ikeda	1967
<i>Rickettsia aeschlimannii</i>	127
<i>Rickettsia africae</i> ESF-5	1229
<i>Rickettsia akari</i> str. Hartford	980

<i>Rickettsia argasii</i> T170-B	1187
<i>Rickettsia australis</i> str. Cutlack	1157
<i>Rickettsia bellii</i> RML369-C	2801
<i>Rickettsia buchneri</i>	1642
<i>Rickettsia canadensis</i> str. McKiel	902
<i>Rickettsia conorii</i> str. Malish 7	1227
<i>Rickettsia</i> endosymbiont of <i>Bemisia tabaci</i>	1522
<i>Rickettsia</i> endosymbiont of <i>Ixodes pacificus</i>	1294
<i>Rickettsia</i> endosymbiont of <i>Ixodes scapularis</i>	1594
<i>Rickettsia felis</i> URRWXCal2	1302
<i>Rickettsia heilongjiangensis</i> 054	1140
<i>Rickettsia helvetica</i> C9P9	1158
<i>Rickettsia honei</i> RB	1171
<i>Rickettsia hoogstraalii</i>	1250
<i>Rickettsia japonica</i> YH	1142
<i>Rickettsia massiliae</i> MTU5	980
<i>Rickettsia massiliae</i> str. AZT80	1207
<i>Rickettsia montanensis</i> str. OSU 85-930	1217
<i>Rickettsia parkeri</i> str. Portsmouth	1318
<i>Rickettsia peacockii</i> str. Rustic	947
<i>Rickettsia philipii</i> str. 364D	1344
<i>Rickettsia prowazekii</i> str. Breinl	920
<i>Rickettsia prowazekii</i> str. BuV67-CWPP	843
<i>Rickettsia prowazekii</i> str. Chernikova	845
<i>Rickettsia prowazekii</i> str. Dachau	839
<i>Rickettsia prowazekii</i> str. RpGvF24	834
<i>Rickettsia rhipicephali</i> str. 3-7-female6-CWPP	1266
<i>Rickettsia rickettsii</i> str. 'Sheila Smith'	1345
<i>Rickettsia rickettsii</i> str. Arizona	1343
<i>Rickettsia rickettsii</i> str. Brazil	1332
<i>Rickettsia rickettsii</i> str. Colombia	1350
<i>Rickettsia rickettsii</i> str. Iowa	1384
<i>Rickettsia sibirica</i>	1519
<i>Rickettsia sibirica</i> 246	1234
<i>Rickettsia slovaca</i> 13-B	1114
<i>Rickettsia slovaca</i> str. D-CWPP	1347
<i>Rickettsia tamurae</i>	1200
<i>Rickettsia typhi</i> str. B9991CWPP	839
<i>Rickettsia typhi</i> str. Wilmington	838
<i>Rickettsiaceae</i> bacterium Os18	1002
<i>Rickettsiales</i> bacterium Ac37b	1780
unclassified <i>Holosporaceae</i>	1562
<i>Wolbachia</i>	1118
<i>Wolbachia</i> endosymbiont of <i>Cimex lectularius</i>	1216
<i>Wolbachia</i> endosymbiont of <i>Culex pipiens molestus</i>	142
<i>Wolbachia</i> endosymbiont of <i>Culex quinquefasciatus</i> Pel	1275
<i>Wolbachia</i> endosymbiont of <i>Dactylopius coccus</i> strain wDacA	1040
<i>Wolbachia</i> endosymbiont of <i>Dactylopius coccus</i> strain wDacB	1246
<i>Wolbachia</i> endosymbiont of <i>Diaphorina citri</i>	856
<i>Wolbachia</i> endosymbiont of <i>Drosophila ananassae</i>	1012
<i>Wolbachia</i> endosymbiont of <i>Drosophila melanogaster</i>	1195
<i>Wolbachia</i> endosymbiont of <i>Drosophila simulans</i> wHa	1009
<i>Wolbachia</i> endosymbiont of <i>Drosophila simulans</i> wNo	1040
<i>Wolbachia</i> endosymbiont of <i>Glossina morsitans morsitans</i>	723

<i>Wolbachia</i> endosymbiont of <i>Muscidifurax uniraptor</i>	679
<i>Wolbachia</i> endosymbiont of <i>Onchocerca ochengi</i>	643
<i>Wolbachia</i> endosymbiont of <i>Onchocerca volvulus</i> str. Cameroon	642
<i>Wolbachia</i> endosymbiont of <i>Operophtera brumata</i>	959
<i>Wolbachia</i> endosymbiont str. valsugana of <i>Drosophila suzukii</i>	1211
<i>Wolbachia</i> endosymbiont strain TRS of <i>Brugia malayi</i>	805
<i>Wolbachia</i> endosymbiont wVitB of <i>Nasonia vitripennis</i>	854
<i>Wolbachia pipientis</i>	945
<i>Wolbachia pipientis</i> wAlbB	2013
<i>Wolbachia pipientis</i> wBoll-b	1131

**Таблица 4.2.** Список использованных геномов *Rickettsiales*.

## 4.2 Результаты

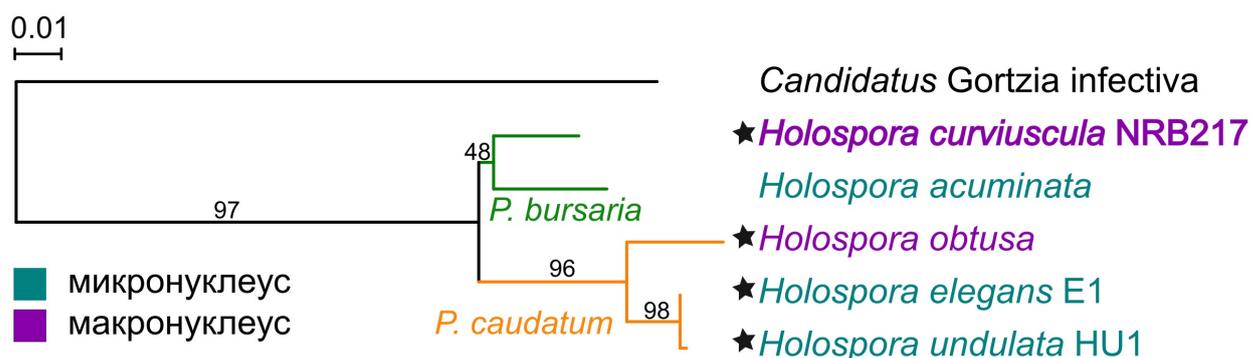
### 4.2.1 Геном *Holospira curviuscula*

Полученная нами сборка генома *H. curviuscula* состоит из 152 скэффолдов (210 контигов) с N50 ~39 тысяч п.о. Общая длина сборки составляет  $1.7 \times 10^9$  нуклеотидов, а GC-состав равен 37.6% (Таблица 3). Самый длинный скэффолд имеет длину 153367 нуклеотидов и кодирует 126 генов. Всего полученная нами геномная сборка содержит 1594 гена, из которых 1555 белок-кодирующие. Нам удалось определить функции 683 генов. Полученная нами сборка содержит все необходимые гены рРНК и 36 разных тРНК, необходимых для распознавания всех двадцати аминокислот. 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 4.1). Геном *H. curviuscula* – самый большой среди отсеквенированных геномов *Holospira* (Табл. 4.3). GC-состав этого генома чуть выше, чем у других облигатных эндосимбионтов [17], но вполне типичный для *Rickettsiales*.

	<i>H. curviuscula</i>	<i>H. obtusa</i>	<i>H. undulata</i>	<i>H. elegans</i>
Длина генома (bp)	1715500	1334837	1402636	1268333
Количество контигов	210	91	203	152
GC состав (%)	37.6%	35.2%	36.1%	36.0%
Количество генов	1594	1117	1224	1212

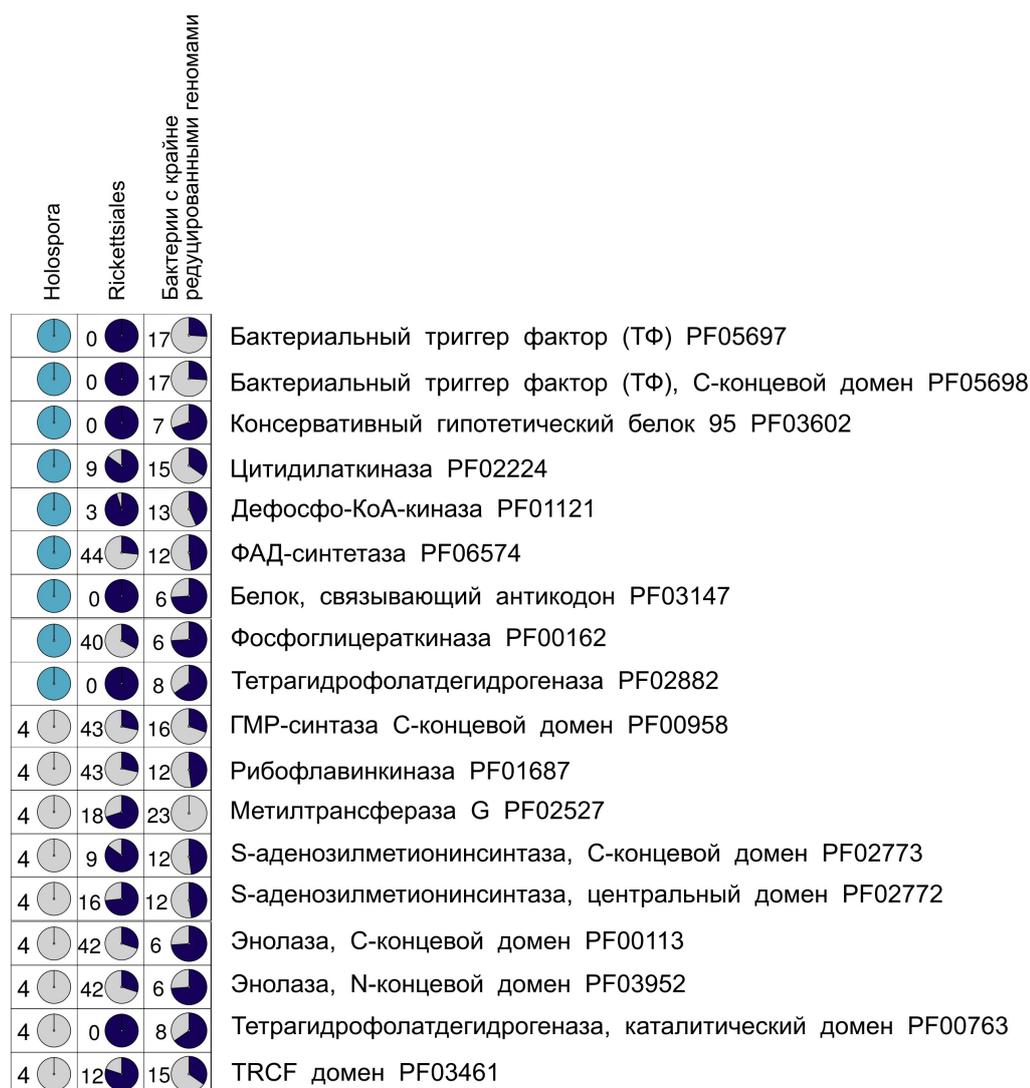
**Таблица 4.3.** Результаты сборки и аннотации генома *H. curviuscula* и сравнение с другими отсеквенированными представителями рода.

Для оценки качества сборки мы использовали набор почти универсальных генов (см. Материалы и методы). Из 139 почти универсальных PFAM доменов в геноме *H. curviuscula* был найден 121. Девять доменов из 18 оставшихся были найдены, но с сходством ниже порогового значения, а ещё 9 не были найдены не только у *Holospora*, но и других облигатных эндосимбионтов и некоторых представителей *Rickettsiales* (Рисунок 4.2), что означает, что они не являются абсолютно необходимыми. Кроме того, мы проверили не находятся ли какие-нибудь из 18 потерянных доменов среди несобранных ридов *H. curviuscula* (порядка 14% от всех ридов) и не нашли никаких похожих доменов.



**Рисунок 4.1.** Филогенетическое древо *Holospora* spp. построено методом наибольшего правдоподобия при помощи программы PhyML (см. раздел 4.1.3). Фиолетовым цветом отмечены эндосимбионты, характерные для макронуклеуса инфузорий, голубым – для микронуклеуса. Для бактерий, отмеченных звёздочками, известны геномные последовательности.

Поскольку нам удалось предсказать функции меньше, чем для половины генов, мы попытались понять, не содержит ли геном *H. curviuscula* большое количество псевдогенов. Мы искали возможные следы псевдогенизации в 615 случайно выбранных открытых рамках считывания (ОРС). Для этого мы присоединяли к нуклеотидной последовательности ОРС ещё 1000 нуклеотидов за стоп кодоном и при помощи blastx искали в базе данных nr гомологичные белки. Лишь в 9 из 615 ОРС последовательность, расположенная ниже стоп-кодона, была похожа на С-концевые участки гомологичных белков. Такая оценка уровня псевдогенизации является консервативной, потому что все представители рода *Holospora* содержат много коротких рамок считывания с неопределёнными функциями, которые на самом деле могут быть остатками более длинных предковых генов. Тем не менее, она свидетельствует о низкой представленности псевдогенов у *Holospora*.



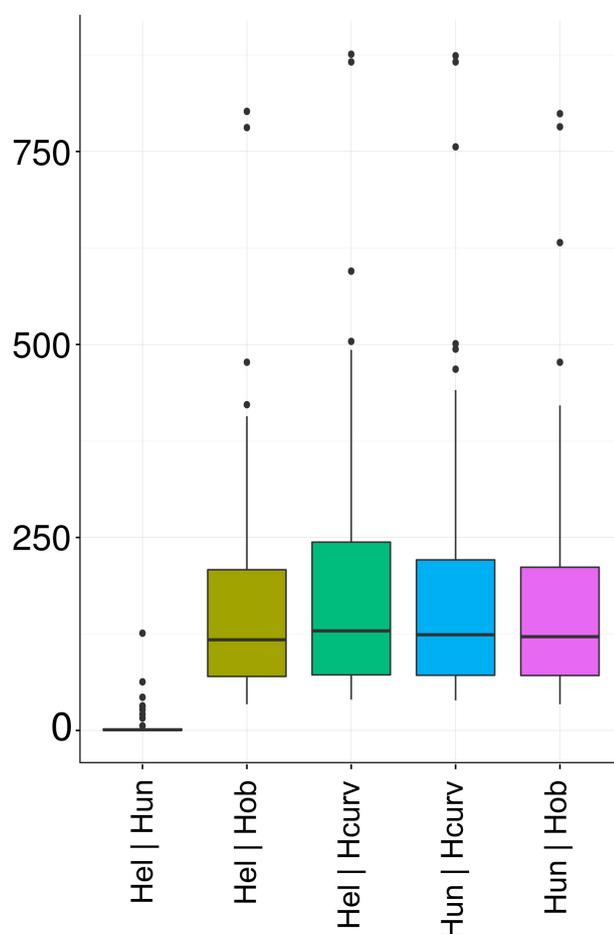
**Рисунок 4.2.** Представленность 18 необходимых PFAM доменов, отсутствующих в роде *Holospira*, у эндосимбиотических бактерий с короткими геномами и среди *Rickettsiales*. Тёмно-синий цвет означает, что домен присутствует, а серый, что домен отсутствует, голубой, что похожие домены были найдены, но ниже порога значимости (см. раздел 4.1.2). Число в ячейке – это количество геномов, в которых данный домен не был найден.

Для всех представителей *Rickettsiales* было описано очень нетипичное устройство оперона, содержащего гены рРНК [202]. Однако, у всех представителей *Holospira* этот оперон имеет типичную для бактерий структуру, когда гены 23S и 16S рРНК расположены рядом друг с другом и разделены лишь несколькими генами тРНК.

#### 4.2.2 Сравнение геномов *H. undulata* и *H. elegans*

Бактерии видов *H. undulata* HU1 и *H. elegans* E1 были выделены из микронуклеуса инфузории *P. caudatum* и определены как разные виды на основе морфологических

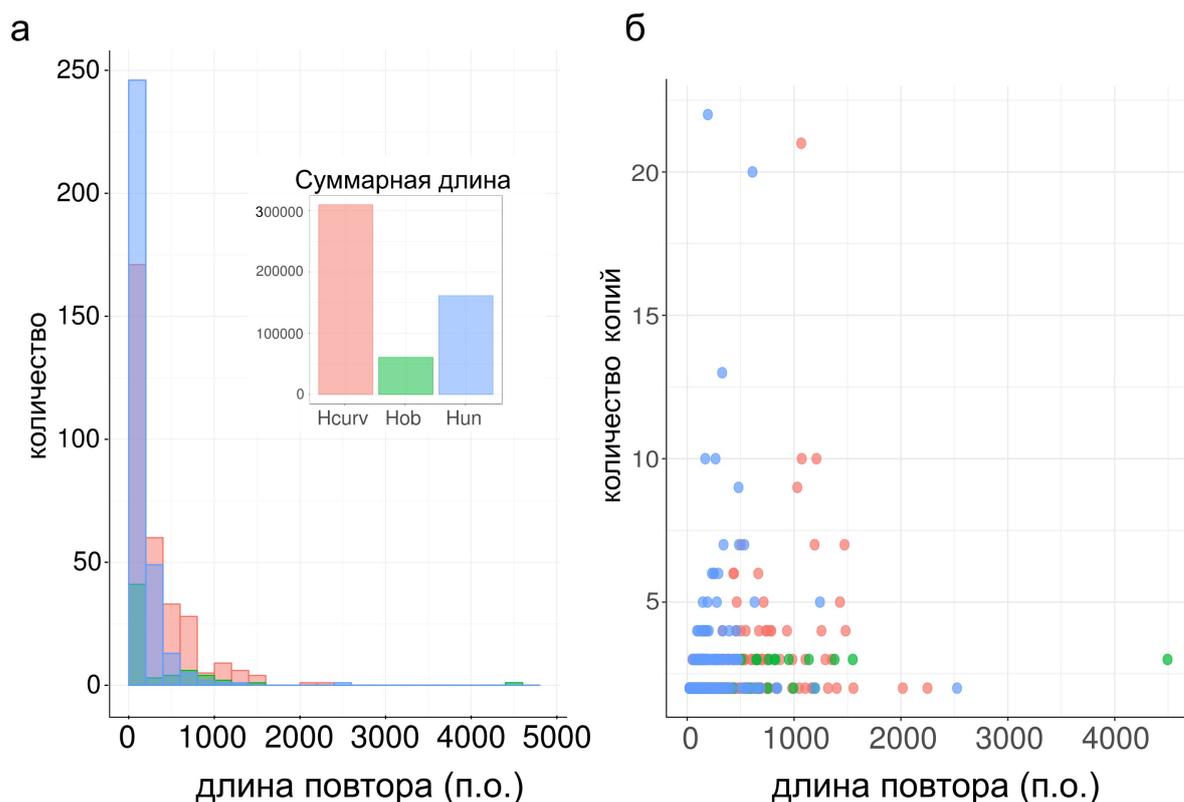
признаков [191]. Однако на основании проведённого нами анализа генов 16S рРНК из геномов этих бактерий (идентификаторы в Genbank NZ\_ARPM00000000 и NZ\_VAUP00000000) показал наличие только одной нуклеотидной замены (Рисунок 4.1). Таким образом, уровень сходства этих последовательностей >99.9%, что значительно выше стандартного порога 97%, используемого в филогенетике для разделения видов [303]. Это дало повод усомниться в том, что опубликованные геномы *H. undulata* и *H. elegans* действительно являются геномами бактерий разных видов. Анализ попарных полногеномных выравниваний также подтвердил это предположение (Рисунок 4.3). Действительно, в паре геномов *H. elegans* – *H. undulata* количество замен приблизительно в сто раз меньше, чем в других парах геномов.



**Рисунок 4.3.** Количество однонуклеотидных отличий в парах геномов *Holospora*. Hcurv – *H. curviuscula* NRB217, Hel – *H. elegans* E1, Hob – *H. obtusa* F1, Hun – *H. undulata* HU1.

