

На правах рукописи

МАЦВАЙ АЛИНА ДМИТРИЕВНА

**РАЗРАБОТКА ТЕХНОЛОГИИ ДЛЯ ИДЕНТИФИКАЦИИ ВИРУСНЫХ
АГЕНТОВ С ИСПОЛЬЗОВАНИЕМ ДНК-МЕТАБАРКОДИНГА И
ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ**

03.01.03 – Молекулярная биология

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата биологических наук

Москва 2021

Работа прошла апробацию в Федеральном государственном автономном образовательном учреждении высшего образования "Московский физико-технический институт (национальный исследовательский университет)"

Научный руководитель: Хафизов Камиль Фаридович, кандидат биологических наук, руководитель научной группы разработки новых методов диагностики на основе технологий секвенирования следующего поколения, ФБУН ЦНИИ эпидемиологии Роспотребнадзора.

Ведущая организация: Федеральное бюджетное учреждение науки «Санкт-Петербургский научно-исследовательский институт эпидемиологии и микробиологии им. Пастера» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека.

Защита состоится «27» декабря 2021 года в 11.00 на заседании диссертационного совета ФБМФ 03.01.03.005, по адресу: 141701, г. Долгопрудный Московской обл., Институтский переулок, д. 9.

С диссертацией можно ознакомиться в библиотеке и на сайте Московского физико-технического института (национальный исследовательский университет) <https://mipt.ru/education/post-graduate/soiskateli-biologicheskie-nauki.php>.

Работа представлена «08» октября 2021 г. в Аттестационную комиссию федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)» для рассмотрения советом по защите диссертаций на соискание ученой степени кандидата наук в соответствии с п .3.1 ст. 4 Федерального закона «О науке и государственной научно - технической политике»

Ученый секретарь диссертационного совета: Зубарев Илья Владимирович, кандидат биологических наук.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Выявление новых видов патогенных микроорганизмов, мониторинг циркуляции известных патогенов и оценка путей их распространения являются одним из наиболее важных аспектов эпидемиологического надзора. Своевременная идентификация этиологического агента и раннее назначение этиотропной терапии значительно снижает риск возникновения осложнений. Однако в настоящее время получить лабораторно подтверждённый диагноз удаётся в относительно небольшой доле случаев инфекционных заболеваний. Например, по данным Всемирной организации здравоохранения, установить выявить возбудителя острого респираторного заболевания удаётся не более чем в 30% случаев. До недавнего времени этой проблеме не уделялось достаточное внимание в связи с тем обстоятельством, что большая часть наиболее часто встречающихся в средних широтах инфекционных заболеваний имеет выраженную клиническую картину с хорошо различимой симптоматикой. Однако в последние годы наблюдается рост случаев инфекционных заболеваний, не имеющих характерной клинической картины; все чаще регистрируются вспышки новых вирусных инфекционных заболеваний, переходящие в эпидемии и пандемии. Например, в 2002-2003 годах наблюдалась вспышка заболеваемости тяжёлым острым респираторным синдромом, вызванным коронавирусом SARS. Были заражены более 8,000 человек, из которых 813 погибли. В 2009 году произошла пандемия нового штамма вируса гриппа А - H1N1. Было зарегистрировано более 250 тысяч случаев заболевания, из которых около 2,600 закончились летальным исходом [1]. В 2014 году наблюдалась вспышка заболеваемости геморрагической лихорадкой Эбола, вызываемой вирусом Эбола, открытый в 1976 году в Судане [2]. Было зарегистрировано более 28 тыс. заболевших, из которых более 11 тыс. погибли [3]. Для предотвращения возникновения вспышек вирусных заболеваний, важен непрерывный эпидемиологический контроль и возможность раннего выявления этиологического фактора заболевания. Так, например, своевременное обнаружение нового штамма *COVID-19*, и дифференцировка его от других вирусов со схожими клиническими симптомами позволила бы предотвратить эпидемию мирового масштаба. На середину июля 2021 года было выявлено заболевших более 188 млн. человек, для более чем четырёх миллионов заболевание закончилось летальным исходом [4].

1. WHO. Pandemic (H1N1) 2009 - update 62 (revised 21 August 2009) URL: http://www.who.int/csr/don/2009_08_21/en/index.html.

2. Na W. et al. Ebola outbreak in Western Africa 2014: what is going on with Ebola virus? // Clin. Exp. Vaccine Res. 2015. Vol. 4, № 1. P. 17–22.

3. Garske T. et al. Heterogeneities in the case fatality ratio in the West African Ebola outbreak 2013-2016 // Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2017. Vol. 372, № 1721.

4. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU).

Описанные вспышки заболеваний демонстрируют необходимость разработки эффективных методов диагностики вирусных патогенов широкого спектра, включая те, что не были достаточно полно описаны ранее. Диагностика в формате широкого скрининга приносит пользу общественному здравоохранению, собирая данные для эпидемиологических исследований, которые позволяют предотвратить вспышки и распространение вирусных заболеваний.

Степень разработанности темы. В настоящее время наиболее распространённым методом лабораторной диагностики вирусных инфекционных заболеваний является полимеразная цепная реакция (ПЦР) в режиме реального времени, предполагающая использование специфических праймеров и ДНК-зондов. На текущий момент было описано значительное количество структур олигонуклеотидов, предназначенных для обнаружения определённых вирусных патогенов. Однако возможность мультиплексирования таких реакций ограничена количеством каналов детектора флуоресценции амплифицирующего прибора (обычно не больше 5). Кроме того, в режиме мультиплекса такие реакции идут менее эффективно из-за различных температур отжига праймеров, неспецифической амплификации и потенциальной самокомплементарности олигонуклеотидов. Таким образом, для видового определения патогенов разных таксономических групп необходимо проводить ряд ПЦР исследований. Для проведения исследований методом специфической ПЦР требуется предварительная гипотеза о наличии конкретных патогенов в образце, в отсутствие которой процесс идентификации патогенов может занять значительное время, что накладывает ограничение на выбор тактики лечения, снижая её эффективность, а также затрудняя проведение соразмерных противоэпидемических мероприятий.

Появление технологий высокопроизводительного секвенирования (или NGS – от англ. Next Generation Sequencing) вызвало настоящую революцию в самых различных областях медицины и биологии. Вследствие устойчиво и постоянно снижавшейся стоимости секвенирования в последние десятилетия, многократно увеличилась доступность этого метода, позволившая получить последовательности нуклеиновых кислот (НК) любого живого организма. Однако, несмотря на постоянное усовершенствование протоколов пробоподготовки, на сегодняшний день этот способ все ещё является достаточно дорогим и трудоемким, в особенности для массового использования в режиме скрининга большого количества образцов.

Применение методов таргетного обогащения NGS библиотек позволяет повысить эффективность анализа данных и существенно снизить стоимость самого исследования. Мультипраймерная полимеразная цепная реакция – один из наиболее распространенных методов создания таргетных библиотек фрагментов который, благодаря его высокой чувствительности и

специфичности, широко используется в различных задачах, таких как идентификация патогенов, скрининговые исследования, количественный анализ образцов, диагностика генетических заболеваний, судебная экспертиза, популяционные генетические исследования и многие другие. Однако, у этого метода есть существенные недостатки, влияющие на качество получаемых данных. Использование большого числа праймерных пар в одной реакционной смеси зачастую приводит к преимущественной амплификации отдельных локусов и перепредставленности их нуклеотидных последовательностей в данных секвенирования. Кроме того, использование больших наборов олигонуклеотидов в режиме мультиплексной ПЦР может приводить к амплификации нецелевых областей, наработке димеров праймеров, образованию химерных последовательностей. Все это значительно снижает качество получаемых данных, и затрудняет их анализ.

Весь объем накопленных данных пока не позволил установить базовые принципы построения генома, однако обнаружил множество небольших фрагментов с различными свойствами. В частности, были исследованы и описаны участки протяженностью, не превышающей несколько тысяч п.н., на основании последовательностей нуклеотидов которых может быть установлена таксономическая принадлежность объектов исследования. Некоторые из таких фрагментов оказались flankированы высококонсервативными участками. Таким образом, при наличии такого универсального локуса в геномах всех представителей выделенной таксономической группы, их вариабельные последовательности могут быть специфическим образом амплифицированы при использовании нескольких праймеров с целью последующей идентификации. Указанные фрагменты называют ДНК-баркодами, а соответствующий метод определения состава компонент в образце – ДНК-баркодингом.

Указанный метод в настоящее время широко применяется в метагеномных исследованиях. Например, для таксономической классификации бактерий используется область гена 16SrRNA, который содержит 9 вариабельных областей; для анализа эукариот используется 18s область или область гена цитохром оксидазы, для грибковых организмов – ITS1,2 области. Однако применение классического метода метабаркодинга для работы с вирусными патогенами весьма затруднительно: геномы вирусов отличаются высокой вариабельностью даже в пределах одной таксономической группы.

Таким образом, используя набор праймеров для разных таксономических групп, возможно из смеси нуклеиновых кислот любых организмов, исключая вирусы, получить смесь коротких ампликонов и идентифицировать представляемые ими организмы в образце. Целевое секвенирование информативных участков сокращает стоимость анализа, обеспечивая возможность работы в потоковом режиме, например, для проведения скрининга пациентов на

присутствие любого из патогенов. Также важно отметить относительно невысокие требования к качеству исходного материала, простоту и непродолжительность пробоподготовки, что вместе с высокой специфичностью амплификации на небольшой панели праймеров, позволяет рассматривать этот метод в качестве основного при идентификации патогенов, не являющихся вирусами.

Стандартный метод ДНК-метабаркодинга позволяет выявлять любой патоген, кроме вирусного, что обусловлено высоким разнообразием геномов последних и как следствие отсутствием вариабельных участков, окружённых высоконсервативными областями у представителей как высокоуровневых таксонов (порядка, класса), так и низкоуровневых (семейства, рода, зачастую и вида). Однако, отсутствие возможности детектирования вирусов – наиболее разнообразных и распространенных патогенов, существенно ограничивает применимость анализа и ценность получаемых результатов. Таким образом, требуется разработать альтернативный (модифицированный) подход для обеспечения эффективного выявления и характеристики вирусных нуклеиновых кислот.

Целью работы является разработка методики для идентификации вирусных агентов с использованием ДНК-метабаркодинга и высокопроизводительного секвенирования. Для достижения этой цели ставились следующие **задачи**:

1. Разработать алгоритм для дизайна первичных структур библиотек олигонуклеотидов для целевого обогащения нуклеиновых кислот вирусных патогенов в режиме мультиплекса;
2. Разработать библиотеки олигонуклеотидов для целевого обогащения нуклеиновых кислот вирусных патогенов целевой группы, а также лабораторную методику получения библиотек фрагментов для высокопроизводительного секвенирования;
3. Провести апробацию разработанного метода на биологических образцах, в том числе с известным содержанием нуклеиновых кислот вирусных патогенов различных таксономических групп;
4. Применить разработанную методику для проведения скрининга широкого спектра вирусных патогенов в рамках исследования вирома диких птиц околоводного комплекса;
5. Провести сравнение разработанного метода с аналогами, в основе которых лежат технологии высокопроизводительного секвенирования.

Научная новизна. В рамках данной работы был разработан новый метод выбора пар праймеров, ключевое отличие которого заключается в отсутствии предварительных построений множественных выравниваний референсных геномов, а также в использовании матричного способа заполнения и хранения данных о выравниваниях олигонуклеотидов на референсные

последовательности. Такой формат позволяет обнаруживать пары праймеров с использованием высокоэффективных параллельных вычислений, осуществляющих математические операции над матрицами и тензорами. Разработанный метод дизайна структур библиотек олигонуклеотидов описан впервые. С использованием разработанной методики было проведено исследование виromа мигрирующих птиц околоводного комплекса, в ходе которого были выявлены часто встречающиеся и релевантные группы вирусных патогенов. С использованием разработанной методики был выявлен неизвестный ранее представитель рода *Atadenovirus*.

