

doi 10.18699/vjgb-26-35

## Анализ вирома растений с помощью высокопроизводительного секвенирования: принципы и подходы

Д.Д. Белкина , С.В. Виноградова  

Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Москва, Россия

 coatprotein@bk.ru

**Аннотация.** В последние годы метагеномный подход, основанный на высокопроизводительном секвенировании, находит все большее применение в диагностике вирусных инфекций растений. Этот метод позволяет изучить видовой состав вирусов, ассоциированных с исследуемым растением, и в том числе обнаружить ранее не описанные виды, охарактеризовать их популяционно-генетическую структуру и разработать генетические тест-системы для рутинной диагностики. Проведение фитосанитарного мониторинга с использованием метагеномного подхода может способствовать определению этиологии неизвестных заболеваний растений, что особенно важно для предотвращения распространения таких патогенов, как вирусы. Кроме того, с учетом невозможности элиминации вирусов растений в полевых условиях комплексная диагностика с помощью высокопроизводительного секвенирования становится эффективным инструментом как в целях соблюдения карантинного законодательства при ввозе импортного материала, так и для получения отечественного посадочного материала высоких категорий. При этом с каждым годом высокопроизводительное секвенирование становится более доступным: расширяется как приборно-техническая, так и аналитическая база. В настоящем обзоре систематизированы ключевые подходы к анализу вирома растений с помощью высокопроизводительного секвенирования. В основной части статьи описаны этапы проведения анализа: от сбора образцов до биоинформатической обработки данных, ее валидации и интерпретации. Подробно рассмотрены особенности современных платформ секвенирования и факторы, влияющие на качество чтения, в том числе контаминация. Охарактеризованы три взаимодополняющих подхода для обработки биоинформатических данных: картирование чтений на референсные последовательности вирусов; сборка и аннотация контигов; таксономическая классификация чтений без дополнительной сборки. Особое внимание уделено необходимости тщательной интерпретации результатов с учетом как биоинформатического анализа, так и валидации идентификации молекулярно-генетическими методами. Обзор будет полезен как для исследователей и специалистов, не имеющих опыта работы с высокопроизводительным секвенированием, так и для тех, кто использовал этот инструмент для выполнения других задач.

**Ключевые слова:** метагеномика; высокопроизводительное секвенирование; вирусы; виром растений; биоинформатика

**Для цитирования:** Белкина Д.Д., Виноградова С.В. Анализ вирома растений с помощью высокопроизводительного секвенирования: принципы и подходы. *Вавиловский журнал генетики и селекции*. 2026;30(2):311-320. doi 10.18699/vjgb-26-35

**Финансирование.** Работа выполнена при поддержке Российского научного фонда (грант 23-16-00232).

## Plant virome analysis by high-throughput sequencing: concepts and approaches

D.D. Belkina , S.V. Vinogradova  

Federal Research Centre "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia

 coatprotein@bk.ru

**Abstract.** The metagenomic approach based on high-throughput sequencing is becoming increasingly prevalent for the detection of viral infections in plants. This method allows us to study the species composition of viruses associated with the plant, including novel species, describe their population genetic structure, and develop genetic test systems for routine diagnostics. A metagenomic approach to phytosanitary monitoring can help to determine the cause of unknown plant diseases, which is particularly important for preventing the spread of pathogens, such as viruses. Furthermore, as it is impossible to eliminate plant viruses in field conditions, comprehensive diagnostics using high-throughput sequencing is becoming an effective tool for complying with quarantine regulations on the import of foreign material, as well as for producing high-quality local planting material. High-throughput sequencing is becoming more affordable every year, with both the instrumentation and analytical capacity improving. This review summarizes key approaches to analyzing plant virome using high-throughput sequencing. The analysis process, from sample collection to bioinformatic data

processing, validation and interpretation, is described in detail. The features of sequencing platforms and the factors affecting sequencing quality, including contamination, are discussed. Three complementary approaches to processing bioinformatic data are described: mapping reads to reference viral sequences; assembling and annotating contigs; taxonomic classification of reads without assembly. The importance of carefully interpreting the results is emphasized, considering the bioinformatic analysis and the validation by molecular genetic methods. This review will be useful for both researchers and specialists who have no experience with high-throughput sequencing, and those who have used this method for other applications.

