

На правах рукописи

**АНТОНЕЦ (СТАРЧЕВСКАЯ)
МАРИЯ ЕВГЕНЬЕВНА**

**ИДЕНТИФИКАЦИЯ И СРАВНИТЕЛЬНЫЙ
АНАЛИЗ ГЕНЕТИЧЕСКИХ
ПОСЛЕДОВАТЕЛЬНОСТЕЙ ВИРУСОВ НА
ОСНОВЕ ГЕНОМНЫХ И ТРАНСКРИПТОМНЫХ
ДАННЫХ КОЛОРАДСКОГО ЖУКА
(LEPTINOTARSA DECEMLINEATA)**

1.5.8 – математическая биология, биоинформатика
(биологические науки)

АВТОРЕФЕРАТ
диссертации на соискание
ученой степени кандидата
биологических наук

Новосибирск - 2025

Работа выполнена в теоретическом отделе ФБУН Государственный научный центр вирусологии и биотехнологии «Вектор» Роспотребнадзора

Научный руководитель: **Антонец Денис Викторович**, кандидат биологических наук, старший научный сотрудник, Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем МГУ имени М.В. Ломоносова, г. Москва

Официальные оппоненты: **Колпаков Федор Анатольевич**, доктор биологических наук, заведующий лабораторией биоинформатики ФИЦ Информационных и вычислительных технологий, Федеральная территория «Сириус»

Игорь Викторович Бабкин, доктор биологических наук, доцент, ведущий научный сотрудник, Институт химической биологии и фундаментальной медицины СО РАН, г. Новосибирск

Ведущая организация: ФГБУН Институт общей генетики им. Н.И. Вавилова РАН, г.Москва

Защита диссертации состоится «___»_____20__ г. на утреннем заседании Диссертационного совета 24.1.239.01 на базе ФГБНУ «Федеральный исследовательский центр Институт цитологии и генетики СО РАН» в конференц-зале Института по адресу: 630090, г. Новосибирск, проспект ак. Лаврентьева, 10, т. (383)363-49-06, факс (383) 333-12-78, e-mail: dissov@bionet.nsc.ru.

С диссертацией можно ознакомиться в библиотеке ИЦиГ СО РАН и на сайте Института <http://www.icgbio.ru/>

Автореферат разослан «___»_____20__г.

Ученый секретарь
диссертационного совета,
доктор биологических наук

Т.М. Хлебодарова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Метагеномный подход оказался высокоэффективным способом обнаружения вирусов в различных средах (Bassi и др. 2022; Roux, Matthijnssens, и Dutilh 2021; Varghese и van Rij 2018). В отличие от традиционных культурально-диагностических методов вирусная метагеномика использует высокопроизводительное секвенирование нового поколения, что позволяет идентифицировать редкие и новые вирусы, а также устанавливать экологические связи между микроорганизмами и их естественной средой обитания (Mohsin и др. 2021; Rosario и Breitbart 2011; Santiago-Rodriguez и Hollister 2022). Передача вирусов насекомыми является одним из наиболее распространенных путей передачи вирусов. Насекомые представляют самую многочисленную группу животных на планете и имеют важное экологическое, сельскохозяйственное и медицинское значение (Varghese и van Rij 2018). Однако наши знания о вирусах насекомых все еще ограничены. Изучение вириона насекомых даст новое понимание экологии и эволюции вирусов и поможет решить проблемы здравоохранения в борьбе с арбовирусными инфекциями. Кроме того, патогенные вирусы насекомых могут быть использованы для биологической борьбы с насекомыми-вредителями, и, таким образом, идентификация новых вирусов насекомых может расширить арсенал потенциальных агентов биоконтроля (Varghese и van Rij 2018; Wagemans и др. 2022).

Одним из самых серьезных вредителей сельского хозяйства является колорадский жук (*Leptinotarsa decemlineata*, семейство Chrysomelidae). Его способность адаптироваться к различным пасленовым растениям, высокая экологическая пластичность и быстрое развитие устойчивости к инсектицидам привели к широкому распространению этого вредителя по всему миру (Cingel и др. 2016). Без использования инсектицидов колорадский жук может уничтожать до 40-80% урожая картофеля (Maharajaya и Vosman 2015). Хотя применение инсектицидов привело к резкому сокращению популяции колорадского жука, у него наблюдается развитие устойчивости ко всем зарегистрированным к настоящему времени химическим препаратам. Увеличение дозировки обеспечивает кратковременное улучшение, но значительно увеличивает скорость развития резистентности (Alyokhin и др. 2008). Несколько исследовательских групп установили причины столь быстрых эволюционных изменений при изучении генетических механизмов возникновения устойчивости колорадского жука к инсектицидам и его адаптации к факторам внешней среды (Schoville и др. 2018). Его высокая способность к акклиматизации связана со специфическими генами, связанными с изменением условий метаболизма, роста и диапаузы у насекомых (Cingel и др. 2016). Механизмы детоксикации растительных соединений и, скорее всего, инсектицидных химических соединений обусловлены широким спектром высоковариабельных генов, включая гены карбоксилэстеразы, глутатион-S-трансферазы, цитохрома P450, пищеварительных ферментов и рецепторов горького вкуса, и связаны со

способностью колорадского жука питаться на различных пасленовых растениях (Schoville и др. 2018). Можно предположить, что в будущем колорадские жуки разовьют устойчивость и ко всем новым инсектицидам (Alyokhin и др. 2008). Кроме того, широкое использование химических пестицидов может привести к серьезным экологическим проблемам из-за неспецифического воздействия на другие виды животных.

Биологические методы являются наиболее экологически безопасными способами борьбы с различными вредителями сельскохозяйственных культур (Bale, van Lenteren и Bigler 2008; van Lenteren и др. 2018). Поиск агентов биологического контроля численности популяций вредителей имеет большой потенциал, но лишь небольшое количество «естественных врагов» можно выращивать в больших масштабах. Кроме того, развитие популяции естественного врага часто происходит медленнее, чем популяции насекомых-мишеней, и для оптимального роста требуются другие условия (Tanwar, Dureja, и Rathore 2012). В настоящее время в для биологической борьбы с колорадским жуком используются патогенные грибы, такие как *Beauveria bassiana*, *Metarhizium sp.* (Alyokhin и Kryukov 2022) и бактерии *Bacillus thuringiensis* (Yu, Yuan и Gao 2016), при этом средства на основе вирусов отсутствуют. Стоит отметить, что энтомопатогенные вирусы в настоящее время используются для борьбы с другими вредителями, например препараты на основе вирусов семейства *Baculoviridae* широко используются для борьбы с насекомыми-вредителями отряда Lepidoptera (Harrison и Hoover 2012).