Научно-практическая значимость. В настоящей работе был разработан подход для выявления нуклеиновых кислот вирусных патогенов, в том числе применимый для расшифровки вирусных инфекционных заболеваний неясной этиологии. Преимущество используемых подходов, по сравнению с рутинно применяющимися в лабораторной диагностике методами, в том, что они лишены необходимости предварительной гипотезы о наличии конкретных таксономических групп вирусов в исследуемом образце и обеспечивают охват широкого спектра патогенов в рамках одного исследования. Разработанная методика, кроме того, менее чувствительна к изменениям в последовательностях геномов патогенов по сравнению со стандартными тест-системами на основе ПЦР. Разработанный подход универсален и может быть применен для различных групп вирусных патогенов. Применение результатов данной разработки в практике лабораторной диагностики направлено на повышение эффективности идентификации возбудителей вирусных инфекций с целью дифференциальной диагностики вирусных инфекций и своевременного назначения или коррекции схемы терапии. Разрабатываемая методика может быть применена как для установления этиологического фактора заболевания в случаях неэффективности стандартных методик, так и для проведения широких скрининговых исследований.

Положения, выносимые на защиту.

1. Разработан и реализован метод анализа широкого спектра вирусных патогенов, включающий алгоритм создания библиотек олигонуклеотидов, использующий матричный способ заполнения, хранения и обработки данных о локальных выравниваниях коротких последовательностей олигонуклеотидов, позволяющий использовать высокоэффективные параллельные вычисления, а также универсальная лабораторная методика проведения исследования;

2. Апробация разработанной методики показала ее эффективность в обнаружении вирусных патогенов различных таксономических групп вирусов на контрольном материале и биологических образцах;
3. Проведено исследование вирома диких птиц околоводного комплекса, выявившее нуклеиновые кислоты более 70 различных видов патогенных для позвоночных вирусов;
4. Выявлен ранее не известный представитель рода *Atadenovirus* (*Sterna Atadenovirus*), известные представители которого являются патогенными для млекопитающих, птиц и пресмыкающихся вирусами.

Личный вклад автора и апробация работы. Автором была проведена работа по созданию наборов референсных геномов для создания и тестирования библиотек олигонуклеотидов. Автором были разработаны программные модули алгоритма дизайна библиотек олигонуклеотидов; программные модули алгоритма *in silico* тестирования полученных библиотек олигонуклеотидов. Автором были разработаны лабораторные методики тестирования библиотек олигонуклеотидов; отработаны лабораторные методики экстракции нуклеиновых кислот, лабораторные методики создания библиотек фрагментов для проведения высокопроизводительного секвенирования. Автором были осуществлены биоинформационический анализ и интерпретация данных высокопроизводительного секвенирования. Результаты работы были представлены на международных и всероссийских научных конференциях:

1. Международная научно-практическая конференция «Молекулярная диагностика 2018» с устным докладом «Технологии NGS для поиска вирусных патогенов в биологических образцах» (2018 г., Минск, Республика Беларусь);
2. «Российский конгресс лабораторной медицины» с устным докладом «Перспективы применения методов NGS в диагностике вирусных инфекций» (2018 г., Москва, Россия);
3. «European Meeting on Molecular Diagnostics» с устным докладом «Metabarcoding-like approach for high throughput detection and identification of viral nucleic acids» (2019 г., Нордвейк, Нидерланды);
4. «63-я Всероссийская научная конференция МФТИ» с устным докладом «Обнаружение нового представителя рода *Atadenovirus* в биологических образцах позвоночных с использованием методики, подобной метабаркодингу» (2020 г., Москва, Россия).
5. «Научная конференция OpenBio 2021» с устным докладом «Разработка и применение универсального алгоритма создания мультипраймерных панелей для детекции широкого спектра вирусных патогенов с использованием методов высокопроизводительного секвенирования» (2021г., Новосибирск, Россия)

Публикации. По теме работы опубликовано 10 печатных работ: из них 4 в научных журналах, индексируемых базами данных Scopus и Web of Science, 6 – в сборниках тезисов международных конференций, индексируемых базой данных РИНЦ. Разработанный алгоритм, реализованный в виде программы для ЭВМ, зарегистрирован в Роспатенте, как объект интеллектуальной собственности.

Структура и объем диссертации. Работа состоит из введения, 6-ти разделов основной части, заключения, списка литературы. Общее количество страниц: 265. Работа содержит 85 иллюстраций, 122 таблицы; список литературы включает 275 позиций.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во введении отражена актуальность работы, сформулирована ее цель и задачи, обсуждена научная новизна результатов и их практическая значимость, представлены выносимые на защиту положения.

Глава 1. Аналитический обзор научно-технической литературы

В данной главе приводится обзор научно-технической литературы по тематике работы. В разделе дается описание существующих методов диагностики вирусных инфекций; приводится более подробное рассмотрение методик выявления вирусных нуклеиновых кислот, основанных на технологиях высокопроизводительного секвенирования, проводится анализ недостатков существующих методик.

На основании литературных данных выделяются актуальные таксономические группы вирусных патогенов, для включения в группу интереса по двум задачам, в рамках которых производилась апробация и применение разработанной методики: исследование этиологии респираторных вирусных инфекций человека; изучение вирома диких птиц околоводного комплекса.

Глава 2. Разработка биоинформационического алгоритма для дизайна структур праймеров и создание библиотек нуклеотидов для целевого обогащения нуклеиновых кислот вирусных патогенов в режиме мультиплекса

В данной главе диссертации описывается процесс разработки биоинформационического алгоритма для дизайна структур праймеров для целевого обогащения нуклеиновых кислот вирусных патогенов в режиме мультиплекса и создания библиотек олигонуклеотидов.

2.1 Получение референсных данных

Так как для работы любого алгоритма подбора праймерных последовательностей в первую очередь необходимы качественные последовательности референсных вирусных геномов, первый подраздел главы содержит информацию о примененных подходах к созданию референсной базы данных нуклеотидных последовательностей геномов. В большинстве из них значимые с медицинской точки зрения и модельные организмы представлены большим количеством одинаковых последовательностей одного или нескольких штаммов, в то время как менее распространенные группы патогенов, клиническая значимость которых изучена недостаточно, представлены в базах данных в значительно меньшей степени. Более того, открытые, не куррируемые базы данных содержат ошибки и неточности в метаданных, приводящие в том числе к записям с ошибочной таксономической классификацией. Такая искаженность

референсных данных может приводить к тому, что при помощи разрабатываемого алгоритма дизайна праймеров будут выбираться не наиболее консервативные участки внутри всего рода вирусов, а специфичные только для конкретного штамма последовательности.

Для получения последовательностей нуклеиновых кислот использовалась общедоступная база данных “The Virus Pathogen Database and Analysis Resource (ViPR)”, разработанная в 2011 году. Наиболее важным преимуществом указанного источника является аутентичность последовательностей нуклеиновых кислот, что обеспечивается постоянным курированием входящих в неё данных специалистами в области вирусологии и биоинформатики. В случае отсутствия целевой группы патогенов в указанной базе данных использовалась нуклеотидная коллекция (NT) NCBI. Так как NT NCBI не курируется вручную, аутентичность последовательностей была проведена с помощью собственных скриптовых программ, производящих выравнивание загруженных последовательностей на полную нуклеотидную коллекцию с целью выявления ошибочно классифицированных последовательностей. Далее для нивелирования перепредставленности отдельных групп хорошо изученных вирусных патогенов, производилась кластеризация последовательностей с пороговыми значениями в 98.0-99.8%. Полученная референсная база данных использовалась для создания библиотек олигонуклеотидов.

2.2 Алгоритм, использующий множественные выравнивания референсных геномов

В данном подразделе описывается схема работы первой версии алгоритма дизайна библиотек олигонуклеотидов, использующий множественные выравнивания последовательностей референсной базы данных. Разработанный алгоритм можно разбить на 5 смысловых блоков: построение множественного выравнивания; выбор консервативных участков; выбор пар

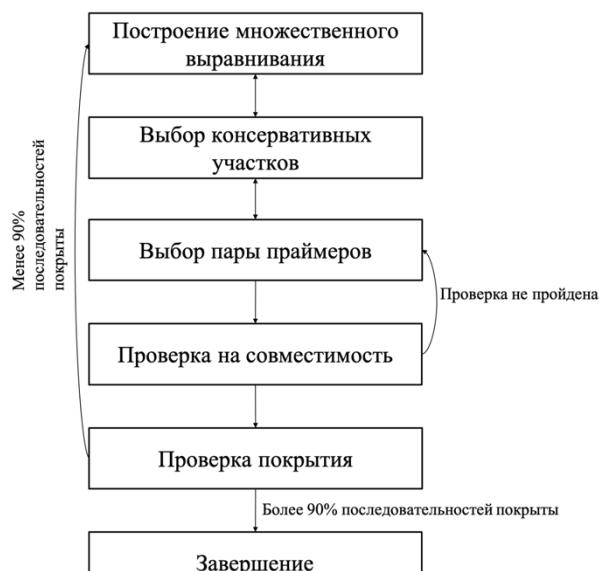


Рисунок 1 — Схема работы алгоритма для выбора праймеров

праймеров; проверка совместимости полученных пар с выбранными ранее; проверка покрытия (Рисунок 1).

На первом этапе проводится построение множественного выравнивания, далее для поиска консервативных участков внутри множественного выравнивания последовательностей используется метод «скользящего окна» с шагом в один нуклеотид и идентифицирует все участки заданной длины, удовлетворяющие заданным критериям: доля неоднозначных позиций (частота встречаемости каждого из $2x$ и более нуклеотидов $>10\%$) $\leq 20\%$; доля уникальных видов/штаммов, не содержащих пробелов в множественном выравнивании $\geq 50\%$; GC-состав усредненной подпоследовательности в интервале 35–65%; отсутствие самокомplementарных участков внутри структуры. После этого осуществляется выбор пар праймеров. В силу ограничений, накладываемых на длину целевого ампликона, праймеры из одной пары должны находиться друг от друга на расстоянии от 200 до 500 п.н, чтобы обеспечивать совместимость панели с наиболее популярными секвенирующими платформами. Таким образом, все полученные на предыдущем шаге консервативные участки объединяются в пары, после чего производится их фильтрация по длине ампликона и покрытию вирусов, последовательности которых были использованы при построении множественного выравнивания. Две последовательности считаются парой, если длина целевого ампликона соответствует введённым ограничениям и пара праймеров покрывает более 90% выровненных последовательностей.

Полученные на предыдущем этапе пары праймеров фильтруются в соответствии с их температурой отжига. Для удовлетворяющих условию пар проверяется отсутствие парного выравнивания на базу данных нуклеотидных последовательностей хозяина. После этого рассчитывается температура формирования гетеродимеров между выбранными на настоящем и предыдущих этапах праймерами. Пары, содержащие последовательности, температура плавления гетеродимеров которых близка к расчётной температуре отжига, удаляются. Пары праймеров сортируются по подсчитанным параметрам, первая из полученных пар праймеров, имеющая наилучшие характеристики, добавляется к выбранным на предыдущих этапах. Затем процесс итерационно повторяется.

2.3 Множественное выравнивание и разнообразие вирусов

Подраздел содержит анализ недостатков и лимитирующих параметров разработанного алгоритма, обсуждает варианты решения выявленных проблем.

Основной лимитирующий фактор, ограничивающий применение алгоритмов, основанных на построении множественных выравниваний геномов, происходит из самой природы вирусных патогенов, геномы которых отличаются высокой вариабельностью и изменчивостью. При этом возможно разбитие референсной базы данных на отдельные кластеры, отобранные на основании

схожести нуклеотидных последовательностей внутри кластеров, и запуск описанного ранее алгоритма на каждом кластере по отдельности. Таким образом, количество пар, отобранных праймеров будет пропорционально количеству кластеров.