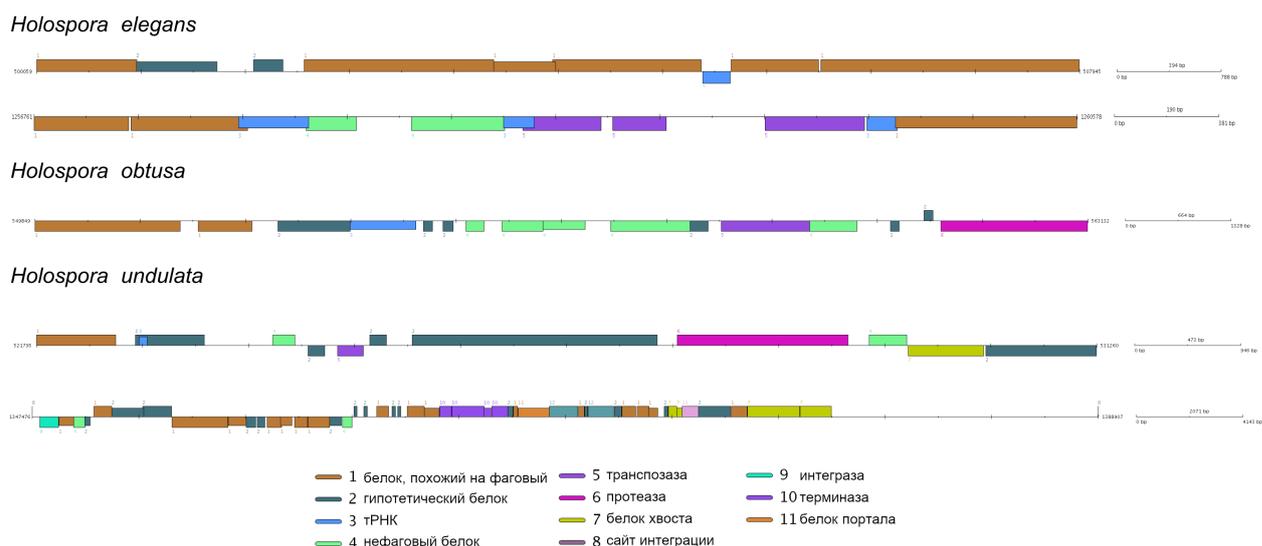
### 4.2.3 Геномы *Holospora* spp. содержат множественные повторы и последовательности профагов

Все имеющиеся на сегодняшний день сборки геномов представителей рода *Holospora* состоят из большого количества контигов, и даже улучшение алгоритмов сборки не приводит к значительному улучшению результата. По имеющимся оценкам, длина генома *H. curviuscula* составляет около 2 Мб (М.С. Раутиан, частное сообщение), тогда как длина полученной нами сборки составляет всего 1.7 Мб. Это позволило нам предположить, что геномы *Holospora* могут содержать значительное количество повторов. Для проверки этой гипотезы мы искали в геномных сборках участки с значительно увеличенным покрытием ридами (см. раздел 4.1.4). В геноме *H. curviuscula* мы нашли около 300 потенциальных повторов общей длиной 300 кб (Рисунок 4.4а). Оценка копийности этих повторов варьировала от 2 до 21 на геном (Рисунок 4.4б). В геномах *H. obtusa* и *H. undulata* оценённая нами длина повторов составила 60 и 161 кб соответственно. Мы не смогли оценить количество повторов в геноме *H. elegans*, так как для этого вида были недоступны исходные данные секвенирования. Вычисленные нами длины повторов – это оценка снизу, так как мы использовали очень жёсткий порог на увеличение покрытия участка. С другой стороны, использование жёстких порогов может означать, что мы переоцениваем количество участков, которые мы считаем повторами, так как один длинный повтор, чуть хуже покрытый ридами в середине, при таком анализе может быть рассмотрен как два независимых повтора. Несмотря на описанные выше трудности, тот факт, что общая длина найденных нами повторов совпадает с длиной нехватящего в сборке фрагмента, говорит, что повторы составляют около 15% от генома *H. curviuscula*. Один из обнаруженных нами повторов – это оперон рибосомных РНК, причем обе его копии, по-видимому, имеют одинаковую структуру. В основном повторы в *H. curviuscula* содержат короткие гены, потенциальные транспозазы и гены, похожие на фаговые.



**Рисунок 4.4.** Распределение длин и количества копий повторов в геномах *Holospora* spp. (а) Длины повторов в геномах *Holospora* spp. (б) Соотношение длины и количества копий повторов. Hcurv – *H. curviuscula* NRB217 (розовый), Hob – *H. obtusa* F1 (зелёный), Hun – *H. undulata* HU1 (голубой).

Для того, чтобы более точно охарактеризовать эти участки, мы решили исследовать все белки профагов в геномах голоспор. Мы нашли две области, содержащие фаговые гены, в геноме *H. elegans*, один – в *H. obtusa* и два – в *H. undulata*. Один из найденных в *H. undulata* локусов оказался интактным профагом, содержащим все необходимые для фаговой репликации гены и инсерционные последовательности (Рисунок 4.5), остальные фрагменты были значительно разрушены и не имели инсерционных последовательностей. Несмотря на то, что полных профагов не было найдено в геноме *H. curviuscula*, в скэффолде 565 мы нашли участок, кодирующий две эндонуклеазы фагового типа и две предположительные транспозазы, а другие скэффолды содержали гены капсида и портала, похожие на найденные в геномах других *Holospora*, и множество генов транспозаз.

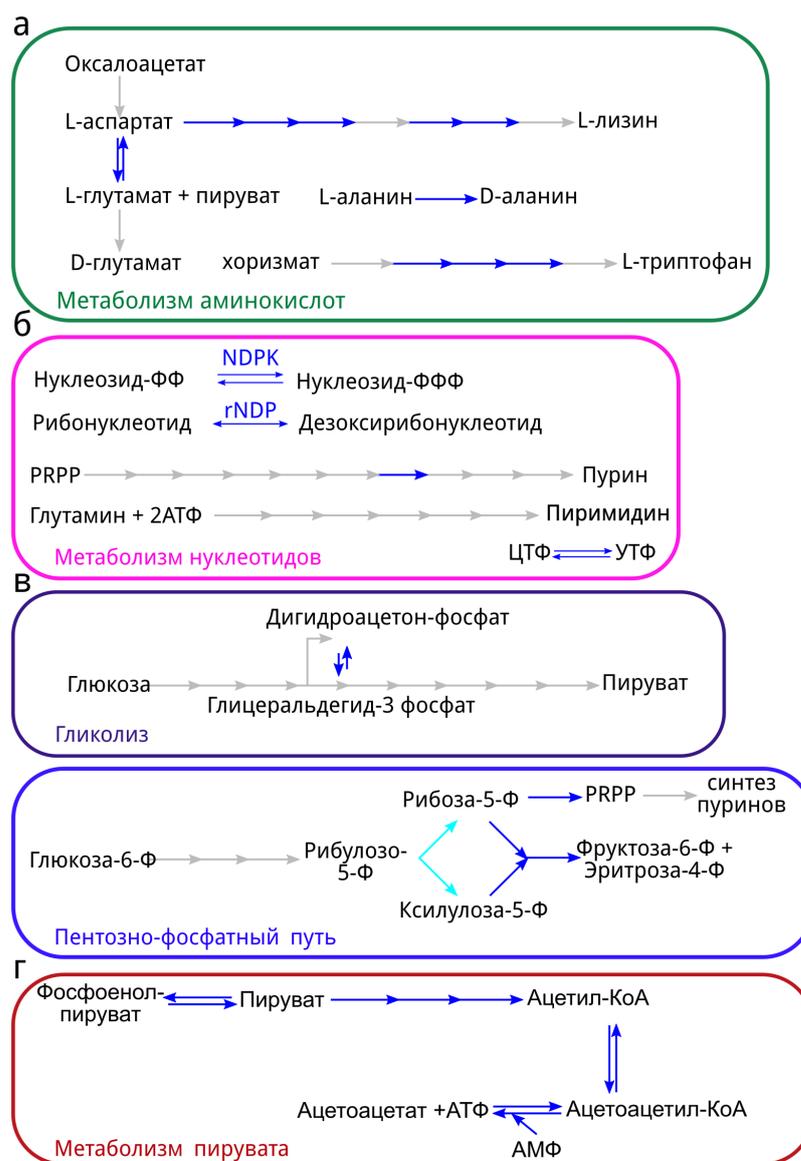


**Рисунок 4.5.** Предсказанные при помощи RAST участки генов *Holospora* spp., содержащие последовательности, похожие на профаги.

#### 4.2.4 *Holospora* spp. не способны синтезировать аминокислоты и некоторые другие важные соединения

Даже по меркам порядка *Rickettsiales*, для представителей которого характерны короткие геномы с малым количеством генов [204], *Holospora* имеет крайне редуцированный метаболизм. В частности, у бактерий из рода *Holospora* нет генов, кодирующих ферменты цикла Кребса, хотя даже у *Rickettsia* сохранились некоторые из этих ферментов, и эти бактерии используют их для превращения различных химических соединений [206].

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 4.6а). Это означает, что все аминокислоты должны транспортироваться извне, т. е. голоспора должна получать их от хозяина. Риккетсии также не способны синтезировать все аминокислоты [206], но эта ситуация не является типичной для представителей *Rickettsiales*, которые обычно способны синтезировать лизин и глутамин [203]. Голоспоры так же, как и риккетсии, не способны синтезировать хоризмат и, видимо, вынуждены импортировать его или его производные извне.



**Рисунок 4.6.** 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 – фосфорибозил пирозин; NDPK – нуклеозид-дифосфат киназа; rNDP – рибонуклеотид редуктаза.

Представители *Holospora*, как и риккетсии, не способны синтезировать пурины и пиримидины (Рисунок 4.6б), 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.

Путь биосинтеза убихинона у голоспор содержит только фрагмент, начинающийся

с диметилаллил дифосфатазы, похожая картина наблюдается у представителей рода *Rickettsia* [206]. При этом *Holospora* в отличие от *Rickettsia* spp., не способны превращать диметилаллил-пирофосфат (DMAPP) в изопентенул-дифосфат, и, видимо, DMAPP должен поступать от хозяина. Единственный метаболический путь, который сохранился полностью – это путь синтеза жирных кислот.

Мы изучили транспортные белки *Holospora* spp., чтобы понять, как соединения, которые голоспоры не способны синтезировать, поступают в клетку. В геноме *H. curviuscula* нам удалось найти 30 генов, кодирующих транспортные белки, среди них есть транспортёры олигопептидов и аминокислот, таких как цистеин/ацетил-серин, глицин, и, видимо, транспортёры для разветвлённых аминокислот. Помимо этого, мы нашли гены, связанные с транспортом магния, железа, рибозы, пуринов, сульфоацетата, L-галактоната или других сахаров и транспортёр путресцина. Идентифицированные транспортные системы явно не позволяют доставить в клетку все необходимые соединения, в частности, не совсем понятно, как в клетку бактерии попадает часть аминокислот, если только они не доставляются в составе олигопептидов.

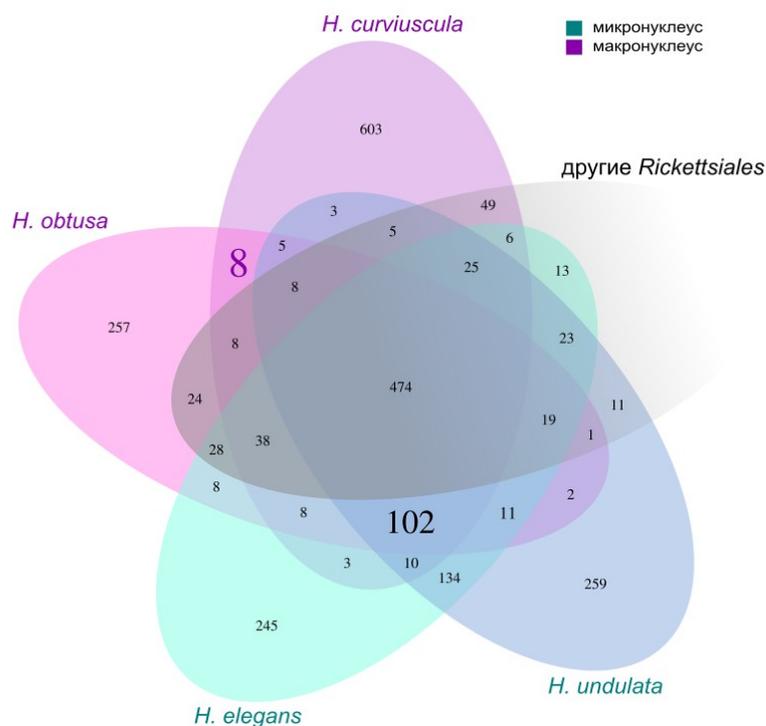
#### **4.2.5 *Holospora* spp. использует нуклеотиды в качестве основного источника энергии**

Известные геномы представителей рода *Holospora* не содержат генов, связанных с получением энергии. В частности, у голоспор отсутствуют ферменты гликолиза, за исключением фосфоглицератмутаза (Рисунок 4.6в), все ферменты цикла Кребса за исключением малатдегидрогеназы, и нет  $F_1F_0$ -АТФазы. Более того, представители рода *Holospora* не способны синтезировать кофермент А (КоА), но, видимо, как и *Rickettsia* [206], способны получать его в виде дефосфо-КоА, т. к. у них есть фермент CoaE (PF01121, Рисунок 4.2). От пентозо-фосфатного пути у голоспор сохранилась только ферменты неокислительного этапа (Рисунок 4.6в), 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 4.6г). 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. Никаких других очевидных источников энергии, кроме нуклеотидов, найдено не было.

#### 4.2.6 Секреторные системы и потенциальные инвазины

Несмотря на то, что большая часть представителей *Rickettsiales* являются паразитами и их геномы значительно редуцированы, они сохраняют многие секреторные системы, такие как пути Tat и Sec, секрецию типа IV, а также белок TolC [304]. Более того, считается, что белок VirB – часть секреторной системы IV – является важным для инвазии хозяина [305,306]. Система секреции у *Holospora* значительно отличается от типичной для *Rickettsiales*, описанной выше, и значительно редуцирована, у них сохранился Sec-путь секреции, дополнительные системы, помогающие протаскивать белки через внешнюю мембрану (LolA, LolD; вероятно, также сохранился LolE) и транспортировать их из клетки (VamA, VamB, VamD и шаперон DegP), а также TolC-подобные белки. Однако нам не удалось найти белки Tat-системы (TatA, TatB и TatC). Так как Tat-белки иногда может быть трудно идентифицировать [304], мы попытались найти в геноме белки, содержащие сигнал транспортировки Tat-системой (twin-arginine signal), и не обнаружили таковых. Тем самым, Tat-система, видимо, действительно отсутствует, и бактерии из рода *Holospora* не способны экспортировать свёрнутые белки.

У риккетсий инвазия ассоциирована с белками RalF, RickA и Sca [304]. Гомологи этих белков отсутствуют у *Holospora*, откуда следует, что эти бактерии используют какие-то другие механизмы для проникновения в клетки хозяина. Кроме этого, у голоспор нет белков с анкириновыми доменами, которые, как предполагается, также необходимы риккетсиям для проникновения в клетки [304]. Несмотря на то, что белки Sec-системы есть у всех представителей *Holospora*, нам удалось найти только один ген в геноме *H. curviuscula* (HCUR\_00103), содержащий домен, характерный для белков-автотранспортёров (PF03797). Гомологи этого гена есть у других представителей *Holospora*, но они, по всей видимости, не содержат этот домен, что может означать, что предсказание HMMeg для *H. curviuscula* неверное. Однако все геномы *Holospora* содержат много (от трех до девяти копий на геном) гомологов гена *ompA*. Это интересно, потому что ранее было показано, что OmpA играет важную роль в патогенезе бактерий из порядка *Rickettsiales* [307].



**Рисунок 4.7.** Группы ортологических белков (ГОБ) *Holospora* spp. и *Rickettsiales* (Таблица 2). Числа в ячейках показывают или количество ГОБ, или, в случае уникальных генов, количество сингтонов. Увеличенным шрифтом отмечены группы, специфичные или для эндосимбионтов макронуклеуса (8) или всех *Holospora* (102).

#### 4.2.7 Гены, специфичные для *Holospora*

Для того, чтобы найти гены, необходимые для выживания в ядре, мы сравнили генный состав всех известных голоспор с генным составом других *Rickettsiales* (Рисунок 4.7). Мы обнаружили 102 ГОБ, которые встречаются только у голоспор, и не встречаются у каких-либо других *Rickettsiales* (далее такие ГОБ называются голоспороспецифичными или ГСГОБ), 97 ГСГОБ не содержат паралогов. Гены, попавшие в ГСГОБ, кодируют транспортные белки, транскрипционные регуляторы и белки с неизвестными функциями. Мы предположили, что белки, необходимые для выживания в ядре и адаптации к нему, должны либо находиться на поверхности бактериальных клеток, либо секретироваться, поэтому мы уделили особое внимание белкам с трансмембранными спиралями и белкам, содержащим сигнальные пептиды. В белках из 37 ГСГОБ удалось предсказать трансмембранные спирали; белки из 12 ГСГОБ содержали классические сигнальные пептиды; представители ещё 18 ГСГОБ, видимо, могут быть экспортированы при помощи неканонических транспортных путей, предсказанных SecretomeP и TatP (см.

Материалы и методы). Одна из таких ГСГОБ содержала периплазматический белок размером 89 кДа, для которого ранее было показано, что он необходим *H. obtusa* для проникновения в клетку хозяина [8]. Одна ГСГОБ содержала белки, похожие на OmpA. Белки, содержащиеся в этой ГСГОБ, присутствуют у всех *Holospira*, и все голоспоры за исключением *H. curviuscula* несут две копии этого гена в геноме.

Геномы *Holospira* кодируют много коротких белков (длиной менее 100 аа) с неизвестными функциями. Было показано, что некоторые короткие белки нужны для целого ряда важных процессов, таких как образование спор, регуляция транспорта и транскрипции, передача сигнала, а также могут обладать антимикробными или токсичными свойствами [308,309]. Из 102 ГСГОБ, семнадцать содержали короткие белки, у пяти из них предсказывается сигнальный пептид. Одна из ГСГОБ содержала белок размером 5.4 кДа, который, как считается, необходим для перехода из репродуктивной в инфекционную форму [9].

Для того, чтобы определить, какие именно белки определяют ядерную специфичность, мы искали такие ГСГОБ, которые встречаются только у двух макронуклеарных видов *Holospira*: *H. curviuscula* и *H. obtusa*, – мы нашли восемь таких ГСГОБ. Из них две группы кодируют альфа/бета гидролазы, одна – белок, похожий на аспаратат/глутамат аспартазу или малат изомеразу, одна – нуклеотид эпимеразу и одна – белок с неизвестной функцией. Три оставшиеся группы кодируют короткие белки, причём белки из одной такой ГСГОБ имеют потенциальный сигнальный пептид.

### 4.3 Обсуждение результатов

До настоящей работы была опубликована, насколько нам известно, единственная работа, непосредственно направленная на исследование геномов *Holospira* spp. [12]. Эта работа носила достаточно общий характер: были описаны только общие особенности метаболизма, такие как отсутствие многих путей. Было выдвинуто предположение, что сохранившийся пируват-дегидрогеназный комплекс, присутствующий у всех *Holospira*, видимо, является остатком от различных исчезнувших метаболических путей, в которых пируват является промежуточным метаболитом [12]. Также было выдвинуто предположение, что голоспоры, видимо, сильно зависят от метаболитов, поставляемых хозяином [12], но детали метаболизма и то, какие именно метаболиты должны экспортироваться, не изучалось. Добавление к выборке отсеквенированного в этой работе генома *H. curviuscula* позволяет провести детальный сравнительный анализ метаболизма

рода *Holospora*.

Несмотря на то, что представители рода *Holospora* имеют относительно большие геномы (Табл. 4.1) по сравнению с другими симбиотическими бактериями, такими как *Buchnera*, *Candidatus* *Baumannia* или *Candidatus* *Carsonella*, *Holospora* не способны синтезировать большую часть необходимых метаболитов. С другой стороны, все доступные геномы голоспор содержат много повторов, которые осложняют сборку геномов. Так как все геномные сборки голоспор состоят из большого количества контигов, можно было бы предположить, что часть ферментов просто отсутствуют в сборке. Однако, по всей видимости, это не так. Во-первых, наши оценки длины сборки с учётом повторов совпадают с экспериментальными оценками длины генома этих бактерий, тем самым, если в сборке и отсутствуют какие-то гены, то их явно немного. Во-вторых, наш анализ основан на сравнении независимыхборок разных видов и трудно предполагать, что у достаточно далёких видов в геномных сборках будут отсутствовать одни и те же гены.

Даже эндосимбионты с меньшими геномами способны сами синтезировать хотя бы какие-то аминокислоты или сохранили какие-то фрагменты центрального метаболизма; представители же рода *Holospora* вынуждены получать все аминокислоты из внешней среды. *Holospora* 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.

