

ОСНОВЫ ПЛАНИРОВАНИЯ НАУЧНОГО ЭКСПЕРИМЕНТА ПРИ ПРОВЕДЕНИИ СИКВЕНС-АНАЛИЗА ПРОКАРИОТИЧЕСКИХ ГЕНОМОВ И БАКТЕРИАЛЬНЫХ СООБЩЕСТВ

С. А. Костюк, А. К. Лямцева, О. С. Полуян

Белорусский государственный медицинский университет, г. Минск, Республика Беларусь

Введение. Любой научный эксперимент требует тщательной предварительной подготовки, при этом в медицинской биологии планирование экспериментальной части исследования имеет большое значение по причине широкой вариабельности свойств, характерной для биологических объектов. В секвенировании прокариотических геномов в настоящее время можно выделить два основных направления: секвенирование отдельных микробов, в том числе некультивируемых и секвенирование микробных биоценозов для изучения структуры сообщества.

Цель исследования. Рассмотреть методологию планирования научного эксперимента при проведении секвенирования прокариотических геномов и бактериальных сообществ.

Материалы и методы. Для поиска источников литературы использовали платформы Web of Science и Scopus. Осуществлен анализ публикаций с описанием оригинальных исследований и научных обзоров.

Результаты. Планирование научного эксперимента – процедура выбора количества и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью. Описаны источники ошибок на этапе пробоподготовки, варианты применения высокопроизводительного секвенирования.

Заключение. Метагеномика, метатранскриптомика и метапротеомика позволяют обнаружить ранее неизвестные некультивируемые микроорганизмы, описать их свойства и функции, обнаружить еще не изученные биомолекулы, а также охарактеризовать обитающие на теле человека микробные сообщества для здоровых людей и пациентов с различными заболеваниями.

Ключевые слова. Научный эксперимент, секвенирование, прокариоты, геном, транскриптом.

Для цитирования: Костюк, С. А. Основы планирования научного эксперимента при проведении сиквенса-анализа прокариотических геномов и бактериальных сообществ / С. А. Костюк, А. К. Лямцева, О. С. Полуян // Биохимия и молекулярная биология. – 2025. – Т. 4, № 1(6). – С. 78–87.

Введение

Появление сиквенса-анализа дало значительный толчок развитию микробиологии и смежным областям. Наблюдающийся в течение последних пяти лет быстрый рост количества практически полных последовательностей геномов бактерий в открытых базах данных позволил по-новому взглянуть на эволюцию прокариот, обнаружить принципиально новые типы генетических событий на уровне возникновения видов живых организмов, с тем чтобы эффективно восстанавливать метаболизм одноклеточных. Для некультивируемых микроорганизмов, например, микроорганизмов, живущих в экстремальных условиях среды, геномные методы становятся важнейшим способом их изучения и отправной точкой для последующей разработки биохимических методов и методов геномной инженерии [1].

Другая область применения секвенирования в микробиологии это медицина. Исследование генома отдельных микробов методами сиквенса-ана-

лиза позволяет быстро выявлять «островки патогенности» и локусы, определяющие устойчивость к антибиотикам, а также изучить микробных сообществ человека, понять взаимосвязь микробиоценозов и состояния здоровья пациента.

Третье важное направление применения секвенирования – биотехнология, применяя бактерии в различных промышленных процессах, на основе геномных данных можно идентифицировать новые ферменты и понять, как модификации генома связаны с функционированием микроорганизмов, что позволяет развиваться биологическим технологиям.

Цель научного обзора – рассмотреть методологию планирования научного эксперимента при проведении секвенирования прокариотических геномов и бактериальных сообществ.

Материалы и методы

Для поиска источников литературы использовали системы Web of Science и Scopus. Осущест-

влен анализ публикаций с описанием оригинальных исследований и научных обзоров.

Результаты и их обсуждение

Сиквенс-анализ имеет значение для молекулярной биологии в целом, в части понимания основных механизмов функционирования живой клетки. На начало 2020 года определено около 1500 полных геномов микроорганизмов, а число частично собранных геномов, представленных в виде множества неупорядоченных фрагментов, составляет уже около десяти тысяч и быстро растет [2].

Секвенирование бактериального генома методом Сенгера производили, создавая сначала библиотеку длинных фрагментов (30–40 т.п.н.) в векторах на основе геномов фагов с получением картированной коллекции фрагментов, покрывающих весь геном бактерии, затем каждый крупный фрагмент субклонировали в плаزمиды фрагментами по 2–3 т.п.н. и секвенировали их насквозь, часто используя не только стандартные праймеры на вектор, но и множество геном-специфичных праймеров [3]. Позднее, с уменьшением стоимости секвенирования за счет выпуска более производительных секвенаторов, но работающих по-прежнему на принципе Сенгера, исследователи стали использовать более простой подход: метод дробовика, когда геном сразу клонируют в плазмидный вектор вставками по 2–3 т.п.н. и секвенируют с многократным покрытием генома по первоначальному данным.

В настоящее время методы высокопроизводительного секвенирования полностью вытеснили принцип Сенгера из области геномного сиквенс-анализа. Однако современный подход очень похож на метод дробовика: получают последовательность коротких фрагментов генома, добываясь сборки 95–98% последовательности генома [4].

Любой эксперимент требует тщательной предварительной подготовки. В медицинской биологии планирование экспериментальной части исследования имеет большое значение по причине широкой вариабельности свойств, характерной для биологических объектов. Эта особенность является основной причиной трудностей при интерпретации результатов, которые могут значительно различаться от опыта к опыту. Перед началом работы исследователь обязан перечислить для себя все элементы исследования, способные исказить результаты, приложив максимальные усилия к уменьшению или устранению таких искажений.

Цель планирования научного эксперимента заключается в создании схемы, позволяющей получить наибольший объем информации при наименьших затратах, а именно выбор количества

и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью.

Среди классических характеристик экспериментального плана выделяют: сравнение (как правило, в научном эксперименте объект сравнивается либо с неким заранее заданным стандартом, либо с контрольным объектом), рандомизацию, повторяемость, однородность и блокировку [5].