**Key words:** metagenomics; high-throughput sequencing; viruses; plant virome; bioinformatics

**For citation:** Belkina D.D., Vinogradova S.V. Plant virome analysis by high-throughput sequencing: concepts and approaches. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed.* 2026;30(2):311-320. doi 10.18699/vjgb-26-35

## Введение

Вирусные заболевания растений наносят значительный ущерб растениеводству по всему миру. В условиях глобализации и изменения климата распространение вирусов может приводить к разрушительным эпифитотиям, угрожающим продовольственной безопасности всего человечества (Jones, 2021). Так, потери урожая томатов и перца из-за инфекции вирусом мозаики томата (tomato mosaic virus, ToMV), по разным оценкам, составляют 25–70 % (Panno et al., 2021), а снижение урожайности винограда от болезни скручивания листьев, вызываемой вирусами из семейства *Closteroviridae*, может достигать 50 % (Atallah et al., 2012). Ключевое значение в предотвращении распространения вирусных инфекций имеют мониторинг, ранняя диагностика и соблюдение карантинных мер.

В настоящее время основными методами идентификации вирусов растений являются серологические и молекулярные, получившие распространение благодаря высоким скорости проведения анализа, чувствительности и специфичности. Однако их применение позволяет идентифицировать только те виды вирусов, для которых разработаны тест-системы (Maina et al., 2024).

До развития метагеномики наше представление о вирусах растений ограничивалось несколькими сотнями видов (Parras et al., 2021), поэтому роль вирусов в микробиоме сельскохозяйственных культур оставалась недооцененной. Метагеномный подход позволяет получить комплексное представление о совокупности вирусов растения, т. е. о его вироме. По состоянию на 2025 г. Международным комитетом по таксономии вирусов зафиксировано 2598 вирусов и вирионов растений, причем 206 из них добавлены в перечень за последний год (Rubino et al., 2025). Секвенирование вирома предоставляет уникальные данные о последовательностях геномов вирусов, на основе которых можно провести популяционные исследования и проследить пути распространения вирусов, связать заражение конкретными штаммами с проявлением симптомов, а также разработать специфичные тест-системы для их быстрого обнаружения.

Анализ вирома растений с помощью высокопроизводительного секвенирования – это активно развивающийся метод. Каждый год публикуются новые рекомендации, протоколы и специализированные программы для обработки данных. Поскольку этот анализ еще не вошел в число методов для рутинной диагностики вирусных инфекций растений, у него нет единой методики. Тем не

менее основные принципы его проведения уже сформировались.

В настоящей работе систематизированы ключевые подходы к анализу вирома растений с помощью высокопроизводительного секвенирования. Этот обзор будет полезен как для исследователей и специалистов, не имеющих опыта работы с высокопроизводительным секвенированием, так и для тех, кто использовал этот инструмент для выполнения других задач.

## Общий план анализа растительного вирома

Структурно анализ вирома можно разделить на два блока. Первый блок включает отбор образцов, пробоподготовку и секвенирование (рис. 1). Эта часть анализа имеет ключевое значение. Если материал для секвенирования будет низкого качества, то на следующих этапах будет невозможно получить достоверные результаты. Во время сбора образцов важно учитывать, что вирусы неравномерно распределены по растительным тканям, поэтому отбирать следует по несколько фрагментов с одного растения. Особенно это актуально для многолетних растений. После сбора образцы должны быть как можно скорее доставлены в лабораторию, чтобы снизить риск деградации нуклеиновых кислот, например, под действием прямых солнечных лучей.

Второй блок анализа растительного вирома включает компьютерную обработку данных секвенирования биоинформатическими программами, валидацию и интерпретацию результатов. В отличие от этапов первой части анализа, биоинформатическую обработку данных в любой момент можно провести повторно и исправить возможные ошибки, применить новые методы или перепроверить результаты других исследований. Поэтому критически важно избегать ошибок при сборе образцов, пробоподготовке и секвенировании библиотек. Для достоверной идентификации вирусов необходимо валидировать результаты биоинформатического анализа молекулярно-генетическими методами.