Несмотря на то, что мы не провели полной метаболической реконструкции, результаты нашего анализа свидетельствуют, что широкий спектр соединений, таких, например, как аминокислоты, DMAPP и производные хоризмата, должен поступать от хозяина. Очень маловероятно, что голоспоры аккумулируют все эти соединения вне ядра, потому что процесс инфекции достаточно короткий, а большую часть жизненного цикла голоспоры проводят именно в ядре. Это значит, что все эти соединения голоспора должна получать, находясь в ядре хозяина. Аминокислоты, вероятно, могут быть получены за счёт разрушения ядерных белков хозяина, и, действительно, было показано, что заражение голоспорами повышает протеолитическую активность в макронуклеусе [310]. Ситуация с другими соединениями более интересная, но все необходимые молекулы достаточно мелкие и могут пассивно проходить через ядерную пору [311]. Более того, было показано, что представители *Holospora* изменяют экспрессию генов хозяина и повышают уровень синтеза РНК [198,310]. Эти изменения могут помогать бактериям получать необходимые соединения. Для того, чтобы понять, как все эти соединения могут попадать в

бактериальную клетку, мы изучили спектр транспортных белков в геномах этих бактерий. Мы нашли часть необходимых транспортёров, но далеко не все, и этот вопрос требует дальнейшего изучения.

Мы предполагаем, что *Holospora* не просто зависит от хозяина в получении энергии, но использует в качестве источника энергии нуклеотиды или рибонуклеотиды. Действительно, эти бактерии способны превращать нуклеотиды в рибонуклеотиды, и, предположительно, имеют транспортёры рибозы или каких-то её производных. Это позволяет предположить, что они способны поглощать и рибонуклеотиды, и нуклеотиды в зависимости от ситуации, и уже в клетке переводить их в необходимые соединения. Мы предполагаем, что рибонуклеотиды являются более вероятным источником, потому что их концентрация в ядре может быть на один-два порядка выше, чем концентрация дезоксирибонуклеотидов [312].

Взаимодействие *Holospora* с хозяином было изучено достаточно детально. Заражение голоспорами может либо увеличивать, либо понижать жизнеспособность инфузории хозяина в зависимости от условий [197–199,313]. Добавление к анализу *Holospora curviuscula* позволило нам изучить не только особенности голоспор в одном определённом хозяине, но и найти белки, которые могут быть важны для заражения макронуклеуса. Несмотря на то, что механизмы инфекции не до конца понятны, поддержание инфекции обеспечивается голоспор-специфичными белками, вероятно, короткими секретлируемыми белками. Обнаруженный ранее у *H. obtusa* белок 89 кДа [8] консервативен у всех представителей рода и, по-видимому, важен для инфекции. Дополнительно мы предполагаем, что белки, похожие на OmpA, также могут играть важную роль в процессе заражения.

В целом, анализ геномов *Holospora* показал, что их метаболические способности крайне скудны, особенно учитывая размер генома этих бактерий, и позволил выявить ряд генов-кандидатов, необходимых для поддержания такого необычного жизненного цикла.

## **Глава 5. Эволюция рибосом у симбиотических бактерий**

Рибосома – важная клеточная органелла, осуществляющая синтез белков. Структура и состав рибосом консервативны у всех живых организмов. Однако, известно, что геномы некоторых бактерий содержат неполный набор генов рибосомных белков. Наименее полные рибосомы были обнаружены у эндосимбионтов насекомых с крайне редуцированными геномами [17]. В этой главе рассматривается устройство рибосом в 214 бактериях с геномами размером менее 1 Мб. Для всех этих бактерий характерен симбиотический образ жизни.

На основании анализа этих геномов нам удалось показать, что состав рибосомы достаточно стабилен, но некоторые рибосомные белки действительно могут отсутствовать у ряда видов. При этом состав большой субъединицы более изменчив, чем состав малой. В настоящей работе нам удалось определить набор рибосомных белков, которые часто отсутствуют (далее «часто пропадающих р-белков»), и описать паттерны их исчезновения. Более того, показано, что часто пропадающие рибосомные белки имеют меньше контактов с другими компонентами рибосомы и расположены в основном на её поверхности. Дополнительно мы показали, что у бактерий часто происходит укорочение рРНК, но в основном за счёт удаления свободных петель. Только в одном случае (при потере белка L24) была показана одновременная потеря участка, с которым связывается белок. Наконец, потеря последовательности Шайна – Дальгарно не скоррелирована с размером генома или с потерей каких-то определённых белков.

### **5.1 Материалы и методы**

#### **5.1.1 Выборка бактериальных геномов**

Список всех полных бактериальных геномов с геномами короче, чем 1 Мбр был составлен на основании записей в базе данных IMG/M [314]. Геномы для всех видов из списка (всего 214 геномов, Табл. 5.1) были скачаны с NCBI FTP; файлы с белковыми последовательностями были скачаны в феврале 2017 года (<ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/>), а файлы с последовательностями рРНК в сентябре 2017 года (<ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/>).

Геномы некоторых штаммов для отобранных видов немного превосходили 1 Мб, но все равно были оставлены в выборке.

Название бактерии	Идентификатор в GenBank
<i>Blochmannia</i> endosymbiont of <i>Camponotus (Colobopsis) obliquus</i> str. 757	CP010049.1
<i>Blochmannia</i> endosymbiont of <i>Polyrhachis (Hedomyrma) turneri</i> str. 675	CP010048.1
<i>Borrelia anserina</i> str. Es	CP013704.1
<i>Borrelia hermsii</i> str. CC1	CP011060.1
<i>Borrelia hermsii</i> str. HS1 isolate Browne Mountain	CP014349.1
<i>Borrelia miyamotoi</i> str. CT14D4	CP010308.1
<i>Borrelia miyamotoi</i> str. LB-2001	CP006647.2
<i>Borrelia parkeri</i> str. HR1	CP007022.1
<i>Borrelia turicatae</i> str. 91E135	CP000049.1
<i>Borrelia afzelii</i> str. HLJ01	CP003882.1
<i>Borrelia afzelii</i> str. K78	CP009058.1
<i>Borrelia afzelii</i> str. PKo	CP002933.1
<i>Borrelia afzelii</i> str. Tom3107	CP009212.1
<i>Borrelia burgdorferi</i> str. B31	AE000783.1
<i>Borrelia burgdorferi</i> str. CA382	CP005925.1
<i>Borrelia burgdorferi</i> str. JD1	CP002312.1
<i>Borrelia burgdorferi</i> str. N40	CP002228.1
<i>Borrelia burgdorferi</i> str. ZS7	CP001205.1
<i>Borrelia garinii</i> str. BgVir	CP003151.1
<i>Borrelia garinii</i> str. CIP 103362 isolate 20047	CP018744.1
<i>Borrelia garinii</i> str. NMJW1	CP003866.1
<i>Borrelia garinii</i> str. SZ	CP007564.1
<i>Borrelia valaisiana</i> str. Tom4006	CP009117.1
<i>Buchnera aphidicola</i> str. 5A ( <i>Acyrtosiphon pisum</i> )	CP001161.1
<i>Buchnera aphidicola</i> str. Ak ( <i>Acyrtosiphon kondoi</i> )	CP002645.1
<i>Buchnera aphidicola</i> str. APS ( <i>Acyrtosiphon pisum</i> )	BA000003.2

Название бактерии	Идентификатор в GenBank
<i>Buchnera aphidicola</i> str. BAg ( <i>Aphis glycines</i> )	CP009253.1
<i>Buchnera aphidicola</i> str. BCc	CP000263.1
<i>Buchnera aphidicola</i> str. BCiconfinis ( <i>Cinara confinis</i> )	LT667503.1
<i>Buchnera aphidicola</i> str. BCifornacula ( <i>Cinara fornacula</i> )	LT667500.1
<i>Buchnera aphidicola</i> str. BCipseudotaxifoliae ( <i>Cinara pseudotaxifoliae</i> )	LT635893.1
<i>Buchnera aphidicola</i> str. Bp ( <i>Baizongia pistaciae</i> )	AE016826.1
<i>Buchnera aphidicola</i> str. BTs ( <i>Thelaxes suberi</i> )	LN890285.1
<i>Buchnera aphidicola</i> str. F009 ( <i>Myzus persicae</i> )	CP002703.1
<i>Buchnera aphidicola</i> str. G002 ( <i>Myzus persicae</i> )	CP002701.1
<i>Buchnera aphidicola</i> str. JF98 ( <i>Acyrtosiphon pisum</i> )	CP002303.1
<i>Buchnera aphidicola</i> str. JF99 ( <i>Acyrtosiphon pisum</i> )	CP002302.1
<i>Buchnera aphidicola</i> str. LL01 ( <i>Acyrtosiphon pisum</i> )	CP002300.1
<i>Buchnera aphidicola</i> str. SC ( <i>Schlechtendalia chinensis</i> )	CP011299.1
<i>Buchnera aphidicola</i> str. Sg ( <i>Schizaphis graminum</i> )	AE013218.1
<i>Buchnera aphidicola</i> str. TLW03 ( <i>Acyrtosiphon pisum</i> )	CP002301.1
<i>Buchnera aphidicola</i> str. Tuc7 ( <i>Acyrtosiphon pisum</i> )	CP001158.1
<i>Buchnera aphidicola</i> str. Ua ( <i>Uroleucon ambrosiae</i> )	CP002648.1
<i>Buchnera aphidicola</i> str. USDA ( <i>Myzus persicae</i> )	CP002697.1
<i>Buchnera aphidicola</i> str. W106 ( <i>Myzus persicae</i> )	CP002699.1
<i>Candidatus</i> Baumannia cicadellincola str. B-GSS	CP011787.1
<i>Candidatus</i> Baumannia cicadellincola str. BGSS	CP008985.1
<i>Candidatus</i> Blochmannia chromaiodes str. 640	CP003903.1
<i>Candidatus</i> Blochmannia floridanus	BX248583.1
<i>Candidatus</i> Blochmannia pennsylvanicus str. BPEN	CP000016.1
<i>Candidatus</i> Blochmannia vafer str. BVAF	CP002189.2
<i>Candidatus</i> Carsonella ruddii str. CE isolate Thao2000	CP003541.1
<i>Candidatus</i> Carsonella ruddii str. CS isolate Thao2000	CP003542.1
<i>Candidatus</i> Carsonella ruddii str. DC	CP003467.1
<i>Candidatus</i> Carsonella ruddii str. HC isolate Thao2000	CP003543.1

Название бактерии	Идентификатор в GenBank
<i>Candidatus Carsonella ruddii</i> str. HT isolate Thao2000	CP003544.1
<i>Candidatus Carsonella ruddii</i> str. PC isolate NHV	CP003545.1
<i>Candidatus Carsonella ruddii</i> str. PV	AP009180.1
<i>Candidatus Doolittlea endobia</i> isolate DEMHIR	LN999833.1
<i>Candidatus Evansia muelleri</i>	LM655252.1
<i>Candidatus Gullanella endobia</i> isolate GEFVIR	LN999832.1
<i>Candidatus Hepatoplasma crinochetorum</i> str. Av	CP006932.1
<i>Candidatus Hoaglandella endobia</i> isolate HETPER	LN999835.1
<i>Candidatus Hodgkinia cicadicola</i> str. Dsem	CP001226.1
<i>Candidatus Kinetoplastibacterium crithidii</i> (ex <i>Angomonas deanei</i> ATCC 30255)	CP003978.1
<i>Candidatus Kinetoplastibacterium crithidii</i> str. TCC036E	CP003804.1
<i>Candidatus Kinetoplastibacterium desouzaii</i> str. TCC079E	CP003803.1
<i>Candidatus Kinetoplastibacterium galatii</i> str. TCC219	CP003806.1
<i>Candidatus Kinetoplastibacterium oncopeltii</i> str. TCC290E	CP003805.1
<i>Candidatus Mikella endobia</i> isolate MEPMAR	LN999831.1
<i>Candidatus Moranella endobia</i> str. PCIT	CP002243.1
<i>Candidatus Moranella endobia</i> str. PCVAL	CP003881.1
<i>Candidatus Mycoplasma girerdii</i> str. VCU_M1	CP007711.1
<i>Candidatus Mycoplasma haemolamae</i> str. Purdue	CP003731.1
<i>Candidatus Nasuia deltocephalinicola</i> str. NAS-ALF	CP006059.1
<i>Candidatus Phytoplasma australiense</i>	AM422018.1
Strawberry lethal yellows phytoplasma (CPA) str. NZSb11	CP002548.1
<i>Candidatus Phytoplasma mali</i> str. AT	CU469464.1
<i>Candidatus Portiera aleyrodidarum</i> str. BT-B	CP003708.1
<i>Candidatus Portiera aleyrodidarum</i> str. BT-Q	CP007563.1
<i>Candidatus Portiera aleyrodidarum</i> str. BT-QVLC	CP003835.1
<i>Candidatus Portiera aleyrodidarum</i> str. TV	CP004358.1
<i>Candidatus Proffrella armatura</i> str. DC	CP003468.1

Название бактерии	Идентификатор в GenBank
<i>Candidatus Proffella armatura</i> str. YCPA	CP012591.1
<i>Candidatus Riesia pediculicola</i> str. USDA	CP001085.1
<i>Candidatus Saccharibacteria oral taxon</i> str. TM7x	CP007496.1
<i>Candidatus Sulcia muelleri</i> str. Sulcia-ALF	CP006060.1
<i>Candidatus Sulcia muelleri</i> str. BGSS	CP008986.1
<i>Candidatus Sulcia muelleri</i> str. CARI	CP002163.1
<i>Candidatus Sulcia muelleri</i> str. DMIN	CP001981.1
<i>Candidatus Sulcia muelleri</i> str. GWSS	CP000770.2
<i>Candidatus Sulcia muelleri</i> str. ML	CP010105.1
<i>Candidatus Sulcia muelleri</i> str. NC	CP016223.1
<i>Candidatus Sulcia muelleri</i> str. PSPU	AP013293.1
<i>Candidatus Sulcia muelleri</i> str. PUNC	CP013212.1
<i>Candidatus Sulcia muelleri</i> str. SMDSEM	CP001605.1
<i>Candidatus Sulcia muelleri</i> str. TETUND	CP007234.1
<i>Candidatus Tremblaya phenacola</i> str. PAVE	CP003982.1
<i>Candidatus Tremblaya princeps</i> str. PCIT	CP002244.1
<i>Candidatus Tremblaya princeps</i> str. PCVAL	CP002918.1
<i>Candidatus Uzinura diaspidicola</i> str. ASNER	CP003263.1
<i>Candidatus Walczuchella monophlebidarum</i> isolate FNIIJ	CP006873.1
<i>Candidatus Zinderia insecticola</i> str. CARI	CP002161.1
<i>Candidatus Kinetoplastibacterium blastocrithidii</i> str. TCC012E	CP003807.1
<i>Candidatus Kinetoplastibacterium blastocrithidii</i> (ex <i>Strigomonas culicis</i> )	CP003733.1
<i>Mesoplasma florum</i> str. L1	AE017263.1
<i>Mesoplasma florum</i> str. W37	CP006778.1
<i>Mycoplasma agalactiae</i> str. 5632	FP671138.1
<i>Mycoplasma agalactiae</i> str. PG2	CU179680.1
<i>Mycoplasma arginini</i> str. HAZ145_1	AP014657.1
<i>Mycoplasma arthritidis</i> str. 158L3-1	CP001047.1
<i>Mycoplasma bovis</i> str. CQ-W70	CP005933.1

Название бактерии	Идентификатор в GenBank
<i>Mycoplasma bovis</i> str. HB0801	CP002058.1
<i>Mycoplasma bovis</i> str. Hubei-1	CP002513.1
<i>Mycoplasma bovis</i> str. NM 2012	CP011348.1
<i>Mycoplasma bovis</i> str. PG45	CP002188.1
<i>Mycoplasma bovoculi</i> str. M165/69	CP007154.1
<i>Mycoplasma californicum</i> str. HAZ160_1	AP013353.1
<i>Mycoplasma californicum</i> str. ST-6	CP007521.1
<i>Mycoplasma canadense</i> str. HAZ360_1	AP014631.1
<i>Mycoplasma canis</i> str. LV	CP011368.1
<i>Mycoplasma canis</i> str. PG 14	CP014281.1
<i>Mycoplasma capricolum</i> subsp. capripneumoniae str. 87001	CP006959.1
<i>Mycoplasma capricolum</i> subsp. capripneumoniae str. 9231-Abomsa	LM995445.1
<i>Mycoplasma capricolum</i> subsp. capricolum str. ATCC 27343	CP000123.1
<i>Mycoplasma capricolum</i> subsp. capripneumoniae str. ILRI181	LN515399.1
<i>Mycoplasma capricolum</i> subsp. capripneumoniae str. M1601	CP017125.1
<i>Mycoplasma capricolum</i> subsp. capripneumoniae str. type strain: F38 = NCTC 10192	LN515398.1
<i>Mycoplasma crocodyli</i> str. MP145	CP001991.1
<i>Mycoplasma cynos</i> str. C142	HF559394.1
<i>Mycoplasma fermentans</i> str. JER	CP001995.1
<i>Mycoplasma fermentans</i> str. M64	CP002458.1
<i>Mycoplasma flocculare</i> ATCC 27399 str. Ms42	CP007585.1
<i>Mycoplasma gallisepticum</i> str. CA06_2006.052-5-2P	CP003512.1
<i>Mycoplasma gallisepticum</i> str. F	CP001873.1
<i>Mycoplasma gallisepticum</i> str. NC06_2006.080-5-2P	CP003511.1
<i>Mycoplasma gallisepticum</i> str. NC08_2008.031-4-3P	CP003513.1
<i>Mycoplasma gallisepticum</i> str. NC95_13295-2-2P	CP003507.1
<i>Mycoplasma gallisepticum</i> str. NC96_1596-4-2P	CP003508.1
<i>Mycoplasma gallisepticum</i> str. NY01_2001.047-5-1P	CP003509.1

Название бактерии	Идентификатор в GenBank
<i>Mycoplasma gallisepticum</i> str. R(high)	CP001872.1
<i>Mycoplasma gallisepticum</i> str. R(low)	AE015450.2
<i>Mycoplasma gallisepticum</i> str. S6	CP006916.3
<i>Mycoplasma gallisepticum</i> str. VA94_7994-1-7P	CP003506.1
<i>Mycoplasma gallisepticum</i> str. WI01_2001.043-13-2P	CP003510.1
<i>Mycoplasma genitalium</i> str. G-37	L43967.2
<i>Mycoplasma genitalium</i> str. M2288	CP003773.1
<i>Mycoplasma genitalium</i> str. M2321	CP003770.1
<i>Mycoplasma genitalium</i> str. M6282	CP003771.1
<i>Mycoplasma genitalium</i> str. M6320	CP003772.1
<i>Mycoplasma haemocanis</i> str. Illinois	CP003199.1
<i>Mycoplasma hominis</i> str. AF1	CP009677.1
<i>Mycoplasma hominis</i> str. ATCC 23114	FP236530.1
<i>Mycoplasma hominis</i> ATCC 27545 str. LBD-4	CP009652.1
<i>Mycoplasma hominis</i> str. Sprott	CP011538.1
<i>Mycoplasma hyopneumoniae</i> str. 168-L	CP003131.1
<i>Mycoplasma hyopneumoniae</i> str. 168	CP002274.1
<i>Mycoplasma hyopneumoniae</i> str. 232	AE017332.1
<i>Mycoplasma hyopneumoniae</i> str. 7422	CP003802.1
<i>Mycoplasma hyopneumoniae</i> str. 7448	AE017244.1
<i>Mycoplasma hyopneumoniae</i> str. J	AE017243.1
<i>Mycoplasma hyorhinae</i> str. DBS 1050	CP006849.1
<i>Mycoplasma hyorhinae</i> str. GDL-1	CP003231.1
<i>Mycoplasma hyorhinae</i> str. HUB-1	CP002170.1
<i>Mycoplasma hyorhinae</i> str. MCLD	CP002669.1
<i>Mycoplasma hyorhinae</i> str. MDBK-IPV	CP016817.1
<i>Mycoplasma hyorhinae</i> str. SK76	CP003914.1
<i>Mycoplasma leachii</i> str. PG50	CP002108.1
<i>Mycoplasma mobile</i> str. 163K	AE017308.1