Рандомизация научного эксперимента представляет собой процесс, используемый для группировки объектов таким образом, чтобы у каждого из них была равная вероятность попасть в контрольную или опытную группу. Другими словами, выбор участников исследования должен происходить случайно, чтобы исследование не было отклонено в сторону «предпочтительного» для исследователя результата. Рандомизация необходима для применимости статистических методов для анализа данных исследования. Она помогает предотвратить смещения, обусловленные причинами, которые не были непосредственно учтены в плане научного эксперимента. Для этого, например, формирование экспериментальных групп лабораторных животных производится случайным образом [6].

Рассмотрим важность рандомизации на конкретном примере. Допустим, вы заинтересованы в поиске различий между двумя образцами ДНК. Для повышения надежности были собраны по шесть повторностей для каждого типа образцов – шесть контрольных и шесть экспериментальных. Предположим, вы секвенировали, используя две дорожки, причем каждая дорожка содержит шесть мультиплексированных повторностей. Неверным является подход с нанесением одинаковых образцов в одну дорожку сиквенсного чипа. При планировании такого эксперимента нанесение обоих типов образцов вперемешку позволяет исключить систематическую ошибку, эффект дорожки, и получить данные даже в случае неудачного прочтения одной из дорожек.

Например, поиск дифференциально представленных транскриптов методом секвенирования транскриптов контрольной и экспериментальной групп образцов. Если одна из дорожек по какой-либо причине даст меньше прочтений, при плохом дизайне можно принять аппаратно возникшую разницу за биологический сигнал. Гораздо надежнее в приведенном примере нанести в каждую из дорожек равное количество образцов из контрольной и экспериментальной групп. Выбор образцов из каждой группы для нанесения в одну дорожку лучше всего производить случайным образом [7, 8].

Для выявления источника вариабельности необходимо провести несколько испытаний повторов. Использование повторов в секвенировании не очень распространено по двум причинам: из-за высокой стоимости исследования и того факта, что один из типов экспериментальных повторов уже встроен в сам алгоритм секвенирования – это число прочтений каждого участка последовательности (степень покрытия). Однако другие типы повторов, такие как биологические повторности, инструментальные повторности и повторности для разных технологий высокопроизводительного секвенирования могут быть крайне информативны и важны для корректной постановки эксперимента [9].

Биологические повторности, когда встречаются по несколько однотипных биологических образцов, являются важной составляющей любого биологического эксперимента, позволяя снизить экспериментальную ошибку – стохастические ошибки секвенирования.

Благодаря возможности кодирования образцов присоединением адаптеров заданной последовательности, исследователь может довольно легко решить задачу инструментальных повторов – секвенирование одного и того же образца несколько раз. Использование разных платформ для секвенирования, особенно основанных на разных принципах секвенирования, также может существенно улучшить результаты. Например, комбинация длинных, но не очень точных прочтений на платформе PacBio и точных, но коротких прочтений на Illumina [9, 10].

С повторениями связана чувствительность и специфичность эксперимента. В частности, повторности могут быть использованы для определения чувствительности и специфичности методов поиска полиморфизма последовательностей ДНК [11].

На этапе планирования научного эксперимента исследователю следует обратить особое внимание на перечисленные ниже источники ошибок во избежание их совершения.

Источники ошибок на этапе пробоподготовки:

1) ошибки лаборанта (например, неправильная маркировка образцов);

2) деградация ДНК и РНК в ходе хранения и транспортировки образцов;

3) загрязнение (контаминация) образца чужеродными (не относящимися к исследованию) нуклеиновыми кислотами;

4) недостаточное количество (или низкая концентрация) образца ДНК.

Источники ошибок на этапе подготовки библиотеки для NGS:

1) ошибки исследователя, например, перекрестное загрязнение (кросс-контаминация) образцов;

2) искажения в ходе ПЦР (разная эффективность амплификации фрагментов, низкая начальная концентрация биологического образца и большое число циклов ПЦР);

3) искажения в ходе обогащения поли-А-фракции РНК;

4) аппаратные ошибки (например, некорректная работа ПЦР-амплификатора);

5) образование химерных прочтений.

Источники ошибок на этапе, собственно, секвенирования:

1) ошибки исследователя (например, перекрестные помехи в получении сигнала от разных микросфер в результате перегрузки чипа);

2) сбой фазы считывания (за счет неполного удлинения или добавления нескольких нуклеотидов вместо одного нуклеотида);

3) сложные для прочтения регионы (GC-богатые или гомополимерные участки);

4) аппаратные ошибки (например, выключение электропитания, выход из строя лазера, помп и приводов, жесткого диска, сбой программного обеспечения).

За последние несколько лет наблюдается большое число публикаций с применением высокопроизводительного секвенирования для различных целей. Наиболее популярные приложения NGS включают в себя: полногеномное секвенирование (*de novo* или повторное); исследование различных РНК (часто называемое RNA-Seq); крупномасштабный анализ метилирования ДНК; изучение ДНК-белковых взаимодействий (ChIP-seq) и ряд других [3, 12].

Полногеномное секвенирование *de novo* применяется для реконструкции работы клетки и организма на молекулярном уровне, решение вопросов эволюционной геномики, поиска генетических вариаций. Метагеномное секвенирование следует использовать для исследования биоценоза, поиска новых видов живых систем; секвенирование транскриптомов – исследования генной экспрессии, аннотации генома; таргетное секвенирование – поиска генетических вариаций; секвенирование обработанной бисульфитом ДНК – исследование профиля метилирования; секвенирование единичных клеток – исследование генной экспрессии, секвенирование некультивируемых бактерий [4, 13, 14].

В ближайшие годы список приложений NGS, несомненно, будет расти благодаря непрерывному усовершенствованию существующих подходов и выходу на рынок принципиально новых технологий, в частности, технологий секвенирования одиночных молекул нуклеиновых кислот [15].

Анализ геномных данных, несомненно, сложнее процесса их получения, это процесс можно

разделить на две составляющие: аннотацию генома или идентификацию генов в последовательности и предсказание их функций, а также биологический анализ, т.е. реконструкцию метаболизма, установление взаимосвязи фенотипа и генотипа и т. д.

Задача аннотирования прокариотических геномов на сегодняшний день имеет ряд неплохих решений, при этом продолжающиеся биоинформатические разработки в этой области сулят появление в скором времени еще лучших алгоритмов. Существующие программные продукты позволяют в автоматическом режиме получить достаточно хорошо аннотированный геном [16].