Однако некоторые группы вирусов, например, род энтеровирусов, характеризуются настолько высокой степенью разнообразия, что при их кластеризации не удается получить удовлетворительного количества кластеров для дальнейшего дизайна праймеров. Кроме того, для некоторых групп вирусов характерно высокое разнообразие и изменчивость генетического материала. В связи с этим, поиск консервативных областей в их геномах с использованием множественного выравнивания будет затруднён (Рисунок 2).

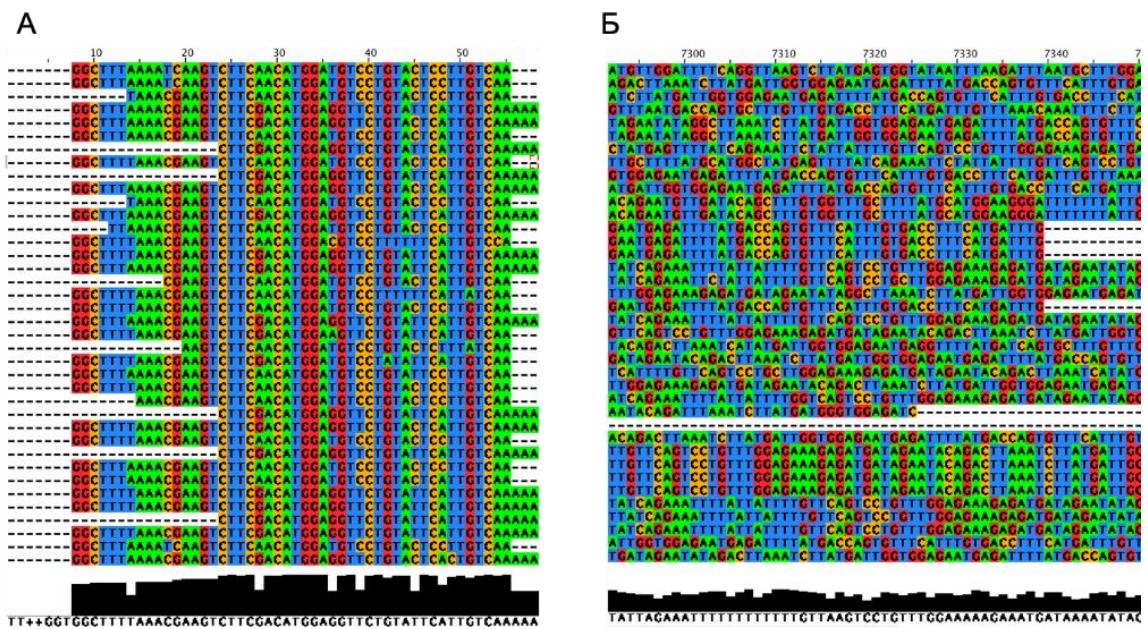


Рисунок 2 – Множественные выравнивания геномов вирусов разных таксономических групп, где (А) пример множественного выравнивания, подходящего для работы версии алгоритма подбора пар праймеров; (Б) пример множественного выравнивания, неподходящего для работы версии алгоритма подбора пар праймеров

Однако, невозможность построения множественного выравнивания не означает отсутствия консервативных областей среди рассматриваемой группы вирусов или среди подгруппы внутри неё. Следует также отметить, что таксономия вирусов сравнительно часто претерпевает существенные изменения. Это во многом связано с тем, что изначально принципы её построения базировались на схожести симптомов, проявляющихся при заражении вирусами, а не на схожести их генетического материала. Поэтому, анализируя некоторые рода вирусов, возможно обнаружить существенные отличия в последовательностях их НК.

Таким образом, необходимо было доработать алгоритм для дизайна первичных структур праймеров с целью расширения его применимости – осуществления поиска консервативных участков внутри интересующих последовательностей без использования множественных выравниваний.

2.4 K-мерный алгоритм с двойной кластеризацией и использованием матричного представления

В подразделе описан вариант модифицированного алгоритма подбора библиотек олигонуклеотидов, использующего подход, не предполагающий построение множественных выравниваний референсных геномов. Основным преимуществом разработанного алгоритма, является поддержка высокоэффективных методов расчетов систем праймеров с использованием матричных представлений. Общая схема алгоритма представлена на рисунке ниже (Рисунок 3).

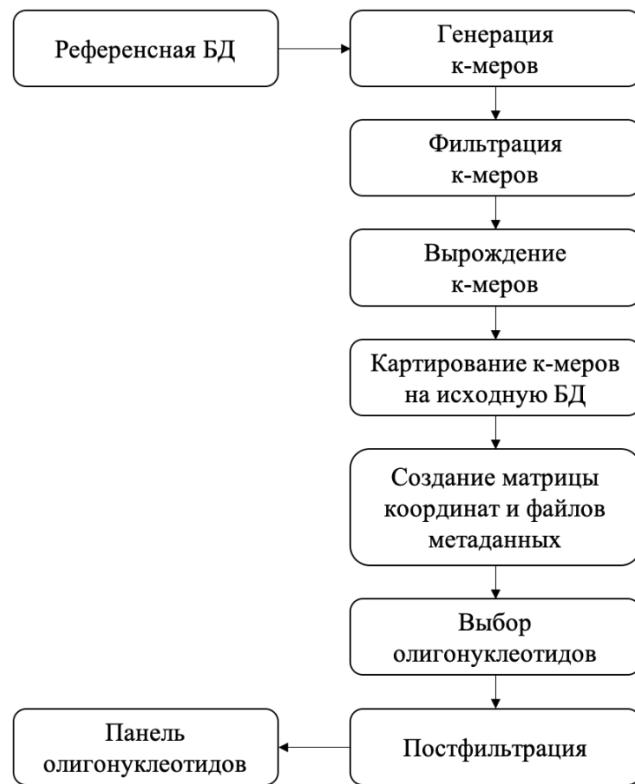


Рисунок 3 – Общая схема разработанного алгоритма

На вход алгоритм принимает базу данных референсных последовательностей в формате fasta, для обогащения которых необходимо разработать панель олигонуклеотидов, а также набор граничных условий, описывающий необходимые свойства панели (Таблица 1).

Таблица 1 – Параметры, принимаемые на вход алгоритмом

Параметр	Типичное значение
Длина олигонуклеотида	18-28 п.н.
Температура плавления Тм	45-58 °C
ГЦ-состав	30-70 %
ГЦ-состав на 3' конце	≤ 3 п.н. (последние 5 п.н.)
Максимальная длина гомополимерной области	≤ 4 п.н.
Динуклеотидные повторы	≤ 4

Параметр	Типичное значение
(например, ATATATAT)	
Длина ампликона	≈ 100 п.н. (РТ-ПЦР) < 500 п.н. (ПЦР)
Разница Т _m между олигонуклеотидами	≤ 5 °C

На выходе алгоритм, в зависимости от заданных условий, выдает один или несколько наборов олигонуклеотидов. Блоки реализованы в виде 4 скриптовых программ для удобства тестирования и разработки алгоритма: скрипт подготовки объединяет блоки генерации, фильтрации, вырождения к-меров и их картирование на исходную БД; скрипт создания матричного представления реализует блок создания матрицы координат и файлов метаданных; скрипт выбора нуклеотидов реализует блок выбора панели с использованием параллельных вычислений на графическом процессоре; скрипт постфильтрации реализует одноименный блок и осуществляет проверку панели и ее дополнительную оптимизацию.

Алгоритм позволяет получать панели олигонуклеотидов, не прибегая к созданию множественного выравнивания, что принципиально важно при работе с разнородными последовательностями, так как в этом случае качественное выравнивание не может быть получено. Вместо этого для поиска лучших последовательностей-кандидатов используется метод локальных попарных выравниваний небольших участков заданной длины (к-меров) и исходных последовательностей БД. Далее проводится фильтрация к-меров для сокращения числа кандидатов, свойства которых не удовлетворяют задаваемым условиям.

Для увеличения количества возможных выравниваний, фрагменты обогащаются за счет введения вырождений. Для этого, производится кластеризация базы данных к-меров. Подобранные условия позволяют объединить в группы схожие к-меры и построить на их основе локальное множественное выравнивание, тем самым сравнив различия между сходными последовательностями. Из каждого кластера вычисляется консенсусная последовательность, в структуру которой вводятся вырождения, количество которых ограничивается входными параметрами. Далее строится распределение термодинамических параметром для каждой из вырожденных структур. К-меры, прошедшие фильтрацию, картируются на БД референсных последовательностей. На основе картирования для каждого из к-меров составляется список референсных последовательностей, на которые он (в случае невырожденной структуры) или набор его уникальных вариантов (в случае вырожденной структуры) выравнивается с указанием начальных координат выравниваний. Алгоритм получения структур кандидатов схематично изображен на рисунке ниже (Рисунок 4). На основании полученных массивов конструируется

матрица. Следует отметить, что если 2 и более представителя одного вырожденного к-мера выравниваются в различные области последовательности, то выбираются координаты лучшего из выравниваний.

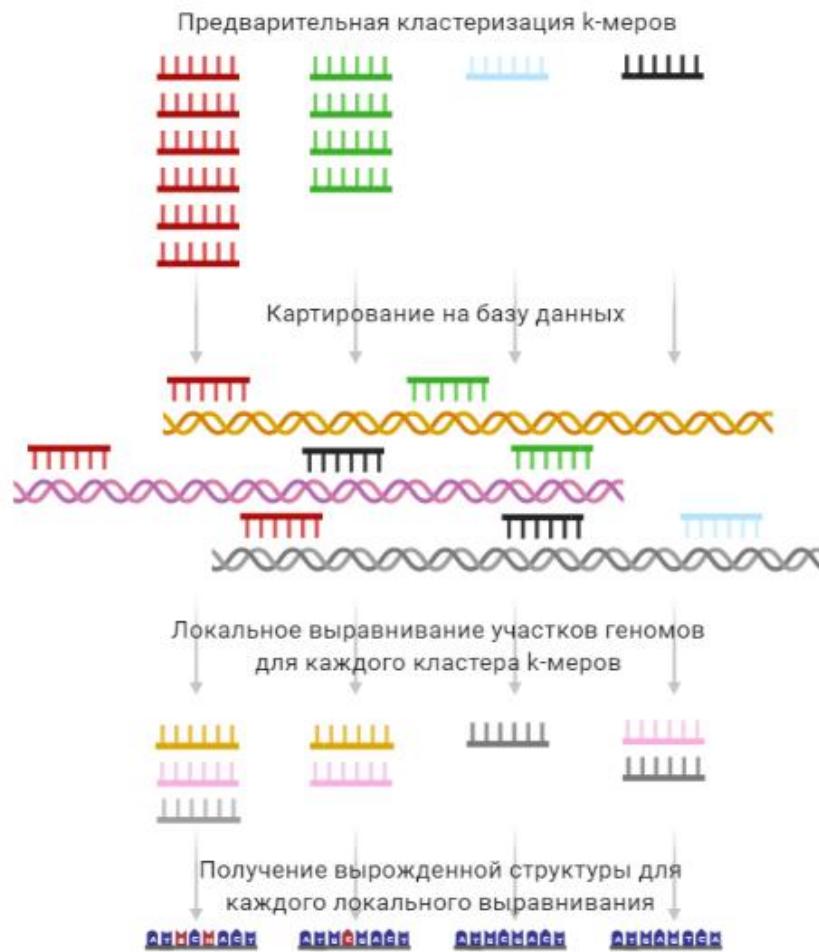


Рисунок 4 – Схема алгоритма получения структур-кандидатов_сМО

Для выбора пар праймеров на основании полученной матрицы необходимо учесть, что в паре праймеров один выступает в качестве форвардного, а второй в качестве реверсного. Таким образом, исходные к-меры должны быть проверены на возможность формирования пары в качестве форвардного и/или реверсного праймера. В случае использования положительной цепи для попарного выравнивания, полученного ранее, необходимо рассматривать 4 варианта положения к-мера относительно референсной последовательности (Рисунок 5).