## Исходный материал для анализа

Вирусы растений отличаются от других фитопатогенов многообразием типов геномов. В качестве носителя генетической информации они используют молекулы как ДНК, так и РНК, которые могут быть одноцепочечными или двуцепочечными, линейными или кольцевыми (Lefeuvre et al., 2019). При этом одноцепочечные РНК-вирусы под-

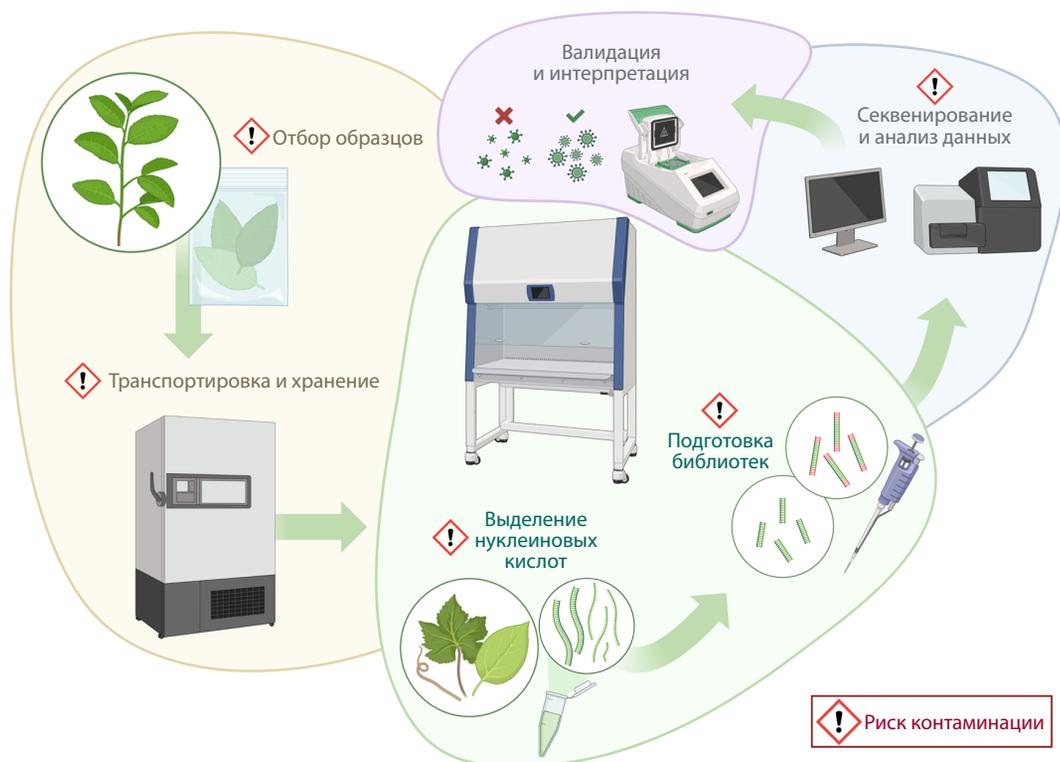


Рис. 1. Основные этапы анализа растительного вирома.

разделяются на вирусы с «положительным» геномом, когда геномная РНК соответствует мРНК, и с «отрицательным» геномом, когда геномная РНК комплементарна мРНК. Идентифицировать все вирусы, инфицирующие растение, с их разнообразными геномами, используя только один подход, довольно сложно (Fitzpatrick et al., 2021). Ниже рассмотрены различные подходы к анализу растительных виромов в зависимости от типа используемой матрицы.

Секвенирование тотальной РНК, выделенной из растительных тканей, – это относительно простой, надежный и широко распространенный метод анализа виромов (Lee et al., 2020; Nabeshima, Abe, 2021; Vinogradova et al., 2023). Он позволяет обнаружить не только вирусы с РНК-геномами, но и транскрипты ДНК-вирусов (Pescan et al., 2017; Cobbin et al., 2021), а также вироиды, что делает этот подход оптимальным для диагностики вирусов с различными типами геномных нуклеиновых кислот.

Однако чувствительность этого метода может значительно снизиться из-за наличия в анализируемом образце РНК растения-хозяина, выделяемой совместно с вирусной РНК в ходе пробоподготовки (Maliogka et al., 2018), что создает трудности в обнаружении вирусов с низким титром (Roossinck et al., 2015). Преодолеть эту проблему позволяет обогащение образца нуклеиновыми кислотами, ассоциированными с вирусами (Gaafar, Ziebell, 2020). Как правило, это достигается путем очистки образца от рибосомальной и транспортной РНК или обогащением его полиА-транскриптами. Первый метод обогащения предпочтительнее, поскольку не все группы вирусов имеют полиаденилированный 3'-конец.