В настоящее время известно более 20 семейств вирусов, поражающих различные группы насекомых (Ros и др. 2022). Многие вирусы, переносимые насекомыми, вызывают арбовирусные инфекции и могут передаваться в том числе человеку, вызывая серьезные заболевания (Jones 2018). Тем не менее большинство вирусов насекомых не опасны для человека. Их используют в качестве модельных объектов, платформ для разработки векторных систем и систем экспрессии, и для борьбы с насекомыми-вредителями (Varghese и van Rij 2018). Предполагается, что жуки-листоеды семейства Chrysomelidae поражаются вирусами семейств *Baculoviridae* и *Iridoviridae* (Selman 1988). Но несмотря на то, что колорадский жук является одним из наиболее серьезных и широко распространенных вредителей сельского хозяйства, его виром в настоящее время практически полностью не изучен, и поэтому изучение вирусов колорадского жука является чрезвычайно важной и актуальной задачей.

Целью работы является обнаружение и изучение разнообразия вирусного генетического материала в геномных и транскриптомных данных колорадского жука, а также идентификация вирусов потенциально патогенных для колорадского жука, которые могли бы стать основой для разработки новых препаратов для биологического контроля численности этого опасного вредителя сельского хозяйства. Для достижения данной цели были поставлены следующие задачи:

1. Сбор опубликованных геномных и транскриптомных данных *L. decemlineata* из базы данных NCBI SRA и создание программного конвейера для поиска и идентификации в них вирусного генетического материала;
2. Поиск, идентификация и анализ в собранном наборе данных высокопроизводительного секвенирования генетического материала, имеющего гомологию с вирусными последовательностями; идентификация потенциальных эндогенных вирусных элементов в геноме *L. decemlineata*;
3. Секвенирование генетического материала, выделенного из стерильных яиц *L. decemlineata*, с помощью платформы секвенирования нового поколения Oxford Nanopore Technologies (ONT, MinION); идентификация в полученных данных браковирусных последовательностей и других эндогенных вирусных элементов;
4. Анализ метатранскриптомных данных, полученных с помощью секвенирования нового поколения материала от личинок *L. decemlineata* IV возраста с характерными внешними проявлениями неизвестной летальной инфекции с целью выявления вирусного генетического материала; сборка геномов обнаруженных вирусов, их аннотация и филогенетический анализ.

Научная новизна. Впервые проведен поиск и идентификация вирусного генетического материала в геномных и транскриптомных образцах колорадского жука. Из анализа были исключены последовательности, отнесенные к бактериофагам и мимивирусам. Обнаруженные вирусные последовательности принадлежат к 32 различным вирусным семействам. Большинство последовательностей были отнесены к вирусам насекомых (*Adintoviridae*, *Ascoviridae*, *Baculoviridae*, *Dicistroviridae*, *Chuviridae*, *Hytrosaviridae*, *Iflaviridae*, *Iridoviridae*, *Nimaviridae*, *Nudiviridae*, *Phasmaviridae*, *Picornaviridae*, *Polydnnaviriformidae*, *Xinmoviridae* и др.), вирусам растений (*Betaflexiviridae*, *Bromoviridae*, *Kitaviridae*, *Potyviridae*) и эндогенным ретровирусным элементам (*Retroviridae*, *Metaviridae*). Высказана гипотеза о том, что значительная часть неаннотированных белков в составе протеома насекомых, и в частности, колорадского жука, могут иметь вирусное происхождение.

Впервые в генетическом материале колорадского жука было показано наличие протяженных фрагментов браковирусных генетических последовательностей и показана их интеграция в геном.

Впервые проведен анализ генетического материала, выделенного из стерильных яиц колорадского жука, с помощью платформы секвенирования ONT. В результате в составе генома колорадского жука было подтверждено наличие интегрированных протяженных браковирусных фрагментов; обнаружено большое количество эндогенных вирусных элементов, происходящих из различных вирусных семейств.

Впервые обнаружены вирусы насекомых, поражающие колорадского жука. В результате метагеномного секвенирования биологического материала, выделенного из предкуколок колорадского жука, погибших от неизвестной инфекции, были собраны практически полногеномные последовательности двух вирусов – *Leptinotarsa iflavivirus 1* (GenBank: OR613011) и *Leptinotarsa solinvi-like virus 1* (GenBank: OR613010). Показано, что *Leptinotarsa iflavivirus 1* может быть связан с летальной инфекцией колорадского жука.

Теоретическая и практическая значимость исследования. Несмотря на чрезвычайно важное значение колорадского жука для сельского хозяйства, его виром в настоящее время не изучен, и данная работа является первой, направленной на поиск и идентификацию вирусного генетического материала в биологических образцах *L. decemlineata*. Высокая теоретическая значимость работы обусловлена тем, что изучение разнообразия вирусов, поражающих колорадского жука, и вирусов, переносимых данным вредителем, позволит углубить наши знания в области его экологии и улучшить понимание его взаимоотношений с естественными врагами, патогенами и кормовыми растениями. Стоит отметить, что программный конвейер, разработанный в рамках представленного исследования, также может быть применен для поиска, идентификации и изучения вирусного генетического материала в геномных и транскриптомных данных других организмов. Помимо важного теоретического значения, работа также имеет и высокую практическую значимость. Обнаружение новых вирусов, потенциально способных вызывать летальную инфекцию у колорадского жука, открывает перспективы расширения арсенала средств биологического контроля численности колорадского жука за счет разработки препаратов на основе энтомопатогенных вирусов, которые в дальнейшем могут помочь не только снизить экономические потери в сельском хозяйстве, но и уменьшить негативные экологические последствия от использования пестицидов.

Положения, выносимые на защиту:

1. Разработан оригинальный программный конвейер uncoVir для поиска генетического материала вирусов в данных высокопроизводительного секвенирования (DNA- и RNA-Seq), позволивший идентифицировать у колорадского жука (*Leptinotarsa decemlineata*) последовательности вирусов насекомых и растений, а также эндогенные вирусные элементы, принадлежащие к более чем 30 различным семействам.
2. Генетические последовательности вирусов, обнаруженные в личинках колорадского жука, погибших от неизвестной инфекции, филогенетически связаны с ифлавирусами и солинви-вирусами и принадлежат новым, ранее не описанным вирусам, аннотированным нами как *Leptinotarsa iflavivirus 1* (GenBank: OR613011) и *Leptinotarsa solinvi-like virus 1* (GenBank: OR613010).