Отдельно следует указать на особенности полногеномного секвенирования *de novo* некультивируемых бактерий. Вследствие малого количества стартовой ДНК ее, как правило, подвергают амплификации по технологии WGA. Наличие этапа амплификации сильно усложняет задачу последующего анализа данных, поскольку покрытие амплифицированного генома оказывается очень неравномерным вследствие искажений амплификации, при использовании инвертированных библиотек с двусторонним прочтением длину вставки гораздо сложнее контролировать. Следует также учитывать, что в процессе амплификации возникают ошибки и химерные прочтения. Существуют специальные программы для сборки таких амплифицированных геномов [17, 18]. Важным является биологический этап, в рамках которого формируется понимание того, как продукты генов работают в системе, как микроорганизм использует имеющийся набор генов в тех или иных условиях. Этот этап пока не подлежит автоматизации и зависит от квалификации и эрудиции исследователя.

Известно, что бактериальный транскриптом по аналогии с транскриптомом эукариотическим это совокупность всех транскриптов данной бактериальной клетки. В среднем на рибосомальную РНК у бактерий приходится 80% от суммарной РНК по массе, на транспортную РНК 15%, на матричную и некодирующие РНК – 4-5% [19].

Сложность бактериального транскриптома долгое время была недооценена, лишь с созданием технологий высокопроизводительного секвенирования удалось доказать, что у бактерий есть практически все те же механизмы регуляции экспрессии, как и у эукариотов. Так, у бактерий были найдены транскодируемые малые РНК. Это участки, расположенные в пространстве между белок-кодирующими генами и способные взаимодействовать с другими РНК либо белками. Найдены у бактерий и цис-кодируемые малые РНК. Они могут занимать 1-2 гена по длине или же иметь протяжен-

ность в несколько генов [20]. Найдены рибопереключатели структуры, чаще всего сенсорные, расположенные в 5'-концевой нетранслируемой области РНК, способные изменять конфигурацию транскрибируемой молекулы и тем самым либо препятствовать, либо помогать экспрессии. У бактерий определены безлидерные РНК или РНК без транслируемой области, в связи с чем меняется схема сборки рибосом на таких транскриптах.

У прокариот обнаружен широкий спектр эпигенетических модификаций, влияющих на спектр транскриптов в транскриптоме. Найдена компартиментализация транскриптов, а именно было доказано, что транскрипты белков, которые в дальнейшем будут взаимодействовать, обычно транскрибируются рядом. Так как у бактерий транскрипция и трансляция сопряжены, время сборки необходимых белковых комплексов у бактерий сокращено. Описаны различные модификации бактериального сплайсинга, полиаденилирования, редактирования РНК и т.д. [19].

Упрощенная схема исследования бактериального транскриптома следующая: выделение РНК из интересующей бактерии, синтез кДНК, создание библиотеки фрагментов и, собственно, секвенирование. На каждой стадии возможны дополнительные манипуляции [21]. Например, можно провести обогащение РНК интересующей последовательностью, отобрать фракцию только малых РНК либо рибосомальных РНК. В конечном итоге после секвенирования исследователь получает набор файлов.

Базовая схема анализа данных секвенирования бактериального транскриптома стандартна: вначале проводят контроль качества полученных файлов, затем осуществляют фильтрацию прочтений и удаление адаптерных последовательностей. После этого выполняют картирование прочтений на референсный геном, при этом могут быть использованы программы Bowtie, BWA, SOAP и др. Файлы, содержащие картированные последовательности, отбирают и проводят качественный либо количественный анализ данных. В ходе качественного анализа транскриптома выполняют визуализацию распределения прочтений по геному бактерии. При количественном анализе определяют функциональные категории или метаболические пути, в которых могут быть представлены найденные гены [22].

В результате качественного анализа бактериальных транскриптов можно найти новые гены как короткие кодирующие белки, так и гены новых некодирующих РНК. В ходе анализа уточняют границы генов, выявляют некодирующие и нетранслируемые участки. Так как у бактерий большинство генов собрано в опероны, можно выявлять

опероны – области покрытия генома прочтениями, включающие сразу несколько генов. Также можно определять антисмысловую РНК – покрытие прочтениями генома по цепи, противоположной смысловой цепи известного гена. Для многих бактерий антисмысловая транскрипция может составлять существенный процент транскриптома. В ряде случаев он может превышать объем транскрипции по смысловой цепи. За счет сопоставления количества прочтений в двух и более библиотеках можно проводить поиск дифференциально представленных транскриптов [23, 24].

Важно отметить, что в ряде биологических задач исследование транскриптома или протеома может быть крайне информативным. Так, было показано, что бактерии *Helicobacter pylori* из разных участков одного и того же пораженного желудка не отличаются генетически, но сильно отличаются по уровню экспрессии генов, т.е. бактерии находятся в разных функциональных состояниях. При этом профиль экспрессии может значительно отличаться до сотни генов с существенно разным уровнем экспрессии. Ряд работ демонстрирует, что даже в культуре каждая бактерия уникальна по профилю экспрессии генов. Анализируя транскриптом в популяции бактерий, исследователь измеряет «среднюю температуру по больнице». Поэтому в перспективе исследование бактериальных транскриптомов методами высокоэффективного секвенирования должно стремиться к возможности определения профиля представленности транскриптов одной клетки [21].

К настоящему времени описано около 10 000 видов бактерий, однако некоторые исследователи полагают, что общее число видов прокариот на Земле достигает миллиона, при этом значительная их часть не культивируется. Истинное разнообразие некультивируемых микроорганизмов стало понятно после появления молекулярно-генетических методов исследования бактериальных сообществ [25].

При исследовании микроорганизмов из различных сред, в том числе живущих в экстремальных условиях, геномные методы стали важнейшим способом их изучения и отправной точкой для последующей разработки биохимических методов и методов геномной инженерии. В этом смысле секвенирование и особенно высокопроизводительное секвенирование стало новым этапом в развитии микробиологии как науки.