Еще один метод анализа растительного вирома – это секвенирование нуклеиновых кислот, ассоциированных с вирионами (Filloux et al., 2015). Преимуществом этого подхода являются секвенирование только вирусных последовательностей и возможность анализа вирусов как с РНК-, так и с ДНК-геномами, в том числе вирусов с низким титром (Moubset et al., 2022). Однако экстракция нуклеиновых кислот, ассоциированных с вирионами, позволяет обнаружить только вирусы, имеющие капсид (который отсутствует у вирусов-сателлитов и вироидов), и требует больших усилий, чем выделение тотальной РНК (Maliogka et al., 2018).

Следующий подход предполагает анализ коротких интерферирующих РНК (киРНК), образующихся в клетках растений в результате иммунного ответа на вирусную инфекцию. Эти молекулы представляют собой расщепленные растительными ферментами фрагменты вирусных геномов длиной по 21–24 нуклеотида (Vivek et al., 2020; Zhuravlyov et al., 2022).

Секвенирование киРНК позволяет обнаруживать РНК- и ДНК-вирусы, но небольшой размер этих молекул затрудняет последующую сборку полных геномов и дифференциацию штаммов вирусов (Pescan et al., 2017; Maliogka et al., 2018; Turco et al., 2018).

Многие вирусы во время репликации в качестве промежуточного продукта образуют двуцепочечную РНК (дцРНК), благодаря чему эти молекулы тоже используются в исследованиях растительного вирома. Секвенирование дцРНК позволяет идентифицировать не только вирусы с дцРНК-геномом, но и большинство одноцепочечных

РНК-вирусов, вириды, а также некоторые ДНК-вирусы (Gallo et al., 2021; Fall et al., 2025). Однако для обнаружения ДНК-вирусов этот метод малоэффективен и показывает положительные результаты только при высоком уровне экспрессии вирусных транскриптов (Gaafar, Ziebell, 2020).

### Платформы секвенирования

В основе всех метагеномных исследований лежит высокопроизводительное секвенирование нуклеиновых кислот (high-throughput sequencing, HTS). Эта группа методов объединяет секвенирование второго и третьего поколений, сменившие секвенирование методом Сэнгера, которое остается золотым стандартом только для прочтения последовательностей до 1000 нуклеотидов длиной (Crossley et al., 2020).

Секвенирование второго поколения (next generation sequencing, NGS), также называемое массовым параллельным, позволяет считывать короткие последовательности длиной до нескольких сотен нуклеотидов. С помощью этого подхода можно быстро секвенировать все последовательности в образце, однако небольшая длина чтения усложняет сборку полных геномных последовательностей вирусов на этапе биоинформатического анализа (Maina et al., 2024).

Лидером рынка NGS платформ является компания Illumina (США), выпускающая секвенаторы с низким уровнем ошибок при высокой производительности (Maina et al., 2024). В последние годы конкуренцию Illumina составляют приборы от китайских компаний MGI, GeneMind и Cugnus Biosciences. Кроме того, недавно разработан и выпущен в продажу секвенатор Нанофор СПС российского производства (Kurochkin et al., 2021). Большинство современных NGS секвенаторов основано на флуоресцентном секвенировании путем синтеза в проточной ячейке (Zubov et al., 2021) и имеет схожие протоколы подготовки библиотек и зачастую совместимые реагенты. Процесс подготовки библиотек обычно включает фрагментацию нуклеиновых кислот (например, с помощью ультразвука или ферментов рестрикции), лигирование коротких олигонуклеотидов (адаптеров) с одним или двумя концами фрагментов и амплификацию этих фрагментов (Kutnjak et al., 2021). Адаптеры различаются в зависимости от NGS платформы и необходимы для запуска процесса секвенирования (Lebas et al., 2022). Важно отметить, что секвенаторы второго поколения способны считывать только молекулы ДНК, поэтому, если в качестве матрицы выбрана РНК, после фрагментации проводится обратная транскрипция для синтеза кДНК.