Название бактерии	Идентификатор в GenBank
<i>Mycoplasma ovis</i> str. Michigan	CP006935.1
<i>Mycoplasma parvum</i> str. Indiana	CP006771.1
<i>Mycoplasma pneumoniae</i> str. 19294	CP010539.1
<i>Mycoplasma pneumoniae</i> str. 309	AP012303.1
<i>Mycoplasma pneumoniae</i> str. 39443	CP010540.1
<i>Mycoplasma pneumoniae</i> str. 51494	CP010541.1
<i>Mycoplasma pneumoniae</i> str. 54089	CP010542.1
<i>Mycoplasma pneumoniae</i> str. 54524	CP010543.1
<i>Mycoplasma pneumoniae</i> str. 85084	CP010544.1
<i>Mycoplasma pneumoniae</i> str. 85138	CP010545.1
<i>Mycoplasma pneumoniae</i> str. C267	CP014267.1
<i>Mycoplasma pneumoniae</i> str. FH 2009	CP017327.1
<i>Mycoplasma pneumoniae</i> str. FH	CP002077.1
<i>Mycoplasma pneumoniae</i> str. M1139	CP010547.1
<i>Mycoplasma pneumoniae</i> str. M129-B7	CP003913.2
<i>Mycoplasma pneumoniae</i> str. M129; ATCC 29342	U00089.2
<i>Mycoplasma pneumoniae</i> str. M2192	CP010548.1
<i>Mycoplasma pneumoniae</i> str. M2592	CP010549.1
<i>Mycoplasma pneumoniae</i> str. M29	CP008895.1
<i>Mycoplasma pneumoniae</i> str. MAC (or Mac)	CP010550.1
<i>Mycoplasma pneumoniae</i> str. PI 1428	CP010538.1
<i>Mycoplasma pneumoniae</i> str. PO1	CP010551.1
<i>Mycoplasma pneumoniae</i> str. S355	CP013829.1
<i>Mycoplasma pulmonis</i> str. UAB CTIP	AL445566.1
<i>Mycoplasma putrefaciens</i> str. KS1	CP003021.1
<i>Mycoplasma putrefaciens</i> str. Mput9231	CP004357.1
<i>Mycoplasma suis</i> str. Illinois	CP002525.1
<i>Mycoplasma suis</i> str. KI3806	FQ790233.1
<i>Mycoplasma synoviae</i> str. 53	AE017245.1

Название бактерии	Идентификатор в GenBank
<i>Mycoplasma synoviae</i> ATCC 25204 str. WVU1853T	CP011096.1
<i>Mycoplasma wenyonii</i> str. Massachusetts	CP003703.1
<i>Mycoplasma yeatsii</i> str. GM274B	CP007520.1
<i>Neorickettsia helminthoeca</i> str. Oregon	CP007481.1
<i>Neorickettsia risticii</i> str. Illinois	CP001431.1
<i>Neorickettsia sennetsu</i> str. Miyayama	CP000237.1
<i>Spiroplasma diminutum</i> str. CUAS-1	CP005076.1
<i>Tropheryma whipplei</i> str. TW08/27	BX072543.1
<i>Tropheryma whipplei</i> str. Twist	AE014184.1
<i>Ureaplasma parvum</i> serovar 3 str. ATCC 27815	CP000942.1
<i>Ureaplasma parvum</i> serovar 3 str. ATCC 700970	AF222894.1
<i>Ureaplasma parvum</i> serovar 3 str. SV3F4	AP014584.1
<i>Ureaplasma urealyticum</i> serovar 10 str. ATCC 33699	CP001184.1
<i>Wigglesworthia glossinidia</i> endosymbiont of <i>Glossina brevipalpis</i>	BA000021.3
<i>Wolbachia</i> endosymbiont of <i>Onchocerca ochengi</i> str. wOo	HE660029.1
<i>Wolbachia</i> endosymbiont of <i>Onchocerca volvulus</i> str. Cameroon	HG810405.1

**Таблица 5.1.** Список использованных штаммов бактерий.

### 5.1.2 Аннотация рибосомных белков

Во всех геномах набор рибосомных белков и триггер-фактор (TF) были переаннотированы (Табл. 5.2) при помощи позиционно-весовых матриц для белковых доменов из базы данных Pfam-A [252]. Матрицы для выбранных доменов использовались для поиска генов в геномах при помощи пакета программ HMMer [288]. Мы считали, что ген присутствует в геноме, если домен был найден в геноме с E-value < 0,001; только лучшая находка оставлялась для дальнейшего анализа. Если параметр HMMer domain bias был того же порядка, что уровень сходства, то проводилась дополнительная фильтрация.

Для каждого домена было построено множественное выравнивание белков, в аннотации которых мы были уверены. Белковые выравнивания были построены при

помощи веб-сервера GUIDANCE2 [250]; в качестве алгоритма выравнивания использовался MAFFT. Белок с высоким значением HMMer domain bias добавлялся к выравниванию, и если его добавление несильно снижало качество выравнивания (значение уверенности в качестве построенного GUIDANCE выравнивания не падало ниже 0.8), то мы считали, что это спорная последовательность действительно содержит данный домен.

Мы считали, что белки, состоящие из двух доменов (L2, L5, L7/L12, L9, L11, S4, S5, TF), присутствуют только в том случае, если оба домена были найдены в одной белковой последовательности.

<b>Р-белок</b>	<b>Идентификатор в базе данных Pfam</b>	<b>Р-белок</b>	<b>Идентификатор в базе данных Pfam</b>
L1	PF00687	L33	PF00471
L2	PF00181, PF03947	L34	PF00468
L3	PF00297	L35	PF01632
L4	PF00573	L36	PF00444
L5	PF00281, PF00673	S1	PF00575
L6	PF00347	S2	PF00318
L7/L12	PF00542, PF16320	S3	PF00189
L9	PF01281, PF03948	S4	PF01479, PF00163
L10	PF00466	S5	PF00333, PF03719
L11	PF00298, PF03946	S6	PF01250
L13	PF00572	S7	PF00177
L14	PF00238	S8	PF00410
L15	PF00828	S9	PF00380
L16	PF00252	S10	PF00338
L17	PF01196	S11	PF00411
L18	PF00861	S12	PF00164
L19	PF01245	S13	PF00416
L20	PF00453	S14	PF00253
L21	PF00829	S15	PF00312
L22	PF00237	S16	PF00886
L23	PF00276	S17	PF00366
L24	PF17136	S18	PF01084
L25	PF01386	S19	PF00203
L27	PF01016	S20	PF01649
L28	PF00830	S21	PF01165
L29	PF00831	S22	PF08136
L30	PF00327	TФ	PF05697, PF05698
L31	PF01197	Thx	PF17070
L32	PF01783		

**Таблица 5.2.** Список рибосомных белков и соответствующих им PFAM доменов.

### 5.1.3 Определение событий потери рибосомных белков

Филогенетическое древо бактерий было построено на основании множественного выравнивания универсальных рибосомных белков (см. раздел 5.1.4). Перед построением древа из выравнивания были удалены все колонки с пробелами. Древо было построено методом наибольшего правдоподобия при помощи программы PhyML v. 3.1 [251] (параметры по умолчанию). Полученное древо использовалось для поиска филогенетических паттернов исчезновения рибосомных белков, т. е. для подсчета количества независимых потерь каждого из белков.

### 5.1.4 Измерение скорости эволюции

Мы определяли скорость эволюции для данного белка как среднюю длину веток на филогенетическом древе данного белка. Для этого для каждого рибосомного белка было построено древо методом максимального правдоподобия при помощи PhyML с параметрами по умолчанию и 100 бутстрэп испытаниями. Так оценённая скорость эволюции для белка сравнивалась с средней скоростью для набора универсальных белков (L5, L14, L15, L20, S5, S7, S8, S9, S11, S13 и S17). Для каждого изучаемого белка средняя скорость эволюции универсальных белков считалась отдельно по поддереву с тем же набором листьев, что и на древе изучаемого белка. Для того, чтобы результаты можно было сравнивать для белков с разными филогенетическими профилями, мы считали относительную скорость эволюции, т.е. скорость эволюции данного белка, разделённая на среднюю скорость эволюции универсальных рибосомных белков.

### 5.1.5 Подсчет количества контактов

Количество контактов оценивалось по попарным расстояниям между атомами в трёхмерных структурах рибосомы из базы PDB (PDB ID: 5H5U и 2D3O). Мы считали, что атомы контактируют, если расстояние между ними было меньше 5Å. Контакты белка с собой не учитывались. Если несколько атомов белка контактировали с одним атомом в рибосоме, учитывался только один такой контакт. Расстояние между атомами считалось

при помощи скрипта `pairwise_dist.py` в программе PyMOL. Для каждого белка считалось суммарное количество контактов.

### 5.1.6 Анализ рРНК

Выравнивания 16S и 23S рРНК были построены при помощи сервиса SINA Alignment Service. Из полученного выравнивания были удалены все общие пробелы.

Мы считали, что у данной бактерии есть последовательно анти-Шайна – Дальгарно, если на 3'-конце 16S рРНК имелась консенсусная последовательность CCUCCU.

### 5.1.7 Статистический анализ и визуализация данных

Весь статистический анализ проводился при помощи средств языка программирования R. Для построения логистической регрессии использовалась функция `glm (family=binomial())`. Зависимость числа рибосомных белков от длины генома (Рисунок 5.3) была построена в R при помощи пакета `ggplot2`. Сглаживание кривой производилось по методу “loess”. Для визуализации филогенетических деревьев использовался сервер iTOL [233].

## 5.2 Результаты

### 5.2.1 У бактерий с короткими геномами отсутствуют некоторые рибосомные белки

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* (Табл. 5.1). Аннотация была произведена при помощи 65 Pfam доменов: 63 домена для канонического набора рибосомных белков и двух доменов для триггер-фактора (TF), ассоциированного с рибосомой шаперона (Табл. 5.2).

Два белка оказались отсутствующими почти во всех штаммах: S22 (отсутствует во всех штаммах) и белок малой субъединицы Thx (есть только у *Candidatus Walczuchella*

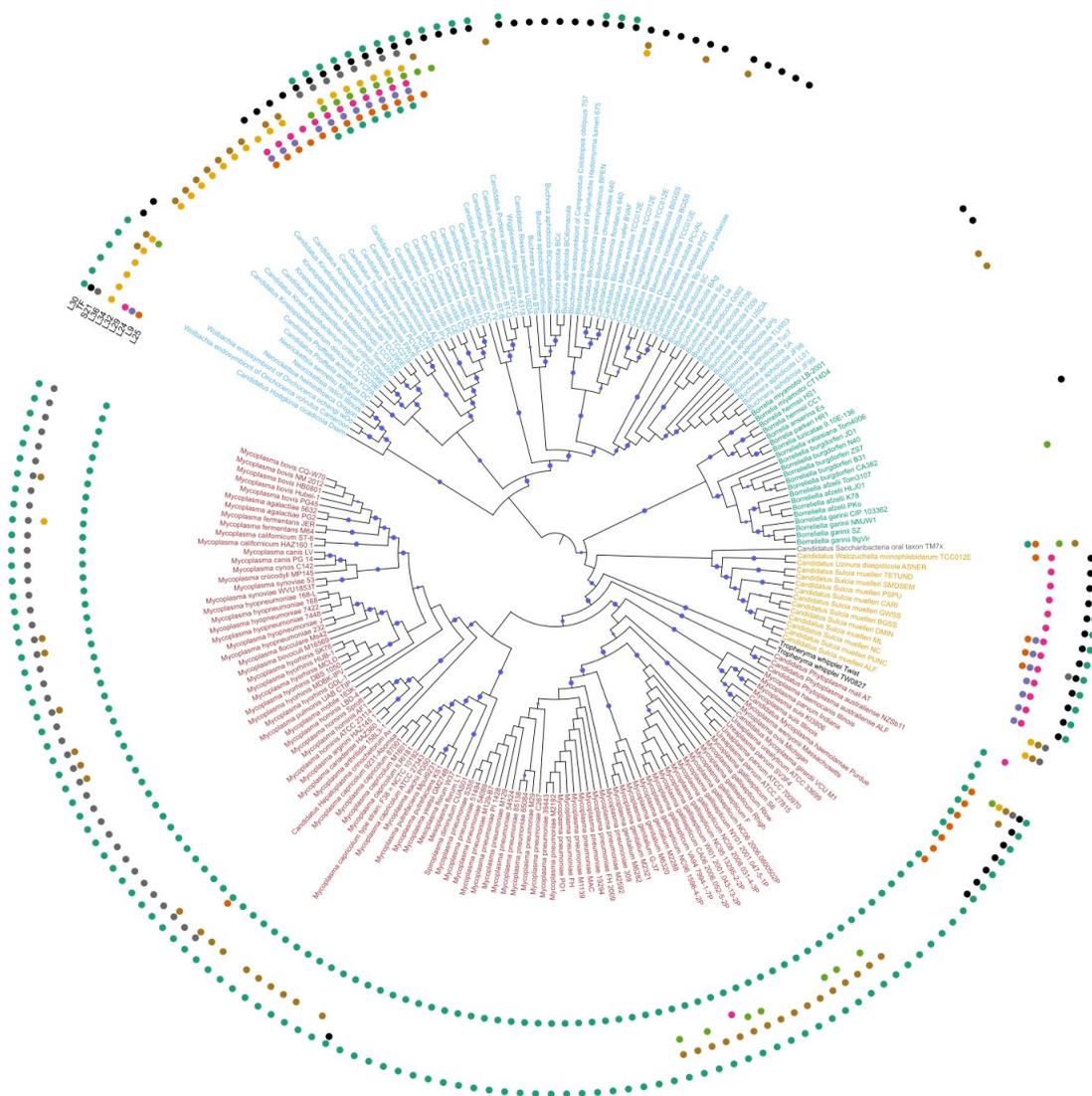
monophlebidarum).



**Рисунок 5.1. 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.

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. Рибосомные белки малой субъединицы пропадают значительно реже, чем белки большой (Рисунок 5.1). Единственный часто пропадающий белок – это S21, он отсутствует в геномах 54 штаммов из семи родов, тогда как самый часто пропадающий белок большой субъединицы, L30, отсутствует в геномах 138 штаммов из 18 родов. Все часто пропадающие рибосомные белки были независимо потеряны несколько раз в разных группах бактерий (Рисунок 5.2). Потери проходили в разных кладах дерева, например, белок L9 был потерян у *Proteobacteria*, *Bacteroidetes* и в ходе двух независимых событий в

*Tenericutes*, а белок L24 был потерян разными *Proteobacteria* в ходе двух независимых событий и исчез у части изучаемых *Bacteroidetes*.

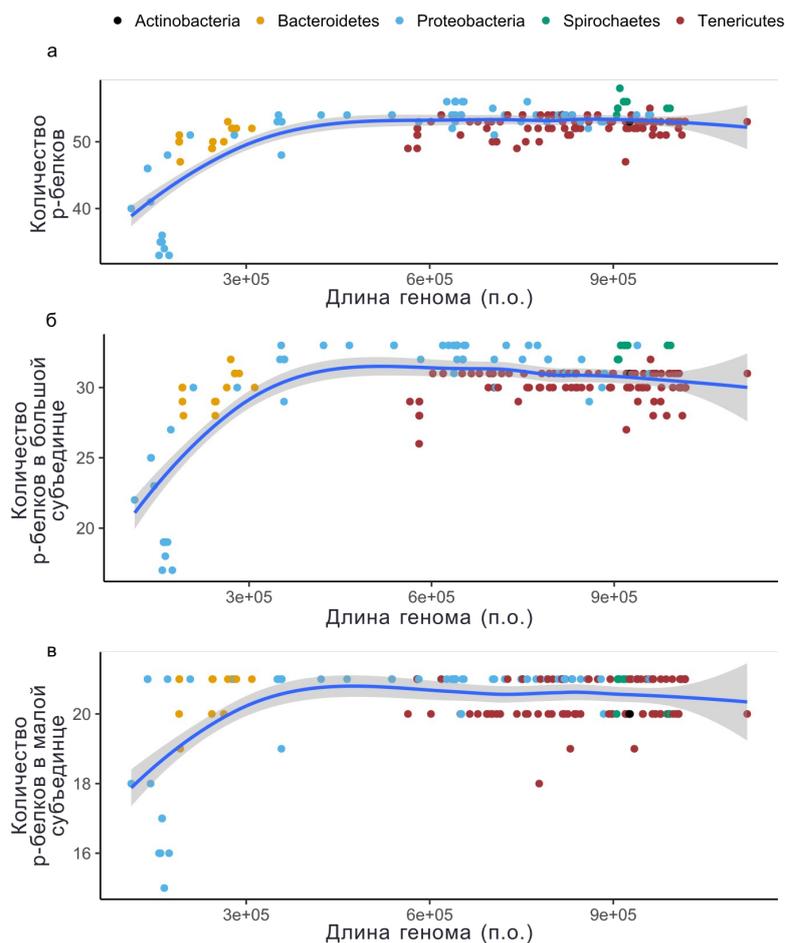


**Рисунок 5.2.** Филогенетическое древо, построенное методом наибольшего правдоподобия на основании выравнивания универсальных рибосомных белков. Древо было построено при помощи PhyML с 100 бутстрэп повторностями. На представленном рисунке длины ветвей не учитываются. Значения бутстрэпа от 0.9 до 1 показаны на рисунке синими кружками. Отсутствие часто пропадающих р-белков (L9, L21, L24, L25, L29, L32, L34, L36, S21, TF) в конкретных штаммах отмечено кружками напротив названия этого штамма. Каждый белок отмечается кружками своего цвета. Листья раскрашены в соответствии с типом, к которому относится данная бактерия: *Actinobacteria* – чёрным, *Bacteroidetes* – жёлтым, *Proteobacteria* – голубым, *Spirochaetes* – зелёным, *Tenericutes* – красным, неклассифицированные бактерии – серым.

Наибольшее количество рибосомных белков отсутствует у *Candidatus Tremblaya princeps*, *Candidatus Hodgkinia cicadicola*, and *Candidatus Carsonella rudii* – бактерий с экстремально редуцированными геномами.

## 5.2.2 Потеря рибосомных белков зависит от степени редукции генома

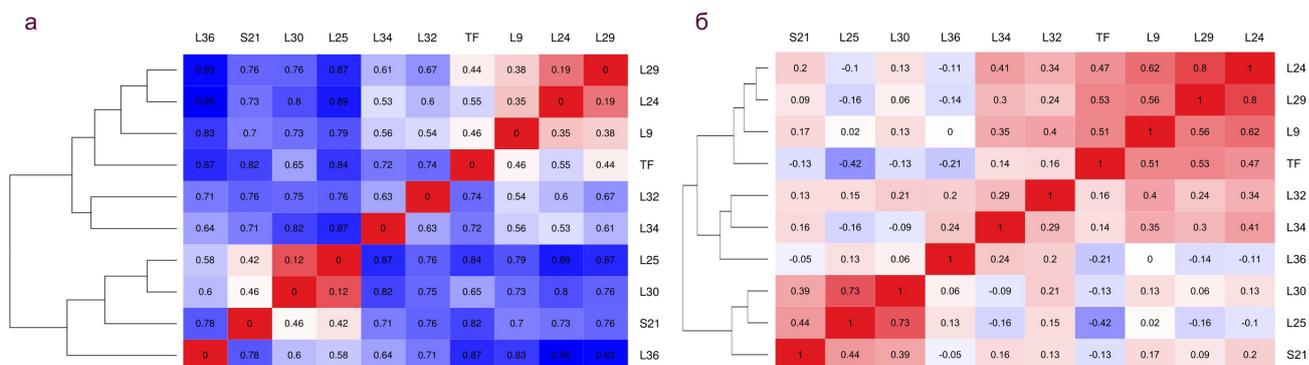
Сравнение количества генов рибосомных белков в геноме и размера генома показало, что эти величины скоррелированы между собой ( $r^2 = 0.6$ ,  $p < 2.2 \times 10^{-16}$ ), эта зависимость особенно ярко видна для крайне редуцированных геномов (Рисунок 5.3а). Действительно, для геномов короче 350 Кб эта корреляция значительно сильнее ( $r^2 = 0.7$ ,  $p = 5.7 \times 10^{-6}$ ), тогда как для более крупных геномов корреляция отсутствует ( $p = 0.53$ ). Это наблюдение верно как для целой рибосомы, так и для отдельных субъединиц (Рисунок 5.3б,в), но более резкое снижение количества рибосомных белков наблюдается для большой субъединицы.



**Рисунок 5.3. 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 сверху 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.