Личный вклад автора. Постановка задач, планирование экспериментов, биоинформационический анализ данных, написание аналитического программного конвейера и секвенирование образцов на приборе MinION (Oxford Nanopore Technologies) выполнены автором самостоятельно. Сбор биологических образцов, подготовка биоматериала, количественный ПЦР выполнены коллегам из ИСиЭЖ СО РАН (У.Н. Роцкой, О.В. Поленовой, Е.С. Косман и В.Ю. Крюковым). Выделение генетического материала, ПЦР и секвенирование образцов с использованием платформ Illumina MiSeq и iSeq и секвенирование по Сэнгеру выполнены коллегами из ФБУН ГНЦ ВБ "Вектор" Роспотребнадзора (С.А. Бодневым, Т.В. Трегубчак и Т.В. Бауэр).

Апробация работы и публикации. Научные результаты, изложенные в данной работе, были представлены на 4 всероссийских и международных конференциях в виде стеновых и устных докладов. А именно: Starchevskaya M.E., Kamanova E.P., Vyatkin Y.V., Tregubchak T.V., Bauer T.V., Rotskaya U.N., Kosman E.S., Antonets D.V. The rich inner world of Colorado potato beetles – a metagenomic survey of viral diversity in public data. 10th Moscow Conference on Computational Molecular Biology MCCMB'21, Moscow, 2021. Proceedings of 10th MCCMB'21, P.184; Starchevskaya M.E., Kamanova E.P., Vyatkin Y.V., Tregubchak T.V., Bauer T.V., Rotskaya U.N., Kosman E.S., Antonets D.V. The metagenomic analysis of Colorado potato beetles public NGS data and the nanopore sequencing of its genetic material. Thirteenth International Multiconference Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2022), Novosibirsk, 2022. Abstracts, P. 62; Старчевская М.Е., Каманова Е.П., Вяткин Ю.В., Трегубчак Т.В., Бауэр Т.В., Роцкая У.Н., Косман Е.С., Антонец Д.В. Метагеномный анализ вирома *Leptinotarsa decemlineata*. III Всероссийская конференция Высокопроизводительное секвенирование в геномике (HSG-2022), Новосибирск, 2022. Сборник тезисов. С. 75; Антонец М.Е., Крюков В.Ю., Боднев С.А., Роцкая У.Н., Косман Е.С., Трегубчак Т.В., Бауэр Т.В., Антонец Д.В. Первые обнаруженные вирусы колорадского жука: перспективы для биологического контроля. 2024. V Всероссийский конгресс по защите растений, 16-19 апреля 2024 г., Санкт-Петербург. Сборник тезисов. С. 61.

Основные результаты были изложены в 2 статьях, опубликованных в рецензируемых научных журналах первого квартриля (Q1) *Viruses* и *Scientific Reports*, индексируемых в международных базах данных Web of Science/Scopus

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, выводов и списка литературы (144 источника). Работа изложена на 128 страницах, проиллюстрирована 16 рисунками, содержит 4 таблицы и 3 приложения.

СОДЕРЖАНИЕ РАБОТЫ

Глава 1. Обзор литературы

Данная глава с одной стороны освещает метагеномные подходы для поиска вирусов в данных высокопроизводительного секвенирования, с другой стороны описывает экологию и генетические особенности колорадского жука. Также охарактеризованы некоторые семейства энтомопатогенных вирусов, такие как *Baculoviridae*, *Iridoviridae*, *Polydnaviriformidae*, *Iflaviridae*, *Dicistroviridae*. Отдельное внимание удалено вирусам растений, которые могут переноситься насекомыми, в частности представителями отряда Coleoptera.

Глава 2. Материалы и методы.

Для подготовки данной работы использовались различные биоинформационные методы и программы, методы статистического анализа. Весь основной код написан на языках программирования Bash и Python, для создания основного программного конвейера использовался язык Snakemake. В ходе работы были освоены и использованы различные методы, программы и программные конвейеры для анализа геномных и транскриптомных данных высокопроизводительного секвенирования. Было освоено секвенирование генетического материала с использованием прибора MinION (Oxford Nanopore Technologies).

Главы 3-4. Результаты и обсуждение.

Разработка аналитического конвейера для метагеномного анализа публичных данных высокопроизводительного секвенирования *L. decemlineata*

Для анализа публичных геномных и транскриптомных данных 297 образцов *L. decemlineata*, полученных из базы данных NCBI SRA, нами был построен аналитический программный конвейер uncoVir (рис. 1). Конвейер, написанный на языке программирования Snakemake, использует несколько подходов для классификации генетических последовательностей и доступен по адресу: <https://github.com/starchevskayamaria17/uncoVir>. Подходы включают анализ спектра k-меров, *de novo* сборку прочтений в контиги с последующей аннотацией и таксономической классификацией, выравнивание прочтений на базу данных вирусных генетических последовательностей. В качестве базы данных была использована база данных собственной сборки BigViralDB, собранная из четырех баз данных: NCBI RefSeq (11 566 вирусных последовательностей; <https://www.ncbi.nlm.nih.gov/refseq/>, июнь 2021), Virxicon (327 932 вирусных последовательностей, июнь 2021) (M. Kudla, K. Gutowska, J. Synak, M. Weber, K. S. Bohsack, P. Lukasiak, T. Villmann, J. Blazewicz 2020), Virosaurus (823 421 вирусная последовательность, апрель

2020) (Gleizes et al. 2020), вирусные последовательности из базы данных NCBI Nucleotide (1 367 485 вирусных последовательностей, июнь 2021).