Для исследования сообществ организмов, обитающих в различных средах, в последние годы все чаще обращаются к «мета»-исследованиям, т.е. секвенированию образцов, получаемых напрямую из окружающей среды. Это могут быть метагеномные, метатранскриптомные и даже метапротеом-

ные данные [26]. Такие исследования позволяют обнаружить ранее неизвестные некультивируемые микроорганизмы, описать их свойства и функции, обнаружить еще не изученные биомолекулы, а также получить более общую картину того, что происходит в той или иной среде.

Основными результатами секвенирования являются таксономический анализ сообщества либо реконструкция типа сообщества, например, по результатам анализа метагенома можно установить, что сообщество хемолитотрофно и доминирующим процессом в нем является сульфаторедукция [27].

Методы высокопроизводительного секвенирования хорошо подходят для исследования бактериальных сообществ, причем как для определения видового разнообразия и количественного состава сообщества т.е. микробиоценоза, так и для секвенирования полных геномов присутствующих в нем некультивируемых организмов. Учитывая, что для прокариотов характерен горизонтальный перенос генов, зачастую для понимания особенностей микробиоценоза важно не только наличие гена, например, устойчивости к антибиотикам у конкретного вида, но и его присутствие в сообществе в целом [28].

Для исследования видового состава прокариот в какой-либо среде обитания можно использовать метод, основанный на секвенировании высококонсервативных регионов генома. Чаще всего для этих целей выбирают ген 16S рибосомальной РНК (16S рРНК). У гена 16S рРНК длиной около 1500 п.н. есть константные и переменные участки. Константные участки позволяют создать «универсальные» праймеры для ПЦР, взаимодействующие с ДНК подавляющего большинства прокариот с примерно равной эффективностью [29]. Методом ПЦР можно наработать расположенные между такими праймерами переменные участки гена 16S рРНК. Секвенирование одного или нескольких переменных регионов позволяет идентифицировать последовательности, принадлежащие разным видам.

До появления технологий высокопроизводительного секвенирования полученные продукты ПЦР клонировали в плазмидные векторы (в *Escherichia coli*), после чего каждый клон секвенировали с помощью метода Сенгера. Использование сенгеровского секвенирования позволяет получить последовательности высокого качества, почти полностью покрывающие ген 16S рРНК, что существенно важно для точной идентификации и различения близких видов. Однако подход с клонированием достаточно дорог, также могут возникнуть дополнительные искажения в конечный результат [30].

Появление технологий высокопроизводительного секвенирования с длиной прочтения в несколько сотен нуклеотидов позволило использовать его для анализа микробных сообществ по гену 16S рРНК без этапа клонирования, что ускоряет и удешевляет исследование.

Недостатком методов высокоэффективного секвенирования в данном применении является меньшая, чем в методе Сенгера, длина прочтения, поскольку секвенирование клонов методом Сенгера с двух сторон дает до 1400 п.н. единой последовательности. По короткому участку гена 16S рРНК в 200–400 п.н. можно определить бактерию лишь до рода. Кроме того, вероятность ошибок для технологий высокоэффективного секвенирования гораздо выше, чем для сенгеровского секвенирования, таким образом, возникает проблема фильтрации ошибок секвенирования. Существуют работы, показывающие, что в ряде метагеномных проектов большое видовое разнообразие сообщества в результате оказалось следствием ошибок секвенирования [31].

Крайне важным является подбор пары праймеров для амплификации регионов гена 16S рРНК. Даже в константных участках этого гена между последовательностями некоторых видов могут встречаться однонуклеотидные вариации, в результате чего эффективность амплификации ДНК разных видов может в значительной мере различаться. Этой проблеме посвящен ряд исследований, предлагающих различные варианты праймеров для получения ампликонов разной длины [32, 33]. В любом случае, при выборе пары праймеров для ПЦР стоит проверить, будут ли с ее помощью амплифицироваться последовательности гена 16S рРНК тех видов, которые ожидаются для данного микробиома, и какие виды при этом могут быть утеряны [34, 35].

Несомненным плюсом использования в качестве маркера гена 16S рРНК при анализе симбиотического микробиоценоза является исключение из анализируемого образца примеси ДНК организма-хозяина за счет наличия этапа амплификации.

Помимо разной эффективности амплификации, при попытке количественно оценить состав микробиоты возникает еще одна сложность: копияемость гена 16S рРНК может варьировать от 2 до 15 копий на гаплоидный геном [36]. Причем она может отличаться даже для микроорганизмов, относящихся к одному виду. Оценка представленности различных видов с учетом числа копий гена 16S рРНК или с помощью референсных генов, копияемость которых постоянна, может улучшить точность получаемых результатов [37, 38].

Для биоинформатической обработки данных высокопроизводительного секвенирования ампликонов используют два подхода [39]:

1. Сравнение полученных последовательностей с различными базами данных. В частности, для гена 16S рРНК можно использовать базы данных: SILVA, Ribosomal Database, NCBI. Базы данных имеют встроенные инструменты поиска. Отдельные базы предоставляют специальные инструменты для анализа данных высокопроизводительного секвенирования, например, SINA. Минусом сравнения с базами данных являются проблемы при аннотации последовательностей, принадлежащих новым или плохо описанным видам некультивируемых бактерий.

2. Сравнение полученных последовательностей между собой, кластеризация и выявление «операционных таксономических единиц» – таксон, аналогичный виду и определяемый на основании степени сходства последовательностей ДНК. Этот подход не опирается на уже известные последовательности и поэтому позволяет обнаруживать неизвестные виды микроорганизмов. Разработан ряд простых в применении инструментов, использующих этот принцип: QIIME, mothur, CLUSTOM.

Путем секвенирования ампликонов можно не только определить таксономический состав микробного сообщества, но и провести количественную оценку содержания того или иного микроорганизма в образце.

Помимо использования для анализа микробного сообщества отдельных консервативных регионов генома таких, как ген 16S рРНК или других филогенетических маркеров, можно напрямую выделить и секвенировать содержащуюся в образце ДНК. Важным преимуществом такого подхода является отсутствие этапа амплификации, вносящего значительные искажения в результаты.

Впервые термин «метагеном» был предложен Хандельсманом с соавт. в 1998 году [40]. Он обозначает суммарный набор генов, выявленных в исследуемом образце, как геном одного псевдоорганизма, используемый для воссоздания свойств исследуемого микробного сообщества.