### 5.2.3 Паттерны потери рибосомных белков

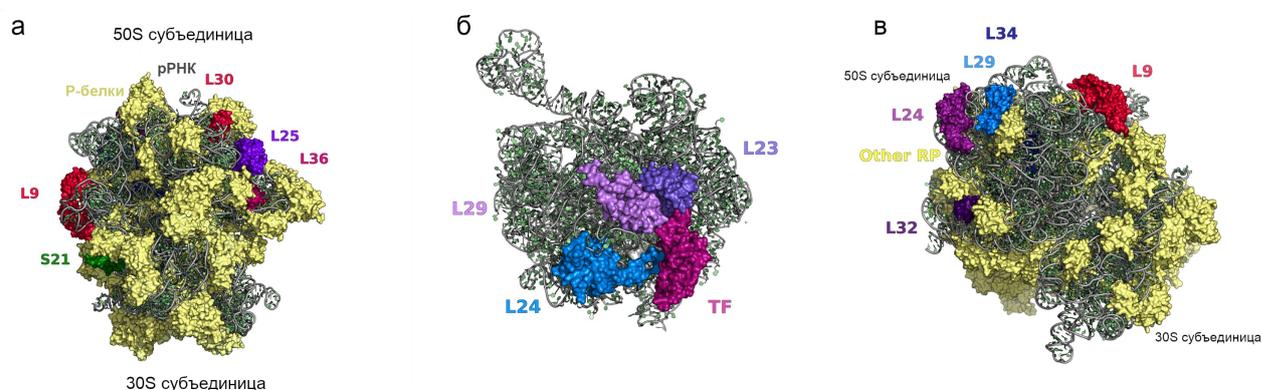
Как уже говорилось выше, десять наиболее часто теряющихся белков могли утрачиваться независимо в ходе событий параллельной эволюции. Для того, чтобы понять, есть ли какая-то связь между потерями р-белков, мы использовали коэффициент сходства Жаккара филогенетических профилей, т.е. векторов присутствия/отсутствия белков во всех исследуемых геномах (Рисунок 5.4а). Нам удалось найти два кластера: L24 + L29 + L9 + TF и L25 + L30 + S21. Данные кластеризуются похожим образом, если использовать коэффициент корреляции Пирсона (Рисунок 5.4б). Так как каждый из часто пропадающих белков был потерян в ходе нескольких независимых событий, мы предполагаем, что наблюдаемые паттерны не являются результатом только филогенетической структуры нашей выборки. Наблюдаемые паттерны частично объясняются положением белков в рибосоме. L25 и L30 расположены рядом (Рисунок 5.5а) и поэтому, вероятно, обычно пропадают вместе.



**Рисунок 5.4. Корреляция филетических паттернов исчезновения разных часто пропадающих рибосомных белков.** (а) Тепловая карта коэффициента сходства Жаккара. Цвета обозначают значения меры от 0 (красный) до 1 (синий). Чем меньше значение, тем более похожи паттерны. (б) Тепловая карта коэффициентов корреляции Пирсона. Цвета обозначают значения меры от 1 (красный) до -1 (синий). Чем больше значение, тем более похожи паттерны.

Совместная потеря L24, L29 и TF может быть объяснена положением TF, который крепится на поверхности рибосомы и взаимодействует с L23, L24 и L29 (Рисунок 5.5б). Потеря TF приводит к потере белков, за счёт которых он крепится к рибосоме. Мы предполагаем, что сначала теряется именно TF, потому что он отсутствует значительно чаще, чем остальные белки из этого паттерна.

Более глобальное наблюдение заключается в том, что часто пропадающие белки по большей части расположены на поверхности рибосомы. Исключение из этого наблюдения составляют белки L34 и L36, которые глубоко погружены в структуру (Рис 5.5а,в).



**Рисунок 5.5. Рибосома *Escherichia coli*.** Цвет белка на структуре и подписи соответствующего белка совпадают. а, в. PDB ID: 5H5U. б. PDB ID: 2D3O.

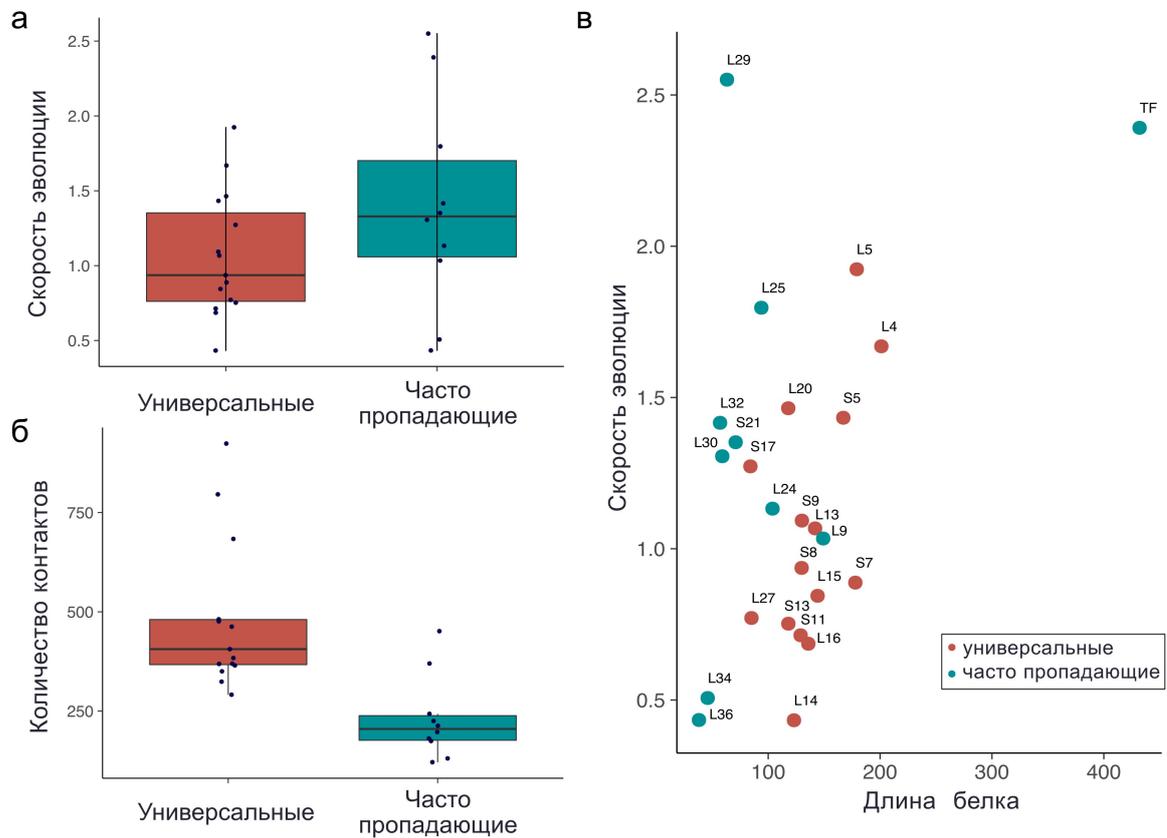
#### 5.2.4 Часто теряющиеся белки меньше контактируют с остальными частями рибосомы по сравнению с универсальными

Дальше 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. Оказалось, что только количество контактов является значимым определяющим фактором ( $p = 0.0211$ ).

Помимо этого, мы сравнили скорости эволюции (Рисунок 5.6а) и количество контактов (Рисунок 5.6б) у универсальных и часто пропадающих белков. Действительно, часто пропадающие белки в среднем меньше контактируют с окружающими белками и рРНК, чем универсальные белки (среднее количество контактов 230.7 против 477.5 соответственно,  $p\text{-value} = 0.00014$ ), так как в основном расположены на поверхности рибосомы.

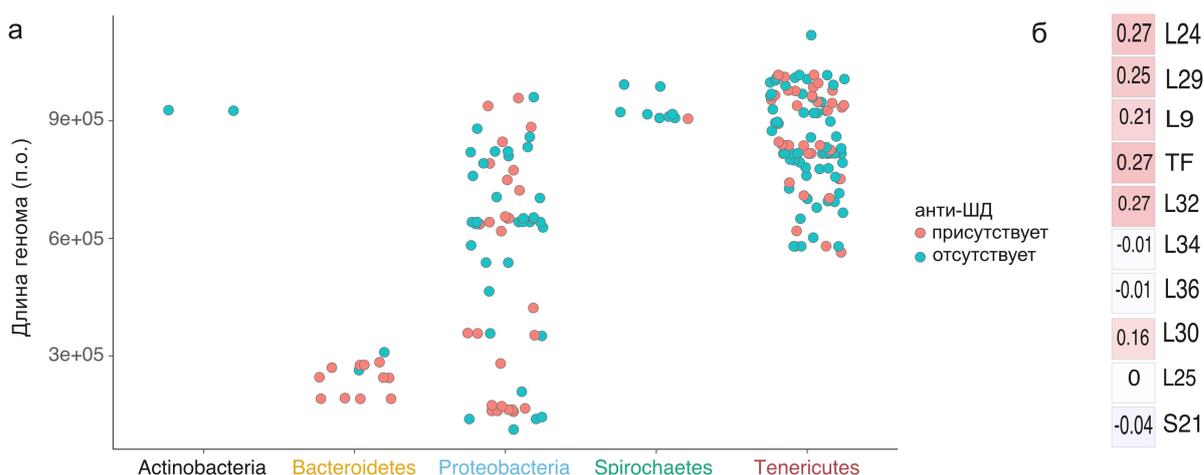
Удивительно, что скорости эволюции универсальных и часто пропадающих белков не отличаются. При этом скорость эволюции и длина белка скоррелированы между собой ( $r^2 = 0.41$ ,  $p = 0.044$ ) и короткие белки эволюционируют медленнее (Рис 5.6в).



**Рисунок 5.6. Сравнение характеристик универсальных и часто пропадающих рибосомных белков.** (а) Разница в скоростях эволюции, (б) разница в количестве контактов между универсальными (отмечены красным) и часто пропадающими (отмечены синим) белками. (в) Зависимость скорости эволюции от длины белка.

### 5.2.5 Укорочение рРНК и потери рибосомных белков

Для того, чтобы изучить связь между исчезновением рибосомных белков и изменениями рРНК, мы проанализировали множественные выравнивания 16S и 23S рРНК и обнаружили девять делеций (четыре – в 23S и пять – в 16S), которые произошли больше чем в двух родах и имели более 10 нуклеотидов в длину хотя бы в одном из видов. Из них одна делеция совпала с сайтом связывания белка L24, и произошла только в таких штаммах, где этот белок отсутствовал. Все другие делеции произошли в участках свободных петель, с которыми не связываются рибосомные белки.



**Рисунок 5.7.** (а) Зависимость между наличием анти-Шайна – Дальгарно (анти-ШД) и длиной генома (в нуклеотидах). Каждая точка – это отдельный геном. Виды, где есть последовательность анти-ШД отмечены синим, виды, где анти-ШД отсутствует, – красным. Точки сгруппированы в колонки по типам. (б) Коэффициент корреляции Пирсона между векторами наличия р-белков и наличия анти-ШД. Красный соответствует значению коэффициента 1, а синий – -1.

### 5.2.6 Потеря анти-ШД происходила много раз в разных таксонах

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 ( $r^2 = 0.24$ ,  $p = 0.00083$ ) (Рисунок 5.7а). Более того, потеря анти-ШД никак не связана с потерей какого-либо из часто пропадающих рибосомных белков (Рисунок 5.7б). Анти-ШД независимо пропадала в разных типах бактерий, и каждый тип содержит как штаммы с этой последовательностью,

так и без неё (Рисунок 5.7а).

### 5.3 Обсуждение результатов

Размер генома бактерий – достаточно лабильная величина, редукция генома происходила независимо в разных таксонах. Наличие достаточно большой выборки редуцированных геномов позволило нам систематически изучить паттерны потери белков, связанные с укорочением генома.

Несмотря на то, что некоторые рибосомные белки могут отсутствовать у бактерий с нередуцированными геномами, сокращение генома приводит к потере и таких рибосомных белков, которые обычно считаются достаточно консервативными. Тем не менее, тот факт, что гены большинства рибосомных белков сохраняются даже в крайне коротких геномах эндосимбионтов говорит о том, что они являются более важными, чем в среднем гены в геноме. Два рибосомных белка, отсутствующие почти во всех геномах в нашей выборке, – это S22 и Thx. Предыдущие исследования показали, что эти белки не очень важны для роста бактерий. Действительно, ген S22 в основном экспрессируется в фазе стационарного роста, и нокаут этого гена не приводит к значительному снижению жизнеспособности мутантов *E. coli* [315]. Thx стабилизирует элементы рРНК в основании головки малой субъединицы у *Thermus thermophilus*, что говорит о том, что этот белок, по всей видимости, необходим для выживания при высоких температурах. Однако отсутствие этих белков, по всей видимости, говорит, не о частой их потере, а, наоборот, о позднем появлении в структуре. S22 встречается только у некоторых энтеробактерий [315], а гомологи Thx найдены только у термофильных бактерий из рода *Thermus* [316,317]. Тем самым, можно полагать, что и S22, и Thx были приобретены некоторыми бактериальными группами уже после расхождения основных групп бактерий.

Список часто пропадающих белков, полученный в этой работе, во многом согласуется с более ранними наблюдениями [18–20,218,219]. Например, из шести неуниверсальных белков, описанных Ютиным и соавт., три (L25, L30, S21) есть в нашем списке часто пропадающих белков, а ещё два (S22 и Thx) практически никогда не встречаются в нашей выборке (см. выше). Однако, тот факт, что наша выборка сильно обогащена симбиотическими бактериями с очень короткими геномами, позволил нам идентифицировать другие белки, которые, по всей видимости, не являются, абсолютно необходимыми. Рибосомные белки L34 и L36, классифицированные ранее [18] как практически универсальные, часто отсутствуют в нашей выборке. Несмотря на то, что эти

белки считаются необходимыми, они могут не влиять на жизнеспособность в некоторых условиях. Было показано, что отсутствие L34 приводит к снижению скорости роста бактерий, однако нормальное функционирование клеток может быть восстановлено, если повысить концентрацию ионов магния в среде [218,318]. Отсутствие белка L36 является критическим только при росте при высоких температурах [319]. Более того, было показано, что L36 отсутствует у бактерий типа *Bacteroidetes* [18]. Рибосомные белки L24 и L29, включённые нами в список часто пропадающих, были потеряны в результате как минимум двух независимых событий, но только у бактерий с крайне редуцированными геномами. Было показано, что нокаутные мутанты *E. coli* по этим генам жизнеспособны [218]. Интересно, что ранее было показано, что белки L31 и S14 отсутствуют у всех *Mollicutes* [20], однако по нашим наблюдениям эти белки отсутствуют только в некоторых штаммах. Расхождение результатов для этих белков, по всей видимости, связано с тем, что мы использовали более чувствительную процедуру для аннотации рибосомных белков.

Итак, какие же характеристики определяют, что белок может быть потерян? Ранее предполагалось, что вероятность потери белка в ходе эволюции тем выше, чем быстрее белок эволюционирует, чем с меньшим количеством других белков он взаимодействует, и чем ниже уровень его экспрессии [320]. Уровень экспрессии разных рибосомных белков должен быть одинаковым, поскольку все эти белки образуют единый комплекс, поэтому мы не включили этот параметр в нашу модель. В этой главе нам удалось показать, что только количество белок-белковых и белок-рРНК контактов определяет вероятность, с которой белок будет потерян. Количество контактов является опосредованной мерой экспонированности белка на поверхность комплекса. И действительно, часто пропадающие рибосомные белки в основном находятся на поверхности, за исключением L34 и L36.

Порядок, в котором рибосомные белки собираются в полный рибосомный комплекс [21], является непрямым свидетельством порядка, в котором белки встраивались в рибосому в ходе эволюции. Тем самым, рибосомные белки, которые характерны только для поздних интермедиатов, скорее всего являются относительно молодыми. Семь из десяти часто пропадающих белков: S21, L9, L25, L29, L32, L34 и L36 – входят только в состав поздних интермедиатов. Таким образом, можно предположить, что часто пропадающие белки появились в составе рибосомы достаточно поздно.

Другой важный вопрос, это пропадают ли рибосомные белки в случайном порядке или есть какие-то паттерны в деградации рибосомы, такие, что исчезновение одного белка ведёт к потере других белков. Наша выборка насыщена геномами с неполным набором

рибосомных белков, что позволило нам изучить такие паттерны потери. Первый идентифицированный паттерн включает потерю целого блока белков вокруг выходного туннеля рибосомы (TF, L9, L24, L29). Триггер-фактор (TF) – это шаперон, ассоциированный с выходным туннелем [321], и он наиболее часто теряется в этом паттерне. TF окружен белками L23, L29 и L24. Было показано, что TF напрямую взаимодействует с L23 и L29 [322]. Роль L9 в этом паттерне не совсем ясна, так как этот белок не взаимодействует напрямую с TF. Ранее было показано, что L9 снижает вероятность сдвига рамки при трансляции [323], что означает, что его потеря приводит к снижению производительности рибосомы. Потеря целого функционального блока кажется достаточно необычным событием, хотя бы потому, что ранее считалось, что бактерии с сильно редуцированными геномами стараются сохранять шапероны [17].

Другой паттерн связан с потерей трёх белков (S21, L25 и L30), и если L25 и L30 расположены близко друг от друга на поверхности рибосомы, то белок S21 достаточно далеко от них отстоит. Множественные независимые потери L25 и L30 и даже S21 наблюдали на других выборках и до этого [18,217]. Этот паттерн остаётся до конца необъясненным, особенно учитывая тот факт, что самый часто пропадающий белок – L30 – крайне консервативен у Archaea и Eucarya [217], где, как предполагается, он необходим для распознавания селеноцистеинового кодона [324]. Роль этого белка у бактерий неясна. Потеря L25 в *E. coli* не является летальной, но приводит к снижению скорости роста [19]. Предполагается, что L25 необходим для того, чтобы взаимодействовать с белком L16, который обеспечивает стабильность рибосомы [325]. S21, по всей видимости, нужен для того, чтобы распознавать мРНК, и во многом его функция напоминает функцию белка S1 [326]. Тем самым, потеря белков в этом паттерне должна приводить к значительному снижению эффективности рибосомы. Таким образом, белки не теряются случайно, есть как минимум два паттерна потерь белков, в каждом из которых происходит совместная потеря белков, расположенных рядом в структуре рибосомы, и один из этих паттернов приводит к потере целого функционального блока.

Таким образом, часто теряющиеся белки меньше контактируют с окружающими структурами, расположены на поверхности рибосомы, и, по всей видимости, являются достаточно молодыми. Отсутствие этих белков снижает эффективность работы рибосом, но не полностью её нарушает, видимо, позволяя им функционировать в условиях стабильной окружающей среды, характерной для симбиотических бактерий, живущих в окружении стабильной среды хозяина.

## 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. Большое количество метагеномных 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 *Holospora* показывает, что приспособление к симбиотическому образу жизни связано с значительным сокращением генома и потерей целого ряда метаболических возможностей. *Holospora* 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 *Holospora* вынуждены получать все аминокислоты из внешней среды, т. е. от хозяина. *Holospora* 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 *Holospora* 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 секвенированной в нашей работе бактерии *Holospora 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 *Holospora* 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.

## **Благодарности**

Я благодарна своему научному руководителю М. С. Гельфанду, который помог мне начать заниматься биоинформатикой. Отдельно мне хотелось бы поблагодарить Н.И. Луховицкую, Е.А. Минину и А.Б. Рахманинову за помощь и поддержку при выборе новой области деятельности. Я признательна своим соавторам – А.Я. Белявской, И. Горянину, М.Д. Казанову, Л. Киселевой, М.Ф. Козну, М.Д. Логачевой, Х. Ма, Д.Б. Малько, М.С. Раутиан, Д.Дж.В. Симпсону и В. Федоровичу. Я благодарю за помощь, поддержку и содержательные обсуждения Д.Д. Николаеву, О.А. Вахрушеву, Д.В. Виноградова, П.В. Шелякина, О.О. Бочкареву, А.Д. Казнадзей и Г.А. Базыкина. Я благодарна за помощь и поддержку при написании текста своим друзьям и семье, а особенно О.А. Мишедченко, М.Г. Кривошеиной и Г.Г. Кривошеиной.