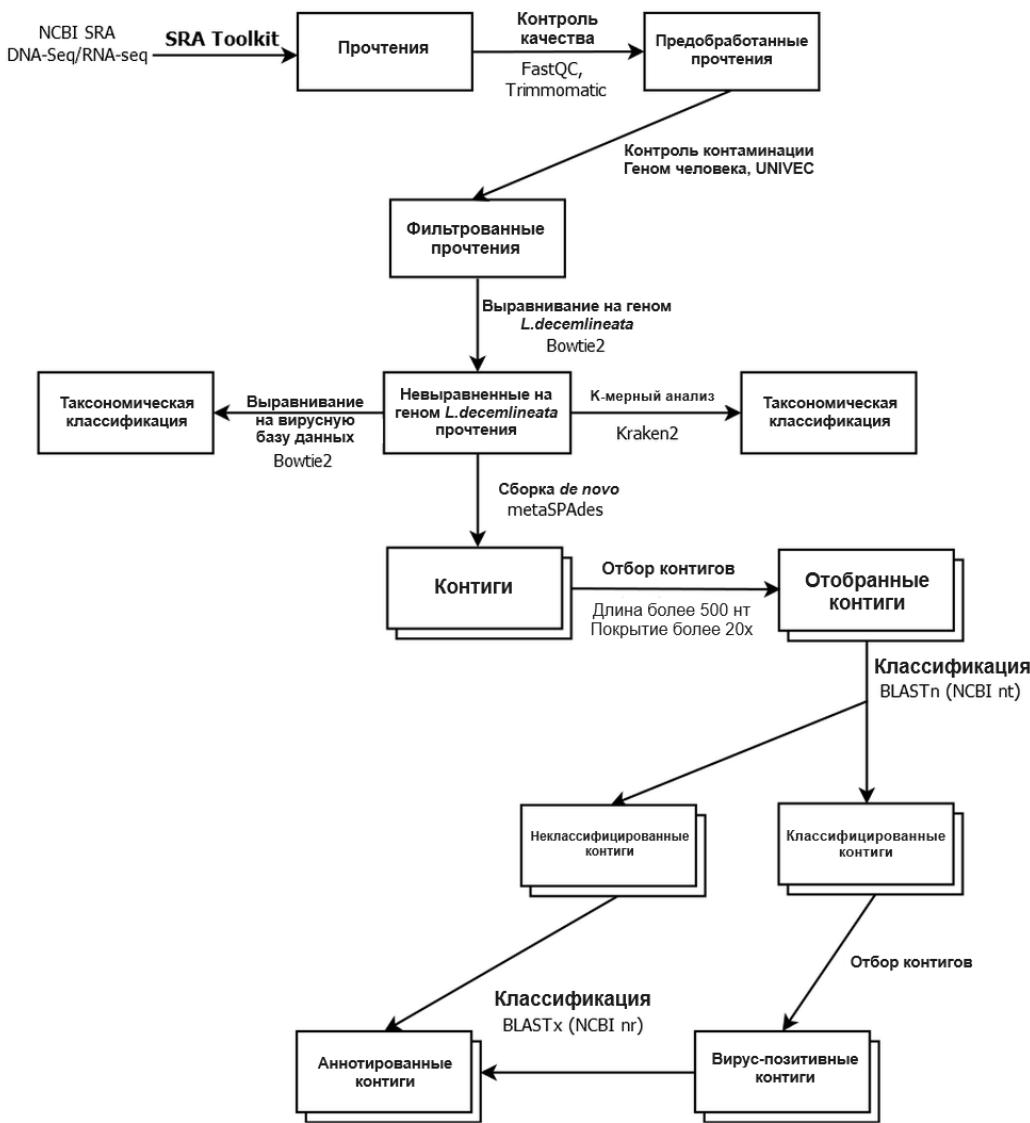


Рис. 1. Программный конвейер, использованный в работе. Указаны этапы анализа, их последовательность и использованные программы. <https://github.com/starchevskayamaria17/uncovir>.

В результате сборки прочтений в контиги с последующей аннотацией были получены 1 652 контига, которые относятся к различным видам вирусов с неустановленным вирусным семейством, и 2 950 контига с установленным вирусным семейством (исключена информация о контигах, белковые продукты которых имеют гомологию с белками бактериофагов). Вирус-позитивные контиги с установленной таксономической принадлежностью вирусного вида принадлежат к 6 типам вирусов: Artverviricota, Kitrinoviricota, Negarnaviricota, Nucleocytoviricota, Pisuviricota, Preplasmiviricota, распределенных по 17 порядкам, 32 вирусным семействам и 97 видам. Также обнаружены по меньшей мере 48 видов вирусов с неопределенным вирусным

семейством (рис. 2).

Среди аннотированных последовательностей контиги, имеющие вирусное происхождение, можно разделить на несколько групп: гомологичные генетическим последовательностям вирусов насекомых (представители семейств *Bornaviridae*, *Caliciviridae*, *Dicistroviridae*, *Iflaviridae*, *Iridoviridae*, *Orthomyxoviridae*, *Partitiviridae*, *Picornaviridae*, *Poxviridae*, *Virgaviridae* и *Xinmoviridae*), вирусам растений (представители семейств *Betaflexiviridae*, *Bromoviridae*, *Kitaviridae* и *Potyviridae*), эндогенным ретровирусным элементам (*Retroviridae*, *Metaviridae*), эндогенным вирусным элементам неретровирусного происхождения (*Adintoviridae*, *Ascoviridae*, *Chuviridae*, *Hytrosaviridae*, *Nimaviridae*, *Nudiviridae*, *Phasmaviridae*, *Polydnnaviriformidae*, *Rhabdoviridae* и *Baculoviridae*), последовательностям бактериофагов и вирусам простейших (не описаны в рамках данного исследования) и многочисленным последовательностям вирусов с неустановленным семейством, поражающих преимущественно насекомых.