Первый эксперимент по метагеномному секвенированию был проведен в 2004 году [41] и позволил обнаружить 148 ранее неизвестных видов бактерий и более 1,2 млн ранее неизвестных генов. Первые работы проводили с помощью клонирования фрагментов ДНК в плазмидных векторах в клетках бактерии кишечной палочки и последующего секвенирования методом Сенгера. Длинные прочтения в значительной степени упрощали сборку и анализ данных, но из-за клонирования

возникали искажения, поскольку некоторые последовательности плохо поддаются клонированию, например, если участок бактериальной ДНК несет ген, токсичный для *E. coli*, то такой фрагмент может быть утерян в ходе клонирования. Эту проблему можно исключить, используя несколько разных организмов для клонирования [42].

Технологии высокопроизводительного секвенирования позволяют избежать клонирования и, соответственно, получить намного больше данных с меньшими финансовыми затратами. Для метагеномного секвенирования используют практически все доступные на сегодняшний день технологии: 454 Life Sciences, Illumina, Ion Torrent. В некоторых случаях такое секвенирование позволяет получить полную последовательность генома для некультивируемых бактерий [43].

Оптимальным представляется комбинированный алгоритм, сочетающий в себе как чтение коротких переменных участков, так и метагеномное секвенирование. Последовательность шагов в таком алгоритме может быть следующей:

1) секвенирование короткого переменного участка (чаще всего это фрагмент или несколько фрагментов 16S рРНК);

2) секвенирование метагенома;

3) анализ прочтений 16S рРНК в метагеноме и их классификация;

4) удаление из результатов всех не классифицируемых данных;

5) анализ сообщества по кратности прочтения контигов с 16S рРНК (чем больше прочтений 16S рРНК, тем больше данного микроба содержится в сообществе);

6) сопоставление результатов двух подходов.

Заключение

Планирование научного эксперимента – процедура выбора количества и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью.

Некоторые биологические задачи предполагают исследование схожих, но территориально разделенных сообществ микроорганизмов. Например, сравнение микробиоты кишечника разных

людей и т.п. После сопоставления видового состава сообщества и процентного соотношения микроорганизмов в ряде задач имеет смысл исследовать генетическое разнообразие отдельных видов бактерий в сообществе, пользуясь нуклеотидным полиморфизмом. Например, для микробиоценозов человека кишечного или урогенитального можно обнаружить, что в разных странах или при разных физиологических состояниях человека бактерии одного вида в сообществе могут быть очень далеки филогенетически.

Биоинформатическое обнаружение гена в последовательности генома далеко не всегда означает, что он работоспособен. Для понимания функциональных возможностей сообщества в дополнение к метагеномным данным целесообразно определение метатранскриптома секвенирование суммарных мРНК сообщества. Исследования метатранскриптомов осложнены нестабильностью молекул РНК и, как следствие, необходимостью быстрой и эффективной фиксации образца. Для экстракции РНК из столь сложных образцов разрабатываются специальные методы.

Поскольку 95% от всех РНК в клетке составляют рибосомальная и транспортная РНК, в процессе исследования их приходится исключать [22]. Чтобы понять, какие гены экспрессируются и в какой степени, необходимо выравнивание полученных данных метатранскриптома на метагеном, аннотация и нормализация. Возможно и выравнивание на референсные геномы из баз данных, но более информативным будет выравнивание и аннотация относительно метагеномных данных для того же образца.

Таким образом, «мета»-исследования сегодня являются одной из быстро развивающихся областей медицины, в том числе благодаря появлению методов высокоэффективного секвенирования. Метагеномика вместе с метатранскриптомикой и метапротеомикой позволяют понять состав и функционирование микробных сообществ, в том числе и симбиотических человеку, охарактеризовать обитающие на теле человека микробные сообщества для здоровых людей и пациентов с различными заболеваниями.

Список использованных источников

1. Lasken, R. S. Recent advances in genomic DNA sequencing of microbial species from single cells / R. S. Lasken, J. S. McLean // *Nat Rev Genet.* – 2014. – Vol. 15, №9. – P.577–584. DOI: 10.1038/nrg3785.
2. Popa, O. From sequence to information / O. Popa, E. Oldenburg, O. Ebenhöf // *Philos Trans R Soc Lond B Biol Sci.* – 2020. – Vol. 375, № 1814. – P. 20190448. DOI: 10.1098/rstb.2019.0448.
3. Long walk to genomics: History and current approaches to genome sequencing and assembly / A. M. Giani, G.R. Gallo, L. Gianfranceschi, G. Formenti // *Comput Struct Biotechnol J.* – 2019. – Vol. 18. – P. 9–19. DOI: 10.1016/j.csbj.2019.11.002.
4. Heather, J. M. The sequence of sequencers: The history of sequencing DNA / J. M. Heather, B. Chain // *Genomics.* – 2016. – Vol. 107, № 1. – P. 1–8. DOI: 10.1016/j.ygeno.2015.11.003.
5. Casler, M. D. Fundamentals of Experimental Design: Guidelines for Designing Successful Experiments /