## Список литературы

1. Loman N.J., Pallen M.J. Twenty years of bacterial genome sequencing // *Nature Reviews Microbiology*. 2015. Vol. 13, № 12. P. 787–794.
2. Land M. et al. Insights from 20 years of bacterial genome sequencing // *Funct Integr Genomics*. 2015. Vol. 15, № 2. P. 141–161.
3. Quainoo S. et al. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis // *Clinical Microbiology Reviews*. 2017. Vol. 30, № 4. P. 1015–1063.
4. Quin J.E. et al. Major transcriptional changes observed in the Fulani, an ethnic group less susceptible to malaria // *eLife Sciences*. 2017. Vol. 6. P. e29156.
5. Kimura N. Chapter 14 - Novel Biological Resources Screened From Uncultured Bacteria by a Metagenomic Method // *Metagenomics* / ed. Nagarajan M. Academic Press, 2018. P. 273–288.
6. Garza D.R., Dutilh B.E. From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems // *Cell Mol Life Sci*. 2015. Vol. 72. P. 4287–4308.
7. Benson D.A. et al. GenBank // *Nucleic Acids Res*. 2018. Vol. 46, № Database issue. P. D41–D47.
8. Iwatani K. et al. Translocation of an 89-kDa periplasmic protein is associated with *Holospora* infection // *Biochemical and Biophysical Research Communications*. 2005. Vol. 337, № 4. P. 1198–1205.
9. Dohra H. et al. Cloning and Sequencing of Gene Coding for a Periplasmic 5.4 kDa Peptide of the Macronucleus-Specific Symbiont *Holospora obtusa* of the Ciliate *Paramecium caudatum* // *Zoological Science*. 1997. Vol. 14, № 1. P. 69–75.
10. Fokin S.I., Görtz H.-D. Diversity of *Holospora* Bacteria in *Paramecium* and Their Characterization // *Endosymbionts in Paramecium* / ed. Fujishima M. Springer Berlin Heidelberg, 2009. P. 161–199.
11. Dohra H. et al. Draft Genome Sequence of *Holospora undulata* Strain HU1, a Micronucleus-Specific Symbiont of the Ciliate *Paramecium caudatum* // *Genome Announc*. 2013. Vol. 1, № 4.
12. Dohra H. et al. Draft genome sequences of three *Holospora* species (*Holospora obtusa*, *Holospora undulata*, and *Holospora elegans*), endonuclear symbiotic bacteria of the ciliate *Paramecium caudatum* // *FEMS Microbiology Letters*. 2014. Vol. 359, № 1. P. 16–18.
13. Ramakrishnan V. Ribosome Structure and the Mechanism of Translation // *Cell*. 2002.

Vol. 108, № 4. P. 557–572.

14. Green R., Noller H.F. Ribosomes and Translation // *Annu. Rev. Biochem.* 1997. Vol. 66, № 1. P. 679–716.
15. Nissen P. et al. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis // *Science*. 2000. Vol. 289, № 5481. P. 920–930.
16. Kaczanowska M., Rydén-Aulin M. Ribosome Biogenesis and the Translation Process in *Escherichia coli* // *Microbiol. Mol. Biol. Rev.* 2007. Vol. 71, № 3. P. 477–494.
17. McCutcheon J.P., Moran N.A. Extreme genome reduction in symbiotic bacteria // *Nature Reviews Microbiology*. 2012. Vol. 10. P. 13–26.
18. Yutin N. et al. Phylogenomics of Prokaryotic Ribosomal Proteins // *PLoS ONE*. 2012. Vol. 7, № 5. P. e36972.
19. Baba T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection // *Mol Syst Biol*. 2006. Vol. 2. P. 2006.0008.
20. Grosjean H. et al. Predicting the Minimal Translation Apparatus: Lessons from the Reductive Evolution of Mollicutes // *PLoS Genetics* / ed. Gojobori T. 2014. Vol. 10, № 5. P. e1004363.
21. Chen S.S., Williamson J.R. Characterization of the Ribosome Biogenesis Landscape in *E. coli* Using Quantitative Mass Spectrometry // *Journal of Molecular Biology*. 2013. Vol. 425, № 4. P. 767–779.
22. Nelson-Sathi S. et al. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea // *Proceedings of the National Academy of Sciences*. 2012. Vol. 109, № 50. P. 20537–20542.
23. Deppenmeier U. et al. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea // *J. Mol. Microbiol. Biotechnol.* 2002. Vol. 4, № 4. P. 453–461.
24. Fournier G.P., Gogarten J.P. Evolution of Acetoclastic Methanogenesis in *Methanosarcina* via Horizontal Gene Transfer from Cellulolytic Clostridia // *Journal of Bacteriology*. 2007. Vol. 190, № 3. P. 1124–1127.
25. Woese C.R., Fox G.E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. // *Proc Natl Acad Sci U S A*. 1977. Vol. 74, № 11. P. 5088–5090.
26. Woese C.R., Kandler O., Wheelis M.L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya // *Proceedings of the National Academy of Sciences*. 1990. Vol. 87, № 12. P. 4576–4579.
27. Woese C.R. There must be a prokaryote somewhere: microbiology's search for itself. //

- Microbiology and Molecular Biology Reviews. 1994. Vol. 58, № 1. P. 1–9.
28. Martin W., Koonin E.V. A positive definition of prokaryotes [Electronic resource]: Comments and Opinion // *Nature*. 2006. URL: <https://www.nature.com/articles/442868c> (accessed: 29.08.2018).
29. Whitman W.B., Coleman D.C., Wiebe W.J. Prokaryotes: The unseen majority // *PNAS*. 1998. Vol. 95, № 12. P. 6578–6583.
30. Hug L.A. et al. A new view of the tree of life // *Nature Microbiology*. 2016. Vol. 1. P. 16048.
31. Staley J.T., Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats // *Annu. Rev. Microbiol.* 1985. Vol. 39. P. 321–346.
32. Mukherjee S. et al. 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life // *Nature Biotechnology*. 2017. Vol. 35, № 7. P. 676–683.
33. Vitousek P.M. et al. Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems // *Philos Trans R Soc Lond B Biol Sci*. 2013. Vol. 368, № 1621.
34. Rother M. Methanogenesis // *Handbook of Hydrocarbon and Lipid Microbiology*. Springer, Berlin, Heidelberg, 2010. P. 481–499.
35. Oren A. Metabolic Diversity in Prokaryotes and Eukaryotes // *Biological Science Fundamentals and Systematics*. 2009. Vol. III.
36. Roumpeka D.D. et al. A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data // *Front. Genet.* 2017. Vol. 8.
37. Handelsman J. et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products // *Chemistry & Biology*. 1998. Vol. 5, № 10. P. R245–R249.
38. Venter J.C. et al. Environmental Genome Shotgun Sequencing of the Sargasso Sea // *Science*. 2004. Vol. 304, № 5667. P. 66–74.
39. Yarza P. et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences // *Nature Reviews Microbiology*. 2014. Vol. 12, № 9. P. 635–645.
40. Mysara M. et al. Reconciliation between operational taxonomic units and species boundaries // *FEMS Microbiol Ecol.* 2017. Vol. 93, № 4.
41. Thompson L.R. et al. A communal catalogue reveals Earth’s multiscale microbial diversity // *Nature*. 2017. Vol. 551, № 7681. P. 457–463.
42. Mason O.U. et al. Metagenome, metatranscriptome and single-cell sequencing reveal

- microbial response to Deepwater Horizon oil spill // ISME J. 2012. Vol. 6, № 9. P. 1715–1727.
43. Lozupone C.A., Knight R. Global patterns in bacterial diversity // PNAS. 2007. Vol. 104, № 27. P. 11436–11440.
44. Lima-Mendez G. et al. Determinants of community structure in the global plankton interactome // Science. 2015. Vol. 348, № 6237. P. 1262073.
45. Stubbendieck R.M., Vargas-Bautista C., Straight P.D. Bacterial Communities: Interactions to Scale // Front. Microbiol. 2016. Vol. 7.
46. Levy R., Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules // PNAS. 2013. Vol. 110, № 31. P. 12804–12809.
47. Roesch L.F.W. et al. Pyrosequencing enumerates and contrasts soil microbial diversity // The ISME Journal. 2007. Vol. 1, № 4. P. 283–290.
48. Carson J.K. et al. Low Pore Connectivity Increases Bacterial Diversity in Soil // Appl. Environ. Microbiol. 2010. Vol. 76, № 12. P. 3936–3942.
49. Barman M. et al. Enteric Salmonellosis Disrupts the Microbial Ecology of the Murine Gastrointestinal Tract // Infect. Immun. 2008. Vol. 76, № 3. P. 907–915.
50. Minot S. et al. The human gut virome: Inter-individual variation and dynamic response to diet // Genome Res. 2011. Vol. 21, № 10. P. 1616–1625.
51. Minot S. et al. Rapid evolution of the human gut virome // Proc Natl Acad Sci U S A. 2013. Vol. 110, № 30. P. 12450–12455.
52. Scarpellini E. et al. The human gut microbiota and virome: Potential therapeutic implications // Digestive and Liver Disease. 2015. Vol. 47, № 12. P. 1007–1012.
53. Altindis E. et al. Viral insulin-like peptides activate human insulin and IGF-1 receptor signaling: A paradigm shift for host–microbe interactions // PNAS. 2018. Vol. 115, № 10. P. 2461–2466.
54. The Human Microbiome Project Consortium et al. Structure, function and diversity of the healthy human microbiome // Nature. 2012. Vol. 486, № 7402. P. 207–214.
55. Arumugam M. et al. Enterotypes of the human gut microbiome // Nature. 2011. Vol. 473, № 7346. P. 174–180.
56. Li J. et al. A metagenomic approach to dissect the genetic composition of enterotypes in Han Chinese and two Muslim groups // Syst. Appl. Microbiol. 2018. Vol. 41, № 1. P. 1–12.
57. Nakayama J. et al. Impact of Westernized Diet on Gut Microbiota in Children on Leyte Island // Front Microbiol. 2017. Vol. 8.
58. Tanaka M., Nakayama J. Development of the gut microbiota in infancy and its impact

- on health in later life // *Allergol Int.* 2017. Vol. 66, № 4. P. 515–522.
59. Mardanov A.V. et al. Metagenomic Analysis of the Dynamic Changes in the Gut Microbiome of the Participants of the MARS-500 Experiment, Simulating Long Term Space Flight // *Acta Naturae.* 2013. Vol. 5, № 3. P. 116–125.
60. Kang C. et al. Healthy Subjects Differentially Respond to Dietary Capsaicin Correlating with Specific Gut Enterotypes // *J. Clin. Endocrinol. Metab.* 2016. Vol. 101, № 12. P. 4681–4689.
61. Wu Q. et al. Fermentation properties of isomaltooligosaccharides are affected by human fecal enterotypes // *Anaerobe.* 2017. Vol. 48. P. 206–214.
62. de Moraes A.C.F. et al. Enterotype May Drive the Dietary-Associated Cardiometabolic Risk Factors // *Front. Cell. Infect. Microbiol.* 2017. Vol. 7.
63. Emoto T. et al. Analysis of Gut Microbiota in Coronary Artery Disease Patients: a Possible Link between Gut Microbiota and Coronary Artery Disease // *J. Atheroscler. Thromb.* 2016. Vol. 23, № 8. P. 908–921.
64. Shortt N. et al. A feasibility study: association between gut microbiota enterotype and antibody response to seasonal trivalent influenza vaccine in adults // *Clin Transl Immunology.* 2018. Vol. 7, № 3.
65. Gorvitovskaia A., Holmes S.P., Huse S.M. Interpreting *Prevotella* and *Bacteroides* as biomarkers of diet and lifestyle // *Microbiome.* 2016. Vol. 4.
66. Yatsunenko T. et al. Human gut microbiome viewed across age and geography // *Nature.* 2012. Vol. 486, № 7402. P. 222–227.
67. Moeller A.H. et al. Cospeciation of gut microbiota with hominids // *Science.* 2016. Vol. 353, № 6297. P. 380–382.
68. Costea P.I. et al. Enterotypes in the landscape of gut microbial community composition // *Nature Microbiology.* 2018. Vol. 3, № 1. P. 8–16.
69. Sunagawa S. et al. Structure and function of the global ocean microbiome // *Science.* 2015. Vol. 348, № 6237. P. 1261359.
70. Biller S.J. et al. *Prochlorococcus*: the structure and function of collective diversity // *Nature Reviews Microbiology.* 2015. Vol. 13, № 1. P. 13–27.
71. Mestre M. et al. Sinking particles promote vertical connectivity in the ocean microbiome // *PNAS.* 2018. Vol. 115, № 29. P. E6799–E6807.
72. Raynaud X., Nunan N. Spatial Ecology of Bacteria at the Microscale in Soil // *PLOS ONE.* 2014. Vol. 9, № 1. P. e87217.
73. Schatz A., Bugle E., Waksman S.A. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria. // *Proceedings of the Society for*

- Experimental Biology and Medicine. 1944. Vol. 55, № 1. P. 66–69.
74. Ramirez O. et al. Analysis of structural diversity in wolf-like canids reveals post-domestication variants // *BMC Genomics*. 2014. Vol. 15. P. 465.
75. Torsvik V., Øvreås L. Microbial diversity and function in soil: from genes to ecosystems // *Current Opinion in Microbiology*. 2002. Vol. 5, № 3. P. 240–245.
76. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome // *Nature Reviews Microbiology*. 2017. Vol. 15, № 10. P. 579–590.
77. Lynch M.D.J., Neufeld J.D. Ecology and exploration of the rare biosphere // *Nature Reviews Microbiology*. 2015. Vol. 13, № 4. P. 217–229.
78. Crowther T.W. et al. Predicting the responsiveness of soil biodiversity to deforestation: a cross-biome study // *Global Change Biology*. 2014. Vol. 20, № 9. P. 2983–2994.
79. Fuhrman J.A., Cram J.A., Needham D.M. Marine microbial community dynamics and their ecological interpretation // *Nature Reviews Microbiology*. 2015. Vol. 13, № 3. P. 133–146.
80. Church M.J. Resource Control of Bacterial Dynamics in the Sea // *Microbial Ecology of the Oceans*. Wiley-Blackwell, 2008. P. 335–382.
81. Chow C.-E.T. et al. Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight // *The ISME Journal*. 2013. Vol. 7, № 12. P. 2259–2273.
82. Abdelgadir A. et al. Characteristics, Process Parameters, and Inner Components of Anaerobic Bioreactors [Electronic resource]: Research article // *BioMed Research International*. 2014. URL: <https://www.hindawi.com/journals/bmri/2014/841573/>.
83. Lee D.-J. et al. Current developments in biotechnology and bioengineering: biological treatment of industrial effluents. 1st ed. New York: Elsevier, 2017. 501 p.
84. Logan B.E. et al. Microbial Fuel Cells: Methodology and Technology // *Environ. Sci. Technol.* 2006. Vol. 40, № 17. P. 5181–5192.
85. Watanabe K. Recent developments in microbial fuel cell technologies for sustainable bioenergy // *J. Biosci. Bioeng.* 2008. Vol. 106, № 6. P. 528–536.
86. Dewan A., Beyenal H., Lewandowski Z. Scaling up Microbial Fuel Cells // *Environ. Sci. Technol.* 2008. Vol. 42, № 20. P. 7643–7648.
87. Logan B.E. Scaling up microbial fuel cells and other bioelectrochemical systems // *Appl. Microbiol. Biotechnol.* 2010. Vol. 85, № 6. P. 1665–1671.
88. Longo S. et al. Monitoring and diagnosis of energy consumption in wastewater treatment plants. A state of the art and proposals for improvement // *Applied Energy*. 2016. Vol. 179. P. 1251–1268.

89. Min B., Logan B.E. Continuous Electricity Generation from Domestic Wastewater and Organic Substrates in a Flat Plate Microbial Fuel Cell // *Environ. Sci. Technol.* 2004. Vol. 38, № 21. P. 5809–5814.
90. Santoro C. et al. Power generation from wastewater using single chamber microbial fuel cells (MFCs) with platinum-free cathodes and pre-colonized anodes. // *Biochemical Engineering Journal.* 2012. Vol. 62. P. 8–16.
91. Li W.-W. et al. Recent advances in the separators for microbial fuel cells // *Bioresour. Technol.* 2011. Vol. 102, № 1. P. 244–252.
92. Wei J., Liang P., Huang X. Recent progress in electrodes for microbial fuel cells // *Bioresour. Technol.* 2011. Vol. 102, № 20. P. 9335–9344.
93. Li W.-W., Yu H.-Q., He Z. Towards sustainable wastewater treatment by using microbial fuel cells-centered technologies // *Energy Environ. Sci.* 2014. Vol. 7, № 3. P. 911–924.
94. Fedorovich V. et al. Novel Electrochemically Active Bacterium Phylogenetically Related to *Arcobacter butzleri*, Isolated from a Microbial Fuel Cell // *Appl. Environ. Microbiol.* 2009. Vol. 75, № 23. P. 7326–7334.
95. Methé B.A. et al. Genome of *Geobacter sulfurreducens*: Metal Reduction in Subsurface Environments // *Science.* 2003. Vol. 302, № 5652. P. 1967–1969.
96. Butler J.E. et al. Comparative genomic analysis of *Geobacter sulfurreducens* KN400, a strain with enhanced capacity for extracellular electron transfer and electricity production // *BMC Genomics.* 2012. Vol. 13, № 1. P. 471.
97. Pham T.H. et al. Metabolites produced by *Pseudomonas* sp. enable a Gram-positive bacterium to achieve extracellular electron transfer // *Appl Microbiol Biotechnol.* 2008. Vol. 77, № 5. P. 1119–1129.
98. Kato S., Hashimoto K., Watanabe K. Iron-Oxide Minerals Affect Extracellular Electron-Transfer Paths of *Geobacter* spp. // *Microbes and Environments.* 2013. Vol. 28, № 1. P. 141–148.
99. Fitzgerald L.A. et al. *Shewanella oneidensis* MR-1 Msh pilin proteins are involved in extracellular electron transfer in microbial fuel cells // *Process Biochemistry.* 2012. Vol. 47, № 1. P. 170–174.
100. Wrighton K.C. et al. Evidence for Direct Electron Transfer by a Gram-Positive Bacterium Isolated from a Microbial Fuel Cell // *Appl. Environ. Microbiol.* 2011. Vol. 77, № 21. P. 7633–7639.
101. Gorby Y.A. et al. Electrically conductive bacterial nanowires produced by *Shewanella oneidensis* strain MR-1 and other microorganisms // *PNAS.* 2006. Vol. 103, № 30. P. 11358–11363.