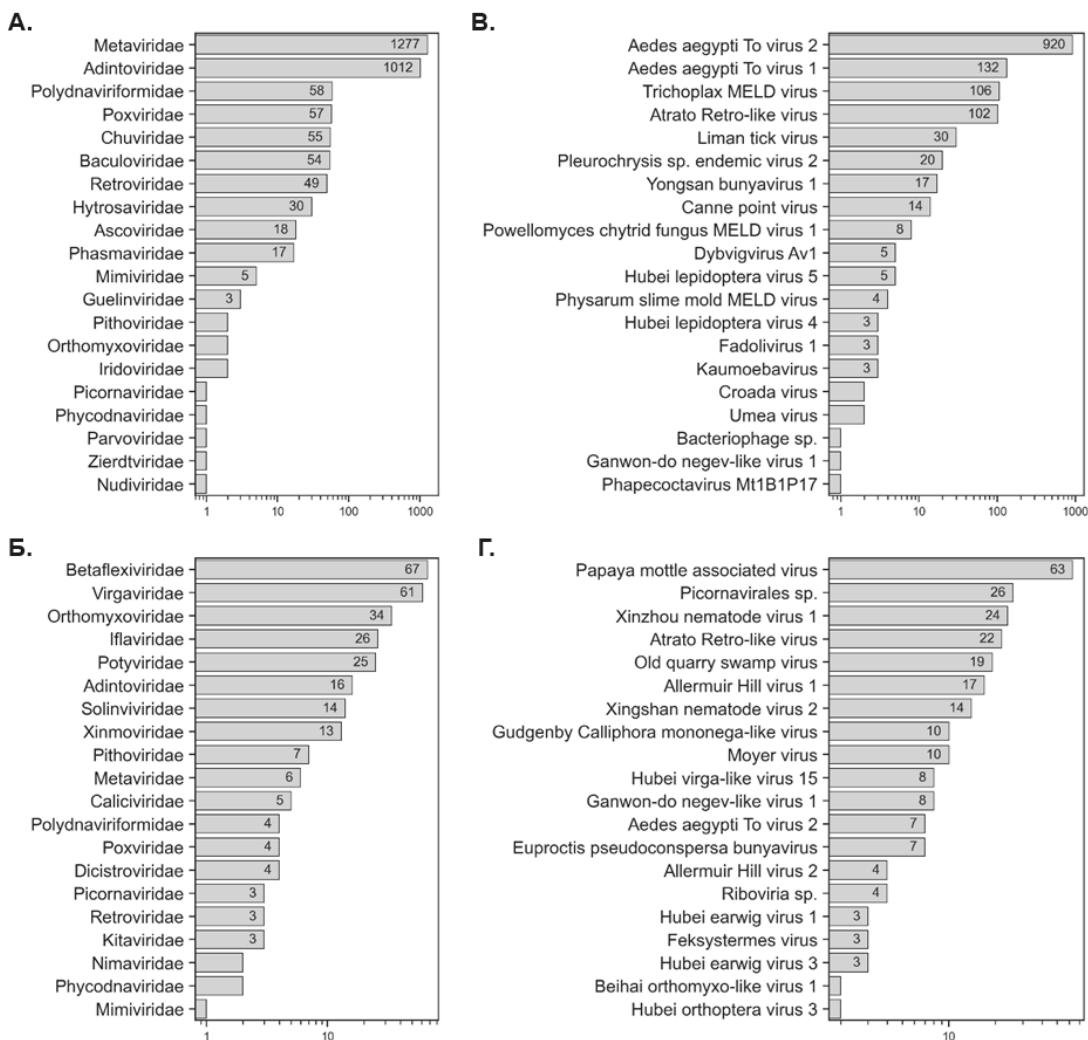


Рис. 2. Количество вирус-позитивных контигов по семействам вирусов в геномных (А) и транскриптомных (Б) образцах. Распределение вирус-позитивных контигов по видам вирусов (семейство которых не определено) в геномных (В) и транскриптомных (Г) образцах.

Анализ геномных данных *L. decemlineata* на предмет эндогенных элементов браковирусного происхождения

В результате k-мерного анализа референсного генома *L. Decemlineata* (сборка GCA_000500325.2 Ldec_2.0) и аннотации собранных контигов, были выявлены многочисленные последовательности, относящиеся к представителям рода *Bracoviriform* (семейство *Polydnaviriformidae*). В результате анализа геномов представителей отряда Coleoptera были найдены многочисленные последовательности гомологичные роду *Bracoviriform* (рис. 3).

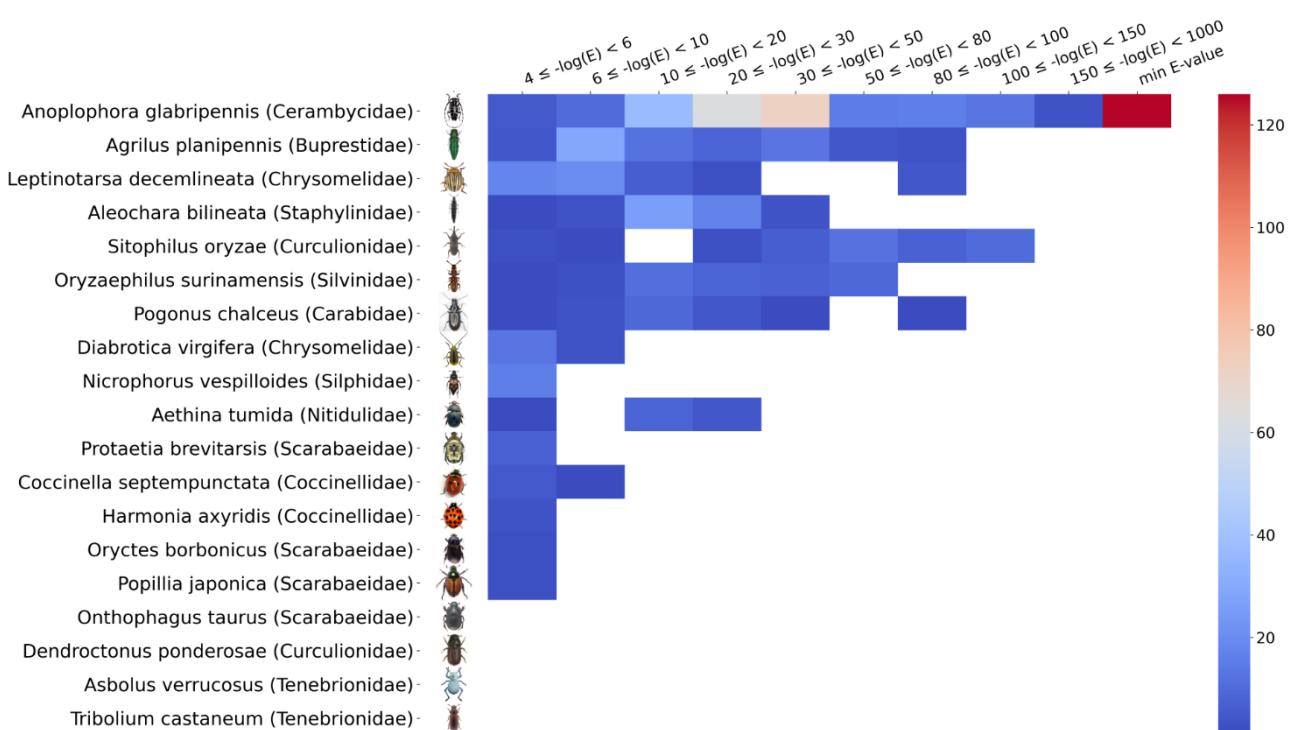


Рис. 3. Распределение браковирусных k-меров в зависимости от E-value в референсных геномах ряда представителей отряда Coleoptera, включая *L. decemlineata*. Цветом обозначено количество найденных браковирусных k-меров, ранжированных по E-value (blastn short). Отсутствие окраски указывает на отсутствие совпадений.

Среди контигов были выбраны 2 последовательности, имеющие гомологию с сегментами браковирусных геномов, гомологичные *Cotesia sesamiae Mombasa bracovirus* (EF710639.1) и *Cotesia vestalis bracovirus segment 35* (HQ009558.1). Эти контиги также выравниваются на геном *L. Decemlineata*, что позволяет предположить встройку вирусной ДНК представителей рода *Bracoviriform* в геном *L. decemlineata*. Были рассчитаны праймеры таким образом, чтобы подтвердить наличие браковирусных фрагментов в генетическом материале *L. decemlineata*, подтвердить встройку в геном *L. decemlineata*, а также исключить наличие фрагментов, содержащих районы, фланкирующие встройку в соответствующем геноме браковируса. В результате ПЦР-анализа ДНК, полученной из различных тканей

L. decemlineata были подтверждено наличие обоих браковирусных фрагментов и показано отсутствие фрагментов, содержащих районы, фланкирующие встройку в геноме браковируса. Нуклеотидная последовательность фрагментов была подтверждена с помощью секвенирования по Сэнгеру. Встройка в геном *L. decemlineata* в результате данного эксперимента подтверждена не была, что может быть обусловлено ошибками сборки или высокой вариабельностью соответствующих районов генома *L. decemlineata*.