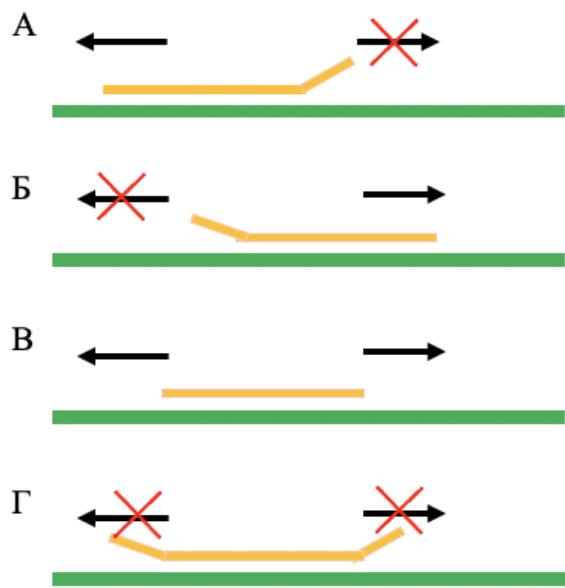


Рисунок 5 – Варианты положения к-мера относительно референсной структуры, где
 (А) – вариант к-мера, подходящего в кандидаты на реверсный праймер, за счет отсутствия несовпадений с референсным геномом в его нуклеотидной последовательности на 5' конце;
 (Б) – вариант к-мера, подходящего в кандидаты на форвардный праймер, за счет отсутствия несовпадений с референсным геномом в его нуклеотидной последовательности на 3' конце;
 (В) – вариант к-мера, подходящего в кандидаты как на форвардный, так и на реверсный праймер, за счет отсутствия несовпадений с референсным геномом в его нуклеотидной последовательности на 3' и 5' концах;
 (Г) – вариант к-мера, неподходящего в кандидаты как на форвардный, так и на реверсный праймер, из-за наличия несовпадений с референсным геномом в его нуклеотидной последовательности на 3' и 5' концах.

Каждому к-меру присваивается показатель возможности его использования в прямом и обратном направлении для формирования пары праймеров. После классификации к-меров на форвардные и реверсные, для каждого из наборов производится расчет температур формирования вторичных структур – шпилек. Если к-мер не удовлетворяет заданным условиям, то он исключается из рассмотрения (при этом, исключение к-мера из списков форвардных и реверсных к-меров производится независимо). Из полученных списков формируются 2 вектора длиной M . Элементы первого вектора принимают значение 0, если соответствующие к-меры не могут рассматриваться в качестве форвардных; 1 – если могут. Аналогично заполняется вектор для реверсных к-меров.

Выбор представления координат олигонуклеотидов в матричном виде позволяет использовать инструменты линейной алгебры для выбора пар праймеров. Так, при вычитании из исходной матрицы столбца P_1 образуется новая матрица той же размерности, элементы которой содержат вычисленные длины всех возможных ампликонов, в которых к-мер P_1 выступает в качестве реверсного праймера, а любой другой – в качестве форвардного (Рисунок 6).

	P_1	P_2	P_3	P_4		P_M
G_1	455	888	1533	434		C_M1
G_2	789	4345	895	566		C_M2
G_3	967	1535	125	331		C_M3
G_4	0	2456	1788	170		C_M4
G_N	C_1N	C_2N	C_3N	C_4N		C_MN

P_1	P_1-P_1	P_1-P_2	P_1-P_3	P_1-P_4		P_1-P_M
455	0	433	1078	-21		C_M1-455
789	0	3556	106	-223		C_M2-789
967	0	568	-842	-636		C_M3-967
0	0	2456	1788	170		C_M4-0
C_1N	C_1N-C_1N	C_2N-C_1N	C_3N-C_1N	C_4N-C_1N		C_MN-C_1N

Рисунок 6 – Вычитание столбца из матрицы

Как видно, повторив для всех столбцов операцию вычитания, получим на выходе M матриц (тензор MxNxM), а задача выбора лучшей пары сводится к поиску столбца, содержащего наибольшее число положительных элементов.

На первом этапе все отрицательные значения получаемой матрицы заменяются нулями. Далее корректируются положительные значения, полученные в результате вычитания из положительных элементов матрицы нулевого элемента столбца. При подсчёте количества покрываемых парой праймеров последовательностей, необходимо учитывать заданные ограничения длин ампликонов. Несоответствующие заданным параметрам значения также преобразуются в нули. После этого задача выбора лучшей пары сводится к суммированию значений каждого столбца и выбору элемента с максимальным значением суммы. Таким образом выбирается лучшая пара в итерации.

Повторив итерацию для каждого столбца исходной матрицы (Рисунок 7) получим вектор размерности Mx1, выбрав из которого элемент с максимальным значением, найдём лучшую пару, покрывающую наибольшее количество референсных последовательностей.

P_1-P_1				P_1-P_2				P_1-P_3				P_1-P_4				P_1-P_M					
G_1	0	1	0	0																	
G_2	0	0	1	0																	
G_3	0	1	0	0																	
G_4	0	0	0	0																	
G_N	C_1N	C_2N	C_3N	C_4N																	
	C_1N	C_2N	C_3N	C_4N																	

P_1-P_1	P_1-P_2	P_1-P_3	P_1-P_4		P_1-P_M
G	0	2	1	0	?

Тензор DD coverage ($M \times 1 \times M$)

P_1-P_2	P_2-P_?	P_3-P_?		P_M-1-P_?	P_M-P_?
C1	C2	C3		C4	C5

DD best coverage ($1 \times M$)

Max (DD best coverage) – первая лучшая пара

Рисунок 7 – Итерационный метод выбора лучшей пары

Однако, стоит обратить внимание на квадратичную зависимость количества операций вычитания ($N \cdot M^2$) от количества столбцов матрицы M . Выбор способа представления координат праймеров в виде матрицы дает возможность применения технологий параллельных вычислений на GPU, которые позволяют увеличить количество входных данных, а, следовательно, дают возможность создавать панели для обогащения разнообразных вирусных патогенов.

Использование GPU, однако, сопряжено с ограничениями в оперативной памяти, избежать которых можно проводя выбор лучших пар итерационно. После получения первой лучшей пары размерность исходной матрицы сокращается за счет удаления строк, соответствующих покрытым ранее последовательностям. Все операции повторяются, пока не будет выбран набор олигонуклеотидов, обеспечивающий заданное значение покрытия базы данных. Важно отметить, что при выборе каждой новой пары, проверяется их совместимость со структурами,ключенными в панель ранее. В случае обнаружения формирования нежелательный структур, производится выбор следующей по покрытию пары.

Реализованный алгоритм выбора праймеров является «жадным», что означает локальный выбор лучшей пары в каждой итерации. Однако, в результате такого метода дизайна панели олигонуклеотидов, в итоговом сете могут присутствовать «лишние» структуры, исключение которых из набора не влияет на итоговое покрытие. С целью уточнения оптимального состава панели, выбранные олигонуклеотиды выравниваются на референсную базу данных, после чего с использованием вычислительной мощности CPU производится выбор пар праймеров, формируемых полученными олигонуклеотидами, что в некоторых случаях приводит к уменьшению итогового сета. Производится расчет всех термодинамических параметров праймеров, формируется fasta-файл и таблицы свойств олигонуклеотидов.

2.5 Создание праймерных панелей для обогащения целевых групп вирусных патогенов

В подразделе описано создание праймерных панелей для обогащения фрагментов нукleinовых кислот вирусов различных групп интереса. Для проверки работоспособности панелей, дизайн которых был осуществлен при помощи разработанных алгоритмов, были созданы версии наборов олигонуклеотидов для тестирования *in vitro*. Далее приведены параметры разработанных библиотек олигонуклеотидов: праймерной панели для исследования вирома мигрирующих птиц; праймерной панели для обогащения коронавирусов; праймерной панели для обогащения энтеровирусов; праймерной панели для обогащения таксономических групп вирусов, вызывающих респираторные заболевания.

Была создана праймерная панель для исследования вирома мигрирующих птиц. Данная панель была создана первой версией алгоритма выбора праймерных последовательностей.

Благодаря экспериментальным данным, полученным с ее помощью, были выявлены и устранены недостатки алгоритма, относящиеся к способу расчета термодинамических параметров отжига праймеров и гетеродимерных структур. Панель состоит из 64 структур олигонуклеотидов и обеспечивает покрытие более 90% референсной базы данных.

Была создана праймерная панель для обогащения коронавирусов. В процессе создания тестовых версий этой панели, для отработки входных параметров алгоритма выбирались по несколько наборов олигонуклеотидов, различающихся задаваемыми значениями: минимальным значением температуры отжига; максимально допустимой температурой формирования вторичных структур и гетеродимеров; максимально допустимым количеством вырожденных позиций. Первый набор олигонуклеотидов cov_45_39 состоит из 20 структур обеспечивает покрытие 99 % последовательностей референсной БД.

Также были созданы праймерные панели для обогащения энтеровирусов. Для тестирования входных значений параметров алгоритма было создано несколько версий панелей: EV-1, состоящая из 6 структур, обеспечивающая покрытие > 99 % последовательностей исходной базы данных; EV-X, состоящая из 7 структур, обеспечивающая покрытие > 99 % последовательностей исходной базы данных; EV-Z состоящая из 7 структур, обеспечивающая покрытие > 99 % последовательностей исходной базы данных. Второй набор олигонуклеотидов на вирусы семейства *Coronaviridae* cov_48_42 состоит из 28 структур, обеспечивает покрытие 99 % последовательностей референсной БД.

Далее была создана праймерная панель для обогащения широкого спектра таксономических групп вирусов, вызывающих респираторные заболевания. Созданная с использованием разработанного алгоритма панель праймеров состоит из 78 структур, обеспечивает покрытие 99.81% последовательностей референсной БД. На рисунке ниже приведены свойства созданной библиотеки олигонуклеотидов (Рисунок 8).

Было проведено сравнение созданной панели с пулом праймеров, собранным из литературных источников, которое показало, что версия праймерной панели, созданная с использованием разработанного алгоритма, позволяет создавать праймеры с меньшим количеством вырождений, которые при этом покрывают большее количество вирусов внутри групп. Таким образом были созданы праймерные панели для обогащения различных групп вирусных патогенов. *In silico* произведен расчет термодинамических параметров выбранных олигонуклеотидов, температуры формирования нежелательных структур внутри каждой панели; проведена оценка покрытия референсных баз данных геномов, показан потенциально широкий охват целевых таксономических групп. Экспериментальные результаты, полученные при их *in silico* тестировании, на примере изучения свойств респираторных панелей, предсказали

значительное улучшение свойств библиотек олигонуклеотидов, созданных с использованием алгоритма, по сравнению с таковыми, собранными из литературных источников что особенно заметно при изучении свойств новых респираторных панелей.