- M. D. Casler // *Agronomy Journal*. – 2015. – Vol. 107. – P. 692–705. DOI:10.2134/agronj2013.0114.
6. Tanious, R. Randomized Single-Case Experimental Designs in Healthcare Research: What, Why, and How? / R. Tanious, P. Ongheña // *Healthcare (Basel)*. 2019. – Vol. 7, №4. – P. 143. DOI: 10.3390/healthcare7040143.
 7. Auer, P. L. Statistical design and analysis of RNA sequencing data / P. L. Auer, R. W. Doerge // *Genetics*. – 2010. – Vol. 185, №2. – P. 405–416. DOI: 10.1534/genetics.110.114983.
 8. Churchill, G. A. Fundamentals of experimental design for cDNA mi-croarrays / G. A. Churchill // *Nat Genet*. – 2002. – Vol. 32. – P. 490–495. DOI: 10.1038/ng1031
 9. Robasky, K. The role of replicates for error mitigation in next-generation sequencing / K. Robasky, N. E. Lewis, G. M. Church // *Nat Rev Genet*. – 2014. – Vol. 15, № 1. – P. 56–62. DOI: 10.1038/nrg3655.
 10. Logsdon, G. A. Long-read human genome sequencing and its applications / G. A. Logsdon, M. R. Vollger, E. E. Eichler // *Nat Rev Genet*. – 2020. – Vol. 21, № 10. – P. 597–614. DOI: 10.1038/s41576-020-0236-x.
 11. Foundational Statistical Principles in Medical Research: Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value / T. F. Monaghan, S.N. Rahman, C.W. Agudelo [et al.] // *Medicina (Kaunas)*. – 2021. – Vol. 57, №5. – P. 503. DOI: 10.3390/medicina57050503.
 12. Next-Generation Sequencing Technology: Current Trends and Advancements / H. Satam, K. Joshi, U. Mangrolia [et al.] // *Biology (Basel)*. – 2023. – Vol. 12, №7. – P. 997. DOI: 10.3390/biology12070997.
 13. Mardis, E. R. Next-generation DNA sequencing methods / E. R. Mardis // *Annu Rev Genomics Hum Genet*. – 2008. – Vol. 9. – P. 387–402. DOI: 10.1146/annurev.genom.9.081307.164359.
 14. Advances in single-cell sequencing technology in microbiome research / Y. Wu, J. Zhuang, Y. Song [et al.] // *Genes Dis*. – 2023. – Vol. 11, №4. – P. 101129. DOI: 10.1016/j.gendis.2023.101129.
 16. AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system / K. Bryson, V. Loux, R. Bossy [et al.] // *Nucleic Acids Res*. – 2006. – Vol. 34, № 12. – P. 3533–3545. DOI: 10.1093/nar/gkl471.
 17. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing / A. Bankevich, S. Nurk, D. Antipov [et al.] // *J Comput Biol*. – 2012. – Vol. 19, №5. – P. 455–477. DOI: 10.1089/cmb.2012.0021.
 18. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products / S. Nark, A. Bankevich, D. Antipov [et al.] // *J Comput Biol*. – 2013. – Vol. 20, № 10. – P. 714–737. DOI: 10.1089/cmb.2013.0084.
 19. Bandyra, K. J. Licensing and due process in the turnover of bacterial RNA / K. J. Bandyra, B. F. Luisi // *RNA Biol*. – 2013. – Vol. 10, №4. – P. 627–635. DOI: 10.4161/rna.24393.
 20. Barorli, D. MicroRNA target and gene validation in viruses and bacteria / D. Barorli, P. Arrigo // *Methods Mol Biol*. – 2014. – Vol. 1107. – P. 223–231. DOI: 10.1007/978-1-62703-748-8_13.
 21. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle / Waldbauer J. R., S. Rodrigue, M.L. Coleman, S.W.Chisholm // *PLoS One*. – 2012. – Vol. 7, №8. – P. e43432. DOI: 10.1371/journal.pone.0043432.
 22. Application of metatranscriptomics to soil environments / L. C. Carvalhais, P.G. Dennis, G.W. Tyson, P.M. Schenk // *Journal of micro-biological methods*. – 2012. – Vol. 91, №2. – P. 246–251. DOI: 10.1016/j.mimet.2012.08.011.
 23. Zerbino, D. R. Progress, Challenges, and Surprises in Annotating the Human Genome / D. R. Zerbino, A. Frankish, P. Flicek // *Annu Rev Genomics Hum Genet*. – 2020. – Vol. 21. – P. 55–79. DOI: 10.1146/annurev-genom-121119-083418.
 24. The Signal and the Noise: Characteristics of Antisense RNA in Complex Microbial Communities / T. Y. Michaelsen, J. Brandt, C.M. Singleton [et al.] // *mSystems*. – 2020. – Vol. 5, № 1. – P. e00587-19. DOI: 10.1128/mSystems.00587-19.
 25. Sleator, R.D. Metagenomics / R. D. Sleator, C. Shortall, C. Hill // *Letters in applied microbiology*. – 2008. – Vol. 47, №5. – P. 361–366. DOI: 10.1111/j.1472-765X.2008.02444.x.
 26. Thomas, T. Metagenomics – a guide from sampling to data analysis / T. Thomas, J. Gilbert, F. Meyer // *Microbial informatics and experimentation*. – 2012. – Vol. 2, № 1. – P. 3. DOI: 10.1186/2042-5783-2-3.
 27. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice / C. Simon, J. Herath, S. Rockstroh, R. Daniel // *Applied and environmental microbiology*. – 2009. – Vol. 75, №9. – P. 2964–2968. DOI: 10.1128/AEM.02644-08.
 28. Environmental Whole-Genome Amplification To Access Microbial Populations in Contaminated Sediments Environmental Whole-Genome Amplification To Access Microbial Populations in Contaminated Sediments / C. B. Abulencia, D.L. Wyborski, J.A. Garcia [et al.] // *Appl Environ Microbiol*. – 2006. – Vol. 72. – P. 3291–3301. DOI: 10.1128/AEM.72.5.3291-3301.2006.
 29. Mizrahi Man, O. Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs / O. Mizrahi Man, E. R. Davenport, Y. Gilad // *PLoS one*. – 2013. – Vol. 8, № 1. – P. e53608. DOI: 10.1371/journal.pone.0053608.
 30. Bacterial community shift in treated periodontitis patients revealed by ion torrent 16S rRNA gene amplicon sequencing / S. Jjinemann, K. Prior, R. Szczepanowski [et al.] // *PLoS one*. – 2012. – Vol. 7, №8. – P. e41606. DOI: 10.1371/journal.pone.0041606.
 31. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates / V. Kanin, A. Engelbrekton, H. Ochman, P. Hugenholtz // *Environmental microbiology*. – 2010. – Vol. 12, № 1. – P. 118–123. DOI: 10.1111/j.1462-2920.2009.02051.x.
 32. Nossa, C.W. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome / C. W. Nossa // *World Journal of Gastro-enterology*. – 2010. – Vol. 16, №33. – P. 4135. DOI: 10.3748/wjg.v16.i33.4135.
 33. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies / A. Klindivorth, E. Priesse, T. Schweer [et al.] // *Nucleic acids research*. – 2013. – Vol. 41, № 1. – P. e1. DOI: 10.1093/nar/gks808.
 34. Conlan, S. Species-level analysis of DNA sequence data from the NIH Human Microbiome Project / S. Conlan, H. H. Kong, J. A. Segre // *PLoS one*. – 2012. – Vol. 7, № 10. – P. e47075. DOI: 10.1371/journal.pone.0047075.
 35. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences / D. Soergel, N. Dey, R. Knight, S.E. Brenner [et al.] // *The ISME journal*. 2012. – Vol. 6, № 7. – P. 1440–1444. DOI: 10.1038/ismej.2011.208.
 36. Crosby, L. D. Understanding bias in microbial community analysis techniques due to rrrn operon copy number heterogeneity / L. D. Crosby, C. S. Criddle // *BioTechniques*. – 2003. – Vol. 34, № 4. – P. 790–794, 796, 798. DOI: 10.2144/03344rr01.
 37. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance / S. W. Kembel, M. Wu, J.A. Eisen, J.L. Green // *PLoS computational biology*. – 2012. – Vol. 8, № 10. – P. e1002743. DOI: 10.1371/journal.pcbi.1002743.
 38. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies / R. J. Case, Y. Boucher, I. Dahllöf [et al.] // *Applied and environmental microbiology*. – 2007. – Vol. 73, № 1. – P. 278–288. DOI: 10.1128/AEM.01177-06.