В результате секвенирования на платформе Oxford Nanopore Technologies (MinION) стерильных яиц колорадского жука были получены 764 879 прочтений средней длинной 5 510 нт и показателем N50 9 912 нт. Были идентифицированы 6 020 прочтений, содержащих открытые рамки трансляции представителей рода Bracoviriform. Средняя длина ОРТ составила 500 а.к.о., максимальная длина – 1 384 а.к.о.

Выявленные в исследовании последовательности, гомологичные сегментам браковирусных геномов, могут указывать на интеграцию браковирусных последовательностей в геном колорадского жука. По нашим данным это первое сообщение об обнаружении браковирусных последовательностей в геноме колорадского жука и вообще у жуков листоедов. По-видимому, присутствие этих фрагментов в геноме колорадского жука также связано с паразитоидными осами, например, осы *Edovium puttleri*, для которой было показано паразитирование на яйцах *L. decemlineata* (Jansson et al. 1987).

Генетический и филогенетический анализ последовательностей вирусов *Leptinotarsa iflavivirus 1*, *Leptinotarsa solinvi-like virus 1* и *Potato virus S*, полученных из метагеномных образцов *L. Decemlineata*

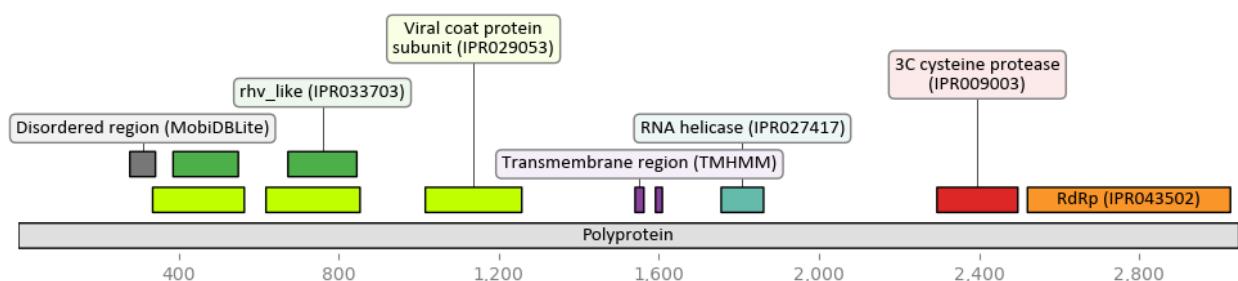
В результате секвенирования образцов колорадского жука, собранных с частных картофельных полей, свободных от обработки ботулотоксином и авермектином (г. Карабук, Новосибирская область) были получены полногеномная последовательность ифлавируса, названного *Leptinotarsa iflavivirus 1* с средней глубиной покрытия 834× (7,1% прочтений) и последовательность *solinvi-like* вируса, названного *Leptinotarsa solinvi-like virus 1* с средней глубиной покрытия 1 116× (10,87% прочтений).

Последовательность генома вируса *Leptinotarsa iflavivirus 1* имеет длину 9 887 нт, исключая поли-А хвост, и содержит одну ОРТ (координаты 561–9 707), 5'-UTR (560 нт) и 3'-UTR (180 нт). ОРТ кодирует полипротеин размером 3 049 а.к.о., который согласно результатам InterProScan, содержит домены, характерные для представителей рода *Iflavirus*. В том числе, два rhv-like домена в двух субъединицах *Picornavirus/Calicivirus*-подобного капсидного белка (IPR033703, E-value: $5,96 \times 10^{-13}$ и $1,15 \times 10^{-21}$), белок из суперсемейства *Calicivirus*-подобных капсидных белков (IPR029053, E-value:

$2,5 \times 10^{-32}$), домен РНК-геликазы (IPR000605, E-value: $1,4 \times 10^{-18}$), домен, принадлежащий к суперсемейству цистеиновых 3С пептидаз (IPR009003, E-value: $2,36 \times 10^{-16}$) и домен РНК-зависимой РНК-полимеразы (IPR043502, E-value: $4,65 \times 10^{-78}$) (рис. 4А). Аминокислотная последовательность полипротеина вириуса *Leptinotarsa iflavivirus 1* обладает наибольшей гомологией с полипротеином вириуса *Apis iflavivirus 2* (UCR92484.1), согласно результатам BLASTx идентичность аминокислотных последовательностей составила 37,72%.

Последовательность генома вириуса *Leptinotarsa solinvi-like virus 1* имеет длину 11 603 нт, исключая поли-А хвост, и содержит одну ОРТ (координаты 62-11 458). ОРТ кодирует полипротеин размером 3 800 а.к.о., который по результатам InterProScan, содержит домен РНК-геликазы (IPR000605, E-value: $7,2 \times 10^{-20}$), домен, принадлежащий к суперсемейству цистеиновых 3С пептидаз (IPR009003, E-value: $7,41 \times 10^{-6}$), домен РНК-зависимой РНК-полимеразы (IPR043502, E-value: $9,39 \times 10^{-55}$) и домен *Calicivirus*-подобного капсидного белка (IPR029053, E-value: $1,5 \times 10^{-18}$) (рис. 4Б). Для аминокислотной последовательности полипротеина вириуса *Leptinotarsa solinvi-like virus* наибольшая гомология найдена с полипротеином вириуса *Hangzhou Solinvi-like virus 1* (UHR49784.1), согласно результатам BLASTx идентичность аминокислотных последовательностей составила 33,32%.

А.



Б.

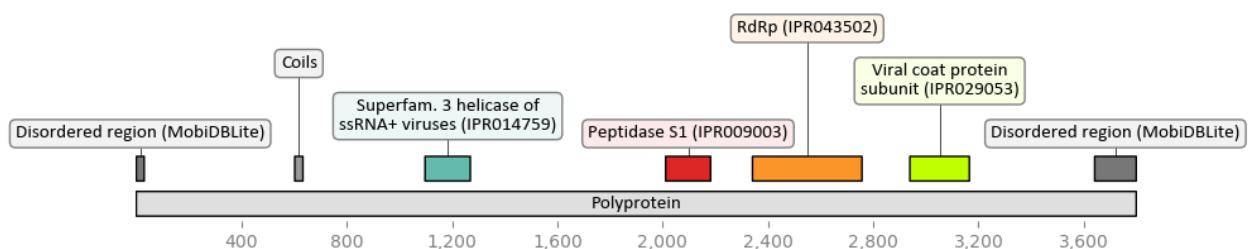


Рис. 4. Результаты предсказания доменов с помощью InterProScan в полипротеине А) *Leptinotarsa iflavivirus 1* и Б) *Leptinotarsa solinvi-like virus 1*.