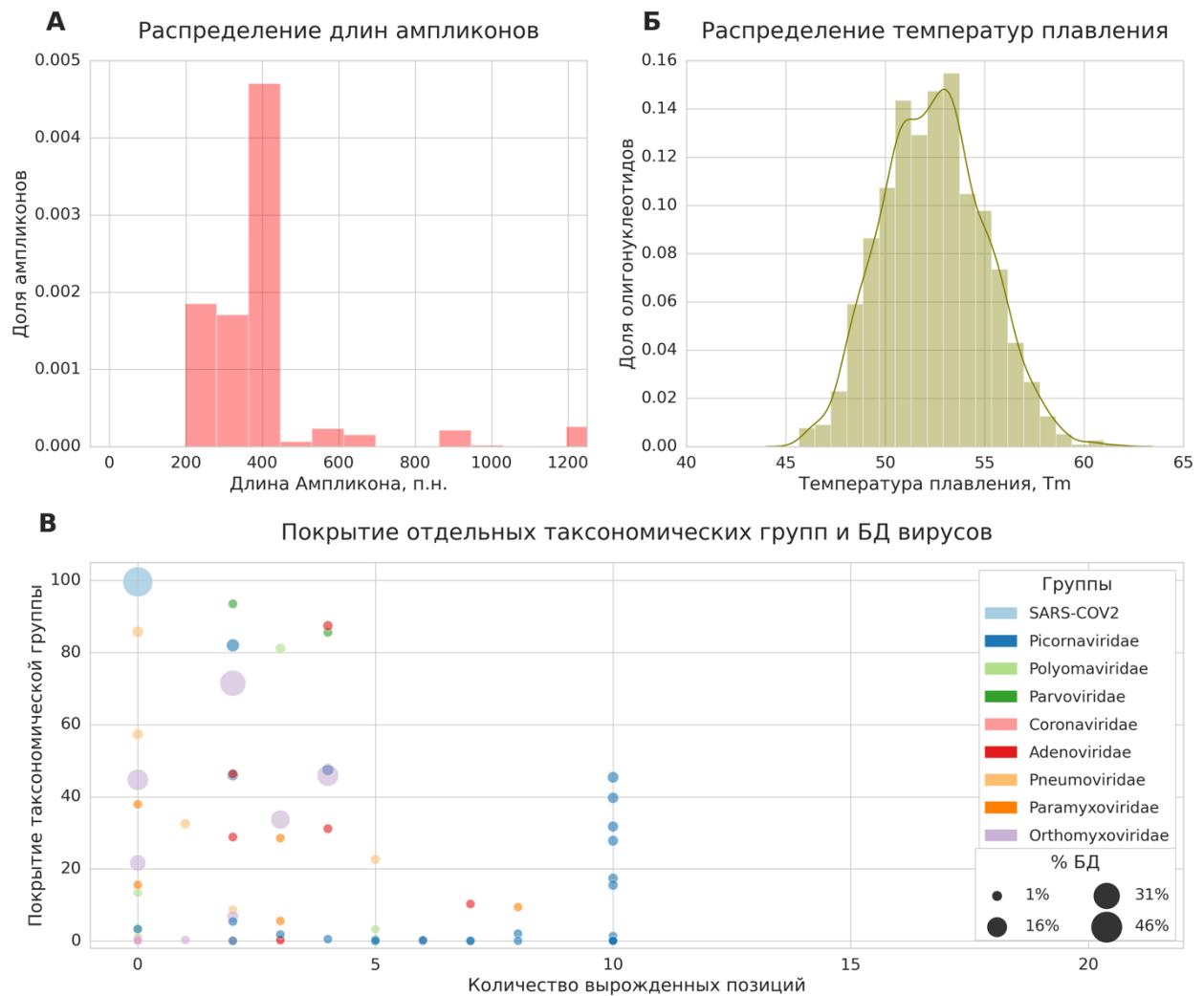


Рисунок 8 – Свойства панели праймеров для обогащения респираторных вирусов
А – распределение длин ампликонов; Б – распределение температур плавления; В – зависимость покрытия группы вирусов (высота расположения) и всей БД (диаметр точки) от кол-ва вырожденных позиций в паре

Глава 3. Разработка методики проведения исследования и ее апробация

В главе описан процесс создания лабораторной методики проведения исследования с использованием разработанных библиотек олигонуклеотидов и их апробация.

Одной из наиболее актуальных проблем диагностических методов, основанных на обнаружении нуклеиновых кислот патогена, является избыток нуклеиновых кислот организма-хозяина. Поскольку данное явление напрямую снижает чувствительность диагностических тестов, определение оптимальных условий выделения нуклеиновых кислот патогена необходимо

для повышения точности диагностических тестов, были проведены эксперименты по определению оптимальной методики экстракции нуклеиновых кислот из различного типа биологических материалов (кала, цельной крови, сыворотки, назофарингеального соскоба). Эксперимент проводился на специально созданных контрольных образцах с достоверной вирусной нагрузкой, одинаковой для каждой из тестируемых биологических проб. При выборе тестируемых наборов учитывались следующие характеристики: одновременное или последовательное выделение ДНК и РНК; высокое качество экстрагируемого материала; Широкий спектр допустимых типов биологического материала; возможность автоматизации процесса. Для оценки эффективности процесса экстракции для каждого типа биологического материала и каждого набора реагентов была проведена сравнительная ПЦР с детекцией сигнала флуоресценции в режиме реального времени. Оценка эффективности была осуществлена путем определения относительного количества вирусных нуклеиновых кислот каждого типа в элюате. Для указанной цели использовались специфичные праймеры, специально разработанные для данной задачи, а также коммерчески доступный интеркалирующий краситель EvaGreen (BIOTINUM). Для качественного анализа результатов амплификации продукты ПЦР были проанализированы методом агарозного электрофореза с использованием 2% геля. Учитывая приоритет специфичности, набор AllPrep ДНК/RNA Mini Kit признан оптимальным инструментом выделения целевых нуклеиновых кислот патогенов для таргетной амплификации вирусных нуклеиновых кислот с использованием специфичных праймеров.

Далее была разработана методика целевого обогащения для одновременного анализа ДНК- и РНК-содержащих вирусов. Для отработки метода использовались контрольные образцы с известной вирусной нагрузкой. В результате экспериментальной работы была определена оптимальная схема проведения амплификации, универсальная для РНК- и ДНК-содержащих вирусов; разработана лабораторная методика, позволяющая производить обогащение библиотек фрагментов целевыми локусами геномов вирусных патогенов с использованием разрабатываемых библиотек олигонуклеотидов.

С использованием контрольного биологического материала (образцов с достоверным содержанием нуклеиновых кислот отдельных патогенов) были отработаны различные режимы амплификации: подобран температурный режим и концентрации компонент реакционной смеси, оптимальные для разрабатываемых олигонуклеотидных панелей. Было показано, что для различных комбинаций пороговых параметров, задаваемых при применении разрабатываемого алгоритма дизайна олигонуклеотидных панелей, требуются различные условия проведения амплификации; определены соответствующие условия. Был отработан протокол приготовления NGS библиотек, оптимизированный для амплификаторов, получаемых с использованием

разрабатываемых панелей олигонуклеотидов, включающий следующие стадии: очистку амплификата, кинирование 5' концов, лигирование универсальных адаптерных последовательностей, индексацию, очистку финальных библиотек, нормализацию концентраций для проведения секвенирования. Эффективность разработанной методики была подтверждена результатами обработки данных секвенирования контрольных образцов.

Глава 4. Исследование клинических образцов, полученных от пациентов с респираторными заболеваниями

Глава содержит данные проведенного исследования клинических образцов, полученных от пациентов с респираторными заболеваниями. Исследовался клинический материал – назофарингеальные мазки, полученные от пациентов с острыми респираторными заболеваниями.

Разработанная методика была апробирована на клиническом материале, полученном от пациентов с острыми респираторными заболеваниями. Всего было исследовано 162 образца, из них 88 с датой поступления до начала пандемии коронавируса SARS-CoV-2.

В исследованной выборке образцов, датируемых до начала пандемии коронавирусной инфекции COVID-19, нуклеиновые кислоты, относимые к патогенным для человека вирусам, были выявлены в 40% образцов. В таблицах ниже приведены данные по частоте обнаружения в исследуемой выборке выявленных таксономических групп вирусных патогенов (Таблица 2, Таблица 3).

Таблица 2 — Данные по частоте обнаружения в исследуемой выборке выявленных таксономических групп вирусных патогенов. Разрешение до рода

Род	Кол-во образцов	Доля от исследованных, %	Доля от положительных, %
<i>Enterovirus</i>	21	23.9%	60.0%
<i>Mastadenovirus</i>	10	11.4%	28.6%
<i>Respirovirus</i>	1	1.1%	2.9%
<i>Alphacoronavirus</i>	1	1.1%	2.9%
<i>Betacoronavirus</i>	1	1.1%	2.9%
<i>Bocavirus</i>	1	1.1%	2.9%

Таблица 3 — Данные по частоте обнаружения в исследуемой выборке выявленных таксономических групп вирусных патогенов. Разрешение до вида

Род	Вид	Кол-во образцов	Доля, %
<i>Enterovirus</i>	<i>Rhinovirus C</i>	4	4.6%
	<i>Echovirus</i>	1	1.1%
	<i>Enterovirus A</i>	4	4.6%
	<i>Rhinovirus B</i>	2	2.3%
	<i>Enterovirus B</i>	2	2.3%
	<i>Rhinovirus A</i>	8	9.1%
<i>Mastadenovirus</i>	<i>Human mastadenovirus E</i>	5	5.7%
	<i>Human mastadenovirus C</i>	1	1.1%
	<i>Human mastadenovirus B</i>	4	4.6%
<i>Respirovirus</i>	<i>Human respirovirus I</i>	1	1.1%
<i>Alphacoronavirus</i>	<i>Human coronavirus 229E</i>	1	1.1%
<i>Betacoronavirus</i>	<i>Betacoronavirus I</i>	1	1.1%
<i>Bocavirus</i>	<i>Human bocavirus</i>	1	1.1%

Чаще всего выявлялись НК различных видов патогенных вирусов рода *Enterovirus*. На первом месте среди выявляемых видов - *Rhinovirus A*, являющийся наиболее частым возбудителем острых респираторных вирусных инфекций. На втором месте по частоте выявления мастаденовирусы (было найдено три различных вида), что также коррелирует со статистическими данными.

В выборке образцов, забор которых был осуществлён после начала пандемии коронавирусной инфекции, в 36% образцов были выявлены нуклеиновые кислоты возбудителя инфекции COVID-19 – вирус SARS-CoV-2. Нуклеиновые кислоты других патогенных для человека вирусов в исследованной выборке выявлено не было.

Полученные данные подтверждены зарегистрированными тест-системами на основе ПЦР в реальном времени. Всего было получено 3 дискордантных результата - для образцов 328 и 355, в которых ПЦР тест-системой были выявлены респираторно-синциальный вирус и адено-вирус человека соответственно; для образца 76, в котором был разработанным методом выявлен *Enterovirus A*. Полученные результаты объясняются большей чувствительностью ПЦР-тест системы относительно целевых адено- и респираторно-синциального вируса, а также отсутствием в используемом ПЦР наборе специфических олигонуклеотидов на *Enterovirus A*.

Таким образом, была показана работоспособность панели и эффективность методики для выявления вирусов 12 из 15 целевых таксономических групп. Оставшиеся три таксономические группы – *Alphainfluenzavirus*, *Gammainfluenzavirus*, *Betapolyomavirus* не были выявлены в исследуемом клиническом материале; контрольные образцы на данные группы вирусных патогенов отсутствовали в распоряжении автора работы.

Глава 5. Анализ вирома мигрирующих птиц околоводного комплекса

В главе описывается применение разработанного подхода для анализа вирома мигрирующих птиц околоводного комплекса в формате широкого скрининга. Приведены экспериментальные данные, данные обработки и интерпретации данных секвенирования.

5.1 Биологический материал

Объектом исследования были клоакальные смывы, ротовые мазки и помет птиц. Сбор полевого материала производился в период с августа 2017 по август 2018 года в Приморском крае: Спасский район, Устье р. Спасовка (образцы от 98-ми различных особей), Хасанский р-н, ур. Протока Пресная под Сюдари (образцы от 16-и различных особей), Хасанский р-н, ур. Островская (образцы от 76-и различных особей); в Сибири: берег р. Енисей, в р-не пос. Мирный (64 образца, от особей *Anser fabalis* (Гусь гуменник) и 45 образцов от особей различных видов воробьинообразных, уток и куликов); в Москве: Московский Зоопарк (собраны образцы от 37 различных особей), Краснодарский край (собраны образцы от 13 различных особей). Всего было собрано более 300 образцов.