Секвенирование третьего поколения, называемое также одномолекулярным, основано на прочтении последовательностей отдельных молекул и не требует их фрагментации и амплификации (Villamor et al., 2019). Максимальная длина чтения в этом случае исчисляется несколькими сотнями тысяч нуклеотидов, что позволяет целиком считывать небольшие геномы, например геномы вирусов (Liu et al., 2025). Важным преимуществом одномолекулярного секвенирования являются возможность прямого чтения

РНК без синтеза кДНК, а также секвенирование в режиме реального времени (Sun et al., 2022).

Ведущие производители секвенаторов третьего поколения – компании Pacific Biosciences (США) и Oxford Nanopore Technologies (Великобритания). Несмотря на то что обе платформы третьего поколения позволяют исключить ошибки, вносимые полимеразой при амплификации, они имеют относительно низкое качество чтения, по сравнению с NGS (Rose et al., 2016; Rang et al., 2018). Впрочем, недавно представленная технология PacBio HiFi обеспечивает точность считывания, превышающую 99.5 % (Han et al., 2024). Нанопорное секвенирование, первоначально разработанное Oxford Nanopore Technologies, тоже с каждым годом становится все более точным и доступным (Javaran et al., 2021). Эту технологию уже освоила китайская компания Qitan Tech (Wang et al., 2022), поэтому можно ожидать, что в будущем одномолекулярное секвенирование придет на смену NGS.

### Другие факторы, влияющие на качество секвенирования растительного виroma

Выше рассмотрено, что выбор типа нуклеиновых кислот и платформы секвенирования играет важную роль в анализе виroma растений. На качество секвенирования и количество чтений оказывает влияние еще ряд факторов.

Контаминация – один из главных негативных факторов любого метагеномного анализа (Piombo et al., 2021; Lebas et al., 2022). Загрязнение образцов может произойти на всех этапах: от сбора растительных тканей до непосредственного процесса секвенирования (см. рис. 1). Для отслеживания контаминации на этапах выделения нуклеиновых кислот и подготовки библиотек рекомендуется использовать отрицательные контроли без добавления матрицы (Fitzpatrick et al., 2021). Кроме того, важно придерживаться физического разделения лабораторных этапов, использовать стерильные реагенты и поддерживать в чистоте посуду, оборудование и рабочие поверхности (Kutnjak et al., 2021; Maina et al., 2024). Исчерпывающее руководство о предотвращении контаминации при использовании HTS в диагностике фитопатогенов приведено в публикации S. Massart с коллегами (2022).

Наряду с контаминацией большое влияние на качество секвенирования виroma оказывают используемые методики, реактивы (особенно ферменты) и оборудование. Так, при подготовке библиотек для секвенирования второго поколения используются обратная транскриптаза и ДНК-полимераза, которые могут вносить ошибки в процессе амплификации (Cholet et al., 2020). Поэтому для уменьшения количества артефактов важно применять высокоточные ферменты.

На точность идентификации вирусов влияет количество чтений в наборе данных. При недостаточном количестве чтений вирусы с низким титром в образце, возможно, обнаружить не удастся. Однако из-за избыточного секвенирования чтений существует риск ложноположительной идентификации (Massart et al., 2014). Для определения оптимального значения может быть полезно изучить опубликованные анализы виroma интересующего растения.

## Биоинформатический анализ

Биоинформатический анализ данных секвенирования растительного виroma – это многоступенчатый процесс, который можно проводить как на персональном компьютере, так и на высокопроизводительных кластерах. Разница в мощности оборудования влияет на скорость анализа. Полученные в результате секвенирования чтения обрабатываются специализированными программами в порядке, адаптированном для анализа конкретного набора данных.

Большая часть биоинформатических программ разработана для операционных систем Linux и не имеет графического интерфейса, поэтому для их использования необходимо работать в командной строке. Для других операционных систем доступны кроссплатформенные приложения, веб-инструменты и коммерческие программные пакеты. Последние особенно удобны для начинающих пользователей благодаря тому, что позволяют выполнять почти весь анализ виroma в одной среде с интуитивно понятным интерфейсом. Популярные примеры таких приложений – Geneious Prime (Новая Зеландия) и CLC Genomics Workbench (США). Несмотря на удобство коммерческих решений, их функционал ограничен встроенным набором инструментов. В целом его достаточно для рутинного анализа набора виромных данных, но больше возможностей обеспечивает работа с командной строкой Linux.