Поскольку последовательности ифлавирусов обладают очень большим разнообразием, для филогенетического анализа были выбраны наиболее консервативные фрагменты полипротеинов, содержащие РНК-зависимую РНК-полимеразу. Для построения множественного выравнивания с помощью BLAST из баз данных NCBI RefSeq и nr были выбраны последовательности, обладающие наибольшей гомологией с РНК-зависимой РНК-полимеразой *Leptinotarsa iflavivirus 1* (с идентичностью более 30%). Филогенетическое дерево было построено методом наибольшего правдоподобия с помощью IQ-Tree (Рис. 5).

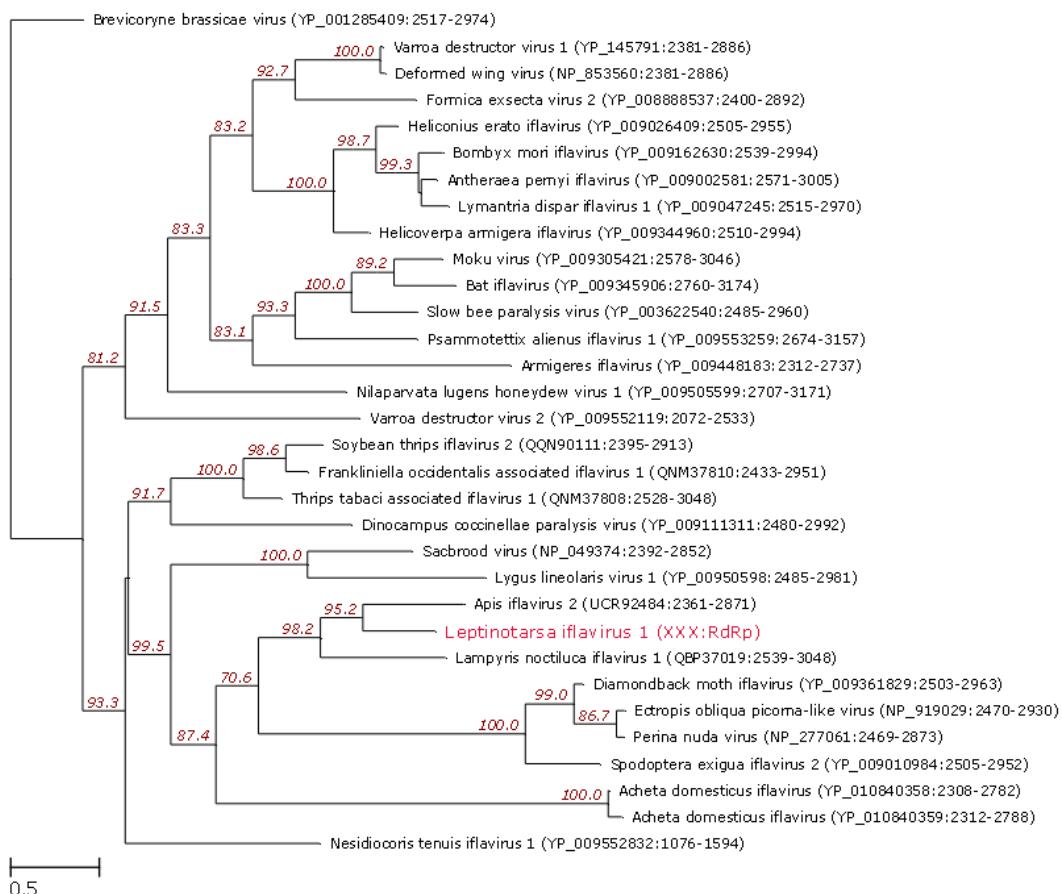


Рис. 5. Филогенетическое дерево RdRp *Leptinotarsa iflavivirus 1* и наиболее близких ифлавирусов и picorna-like вирусов. Дерево построено методом максимального правдоподобия. Красным цветом отмечен *Leptinotarsa iflavivirus 1*. На дереве обозначены индексы поддержки ($\geq 70\%$ при 1000 репликаций).

Наиболее близкими к *Leptinotarsa iflavivirus 1* оказались *Apis iflavivirus 2* (UCR92484), полученный в результате метагеномного анализа образца генетического материала, выделенного из медоносной пчелы в 2017 г. в Китае в провинции Хенань, и *Lampyris noctiluca iflavivirus 1* (QBP37019), полученный в результате метагеномного анализа образца, выделенного из светлячка в Финляндии в 2017 г.

Для построения филогенетического дерева *Leptinotarsa solinvi-like virus 1* было использовано выравнивание аминокислотных последовательностей RdRp (<https://ictv.global/sites/default/files/inline-images/Figure3-1.v2.aa.alignment.fst>) с портала ICTV; также в выравнивание были добавлены наиболее близкие последовательности RdRp *solinvi-like* вирусов (по результатам BLAST) с идентичностью более 30%. На рис. 6 представлен фрагмент полученной филограммы.