Сбор материала был осуществлен с соблюдением этических принципов обращения с подопытными животными: биологический материал был собран квалифицированными учеными-зоологами, без причинения животным боли, физических страданий и повреждений.

Всё пробы биологического материала замораживались и транспортировались в лабораторию в замороженном виде. В лаборатории, вплоть до использования в эксперименте, образцы хранились в низкотемпературном холодильнике (- 82 °C).

5.2 Пробоподготовка

Пробоподготовка была проведена согласно разработанной схеме. Всего было исследовано 349 биологических образцов, взятых от 323-х индивидуальных особей.

5.3 Результаты анализа данных

В 64% исследованных образцов выявлены нуклеиновые кислоты вирусов, патогенных для позвоночных. Всего было выявлено 72 различных видов вирусов, относящихся к 25 родам и 19 различным семействам; в 12 образцах были выявлены фрагменты нуклеиновых кислот вирусов, неклассифицированных к определенной таксономической группе (роду и/или виду). Разработанная методика также позволила выявить случаи ко-инфекций (до 6 различных таксономических групп вирусов в одном образце). Более подробно таксономическая структура выявленных в образцах вирусов с частотой выявления более 1% от всех положительных образцов, представлена в таблицах ниже (Таблица 4).

Таблица 4 — Частота выявления НК вирусов различных таксономических групп вирусов, способных к инфицированию позвоночных

Таксон	Количество образцов	Доля от всех положительных образцов, %
<i>Hepadnaviridae/Avihepadnavirus/Duck hepatitis B virus/</i>	135	59.7%
<i>Adenoviridae/Aviadenovirus/Goose aviadenovirus A/</i>	42	18.6%
<i>Coronaviridae/Gammacoronavirus/Avian coronavirus/</i>	19	8.4%
<i>Adenoviridae/Atadenovirus/Duck atadenovirus A/</i>	8	3.6%
<i>Parvoviridae/Dependoparvovirus/Adeno-associated virus/</i>	7	3.1%
<i>Circoviridae/Circovirus/Tick circovirus/</i>	7	3.1%
<i>Astroviridae/Avastrovirus/Chicken astrovirus/</i>	6	2.7%
<i>Coronaviridae/Gammacoronavirus/Gammacoronavirus sp./</i>	6	2.7%
<i>Circoviridae//Circoviridae sp./</i>	5	2.2%
<i>Parvoviridae//Simian parvo-like virus 3/</i>	5	2.2%
<i>Adenoviridae/Aviadenovirus/Duck aviadenovirus B/</i>	5	2.2%
<i>Retroviridae/Alpharetrovirus/Avian leukosis virus/</i>	5	2.2%
<i>//Beihai noda-like virus 14/</i>	5	2.2%
<i>Orthomyxoviridae/Alphainfluenzavirus/Influenza A virus/</i>	4	1.8%
<i>Astroviridae/Avastrovirus/Duck astrovirus GII.A/</i>	4	1.8%
<i>Caliciviridae//Duck calicivirus/</i>	4	1.8%
<i>Iridoviridae//Cherax quadricarinatus iridovirus/</i>	3	1.3%
<i>Picornaviridae/Megrivirus/Goose megrivirus/</i>	3	1.3%
<i>Picornaviridae/Aalivirus/Aalivirus A/</i>	3	1.3%
<i>Coronaviridae//Bird droppings coronavirus/</i>	3	1.3%

Таксон	Количество образцов	Доля от всех положительных образцов, %
<i>Caliciviridae//Goose calicivirus/</i>	3	1.3%
<i>Circoviridae/Circovirus/European catfish circovirus/</i>	3	1.3%

Были выявлены ко-инфекции. Чаще всего выявлялась двойная инфекция *Aviadenovirus* и *Avihepadnavirus* (в 7% положительных образцов). Также часто встречалась комбинация *Avihepadnavirus* с *Gammacoronavirus* (3% образцов) и ассоциация *Aviadenovirus* с аденоассоциированным вирусом группы *Dependoparvovirus* (в 2% исследованных образцов, в которых были вирусные выявлены нуклеиновые кислоты).

Наибольшее количество локаций было проанализировано в Приморском крае. При этом образцы были забраны из трех локаций Хасанского района (сбор осуществлен в октябре 2017 года) и одной в Спасском районе (датируемые апрелем 2018 года). Спектры выявленных патогенных для животных вирусов по обеим локациям приведены в таблицах ниже.

Таблица 5 — Образцы локации: Приморский край, Хасанский р-н, ур. Островская, приведены таксономические группы вирусных патогенов, выявленные более чем в одном образце

Таксон	Количество положительных образцов	Доля от исследованных образцов из данного региона
<i>Hepadnaviridae/Avihepadnavirus/Duck hepatitis B virus/</i>	4	36.4%
<i>Coronaviridae/Gammacoronavirus/Avian coronavirus/</i>	2	18.2%
<i>Retroviridae/Alpharetrovirus/Avian leukosis virus/</i>	2	18.2%
<i>Adenoviridae/Aviadenovirus/Goose aviadenovirus A/</i>	2	18.2%

Таблица 6 — Образцы локации: Приморский край, Хасанский р-н, ур. Пресняк Сюдари-Тальми, приведены таксономические группы вирусных патогенов, выявленные более чем в одном образце

Таксон	Количество положительных образцов	Доля от исследованных образцов из данного региона
<i>Hepadnaviridae/Avihepadnavirus/Duck hepatitis B virus/</i>	17	26.2%
<i>Coronaviridae/Gammacoronavirus/Avian coronavirus/</i>	6	9.2%
<i>//Beihai noda-like virus 14/</i>	5	7.7%
<i>Circoviridae//Circoviridae sp./</i>	3	4.6%

Таксон	Количество положительных образцов	Доля от исследованных образцов из данного региона
<i>Orthomyxoviridae/Alphainfluenzavirus/Influenza A virus/</i>	3	4.6%
<i>Adenoviridae/Aviadenovirus/Goose aviadenovirus A/</i>	3	4.6%
<i>Paramyxoviridae/Avulavirus/Avian avulavirus 4/</i>	2	3.1%
<i>Adenoviridae/Aviadenovirus/Aviadenovirus sp./</i>	2	3.1%
<i>Paramyxoviridae/Avulavirus/Avian avulavirus 6/</i>	2	3.1%
<i>Birnaviridae/Avibirnavirus/Infectious bursal disease virus/</i>	2	3.1%
<i>Coronaviridae/Gammacoronavirus/Gammacoronavirus sp./</i>	2	3.1%
<i>Astroviridae/Avastrovirus/Duck astrovirus CPH/</i>	2	3.1%

Таблица 7 — Образцы локации: Приморский край, Хасанский р-н, ур. Протока Пресная под Сюдари, приведены таксономические группы вирусных патогенов, выявленные более чем в одном образце

Таксон	Количество положительных образцов	Доля от исследованных образцов из данного региона
<i>Hepadnaviridae/Avihepadnavirus/Duck hepatitis B virus/</i>	8	50.0%
<i>Coronaviridae/Gammacoronavirus/Avian coronavirus/</i>	6	37.5%
<i>Astroviridae/Avastrovirus/Chicken astrovirus/</i>	4	25.0%
<i>Coronaviridae/Gammacoronavirus/Gammacoronavirus sp./</i>	2	12.5%

Таблица 8 — Образцы локации: Приморский край, Спасский район, Устье р. Спасовка, приведены группы вирусных патогенов, выявленные более чем в одном образце

Таксон	Количество положительных образцов	Доля от исследованных образцов из данного региона
<i>Hepadnaviridae/Avihepadnavirus/Duck hepatitis B virus/</i>	76	77.6%
<i>Picornaviridae/Aalivirus/Aalivirus A/</i>	3	3.1%
<i>Picornaviridae/Megrivirus/Goose megrivirus/</i>	2	2.0%
<i>Coronaviridae/Gammacoronavirus/Gammacoronavirus sp./</i>	2	2.0%

По таблицам видно, что образцы, собранные в различных районах Приморского края и различных временных промежутках, отличаются по видовому составу вирома, однако в каждой

из проанализированных групп образцов в значительном количестве образцов (от 26 до 50% для Хасанского района и более 77% образцов из Спасского района) выявлялись вирусные нуклеиновые кислоты, относимы алгоритмом таксономической идентификации к *Duck hepatitis B virus*. Увеличение доли зараженных особей в образцах, собранных в 2018 году, по отношению к образцам 2017 года указывает на развивающийся характер поражения популяций птиц данным вирусом.

Наибольшим вирусным разнообразием отличались образцы из Сибири, сбор которых осуществлялся в августе 2017 года. Сбор образцов в данной области осуществлялся в 2-х локациях на берегу р. Енисей. Результаты анализа спектра вирусных патогенов по обеим локациям приведены в таблицах ниже.

Таблица 9 — Образцы, собранные после ночной стоянки перелетных гусей, приведены таксономические группы вирусных патогенов, выявленные более чем в двух образцах

Таксон	Количество положительных образцов	Доля от исследованных образцов из данного региона
<i>Adenoviridae/Aviadenovirus/Goose aviadenovirus A/</i>	36	56.3%
<i>Hepadnaviridae/Avihepadnavirus/Duck hepatitis B virus/</i>	26	40.6%
<i>Parvoviridae/Dependoparvovirus/Adeno-associated virus/</i>	7	10.9%
<i>Parvoviridae//Simian parvo-like virus 3/</i>	4	6.3%
<i>Coronaviridae/Gammacoronavirus/Avian coronavirus/</i>	3	4.7%
<i>Coronaviridae//Bird droppings coronavirus/</i>	3	4.7%

Таблица 10 — Образцы, собранные от различных видов воробьинообразных, уток и куликов, обитающих в районе п. Мирный

Таксон	Количество положительных образцов	Доля от исследованных образцов из данного региона
<i>Circoviridae/Circovirus/Tick circovirus/</i>	7	15.6%
<i>Adenoviridae/Atadenovirus/Duck atadenovirus A/*</i>	6	13.3%
<i>Hepadnaviridae/Avihepadnavirus/Duck hepatitis B virus/</i>	4	8.9%
<i>Circoviridae/Circovirus/European catfish circovirus/</i>	3	6.7%
<i>Iridoviridae/Megalocytivirus/Scale drop disease virus/</i>	2	4.4%
<i>Adenoviridae/Aviadenovirus/Duck aviadenovirus B/</i>	2	4.4%
<i>Iridoviridae//Cherax quadricarinatus iridovirus/</i>	2	4.4%
<i>Caliciviridae//Wuhan carp calicivirus 2/</i>	2	4.4%

Из таблиц видно, что таксономический состав вирома, выделенного из образцов от перелетных птиц отличается от такового, полученного из смешанного типа образцов от птиц, обитающих на данной территории. Тогда как в составе вирома перелетных птиц превалируют авидено- и ассоциированные с ними вирусы, виром, полученный от смешанной группы местных птиц представлен в основном цирковирусами, а также неизвестным вирусом, определяемым по кодирующими последовательностям, как близкий к Duck adenovirus A (отмечен знаком *). Кроме того, в составе вирома перелетных птиц наблюдалось значительное количество вирусных нуклеиновых кислот, идентифицируемых, как принадлежащие утиному вирусу гепатита В (более чем в 40% образцов). В то время как данный вирус был выявлен и среди местных птиц, однако в значительно меньшей доле образцов (около 9%), указывая на внешний источник повышения носительства вируса гепатита В в данном регионе.