102. Ishii S. et al. Characterization of a filamentous biofilm community established in a cellulose-fed microbial fuel cell // *BMC Microbiol.* 2008. Vol. 8. P. 6.
103. Kouzuma A. et al. Comparative Metagenomics of Anode-Associated Microbiomes Developed in Rice Paddy-Field Microbial Fuel Cells // *PLoS One.* 2013. Vol. 8, № 11.
104. Yates M.D. et al. Convergent development of anodic bacterial communities in microbial fuel cells // *ISME J.* 2012. Vol. 6, № 11. P. 2002–2013.
105. Pirbadian S. et al. *Shewanella oneidensis* MR-1 nanowires are outer membrane and periplasmic extensions of the extracellular electron transport components // *Proc. Natl. Acad. Sci. U.S.A.* 2014. Vol. 111, № 35. P. 12883–12888.
106. Lovley D.R. Electromicrobiology // *Annual Review of Microbiology.* 2012. Vol. 66, № 1. P. 391–409.
107. Commault A.S. et al. Influence of anode potentials on selection of *Geobacter* strains in microbial electrolysis cells. // *Bioresour Technol.* 2013. Vol. 139. P. 226–234.
108. Franks A.E. et al. Microtoming coupled to microarray analysis to evaluate the spatial metabolic status of *Geobacter sulfurreducens* biofilms // *ISME J.* 2010. Vol. 4, № 4. P. 509–519.
109. Sun D. et al. Temporal-Spatial Changes in Viabilities and Electrochemical Properties of Anode Biofilms // *Environ. Sci. Technol.* 2015. Vol. 49, № 8. P. 5227–5235.
110. Yang Y. et al. Bacterial extracellular electron transfer in bioelectrochemical systems // *Process Biochemistry.* 2012. Vol. 47, № 12. P. 1707–1714.
111. Reguera G. et al. Extracellular electron transfer via microbial nanowires // *Nature.* 2005. Vol. 435, № 7045. P. 1098–1101.
112. Tremblay P.-L. et al. A genetic system for *Geobacter metallireducens*: role of the flagellin and pilin in the reduction of Fe(III) oxide // *Environ Microbiol Rep.* 2012. Vol. 4, № 1. P. 82–88.
113. Jiang X. et al. Probing electron transfer mechanisms in *Shewanella oneidensis* MR-1 using a nanoelectrode platform and single-cell imaging // *PNAS.* 2010. Vol. 107, № 39. P. 16806–16810.
114. Choi O., Sang B.-I. Extracellular electron transfer from cathode to microbes: application for biofuel production // *Biotechnol Biofuels.* 2016. Vol. 9.
115. Fedorovich V. et al. Multi-electrode microbial fuel cell with horizontal liquid flow // *Water Sci. Technol.* 2009. Vol. 60, № 2. P. 347–355.
116. Rabaey K. et al. Biofuel cells select for microbial consortia that self-mediate electron transfer // *Appl. Environ. Microbiol.* 2004. Vol. 70, № 9. P. 5373–5382.
117. Ross D.E., Brantley S.L., Tien M. Kinetic Characterization of OmcA and MtrC,

- Terminal Reductases Involved in Respiratory Electron Transfer for Dissimilatory Iron Reduction in *Shewanella oneidensis* MR-1 // *Appl. Environ. Microbiol.* 2009. Vol. 75, № 16. P. 5218–5226.
118. von Canstein H. et al. Secretion of flavins by *Shewanella* species and their role in extracellular electron transfer // *Appl. Environ. Microbiol.* 2008. Vol. 74, № 3. P. 615–623.
119. Hoppe H.-G. et al. Degradation of Macromolecular Organic Compounds in a Tropical Lagoon (Ciénaga Grande, Colombia) and its Ecological Significance // *Internationale Revue der gesamten Hydrobiologie und Hydrographie.* 1983. Vol. 68, № 6. P. 811–824.
120. Borole A.P. et al. Improving power production in acetate-fed microbial fuel cells via enrichment of exoelectrogenic organisms in flow-through systems // *Biochemical Engineering Journal.* 2009. Vol. 48, № 1. P. 71–80.
121. Khater D.Z., El-Khatib K.M., Hassan H.M. Microbial diversity structure in acetate single chamber microbial fuel cell for electricity generation // *Journal of Genetic Engineering and Biotechnology.* 2017. Vol. 15, № 1. P. 127–137.
122. Liu H., Cheng S., Logan B.E. Production of Electricity from Acetate or Butyrate Using a Single-Chamber Microbial Fuel Cell // *Environ. Sci. Technol.* 2005. Vol. 39, № 2. P. 658–662.
123. Stager J.L., Zhang X., Logan B.E. Addition of acetate improves stability of power generation using microbial fuel cells treating domestic wastewater // *Bioelectrochemistry.* 2017. Vol. 118. P. 154–160.
124. Bond D.R., Lovley D.R. Electricity Production by *Geobacter sulfurreducens* Attached to Electrodes // *Appl. Environ. Microbiol.* 2003. Vol. 69, № 3. P. 1548–1555.
125. Harden A. LXIV.–The chemical action of *Bacillus coli communis* and similar organisms on carbohydrates and allied compounds // *J. Chem. Soc., Trans.* 1901. Vol. 79, № 0. P. 610–628.
126. Wolfe A.J. The Acetate Switch // *Microbiol. Mol. Biol. Rev.* 2005. Vol. 69, № 1. P. 12–50.
127. Castaño-Cerezo S. et al. An insight into the role of phosphotransacetylase (pta) and the acetate/acetyl-CoA node in *Escherichia coli* // *Microbial Cell Factories.* 2009. Vol. 8, № 1. P. 54.
128. Somerville G.A. et al. Correlation of Acetate Catabolism and Growth Yield in *Staphylococcus aureus*: Implications for Host-Pathogen Interactions // *Infect Immun.* 2003. Vol. 71, № 8. P. 4724–4732.
129. Kakuda H., Hosono K., Ichihara S. Identification and Characterization of the ackA (Acetate Kinase A)-pta (Phosphotransacetylase) Operon and Complementation Analysis of Acetate Utilization by an ackA-pta Deletion Mutant of *Escherichia coli* // *J Biochem.* 1994. Vol. 116, № 4. P. 916–922.
130. Kumari S. et al. Cloning, characterization, and functional expression of *acs*, the gene

- which encodes acetyl coenzyme A synthetase in *Escherichia coli*. // *Journal of Bacteriology*. 1995. Vol. 177, № 10. P. 2878–2886.
131. Enjalbert B. et al. Acetate fluxes in *Escherichia coli* are determined by the thermodynamic control of the Pta-AckA pathway // *Scientific Reports*. 2017. Vol. 7. P. 42135.
132. Ishii S. et al. Functionally Stable and Phylogenetically Diverse Microbial Enrichments from Microbial Fuel Cells during Wastewater Treatment // *PLOS ONE*. 2012. Vol. 7, № 2. P. e30495.
133. Zhang H. et al. Phylogenetic and Metagenomic Analyses of Substrate-Dependent Bacterial Temporal Dynamics in Microbial Fuel Cells // *PLOS ONE*. 2014. Vol. 9, № 9. P. e107460.
134. Martínez J.L. Antibiotics and Antibiotic Resistance Genes in Natural Environments // *Science*. 2008. Vol. 321, № 5887. P. 365–367.
135. Ochman H., Lawrence J.G., Groisman E.A. Lateral gene transfer and the nature of bacterial innovation // *Nature*. 2000. Vol. 405, № 6784. P. 299–304.
136. Omelchenko M.V. et al. Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ // *Genome Biol*. 2003. Vol. 4, № 9. P. R55.
137. Wozniak R.A.F., Waldor M.K. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow // *Nature Reviews Microbiology*. 2010. Vol. 8, № 8. P. 552–563.
138. Babic A. et al. Efficient Gene Transfer in Bacterial Cell Chains // *mBio*. 2011. Vol. 2, № 2. P. e00027-11-e00027-11.
139. Sun D. Pull in and Push Out: Mechanisms of Horizontal Gene Transfer in Bacteria // *Front. Microbiol*. 2018. Vol. 9.
140. Thomas C.M., Nielsen K.M. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria // *Nature Reviews Microbiology*. 2005. Vol. 3, № 9. P. 711–721.
141. Polz M.F., Alm E.J., Hanage W.P. Horizontal gene transfer and the evolution of bacterial and archaeal population structure // *Trends in Genetics*. 2013. Vol. 29, № 3. P. 170–175.
142. Popa O., Dagan T. Trends and barriers to lateral gene transfer in prokaryotes // *Current Opinion in Microbiology*. 2011. Vol. 14, № 5. P. 615–623.
143. Sjostrand J. et al. A Bayesian Method for Analyzing Lateral Gene Transfer // *Systematic Biology*. 2014. Vol. 63, № 3. P. 409–420.
144. Gogarten J.P., Townsend J.P. Horizontal gene transfer, genome innovation and evolution // *Nature Reviews Microbiology*. 2005. Vol. 3, № 9. P. 679–687.
145. Choi I.-G., Kim S.-H. Global extent of horizontal gene transfer // *Proceedings of the*

- National Academy of Sciences. 2007. Vol. 104, № 11. P. 4489–4494.
146. Koonin E.V. Horizontal gene transfer: the path to maturity // *Molecular Microbiology*. 2003. Vol. 50, № 3. P. 725–727.
147. Novichkov P.S. et al. Genome-Wide Molecular Clock and Horizontal Gene Transfer in Bacterial Evolution // *Journal of Bacteriology*. 2004. Vol. 186, № 19. P. 6575–6585.
148. Nelson-Sathi S. et al. Origins of major archaeal clades correspond to gene acquisitions from bacteria // *Nature*. 2015. Vol. 517(7532). P. 77–80.
149. Creevey C.J. et al. Universally Distributed Single-Copy Genes Indicate a Constant Rate of Horizontal Transfer // *PLoS ONE* / ed. Liberles D. 2011. Vol. 6, № 8. P. e22099.
150. Jain R., Rivera M.C., Lake J.A. Horizontal gene transfer among genomes: the complexity hypothesis // *Proceedings of the National Academy of Sciences*. 1999. Vol. 96, № 7. P. 3801–3806.
151. Wellner A., Lurie M.N., Gophna U. Complexity, connectivity, and duplicability as barriers to lateral gene transfer // *Genome Biol*. 2007. Vol. 8, № 8. P. R156.
152. Medrano-Soto A. Successful Lateral Transfer Requires Codon Usage Compatibility Between Foreign Genes and Recipient Genomes // *Molecular Biology and Evolution*. 2004. Vol. 21, № 10. P. 1884–1894.
153. Sorek R. et al. Genome-Wide Experimental Determination of Barriers to Horizontal Gene Transfer // *Science*. 2007. Vol. 318, № 5855. P. 1449–1452.
154. Wolf Y.I. et al. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer // *Biology direct*. 2012. Vol. 7, № 1. P. 1–15.
155. Allen M.A. et al. The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation // *The ISME journal*. 2009. Vol. 3, № 9. P. 1012–1035.
156. Chen Z. et al. The genome and transcriptome of a newly described psychrophilic archaeon, *Methanolobus psychrophilus* R15, reveal its cold adaptive characteristics: Cold adaptation of a psychrophilic methanogenic archaeon // *Environmental Microbiology Reports*. 2012. P. n/a-n/a.
157. Lomans B.P. et al. Isolation and Characterization of *Methanomethylovorans hollandica* gen. nov., sp. nov., Isolated from Freshwater Sediment, a Methylophilic Methanogen Able To Grow on Dimethyl Sulfide and Methanethiol // *Appl Environ Microbiol*. 1999. Vol. 65, № 8. P. 3641–3650.
158. Anderson I. et al. Genomic Characterization of Methanomicrobiales Reveals Three

Classes of Methanogens // PLoS ONE. 2009. Vol. 4, № 6. P. e5797.

159. Fournier G.P., Huang J., Gogarten J.P. Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life // Philosophical Transactions of the Royal Society B: Biological Sciences. 2009. Vol. 364, № 1527. P. 2229–2239.

160. Galagan J.E. The Genome of *M. acetivorans* Reveals Extensive Metabolic and Physiological Diversity // Genome Research. 2002. Vol. 12, № 4. P. 532–542.

161. Maeder D.L. et al. The *Methanosarcina barkeri* Genome: Comparative Analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* Reveals Extensive Rearrangement within Methanosarcinal Genomes // Journal of Bacteriology. 2006. Vol. 188, № 22. P. 7922–7931.

162. Lind A.E. et al. Genomes of two archaeal endosymbionts show convergent adaptations to an intracellular lifestyle // The ISME Journal. 2018. Vol. 12, № 11. P. 2655.

163. Podar M. et al. Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park // Biol Direct. 2013. Vol. 8. P. 9.

164. Becker M.H. et al. Phylogenetic distribution of symbiotic bacteria from Panamanian amphibians that inhibit growth of the lethal fungal pathogen *Batrachochytrium dendrobatidis* // Mol Ecol. 2015. Vol. 24, № 7. P. 1628–1641.

165. Bennett G.M., Moran N.A. Small, Smaller, Smallest: The Origins and Evolution of Ancient Dual Symbioses in a Phloem-Feeding Insect // Genome Biol Evol. 2013. Vol. 5, № 9. P. 1675–1688.

166. Kawaguchi M., Minamisawa K. Plant–Microbe Communications for Symbiosis // Plant Cell Physiol. 2010. Vol. 51, № 9. P. 1377–1380.

167. McCutcheon J.P., Moran N.A. Functional convergence in reduced genomes of bacterial symbionts spanning 200 million years of evolution // Genome Biology and Evolution. 2010.

168. Moran N.A. et al. The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes // PNAS. 2005. Vol. 102, № 47. P. 16919–16926.

169. Wexler A.G. et al. Human symbionts inject and neutralize antibacterial toxins to persist in the gut // Proceedings of the National Academy of Sciences. 2016. Vol. 113, № 13. P. 3639–3644.

170. Wu M. et al. Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut *Bacteroides* // Science. 2015. Vol. 350, № 6256. P. aac5992–aac5992.

171. Zipfel C., Oldroyd G.E.D. Plant signalling in symbiosis and immunity // Nature. 2017. Vol. 543, № 7645. P. 328–336.

172. Oulhen N., Schulz B.J., Carrier T.J. English translation of Heinrich Anton de Bary's 1878 speech, '*Die Erscheinung der Symbiose*' ('*De la symbiose*') // *Symbiosis*. 2016. Vol. 69, № 3. P. 131–139.
173. Archibald J.M. Endosymbiosis and Eukaryotic Cell Evolution // *Current Biology*. 2015. Vol. 25, № 19. P. R911–R921.
174. Sabater-Muñoz B. et al. Chance and necessity in the genome evolution of endosymbiotic bacteria of insects // *ISME J*. 2017.
175. Dale C., Welburn S.C. The endosymbionts of tsetse flies: manipulating host–parasite interactions // *International Journal for Parasitology*. 2001. Vol. 31, № 5. P. 628–631.
176. Hedges L.M. et al. *Wolbachia* and Virus Protection in Insects // *Science*. 2008. Vol. 322, № 5902. P. 702–702.
177. Weiss B.L., Wang J., Aksoy S. Tsetse Immune System Maturation Requires the Presence of Obligate Symbionts in Larvae // *PLOS Biology*. 2011. Vol. 9, № 5. P. e1000619.
178. Kusch J. et al. The toxic symbiont *Caedibacter caryophila* in the cytoplasm of *Paramecium novaurelia* // *Microbial ecology*. 2000. Vol. 40, № 4. P. 330–335.
179. Vannini C. et al. Betaproteobacterial symbionts of the ciliate *Euplotes*: origin and tangled evolutionary path of an obligate microbial association // *Environmental Microbiology*. 2012. Vol. 14, № 9. P. 2553–2563.
180. Vannini C. et al. High degree of specificity in the association between symbiotic betaproteobacteria and the host *Euplotes* (Ciliophora, Euplotia) // *European Journal of Protistology*. 2017. Vol. 59. P. 124–132.
181. Lu M., Hulcr J., Sun J. The Role of Symbiotic Microbes in Insect Invasions // *Annual Review of Ecology, Evolution, and Systematics*. 2016. Vol. 47, № 1. P. 487–505.
182. Yañez O. et al. Endosymbiotic bacteria in honey bees: *Arsenophonus* spp. are not transmitted transovarially // *FEMS Microbiol Lett*. 2016. Vol. 363, № 14.
183. Batut B. et al. Reductive genome evolution at both ends of the bacterial population size spectrum // *Nat Rev Micro*. 2014. Vol. 12, № 12. P. 841–850.
184. Martinez-Cano D.J. et al. Evolution of small prokaryotic genomes // *Frontiers in Microbiology*. 2015. Vol. 5.
185. Boscaro V. et al. Parallel genome reduction in symbionts descended from closely related free-living bacteria // *Nature Ecology & Evolution*. 2017. Vol. 1, № 8. P. 1160–1167.
186. Marais G.A.B., Calteau A., Tenaille O. Mutation rate and genome reduction in endosymbiotic and free-living bacteria // *Genetica*. 2008. Vol. 134, № 2. P. 205–210.
187. Horst J.-P. et al. *Escherichia coli* mutator genes // *Trends in Microbiology*. 1999. Vol. 7,

№ 1. P. 29–36.

188. Long H. et al. Evolutionary determinants of genome-wide nucleotide composition // *Nature Ecology & Evolution*. 2018. Vol. 2, № 2. P. 237–240.
189. Ferla M.P. et al. New rRNA Gene-Based Phylogenies of the Alphaproteobacteria Provide Perspective on Major Groups, Mitochondrial Ancestry and Phylogenetic Instability // *PLOS ONE*. 2013. Vol. 8, № 12. P. e83383.
190. Szokoli F. et al. Disentangling the taxonomy of Rickettsiales and description of two novel symbionts (“*Candidatus* Bealeia paramacronuclearis” and “*Candidatus* Fokinia cryptica”) sharing the cytoplasm of the ciliate protist *Paramecium biaurelia* // *Appl. Environ. Microbiol.* 2016. P. AEM.02284-16.
191. Görtz H.-D., Schmidt H.J. *Holospora* // *Bergey’s Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd, 2015.
192. Gromov B.V., Ossipov D.V. *Holospora* (ex Hafkine 1890) nom. rev., a Genus of Bacteria Inhabiting the Nuclei of Paramecia // *International Journal of Systematic and Evolutionary Microbiology*. 1981. Vol. 31, № 3. P. 348–352.
193. Schweikert M., Fujishima M., Görtz H.-D. Symbiotic Associations Between Ciliates and Prokaryotes // *The Prokaryotes*. Springer, Berlin, Heidelberg, 2013. P. 427–463.
194. Sabaneyeva E.V. et al. Actin-based mechanism of *Holospora obtusa* trafficking in *Paramecium caudatum* // *Protist*. 2009. Vol. 160, № 2. P. 205–219.
195. Abamo F., Dohra H., Fujishima M. Fate of the 63-kDa periplasmic protein of the infectious form of the endonuclear symbiotic bacterium *Holospora obtusa* during the infection process // *FEMS Microbiol Lett.* 2008. Vol. 280, № 1. P. 21–27.
196. Fujishima M., Dohra H., Kawai M. Quantitative Changes in Periplasmic Proteins of the Macronucleus-Specific Bacterium *Holospora obtusa* in the Infection Process of the Ciliate *Paramecium caudatum* // *Journal of Eukaryotic Microbiology*. 1997. Vol. 44, № 6. P. 636–642.
197. Fujishima M., Kawai M., Yamamoto R. *Paramecium caudatum* acquires heat-shock resistance in ciliary movement by infection with the endonuclear symbiotic bacterium *Holospora obtusa* // *FEMS Microbiology Letters*. 2005. Vol. 243, № 1. P. 101–105.
198. Fujishima M. Infection and Maintenance of *Holospora* Species in *Paramecium caudatum* // *Endosymbionts in Paramecium* / ed. Fujishima M. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. Vol. 12. P. 201–225.
199. Bella C. et al. Fitness Impact of Obligate Intranuclear Bacterial Symbionts Depends on Host Growth Phase // *Front Microbiol.* 2016. Vol. 7.
200. Görtz H.D., Fujishima M. Conjugation and meiosis of *Paramecium caudatum* infected

- with the micronucleus-specific bacterium *Holospora elegans* // *Eur. J. Cell Biol.* 1983. Vol. 32, № 1. P. 86–91.
201. Restif O., Kaltz O. Condition-dependent virulence in a horizontally and vertically transmitted bacterial parasite // *Oikos*. 2006. Vol. 114, № 1. P. 148–158.
202. Andersson S.G.E. et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria // *Nature*. 1998. Vol. 396, № 6707. P. 133–140.
203. Fuxelius H.-H. et al. The genomic and metabolic diversity of *Rickettsia* // *Research in Microbiology*. 2007. Vol. 158, № 10. P. 745–753.
204. Georgiades K. et al. Gene gain and loss events in *Rickettsia* and *Orientia* species // *Biology Direct*. 2011. Vol. 6, № 1. P. 6.
205. Hotopp J.C.D. et al. Comparative Genomics of Emerging Human Ehrlichiosis Agents // *PLOS Genetics*. 2006. Vol. 2, № 2. P. e21.
206. Driscoll T.P. et al. Wholly *Rickettsia*! Reconstructed Metabolic Profile of the Quintessential Bacterial Parasite of Eukaryotic Cells // *mBio*. 2017. Vol. 8, № 5. P. e00859-17.
207. Fraser C.M. et al. The Minimal Gene Complement of *Mycoplasma genitalium* // *Science*. 1995. Vol. 270, № 5235. P. 397–404.
208. Nakabachi A. et al. The 160-Kilobase Genome of the Bacterial Endosymbiont *Carsonella* // *Science*. 2006. Vol. 314, № 5797. P. 267–267.
209. Hutchison C.A. et al. Design and synthesis of a minimal bacterial genome // *Science*. 2016. Vol. 351, № 6280. P. aad6253.
210. Kurland C.G. Structure and Function of the Bacterial Ribosome // *Annual Review of Biochemistry*. 1972. Vol. 41, № 1. P. 377–408.
211. Schmeing T.M., Moore P.B., Steitz T.A. Structures of deacylated tRNA mimics bound to the E site of the large ribosomal subunit // *RNA*. 2003. Vol. 9, № 11. P. 1345–1352.
212. Khaitovich P. et al. Characterization of functionally active subribosomal particles from *Thermus aquaticus* // *PNAS*. 1999. Vol. 96, № 1. P. 85–90.
213. Schuwirth B.S. et al. Structures of the Bacterial Ribosome at 3.5 Å Resolution // *Science*. 2005. Vol. 310, № 5749. P. 827–834.
214. Aseev L.V., Boni I.V. Extraribosomal functions of bacterial ribosomal proteins // *Mol Biol*. 2011. Vol. 45, № 5. P. 739.
215. Roberts E. et al. Molecular signatures of ribosomal evolution // *PNAS*. 2008. Vol. 105, № 37. P. 13953–13958.
216. Smith T.F. et al. The origin and evolution of the ribosome // *Biol Direct*. 2008. Vol. 3, № 1. P. 16.