39. Bioinformatics for the Human Microbiome Project / D. Gevers M. Pop, P.D. Schloss, C. Huttenhower // *PLoS computational biology*. – 2012. – Vol. 8, № 11. – P. e1002779. DOI: 10.1371/journal.pcbi.1002779.
40. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products / J. Handelsman, M.R. Rondon, S.F. Brady [et al.] // *Chemistry & biology*. – 1998. – Vol. 5, № 10. – P. 245–249. DOI: 10.1016/s1074-5521(98)90108-9.
41. Environmental genome shotgun sequencing of the Sargasso Sea / J. C. Venter, K. Remington, J.F. Heidelberg [et al.] // *Science*. – 2004. – Vol. 304, № 5667. – P. 66–74. DOI: 10.1126/science.1093857.
42. Wooley, J. C. A primer on metagenomics / J. C. Wooley, A. Godzik, I. Friedberg // *PLoS computational biology*. – 2010. – Vol. 6, № 2. – P. e1000667. DOI: 10.1371/journal.pcbi.1000667.
43. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform / A. S. Whiteley, S. Jenkins, I. Waite [et al.] // *Journal of microbiological methods*. – 2012. – Vol. 91, № 1. – P. 80–88. DOI: 10.1016/j.mimet.2012.07.008.
- 13 Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*. 2008;9:387–402. DOI: 10.1146/annurev.genom.9.081307.164359.
- 14 Wu Y, Zhuang J, Song Y, Gao X, Chu J, Han S. Advances in single-cell sequencing technology in microbiome research. *Genes Dis*. 2023;11(4):101129. DOI: 10.1016/j.gendis.2023.101129.
- 16 Bryson K, Loux V, Bossy R, Nicolas P, Chaillou S, van de Guchte M, Penaud S, Maguin E, Hoebeke M, Bessières P, Gibrat JF. AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system. *Nucleic Acids Res*. 2006;34(12):3533–3545. DOI: 10.1093/nar/gkl471.
- 17 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–477. DOI: 10.1089/cmb.2012.0021.
- 18 Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol*. 2013;20(10):714–737. DOI: 10.1089/cmb.2013.0084.
- 19 Bandyra KJ, Luisi BF. Licensing and due process in the turnover of bacterial RNA. *RNA Biol*. 2013;10(4):627–635. DOI: 10.4161/rna.24393.
- 20 Baroni D, Arrigo P. MicroRNA target and gene validation in viruses and bacteria. *Methods Mol Biol*. 2014;1107:223–231. DOI: 10.1007/978-1-62703-748-8_13.
- 21 Waldbauer JR, Rodrigue S, Coleman ML, Chisholm SW. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS One*. 2012;7(8):e43432. DOI: 10.1371/journal.pone.0043432.
- 22 Carvalhais LC, Dennis PG, Tyson GW, Schenk PM. Application of metatranscriptomics to soil environments. *J Microbiol Methods*. 2012;91(2):246–251. DOI: 10.1016/j.mimet.2012.08.011.
- 23 Zerbino DR, Frankish A, Flicek P. Progress, Challenges, and Surprises in Annotating the Human Genome. *Annu Rev Genomics Hum Genet*. 2020;21:55–79. DOI: 10.1146/annurev-genom-121119-083418.
- 24 Michaelsen TY, Brandt J, Singleton CM, Kirkegaard RH, Wiesinger J, Segata N, Albertsen M. The Signal and the Noise: Characteristics of Antisense RNA in Complex Microbial Communities. *mSystems*. 2020;5(1):e00587-19. DOI: 10.1128/mSystems.00587-19.
- 25 Sleator RD, Shortall C, Hill C. Metagenomics. *Letters in applied microbiology*. 2008;47(5):361–366. DOI: 10.1111/j.1472-765X.2008.02444.x.
- 26 Thomas T, Gilbert J, Meyer F. Metagenomics – a guide from sampling to data analysis. *Microb Inform Exp*. 2012;2(1):3. DOI: 10.1186/2042-5783-2-3.
- 27 Simon C, Herath J, Rockstroh S, Daniel R. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol*. 2009;75(9):2964–2968. DOI: 10.1128/AEM.02644-08.
- 28 Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, Chang HW, Watson D, Brodie EL, Hazen TC, Keller M. Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl Environ Microbiol*. 2006;72(5):3291–3301. DOI: 10.1128/AEM.72.5.3291-3301.2006.
- 29 Mizrahi-Man O, Davenport ER, Gilad Y. Taxonomic classification of bacterial 16S rRNA genes using short