В образцах, полученных из Москвы и Краснодарского края нуклеиновые кислоты вирусных патогенов были выявлены в единичных образцах. Таксономический состав выявленных вирусов приведен в таблицах ниже.

Таблица 11 — Образцы, собранные в г. Москва, Московский Зоопарк

Таксон	Количество положительных образцов	Доля от исследованных образцов из данного региона
<i>Picornaviridae/Megrivirus/Duck megrivirus/</i>	1	2.7%
<i>Papillomaviridae//Duck papillomavirus/</i>	1	2.7%
<i>Genomoviridae//Lynx canadensis faeces associated genomavirus CL1 58/</i>	1	2.7%
<i>Astroviridae/Avastrovirus/Duck astrovirus GII.A/</i>	1	2.7%
<i>Adenoviridae/Aviadenovirus/Pigeon adenovirus 2a/</i>	1	2.7%

Таблица 12 — Образцы, полученные из Краснодарского края

Таксон	Количество положительных образцов	Доля от исследованных образцов из данного региона
<i>Adenoviridae/Aviadenovirus/Duck aviadenovirus B/</i>	2	15.4%
<i>Coronaviridae/Gammacoronavirus/Avian coronavirus/</i>	1	7.7%
<i>Retroviridae//Corvus frugilegus retrovirus/</i>	1	7.7%
<i>Nodaviridae//Bat nodavirus/</i>	1	7.7%

Вируса гепатита В в данных выборках обрзцов выявлено не было.

5.4 Исследование нового представителя рода *Atadenovirus*

В шести биологических образцах, полученных из одной локации в Сибири, в формате широкого скрининга были выявлены НК, отнесенные алгоритмом таксономической идентификации к роду *Atadenovirus*, наиболее близкий вид - *Duck atadenovirus A*. При этом параметр идентичности при сравнении белковых последовательностей был равен 68,4%, что ниже среднего значения схожести кодирующих последовательностей внутри видов рода *Atadenovirus* (указанные параметры приведены в тексте далее).

Самым высоким содержанием вирусных НК отличался образец №81 и, следовательно, данный образец был подвергнут более тщательному анализу с целью исследования описаной находки. Было осуществлено повторное выделение вируной ДНК, проведено метагеномное секвенирование с увеличенной глубиной покрытия. Осуществлялась сборка *de novo*. Всего было получено 24465 контигов длиной от 128 до 14540 нуклеотидов, из них 32 – длиной более 1000 нуклеотидов; была проведена таксономическая классификация транслируемых нуклеотидных, а также транслированных белковых последовательностей полученных контигов, которая показала, что 15 из них могут быть отнесены к представителю семейства *Adenoviridae*. Эти контиги (максимальная длина 14540 нуклеотидов) были использованы для получения полногеномной сборки. Финальный скаффолд был получен вручную, путем визуальной валидации множественного выравнивания, включающего полученные промежуточные скаффолды по каждому из референсных последовательностей и всех представителей рода *Atadenovirus*. Для проверки качества сборки и коррекции ошибок было проведено картирование исходных прочтений на получившуюся сборку. Сборка была проверена на предмет однонуклеотидных ошибок и точек разрыва (стеков soft-, hard-clipped прочтений) как визуально, так и с помощью собственных скриптовых программ. Отдельно было оценено качество и целостность 3' конца сборки. Длина финальной сборки составила 31334 нуклеотида.

Для построения филогенетических деревьев последовательностей коровых белков, были взяты 63 представителя различных видов семейства *Adenoviridae*. На рисунках ниже приведены филогенетические деревья максимального правдоподобия, построенные на основе конкатенированных аминокислотных последовательностей 12 коровых генов, а также отдельно – по последовательности гена ДНК-полимеразы, отражающие эволюционные взаимоотношения представителей семейства. Жёлтым на цветных рисунках выделен род *Atadenovirus*, голубым – род *Aviadenovirus*, красным – род *Siadenovirus*, фиолетовым – род *Ichtadenovirus*, белым – род *Mastadenovirus*. Узел, соответствующий сборке 81, выделен зелёным квадратом. На отдельных рисунках с большим разрешением показан род *Atadenovirus*.

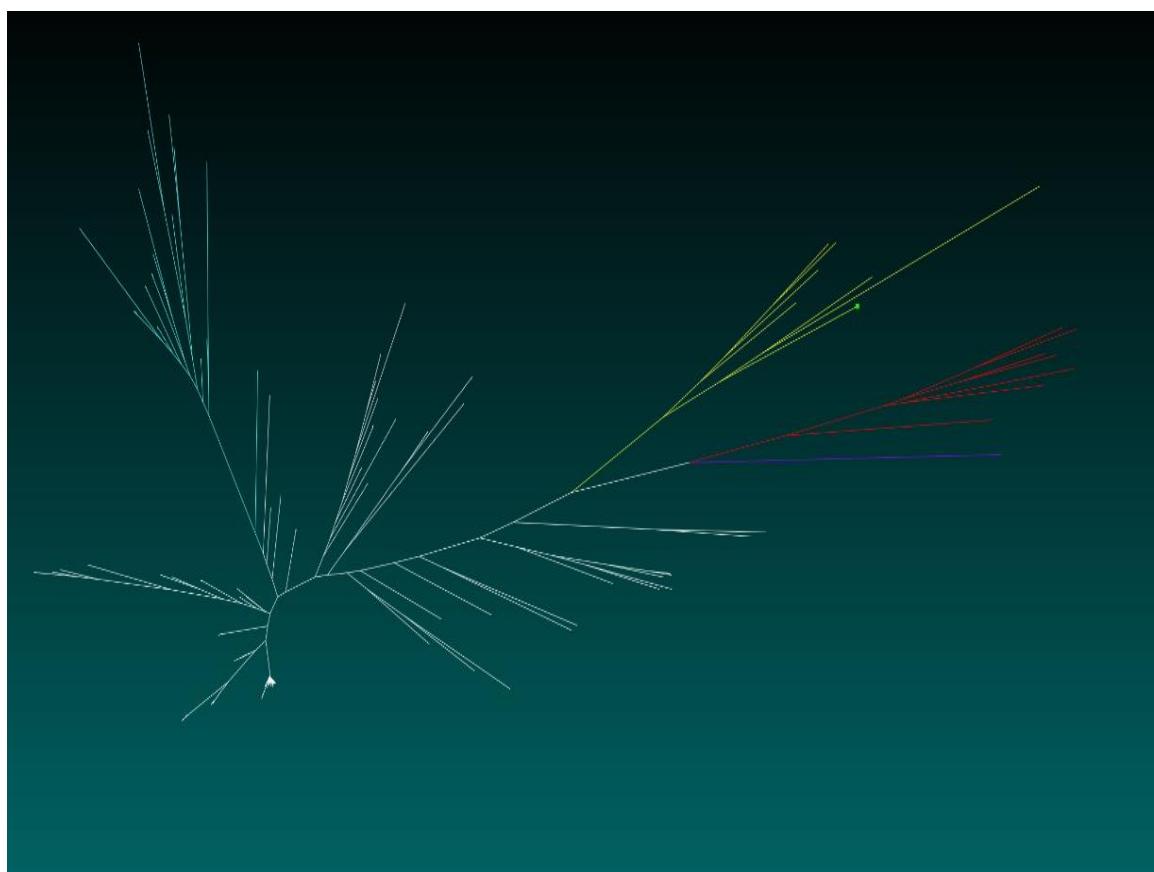


Рисунок 9 — Филогенетическое дерево, построенное по гену ДНК полимеразы для всех известных представителей семейства *Adenoviridae*

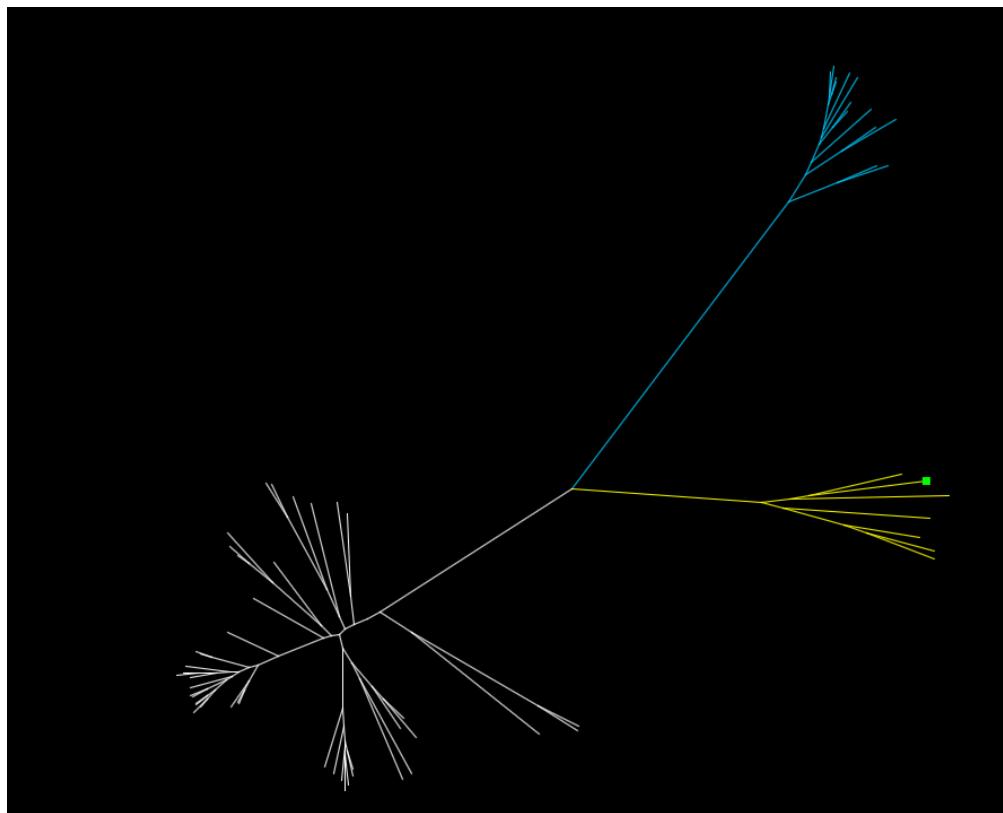


Рисунок 10 — Филогенетическое дерево, построенное по коровому геному для всех известных представителей семейства *Adenoviridae*



Рисунок 11 — Филогенетическое дерево, построенное по коровому геному для всех известных представителей рода *Atadenovirus*

В результате филогенетического анализа установлено, что исследуемый геном расположен в кладе филогенетического дерева, соответствующего роду *Atadenovirus*, что соответствует результатам, полученным при широком скрининге на разработанной панели олигонуклеотидов.

Для оценки видовой принадлежности нового представителя рода *Atadenovirus*, для всех известных представителей рода была рассчитана матрица схожести белковых последовательностей гена ДНК полимеразы, стандартно использующегося для таксономической дифференциации (где схожесть 1 - филогенетическая дистанция, полученная из матрицы дистанций, полученной программным обеспечением RAxML). Для видов, для которых известно несколько представителей (штаммов, серотипов), в качестве расчётного значение бралось среднее всех значений, полученных для каждого из представителей. При этом степень дивергентности последовательностей внутри видов не превышала 2% (Таблица 13).

Таблица 13 — Схожесть белковых последовательностей ДНК полимеразы внутри видов, для которых известно несколько представителей

Вид рода Atadenovirus	Средняя схожесть внутри вида
Duck atadenovirus A	99.88%
Deer atadenovirus A	99.95%
Psittacine atadenovirus A	98.07%

В таблице ниже (Таблица 14) приведены параметры схожести (максимальной, минимальной и медианной) каждого вида рода *Atadenovirus* ко всем остальным родам, а также к последовательности сборки 81.