Как правило, программы, предназначенные для Linux-систем, публикуются в открытом бесплатном доступе. Для пользователей Windows можно воспользоваться подсистемой Windows для Linux (WSL) или виртуальной машиной (например, VirtualBox). Существуют автоматизированные наборы программ, предназначенные для идентификации вирусов и других фитопатогенов: VirusDetect (Zheng et al., 2017), Virtool (Rott et al., 2017), Kodoja (Baizan-Edge et al., 2019), PhytoPipe (Hu et al., 2023) и др. Такие решения позволяют достаточно быстро получить результаты, но могут не учитывать особенности конкретного набора данных. Лучших результатов можно достичь, используя навыки программирования для написания или редактирования собственных программ. В области обработки биологических данных наибольшей популярностью пользуются языки Python и R.

Результаты секвенирования чаще всего предоставляют в виде файлов формата FastQ, в которых содержатся чтения с указанием качества прочтения каждого нуклеотида. Для визуализации этой информации доступна бесплатная программа FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>), формирующая подробные отчеты о качестве секвенирования.

Вне зависимости от типа анализируемых нуклеиновых кислот и платформ секвенирования важным этапом биоинформатического анализа виroma является препроцессинг – набор операций, призванных повысить качество данных. Препроцессинг оказывает определяющее влияние на весь последующий анализ. Некачественная подготовка данных может привести к ошибкам во время сборки геномов вирусов и препятствовать картированию чтений, вследствие чего результаты анализа будут искажены.

В большинстве случаев препроцессинг включает следующие шаги.

- Фильтрация чтений. Этот шаг включает обрезку (тримминг) и удаление низкокачественных и слишком коротких чтений. Необходимый порог качества и минимальная длина чтения задаются пользователем. При обработке данных Illumina часто используются такие значения качества, как Q20 или Q30, что означает точность считывания 99 и 99.9 % соответственно (Kutnjak et al., 2021). Слишком строгие настройки фильтрации могут привести к частичной потере информативных данных, слишком мягкие – к ошибкам во время картирования и сборки геномов. Наряду с удалением низкокачественных чтений, фильтрация включает в себя очистку данных от адаптерных последовательностей, которые были искусственно добавлены в библиотеки для запуска секвенирования. Выполнить эти операции можно с помощью таких программ, как BBDuk (<https://sourceforge.net/projects/bbmap>), Trimmomatic (Bolger et al., 2014) и другие.
- Объединение парных чтений на основе перекрывающейся области. Этот шаг необходим только для секвенирования с парным чтением концов.
- Удаление дублирующихся чтений для уменьшения объема данных и повышения скорости анализа.
- Еще одним шагом препроцессинга может стать удаление чтений, картирующихся на геном организма-хозяина. Этот шаг позволяет значительно сократить объем и повысить качество данных. Однако следует помнить, что при этом будут удалены последовательности вирусов, интегрированных в геном хозяина (Pappas et al., 2021). Кроме того, сборка генома растения-хозяина может содержать ошибки. Если последовательности некоторых вирусов случайно попали в сборку, удаление растительных чтений исключит такие вирусы из набора данных. Поэтому удаление последовательностей растения-хозяина следует проводить с осторожностью.

Идентификация вирусов в данных секвенирования всегда опирается на сравнение последовательностей с базами данных (Lebas et al., 2022). Качество идентификации прямо пропорционально качеству используемых баз данных. Чем полнее база данных и точнее аннотация последовательностей, тем выше вероятность точной идентификации вирусов. По мере пополнения базы данных новыми последовательностями любой набор данных можно проанализировать повторно.

Существует три основных подхода, которые при совместном использовании могут взаимодополнять друг друга (рис. 2): картирование чтений на референсные последовательности вирусов; сборка и аннотация контигов; таксономическая классификация чтений без дополнительной сборки.

Наиболее быстрый и простой подход к идентификации вирусов в данных HTS состоит в картировании чтений на референсные последовательности вирусов. Картирование – это одна из базовых операций при анализе результатов секвенирования второго поколения. Алгоритм