217. Lecompte O. et al. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale // *Nucleic Acids Res.* 2002. Vol. 30, № 24. P. 5382–5390.
218. Shoji S. et al. Systematic chromosomal deletion of bacterial ribosomal protein genes // *J Mol Biol.* 2011. Vol. 413, № 4. P. 751–761.
219. Akanuma G. et al. Inactivation of Ribosomal Protein Genes in *Bacillus subtilis* Reveals Importance of Each Ribosomal Protein for Cell Proliferation and Cell Differentiation // *J. Bacteriol.* 2012. Vol. 194, № 22. P. 6282–6291.
220. Bokov K., Steinberg S.V. A hierarchical model for evolution of 23S ribosomal RNA // *Nature.* 2009. Vol. 457, № 7232. P. 977–980.
221. Margulies M. et al. Genome sequencing in microfabricated high-density picolitre reactors // *Nature.* 2005. Vol. 437, № 7057. P. 376–380.
222. Meyer F. et al. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes // *BMC Bioinformatics.* 2008. Vol. 9, № 1. P. 386.
223. Quast C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools // *Nucleic Acids Research.* 2012. Vol. 41, № D1. P. D590–D596.
224. Sayers E.W. et al. Database resources of the National Center for Biotechnology Information // *Nucleic Acids Res.* 2009. Vol. 37, № Database issue. P. D5–D15.
225. Tatusov R.L. et al. The COG database: a tool for genome-scale analysis of protein functions and evolution // *Nucleic Acids Res.* 2000. Vol. 28, № 1. P. 33–36.
226. Kanehisa M., Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes // *Nucleic Acids Res.* 2000. Vol. 28, № 1. P. 27–30.
227. Caspi R. et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases // *Nucleic Acids Res.* 2012. Vol. 40, № D1. P. D742–D753.
228. Yang Y. et al. Bacterial extracellular electron transfer in bioelectrochemical systems // *Process Biochemistry.* 2012. Vol. 47, № 12. P. 1707–1714.
229. Bouhenni R.A. et al. The Role of *Shewanella oneidensis* MR-1 Outer Surface Structures in Extracellular Electron Transfer // *Electroanalysis.* 2010. Vol. 22, № 7–8. P. 856–864.
230. Carlson H.K. et al. Surface multiheme c-type cytochromes from *Thermincola potens* and implications for respiratory metal reduction by Gram-positive bacteria // *Proc. Natl. Acad. Sci. U.S.A.* 2012. Vol. 109, № 5. P. 1702–1707.
231. Carmona-Martinez A.A. et al. Cyclic voltammetric analysis of the electron transfer of

- Shewanella oneidensis* MR-1 and nanofilament and cytochrome knock-out mutants // Bioelectrochemistry. 2011. Vol. 81, № 2. P. 74–80.
232. Okamoto A., Nakamura R., Hashimoto K. In-vivo identification of direct electron transfer from *Shewanella oneidensis* MR-1 to electrodes via outer-membrane OmcA–MtrCAB protein complexes // Electrochimica Acta. 2011. Vol. 56, № 16. P. 5526–5531.
233. Letunic I., Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy // Nucl. Acids Res. 2011. Vol. 39, № suppl 2. P. W475–W478.
234. Kiseleva L. et al. Taxonomic and functional metagenomic analysis of anodic communities in two pilot-scale microbial fuel cells treating different industrial wastewaters // J Integr Bioinform. 2015. Vol. 12, № 3. P. 1–15.
235. Klocke M. et al. Microbial community analysis of a biogas-producing completely stirred tank reactor fed continuously with fodder beet silage as mono-substrate // Syst. Appl. Microbiol. 2007. Vol. 30, № 2. P. 139–151.
236. Kampmann K. et al. Unexpected Stability of *Bacteroidetes* and *Firmicutes* Communities in Laboratory Biogas Reactors Fed with Different Defined Substrates // Appl. Environ. Microbiol. 2012. Vol. 78, № 7. P. 2106–2119.
237. Logan B.E. Exoelectrogenic bacteria that power microbial fuel cells // Nature Reviews Microbiology. 2009. Vol. 7, № 5. P. 375–381.
238. Lu L. et al. Syntrophic interactions drive the hydrogen production from glucose at low temperature in microbial electrolysis cells // Bioresour. Technol. 2012. Vol. 124. P. 68–76.
239. Lanthier M., Gregory K.B., Lovley D.R. Growth with high planktonic biomass in *Shewanella oneidensis* fuel cells // FEMS Microbiology Letters. 2008. Vol. 278, № 1. P. 29–35.
240. Malvankar N.S. et al. Tunable metallic-like conductivity in microbial nanowire networks // Nature Nanotechnology. 2011. Vol. 6, № 9. P. 573–579.
241. Spring S. et al. The Genome Sequence of *Methanohalophilus mahii* SLPT Reveals Differences in the Energy Metabolism among Members of the *Methanosarcinaceae* Inhabiting Freshwater and Saline Environments // Archaea. 2010. Vol. 2010. P. 1–16.
242. Kawarabayasi Y. et al. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3 // DNA Res. 1998. Vol. 5, № 2. P. 55–76.
243. Cohen G.N. et al. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi* // Mol. Microbiol. 2003. Vol. 47, № 6. P. 1495–1512.
244. Robb F.T. et al. Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology // Meth. Enzymol. 2001. Vol. 330. P. 134–157.

245. Swithers K.S. et al. Genome sequence of *Thermotoga* sp. strain RQ2, a hyperthermophilic bacterium isolated from a geothermally heated region of the seafloor near Ribeira Quente, the Azores // *J. Bacteriol.* 2011. Vol. 193, № 20. P. 5869–5870.
246. Zhaxybayeva O. et al. On the chimeric nature, thermophilic origin, and phylogenetic placement of the *Thermotogales* // *Proceedings of the National Academy of Sciences.* 2009. Vol. 106, № 14. P. 5865–5870.
247. Swithers K.S. et al. Genome sequence of *Kosmotoga olearia* strain TBF 19.5.1, a thermophilic bacterium with a wide growth temperature range, isolated from the Troll B oil platform in the North Sea // *J. Bacteriol.* 2011. Vol. 193, № 19. P. 5566–5567.
248. Altschul S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs // *Nucleic Acids Res.* 1997. Vol. 25, № 17. P. 3389–3402.
249. Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput // *Nucleic Acids Res.* 2004. Vol. 32, № 5. P. 1792–1797.
250. Penn O. et al. GUIDANCE: a web server for assessing alignment confidence scores // *Nucleic Acids Research.* 2010. Vol. 38, № Web Server. P. W23–W28.
251. Guindon S. et al. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0 // *Syst Biol.* 2010. Vol. 59, № 3. P. 307–321.
252. Finn R.D. et al. The Pfam protein families database: towards a more sustainable future // *Nucleic Acids Res.* 2016. Vol. 44, № D1. P. D279–D285.
253. Arakaki A.K., Huang Y., Skolnick J. EFICAz2: enzyme function inference by a combined approach enhanced by machine learning // *BMC Bioinformatics.* 2009. Vol. 10, № 1. P. 107.
254. Simankova M.V. et al. *Methanosarcina lacustris* sp. nov., a new psychrotolerant methanogenic archaeon from anoxic lake sediments // *Syst. Appl. Microbiol.* 2001. Vol. 24, № 3. P. 362–367.
255. Rouvière P. et al. A detailed phylogeny for the *Methanomicrobiales* // *Syst. Appl. Microbiol.* 1992. Vol. 15, № 3. P. 363–371.
256. Gao B., Gupta R.S. Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis // *BMC Genomics.* 2007. Vol. 8, № 1. P. 86.
257. Nelson K.E. et al. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima* // *Nature.* 1999. Vol. 399, № 6734. P. 323–329.
258. Logsdon J.M., Faguy D.M. Evolutionary genomics: *Thermotoga* heats up lateral gene

- transfer // *Current Biology*. 1999. Vol. 9, № 19. P. R747–R751.
259. Nesbo C.L. et al. Phylogenetic analyses of two “archaeal” genes in *thermotoga maritima* reveal multiple transfers between archaea and bacteria // *Mol. Biol. Evol.* 2001. Vol. 18, № 3. P. 362–375.
260. Koonin E.V., Wolf Y.I. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world // *Nucleic Acids Res.* 2008. Vol. 36, № 21. P. 6688–6719.
261. Gordienko E.N., Kazanov M.D., Gelfand M.S. Evolution of Pan-Genomes of *Escherichia coli*, *Shigella* spp., and *Salmonella enterica* // *J Bacteriol.* 2013. Vol. 195, № 12. P. 2786–2792.
262. Tang Y.-Q. et al. Anaerobic treatment performance and microbial population of thermophilic upflow anaerobic filter reactor treating awamori distillery wastewater // *Journal of Bioscience and Bioengineering*. 2007. Vol. 104, № 4. P. 281–287.
263. Patil S.S., Kumar M.S., Ball A.S. Microbial community dynamics in anaerobic bioreactors and algal tanks treating piggery wastewater // *Applied Microbiology and Biotechnology*. 2010. Vol. 87, № 1. P. 353–363.
264. Ren Q., Paulsen I.T. Comparative Analyses of Fundamental Differences in Membrane Transport Capabilities in Prokaryotes and Eukaryotes // *PLoS Computational Biology*. 2005. Vol. 1, № 3. P. e27.
265. Dibrova D.V., Galperin M.Y., Mulkijanian A.Y. Characterization of the N-ATPase, a distinct, laterally transferred Na<sup>+</sup>-translocating form of the bacterial F-type membrane ATPase // *Bioinformatics*. 2010. Vol. 26, № 12. P. 1473–1476.
266. Hovey R. et al. DNA microarray analysis of *Methanosarcina mazei* Gö1 reveals adaptation to different methanogenic substrates // *Molecular Genetics and Genomics*. 2005. Vol. 273, № 3. P. 225–239.
267. Veit K. et al. Global transcriptional analysis of *Methanosarcina mazei* strain Gö1 under different nitrogen availabilities // *Molecular Genetics and Genomics*. 2006. Vol. 276, № 1. P. 41–55.
268. Pfluger K. et al. Identification of genes involved in salt adaptation in the archaeon *Methanosarcina mazei* Gö1 using genome-wide gene expression profiling // *FEMS Microbiology Letters*. 2007. Vol. 277, № 1. P. 79–89.
269. Weidenbach K. et al. Deletion of the archaeal histone in *Methanosarcina mazei* Gö1 results in reduced growth and genomic transcription // *Molecular Microbiology*. 2008. Vol. 67, № 3. P. 662–671.
270. Garushyants S.K., Kazanov M.D., Gelfand M.S. Horizontal gene transfer and genome

- evolution in *Methanosarcina* // BMC Evol. Biol. 2015. Vol. 15. P. 102.
271. Li Q. et al. Proteome of *Methanosarcina acetivorans* Part I: An Expanded View of the Biology of the Cell // Journal of Proteome Research. 2005. Vol. 4, № 1. P. 112–128.
272. Li Q. et al. Proteome of *Methanosarcina acetivorans* Part II: Comparison of Protein Levels in Acetate- and Methanol-Grown Cells // Journal of Proteome Research. 2005. Vol. 4, № 1. P. 129–135.
273. Li L. et al. Quantitative Proteomic and Microarray Analysis of the Archaeon *Methanosarcina acetivorans* Grown with Acetate versus Methanol // Journal of Proteome Research. 2007. Vol. 6, № 2. P. 759–771.
274. Zhang W. et al. DNA microarray analysis of anaerobic *Methanosarcina barkeri* reveals responses to heat shock and air exposure // Journal of Industrial Microbiology & Biotechnology. 2006. Vol. 33, № 9. P. 784–790.
275. Cohan F.M., Koeppel A.F. The origins of ecological diversity in prokaryotes // Curr. Biol. 2008. Vol. 18, № 21. P. R1024-1034.
276. Datta N., Kontomichalou P. Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae // Nature. 1965. Vol. 208, № 5007. P. 239–241.
277. Lang A.S., Beatty J.T. Importance of widespread gene transfer agent genes in  $\alpha$ -proteobacteria // Trends in microbiology. 2007. Vol. 15, № 2. P. 54–62.
278. McDaniel L.D. et al. High Frequency of Horizontal Gene Transfer in the Oceans // Science. 2010. Vol. 330, № 6000. P. 50–50.
279. Crisp A. et al. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes // Genome Biology. 2015. Vol. 16, № 1.
280. Mongodin E.F. et al. Gene Transfer and Genome Plasticity in *Thermotoga maritima*, a Model Hyperthermophilic Species // Journal of Bacteriology. 2005. Vol. 187, № 14. P. 4935–4944.
281. Schliep K. et al. Harvesting Evolutionary Signals in a Forest of Prokaryotic Gene Trees // Molecular Biology and Evolution. 2010. Vol. 28, № 4. P. 1393–1405.
282. Kimelman A. et al. A vast collection of microbial genes that are toxic to bacteria // Genome Res. 2012. Vol. 22, № 4. P. 802–809.
283. Puigbò P., Wolf Y.I., Koonin E.V. Seeing the Tree of Life behind the phylogenetic forest // BMC biology. 2013. Vol. 11, № 1. P. 46.
284. Maslov S. et al. Toolbox model of evolution of prokaryotic metabolic networks and their regulation // Proceedings of the National Academy of Sciences. 2009. Vol. 106, № 24. P. 9743–9748.

285. Rautian M.S., Wackerow-Kouzova N.D. Phylogenetic placement of two previously described intranuclear bacteria from the ciliate *Paramecium bursaria* (Protozoa, Ciliophora): “*Holospora acuminata*” and “*Holospora curviuscula*” // *Int. J. Syst. Evol. Microbiol.* 2013. Vol. 63, № Pt 5. P. 1930–1933.
286. Leggett R.M. et al. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries // *Bioinformatics.* 2014. Vol. 30, № 4. P. 566–568.
287. Rinke C. et al. Insights into the phylogeny and coding potential of microbial dark matter // *Nature.* 2013. Vol. 499, № 7459. P. 431–437.
288. Mistry J. et al. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions // *Nucleic Acids Res.* 2013. Vol. 41, № 12. P. e121–e121.
289. Seemann T. Prokka: rapid prokaryotic genome annotation // *Bioinformatics.* 2014. Vol. 30, № 14. P. 2068–2069.
290. Overbeek R. et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST) // *Nucleic Acids Res.* 2014. Vol. 42, № Database issue. P. D206–D214.
291. Caspi R. et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases // *Nucleic Acids Res.* 2016. Vol. 44, № D1. P. D471–D480.
292. Wang G., Li X., Wang Z. APD3: the antimicrobial peptide database as a tool for research and education // *Nucleic Acids Res.* 2016. Vol. 44, № Database issue. P. D1087–D1093.
293. van Heel A.J. et al. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides // *Nucleic Acids Res.* 2013. Vol. 41, № Web Server issue. P. W448–W453.
294. Weber T. et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters // *Nucl. Acids Res.* 2015. Vol. 43, № W1. P. W237–W243.
295. Saier M.H. et al. The Transporter Classification Database (TCDB): recent advances // *Nucleic Acids Res.* 2016. Vol. 44, № D1. P. D372–D379.
296. Bendtsen J.D. et al. Non-classical protein secretion in bacteria // *BMC Microbiol.* 2005. Vol. 5. P. 58.
297. Nielsen H. Predicting Secretory Proteins with SignalP // *Methods Mol. Biol.* 2017. Vol. 1611. P. 59–73.
298. Bendtsen J.D. et al. Prediction of twin-arginine signal peptides // *BMC Bioinformatics.* 2005. Vol. 6. P. 167.
299. Zhou Y. et al. PHAST: A Fast Phage Search Tool // *Nucleic Acids Res.* 2011. Vol. 39, №

suppl\_2. P. W347–W352.

300. Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2 // Nat Meth. 2012. Vol. 9, № 4. P. 357–359.
301. Darling A.E., Mau B., Perna N.T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement // PLoS ONE. 2010. Vol. 5, № 6. P. e11147.
302. Li L., Stoeckert C.J., Roos D.S. OrthoMCL: identification of ortholog groups for eukaryotic genomes // Genome Res. 2003. Vol. 13, № 9. P. 2178–2189.
303. Janda J.M., Abbott S.L. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls // J. Clin. Microbiol. 2007. Vol. 45, № 9. P. 2761–2764.
304. Gillespie J.J. et al. Secretome of obligate intracellular *Rickettsia* // FEMS Microbiol Rev. 2015. Vol. 39, № 1. P. 47–80.
305. Gillespie J.J. et al. The *Rickettsia* type IV secretion system: unrealized complexity mired by gene family expansion // Pathog Dis. 2016. Vol. 74, № 6.
306. Rennoll-Bankert K.E. et al. Which Way In? The RalF Arf-GEF Orchestrates *Rickettsia* Host Cell Invasion // PLOS Pathogens. 2015. Vol. 11, № 8. P. e1005115.
307. Palmer G.H., Noh S.M. Rickettsial Entry into Host Cells: Finding the Keys To Unlock the Doors // Infect. Immun. 2012. Vol. 80, № 11. P. 3746–3747.
308. Storz G., Wolf Y.I., Ramamurthi K.S. Small Proteins Can No Longer Be Ignored // Annu Rev Biochem. 2014. Vol. 83. P. 753–777.
309. Wang F. et al. A Systematic Survey of Mini-Proteins in Bacteria and Archaea // PLOS ONE. 2008. Vol. 3, № 12. P. e4027.
310. Freiburg M. Isolation and characterization of macronuclei of *Paramecium caudatum* infected with the macronucleus-specific bacterium *Holospora obtusa* // Journal of Cell Science. 1985. Vol. 73, № 1. P. 389–398.
311. Knockenhauer K.E., Schwartz T.U. The Nuclear Pore Complex as a Flexible and Dynamic Gate // Cell. 2016. Vol. 164, № 6. P. 1162–1171.
312. Traut T.W. Physiological concentrations of purines and pyrimidines // Mol Cell Biochem. 1994. Vol. 140, № 1. P. 1–22.
313. Hori M., Fujii K., Fujishima M. Micronucleus-Specific Bacterium *Holospora elegans* Irreversibly Enhances Stress Gene Expression of the Host *Paramecium caudatum* // Journal of Eukaryotic Microbiology. 2008. Vol. 55, № 6. P. 515–521.
314. Chen I.-M.A. et al. IMG/M: integrated genome and metagenome comparative data analysis system // Nucleic Acids Res. 2017. Vol. 45, № Database issue. P. D507–D516.

315. Izutsu K. et al. Escherichia coli ribosome-associated protein SRA, whose copy number increases during stationary phase // J. Bacteriol. 2001. Vol. 183, № 9. P. 2765–2773.
316. Wimberly B.T. et al. Structure of the 30S ribosomal subunit // Nature. 2000. Vol. 407, № 6802. P. 327–339.
317. Choli T. et al. Isolation and Characterization of a New Ribosomal Protein from the Thermophilic Eubacteria, *Thermus thermophilus*, *T. aquaticus* and *T. flavus* // Biological Chemistry Hoppe-Seyler. 1993. Vol. 374, № 1–6. P. 377–384.
318. Akanuma G. et al. Defect in the Formation of 70S Ribosomes Caused by Lack of Ribosomal Protein L34 Can Be Suppressed by Magnesium // J. Bacteriol. 2014. Vol. 196, № 22. P. 3820–3830.
319. Ikegami A. et al. Disruption of rpmJ Encoding Ribosomal Protein L36 Decreases the Expression of secY Upstream of the spc Operon and Inhibits Protein Translocation in Escherichia coli // Bioscience, Biotechnology, and Biochemistry. 2005. Vol. 69, № 8. P. 1595–1602.
320. Krylov D.M. et al. Gene Loss, Protein Sequence Divergence, Gene Dispensability, Expression Level, and Interactivity Are Correlated in Eukaryotic Evolution // Genome Res. 2003. Vol. 13, № 10. P. 2229–2235.
321. Hoffmann A., Bukau B., Kramer G. Structure and function of the molecular chaperone Trigger Factor // Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2010. Vol. 1803, № 6. P. 650–661.
322. Ferbitz L. et al. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins // Nature. 2004. Vol. 431, № 7008. P. 590–596.
323. Dunkle J.A. et al. Structures of the *Escherichia coli* ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action // PNAS. 2010. Vol. 107, № 40. P. 17152–17157.
324. Chavatte L., Brown B.A., Driscoll D.M. Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes // Nature Structural & Molecular Biology. 2005. Vol. 12, № 5. P. 408–416.
325. Anikaev A.Y. et al. Role of protein L25 and its contact with protein L16 in maintaining the active state of *Escherichia coli* ribosomes *in vivo* // Biochemistry Moscow. 2016. Vol. 81, № 1. P. 19–27.
326. Duin J., Wijnands R. The Function of Ribosomal Protein S21 in Protein Synthesis // European Journal of Biochemistry. 1981. Vol. 118, № 3. P. 615–619.