Таблица 14 — Схожесть каждого вида рода *Atadenovirus* ко всем остальным родам, а также с последовательностью сборки 81

Вид рода Atadenovirus	K 81_assembly	Максимальная	Минимальная	Медианная
Bearded dragon adenovirus 1	69.70%	78.76%	67.55%	77.60%
Bovine adenovirus D	73.63%	93.23%	69.01%	81.91%
Bovine adenovirus E	73.55%	93.23%	68.68%	81.61%
Deer atadenovirus A	73.66%	87.45%	68.86%	81.79%
Duck adenovirus A	74.13%	82.50%	68.58%	81.77%
Lizard adenovirus 2	73.37%	88.18%	68.53%	82.07%
Ovine adenovirus D	73.41%	87.30%	68.86%	81.69%
Passerine adenovirus 1	63.00%	69.35%	63.00%	68.63%
Psittacine adenovirus A	71.63%	81.69%	69.35%	80.45%

Вид рода <i>Atadenovirus</i>	K 81_assembly	Максимальная	Минимальная	Медианная
Snake adenovirus 1	72.10%	88.18%	67.88%	80.98%

Значение схожести, рассчитанное для сборки 81 и каждого из известных видов рода *Atadenovirus*, лежит в пределах от минимального до медианного значения разброса этого параметра внутри рода для данного вида. А, следовательно, полученный геном вируса характеризуется степенью дивергенции от известных последовательностей, превышающей среднюю межвидовую дистанцию внутри рода.

Таким образом филогенетический анализ белковых последовательностей отдельных генов и полученной сборки корового генома, позволил подтвердить таксономическую принадлежность нового представителя к роду *Atadenovirus* со степенью дивергенции от известных последовательностей, превышающей среднюю межвидовую дистанцию внутри рода.

Глава 6. Сравнение разработанного метода с аналогами

В данной главе проводится сравнение разработанного метода с аналогами. В качестве методов сравнения были использованы метагеномный подход и коммерчески доступная система VirCapSeq-VERT. Сравнение проводилось с использованием контрольных образцов, для которых содержание вирусных нуклеиновых кислот было подтверждено предыдущими экспериментами. Степень обогащения рассчитывалась относительно результатов, полученных при метагеномном секвенировании, как референсного метода; относительно результатов VirCapSeq-VERT, как метода сравнения.

Разработанная методика показала лучшие или сравнимые результаты (в качестве критерия сравнения использовался следующий показатель: доля прочтений, классифицированная к целевой таксономической группе, рассчитанная относительно общего количества прочтений, полученных для образца) с системой VirCapSeq-VERT и метагеномным секвенированием для всех исследованных таксономических групп вирусов, кроме одной (*Alphainfluenzavirus*). Было показано, что алгоритм пробоподготовки, включающий целевую ПЦР на разработанной нами праймерной панели, позволяет добиться значительного обогащения фрагментов вирусных НК образцов по сравнению с метагеномным исследованием (до 51000 раз). Кроме того, амплификация позволяет обнаруживать вирусные НК, полностью отсутствующие в результатах метагеномного анализа.

Дополнительным преимуществом обогащения с помощью целевой ПЦР является более точное определение некоторых таксономических групп. В частности, видовой, принадлежности организмов, которым принадлежат амплифицируемые фрагменты нуклеиновых кислот

аденовирусов. Так, например, в метагеномном исследовании образца 60 автоматическим анализом были определены нуклеиновые кислоты, отнесенные алгоритмом классификации к роду *Mastadenovirus* – вирусам, вызывающим респираторные заболевания у млекопитающих, в том числе у человека. В то время как в исследовании разработанным методом с применением того же алгоритма анализа был выявлен *Goose adenovirus A*, что с большей вероятностью соответствует действительности. Рассмотренный эффект вероятно связан с ошибочной таксономической классификацией консервативных фрагментов генома адено-вируса, анализируемых в рамках метагеномного исследования, тогда как получаемый после целевой ПЦР фрагмент обладает большим таксономическим разрешением, позволяя проводить однозначную классификацию нуклеиновых кислот до вида.

Целевая ПЦР на разработанной панели праймеров позволяет обнаруживать нуклеиновые кислоты релевантных вирусов, значительно снижая требуемую глубину их секвенирования и сокращая количество артефактов. Кроме того, было показано, что выбранные локусы обеспечивают сравнимую или большую точность таксономической классификации, поскольку возрастает качество выходных данных за счёт многократного прочтения целевых областей. Таким образом, удается избежать множества нецелевых прочтений, без потерь качества анализа.

ЗАКЛЮЧЕНИЕ

В соответствии с задачами исследования, были достигнуты следующие результаты:

1. Разработан и реализован алгоритм создания библиотек олигонуклеотидов для проведения целевой амплификации нукleinовых кислот вирусов широкого спектра таксономических групп, использующий матричный способ заполнения, хранения и обработки данных о локальных выравниваниях коротких последовательностей олигонуклеотидов, позволяющий использовать высокоэффективные параллельные вычисления;
2. Разработаны наборы библиотек олигонуклеотидов для целевого обогащения нукleinовых кислот вирусных патогенов в режиме мультиплекса, а также универсальная лабораторная методика, позволяющая, с использованием созданных библиотек олигонуклеотидов и технологий высокопроизводительного секвенирования, проводить анализ широкого спектра вирусных патогенов в рамках одного исследования;
3. Проведена апробация разработанной методики, показана ее эффективность в обнаружении различных таксономических групп вирусов на контрольном материале и биологических образцах; полученные результаты подтверждены с использованием ПЦР тест систем;
4. С использованием разработанной методики проведено исследование вирома мигрирующих птиц, в результате которого были выявлены нукleinовые кислоты, относящиеся к более 70-и различным видам вирусов, патогенных для позвоночных;
5. С использованием разработанной методики выявлен новый представитель рода *Atadenovirus* (*Sterna Atadenovirus*), известные представители которого являются патогенными для млекопитающих, птиц и пресмыкающихся; показана возможность выявления с использованием разработанной методики вирида, степень дивергенции кодирующих последовательностей которого превышает межвидовую дистанцию внутри рода;
6. Осуществлено сравнение разработанного метода с аналогами (метагеномным секвенированием и коммерчески доступной системой VirCapSeq-VERT), в основе которых лежат технологии высокопроизводительного секвенирования, в котором разработанная методика показала лучшие или сравнимые результаты для большинства исследуемых таксономических групп, кроме того на порядки снижая требуемую глубину секвенирования и сокращая количество артефактов.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ РАБОТЫ

Статьи в рецензируемых научных журналах:

1. Ayginin A.A., Pimkina E.V., Matsvay A.D., Speranskaya A.S., Safonova M.V., Blinova E.A., Dedkov V.G., Shipulin G.A., Khafizov K., Artyushin I.V. THE STUDY OF VIRAL RNA DIVERSITY IN BIRD SAMPLES USING DE NOVO DESIGNED MULTIPLEX GENUS-SPECIFIC PRIMER PANELS // Advances in Virology. 2018. T. 2018. C. 3248285. DOI: 10.1155/2018/3248285
2. Kiselev D., Matsvay A. Abramov I., Dedkov V., Shipulin G., Khafizov K. Current Trends in Diagnostics of Viral Infections of Unknown Etiology // Viruses. 2020. Vol. 12, № 2. P. 211. DOI:10.3390/v12020211
3. Khafizov K. F., Speranskaya A. S., Matsvay A. D., Shipulin G. A., Dedkov V. G. ADVANCED TECHNOLOGIES IN DIAGNOSTICS OF VIRAL DISEASES OF UNKNOWN ETIOLOGY // Инфекция и иммунитет. 2020. Т. 10. № 1. С. 9-25. DOI:10.15789/2220-7619-ATI-82
4. Makenov MT, Toure AH, Bayandin RB, Gladysheva AV, Shipovalov AV, Sanaba B, Korneev MG, Yakovlev SA, Zhurenkova OB, Grigoreva YaE, Fyodorova MV, Radyuk EV, Morozkin ES, Boiro MY, Khafizov K, Matsvay A, Karan LS NGARI VIRUS (ORTHOBUNYAVIRUS, PERIBUNYAVIRIDAE) IN IXODID TICKS COLLECTED FROM CATTLE IN GUINEA // Acta Tropica. 2021. Vol. 214. P. 105790. DOI: 10.1016/j.actatropica.2020.105790

Результаты интеллектуальной деятельности:

5. Мацвай А. Д., Айгинин А. А., Хафизов К. Ф. Шипулин Г.А. Программа для дизайна первичных структур библиотек олигонуклеотидов для целевого обогащения нуклеиновых кислот вирусных патогенов в режиме мультиплекса // Свидетельство на программу для ЭВМ. (Номер свидетельства: 2021660832 от 01.07.2021 г.)

Тезисы конференций:

6. Мацвай А.Д., Пимкина Е.В., Альборова И.Э., Хафизов К.Ф., Мустафин Х.Х. РАЗРАБОТКА МЕТОДА ОБОГАЩЕНИЯ NGS БИБЛИОТЕК ЦЕЛЕВЫМИ ЛОКУСАМИ, ОСНОВАННОЙ НА ДНК-ДНК ГИБРИДИЗАЦИИ // В сборнике: Молекулярная диагностика 2018 Сборник трудов Международной научно-практической конференции. 2018. С. 388-389. ISBN: 978-985-7091-99-7
7. Matsvay A.D., Ayginin A.A., Pimkina E.V., Speranskaya A.S., Safonova M.V., Dedkov V.G., Shipulin G.A., Khafizov K. AVOIDING MULTIPLE SEQUENCE ALIGNMENTS FOR THE EFFICIENT DESIGN OF NGS PANELS FOR THE DETECTION OF VIRUSES // В сборнике:

Молекулярная диагностика 2018 Сборник трудов Международной научно-практической конференции. 2018. С. 379-380. ISBN: 978-985-7091-99-7

8. Matsvay, A.; Kiselev, D.; Ayginin, A.; Abramov, I.; Dedkov, V.; Shipulin, G.; Khafizov, K. Metabarcoding-Like Approach for High Throughput Detection and Identification of Viral Nucleic Acids. *Proceedings* 2020, 50, 136. DOI:10.3390/proceedings2020050136
9. Khafizov K., Ayginin A.A., Pimkina E.V., Matsvay A.D., Speranskaya A.S., Safonova M.V., Artyushin I.V., Dedkov V.G., Shipulin G.A. THE STUDY OF VIRAL RNA DIVERSITY IN BIRD SAMPLES USING DE NOVO DESIGNED MULTIPLEX GENUS-SPECIFIC PRIMER PANELS // В сборнике: Молекулярная диагностика 2018 Сборник трудов Международной научно-практической конференции. 2018. С. 384-385. ISBN: 978-985-7091-99-7
10. KHAFIZOV K., AYGININ A.A., PIMKINA E.V., MATSVAY A.D., SPERANSKAYA A.S., SAFONOVA M.V., BLINOVA E.A., ARTYUSHIN I.V., DEDKOV V.G., SHIPULIN G.A. Design of genus-specific primer panel for detection and identification of viral DNA in environmental samples using next-generation sequencing // BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE\SYSTEMS BIOLOGY (BGRS\SB-2018) 2018. С. 49–49. DOI: 10.18699/BGRSSB-2018-026
11. Korneenko E.V., Samoilov A.E., Artyushin I.V., Dudorova A.V., Pimkina E.V., Dedkov V.G., Safonova M.V., Matsvay A.D., Speranskaya A.S. DETECTION OF ALPHACORONAVIRUS IN BAT FECAL SAMPLES FROM VOLGOGRAD REGION // BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE/SYSTEMS BIOLOGY (BGRS/SB-2020). The Twelfth International Multiconference Abstracts. 2020. С. 47. DOI: 10.18699/BGRS/SB-2020-026