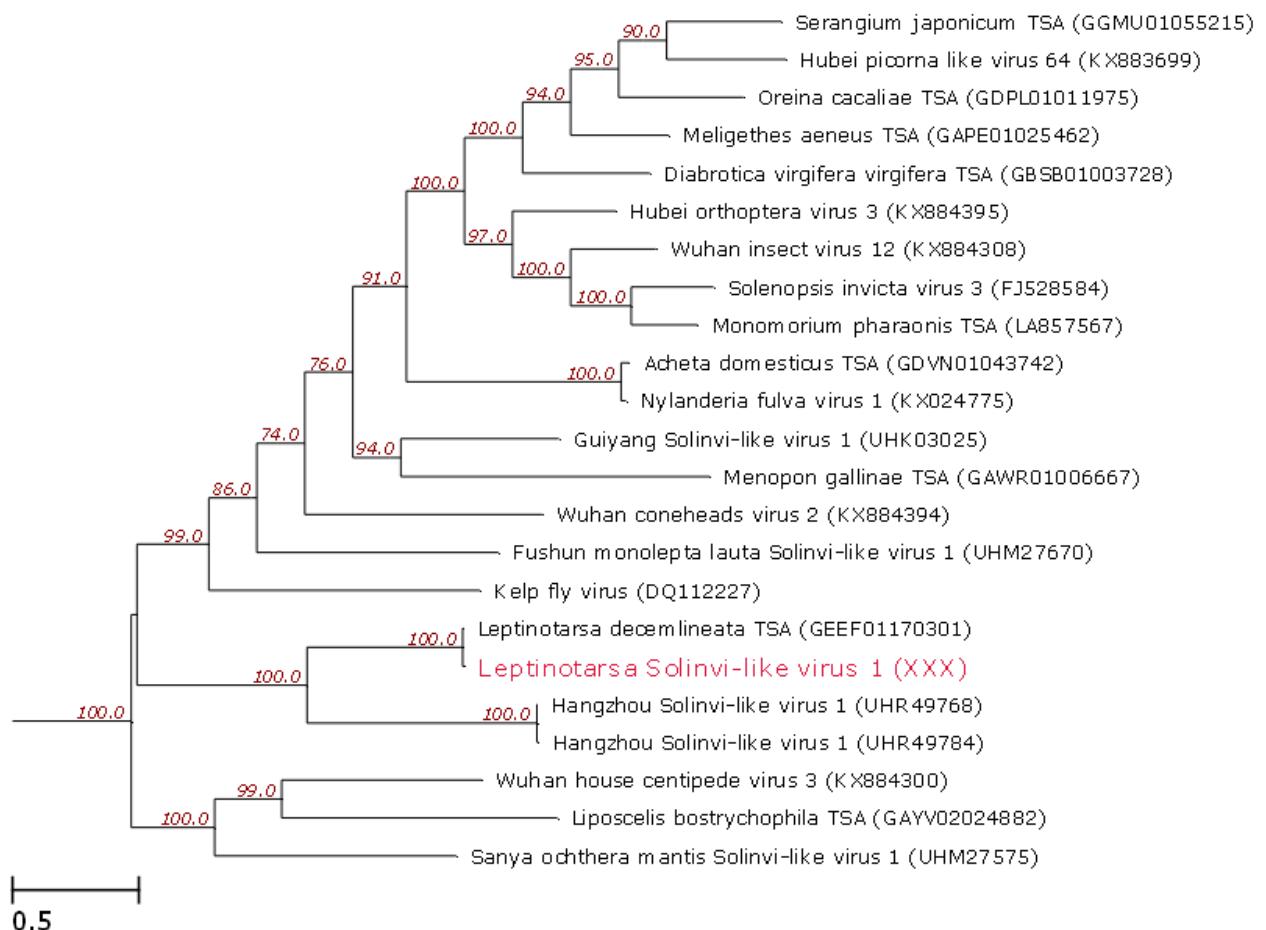


Рис. 6. Филогенетическое дерево RdRp *Leptinotarsa solinvi-like virus 1* и наиболее близких *solinvi-like* вирусов. Дерево построено методом максимального правдоподобия. Красным цветом отмечен *Leptinotarsa solinvi-like virus 1*. На дереве обозначены индексы поддержки ($\geq 70\%$ при 1000 репликаций).

ЗАКЛЮЧЕНИЕ

Метагеномные исследования данных высокопроизводительного секвенирования являются важным источником информации для поиска как известных, так и новых вирусов. Представленная работа подчеркивает важность метагеномных исследований геномных и транскриптомных данных, полученных от насекомых вредителей, для более глубокого понимания их экологии и эволюции. Обнаружение генетических последовательностей, потенциально принадлежащих поражающих насекомых вирусам, указывает на возможность расширения арсенала биологических методов контроля численности колорадского жука на основе энтомопатогенных вирусов.

Проведенное исследование позволило получить важные новые данные о вирусах, ассоциированных с колорадским жуком. В результате секвенирования и *de novo* сборки геномов двух вирусов — *Leptinotarsa iflavivirus 1* и *Leptinotarsa solinvi-like virus 1* — удалось установить их полные генетические последовательности и провести детальный анализ их структурных доменов. Эти результаты существенно расширяют наше понимание вирусных сообществ, связанных с колорадским жуком, и их возможного влияния на экосистему и сельскохозяйственные культуры. Кроме того, анализ генетических последовательностей вирусов растений, так как potato virus S, выявленных в образцах колорадского жука, указывают на его возможную роль в передаче данного вируса и в целом подчеркивает роль растительноядных насекомых в распространении вирусов растений.

Таким образом, полученные результаты предоставляют новые данные о вирусах, связанных с колорадским жуком, и подчеркивают важность дальнейших исследований для понимания их биологии, эволюции и потенциального воздействия на агроэкосистемы.

ВЫВОДЫ

1. Разработан программный конвейер для поиска и идентификации вирусного генетического материала в геномных и транскриптомных данных высокопроизводительного секвенирования. Конвейер написан на языке программирования Snakemake и доступен по адресу: <https://github.com/starchevskayamaria17/uncoVir>.
2. С помощью разработанного программного конвейера в результате анализа геномных и транскриптомных образцов колорадского жука, опубликованных в базе данных NCBI SRA, обнаружены генетические последовательности вирусов, принадлежащие к более чем 32 различным семействам. Большинство последовательностей были отнесены к вирусам насекомых (*Adintoviridae*, *Ascoviridae*, *Baculoviridae*, *Dicistroviridae*, *Chuviridae*, *Hytrosaviridae*, *Iflaviridae*, *Iridoviridae*, *Nimaviridae*, *Nudiviridae*, *Phasmaviridae*, *Picornaviridae*, *Polydnnaviriformidae*, *Xinmoviridae* и др.), вирусам растений (*Betaflexiviridae*, *Bromoviridae*, *Kitaviridae*, *Potyviridae*) и эндогенным ретровирусным элементам (*Retroviridae*, *Metaviridae*).
3. В результате анализа генетического материала колорадского жука, выделенного из стерильных яиц, в составе его генома впервые было показано наличие протяженных фрагментов последовательностей браковирусов (сем. *Polydnnaviriformidae*).
4. В образцах личинок колорадского жука, погибших от неизвестной инфекции, обнаружены последовательности ранее неописанных вирусов, названных нами *Leptinotarsa iflavivirus 1* и *Leptinotarsa solinvi-like virus 1*; собраны и аннотированы практически полные последовательности их геномов. Последовательности депонированы в международную базу данных NCBI GenBank (*OR613011* и *OR613010* соответственно).

СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИОННОЙ РАБОТЫ:

1. **Starchevskaya (Antonets) M.**, Kamanova E., Vyatkin Yu., Tregubchak T., Bauer T., Bodnev S., Rotskaya U., Polenogova O., Kryukov V., Antonets D. The metagenomic analysis of viral diversity in Colorado potato beetle public NGS data. *Viruses*. 2023, V. 15, P. 395.
2. **Antonets (Starchevskaya) M.**, Bodnev S., Rotskaya U., Kosman E., Tregubchak T., Bauer T., Azaev M., Kryukov V., Antonets D. Nearly complete genome sequences of the first two identified Colorado potato beetle viruses. *Scientific Reports*. 2024, V. 14, P. 352.