References

- 1 Lasken RS, McLean JS. Recent advances in genomic DNA sequencing of microbial species from single cells. *Nat Rev Genet*. 2014;15(9):577–584. DOI: 10.1038/nrg3785.
- 2 Popa O, Oldenburg E, Ebenhöf O. From sequence to information. *Philos Trans R Soc Lond B Biol Sci*. 2020;375(1814):20190448. DOI: 10.1098/rstb.2019.0448.
- 3 Giani AM, Gallo GR, Gianfranceschi L, Formenti G. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput Struct Biotechnol J*. 2019;18:9–19. DOI: 10.1016/j.csbj.2019.11.002.
- 4 Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2016;107(1):1–8. DOI: 10.1016/j.ygeno.2015.11.003.
- 5 Casler M.D. Fundamentals of Experimental Design: Guidelines for Designing Successful Experiments. *Agronomy Journal*. 2015;107:692–705. DOI: 10.2134/agronj2013.0114.
- 6 Tanious R, Onghena P. Randomized Single-Case Experimental Designs in Healthcare Research: What, Why, and How? *Healthcare (Basel)*. 2019;7(4):143. DOI: 10.3390/healthcare7040143.
- 7 Auer PL, Doerge RW. Statistical design and analysis of RNA sequencing data. *Genetics*. 2010;185(2):405–416. DOI: 10.1534/genetics.110.114983.
- 8 Churchill GA. Fundamentals of experimental design for cDNA mi-croarrays. *Nat Genet*. 2002;32:490–495. DOI: 10.1038/ng1031.
- 9 Robasky K, Lewis NE, Church GM. The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet*. 2014;15(1):56–62. DOI: 10.1038/nrg3655.
- 10 Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet*. 2020;21(10):597–614. DOI: 10.1038/s41576-020-0236-x.
- 11 Monaghan TF, Rahman SN, Agudelo CW, Wein AJ, Lazar JM, Everaert K, Dmochowski RR. Foundational Statistical Principles in Medical Research: Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value. *Medicina (Kaunas)*. 2021;57(5):503. DOI: 10.3390/medicina57050503.
- 12 Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare RP, Banday S, Mishra AK, Das G, Malonia SK. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology (Basel)*. 2023;12(7):997. DOI: 10.3390/biology12070997.

- sequencing reads: evaluation of effective study designs. *PLoS One*. 2013;8(1):e53608. DOI: 10.1371/journal.pone.0053608.
- 30 Jünemann S, Prior K, Szczepanowski R, Harks I, Ehmke B, Goesmann A, Stoye J, Harmsen D. Bacterial community shift in treated periodontitis patients revealed by ion torrent 16S rRNA gene amplicon sequencing. *PLoS One*. 2012;7(8):e41606. DOI: 10.1371/journal.pone.0041606.
- 31 Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol*. 2010;12(1):118–123. DOI: 10.1111/j.1462-2920.2009.02051.x.
- 32 Nossa CW, Oberdorf WE, Yang L, Aas JA, Paster BJ, Desantis TZ, Brodie EL, Malamud D, Poles MA, Pei Z. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J Gastroenterol*. 2010;16(33):4135–44. DOI: 10.3748/wjg.v16.i33.4135.
- 33 Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. 2013;41(1):e1. DOI: 10.1093/nar/gks808.
- 34 Conlan S, Kong HH, Segre JA. Species-level analysis of DNA sequence data from the NIH Human Microbiome Project. *PLoS One*. 2012;7(10):e47075. DOI: 10.1371/journal.pone.0047075.
- 35 Soergel DA, Dey N, Knight R, Brenner SE. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J*. 2012;6(7):1440–1444. DOI: 10.1038/ismej.2011.208.
- 36 Crosby LD, Criddle CS. Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. *Biotechniques*. 2003;34(4):790–794, 796, 798. DOI: 10.2144/03344rr01.
- 37 Kembel SW, Wu M, Eisen JA, Green JL. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Comput Biol*. 2012;8(10):e1002743. DOI: 10.1371/journal.pcbi.1002743.
- 38 Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol*. 2007;73(1):278–288. DOI: 10.1128/AEM.01177-06.
- 39 Gevers D, Pop M, Schloss PD, Huttenhower C. Bioinformatics for the Human Microbiome Project. *PLoS Comput Biol*. 2012;8(11):e1002779. DOI: 10.1371/journal.pcbi.1002779.
- 40 Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol*. 1998;5(10):R245–249. DOI: 10.1016/s1074-5521(98)90108-9.
- 41 Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 2004;304(5667):66–74. DOI: 10.1126/science.1093857.
- 42 Wooley JC, Godzik A, Friedberg I. A primer on metagenomics. *PLoS Comput Biol*. 2010;6(2):e1000667. DOI: 10.1371/journal.pcbi.1000667.
- 43 Whiteley AS, Jenkins S, Waite I, Kresoje N, Payne H, Mullan B, Allcock R, O'Donnell A. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. *J Microbiol Methods*. 2012;91(1):80–88. DOI: 10.1016/j.mimet.2012.07.008.

BASICS OF SCIENTIFIC EXPERIMENT PLANNING FOR SEQUENCING ANALYSIS OF PROKARYOTIC GENOMES AND BACTERIAL COMMUNITIES

S. A. Kostiuk, A. K. Lyamtseva, O. S. Poluyan
Belarusian State Medical University, Minsk, Republic of Belarus

Background. Any scientific experiment requires careful preliminary preparation, and in medical biology planning of the experimental part of the study is of great importance due to the wide variability of properties characteristic of biological objects. In sequencing of prokaryotic genomes two main directions can be distinguished at present: sequencing of individual microbes, including uncultivated ones, and sequencing of microbial biocenoses to study the community structure.

Objective. To consider the methodology of scientific experiment planning when sequencing prokaryotic genomes and bacterial communities.

Material and Methods. The Web of Science and Scopus platforms were used to search for literature sources. Publications describing original studies and scientific reviews were analyzed.

Results. Planning a scientific experiment is a procedure for selecting the number and conditions of experiments necessary and sufficient to solve the problem with the required accuracy. The sources of errors at the stage of sample preparation, variants of high-throughput sequencing application are described.

Conclusions. Metagenomics, metatranscriptomics and metaproteomics make it possible to discover previously unknown unculturable microorganisms, describe their properties and functions, discover not yet studied biomolecules, and characterize microbial communities inhabiting the human body for healthy people and patients with various diseases.

Keywords. Scientific experiment, sequencing, prokaryotes, genome, transcriptome.

For citation: Kostiuk SA, Lyamtseva AK, Poluyan OS. Basics of scientific experiment planning for sequencing analysis of prokaryotic genomes and bacterial communities. *Biochemistry and Molecular Biology*. 2025, vol. 4, no. 1(6). pp. 78–87 (in Russian).

Поступила 11.10.2024