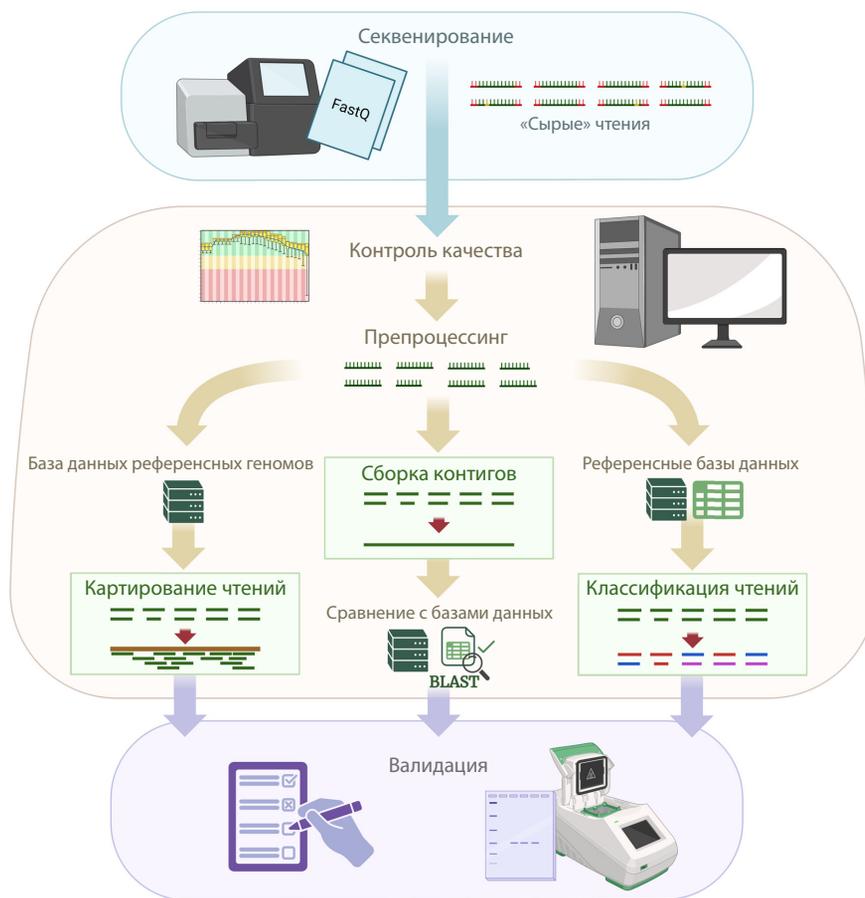


Рис. 2. Схема биоинформатического анализа виромных данных.

выравнивает чтения с референсной последовательностью с учетом заданного порога сходства. Чтения, не удовлетворяющие порогу, отбрасываются, а оставшиеся занимают такие позиции на референсной последовательности, в которых наиболее с ней совпадают.

Чтобы обнаружить вирусы растений с помощью этого подхода, необходимо предварительно подготовить базу данных референсных последовательностей, которая должна содержать по меньшей мере по одному полному геному каждого вируса, способного заражать исследуемую культуру.

Затем препроцессированные чтения картируют на эти референсные последовательности. В результате пользователь получает информацию о доле покрытия референсной последовательности чтениями, глубине покрытия, проценте сходства между референсом и картированными чтениями и другие показатели (Lebas et al., 2022). При достаточном покрытии картированные чтения образуют непрерывную последовательность по всей длине референсного генома.

Для картирования чтений используют такие программы, как BWA (Li, Durbin, 2009), Bowtie2 (Langmead, Salzberg, 2012), Minimap2 (Li, 2018), BBmap (<https://sourceforge.net/projects/bbmap>), STAR (Dobin, Gingeras, 2015), GraphMap (Sović et al., 2016) и Hisat2 (Kim et al., 2019). Настройки и пороговые значения этих программ имеют большое зна-

чение. Следует понимать, что последовательность генома вируса, заразившего исследуемое растение, может значительно отличаться от референсной, поскольку вирусы известны быстрыми темпами накопления мутаций. При слишком строгих настройках картирования будут обнаружены только изоляты, близкие к референсной последовательности (Kutnjak et al., 2021). С другой стороны, если настройки картирования будут недостаточно строгими, на последовательность генома одного вида вируса будут картированы чтения другого, что приведет к ложноположительной идентификации (Roy et al., 2018).

Картирование позволяет эффективно обнаруживать вирусы даже в тех случаях, когда число чтений вируса в анализируемом наборе данных невелико. Однако важно с осторожностью интерпретировать результаты и помнить о риске контаминации во время пробоподготовки и секвенирования. По этой причине рекомендуется устанавливать пороговые значения в количестве чтений или доле покрытия референсного генома, ниже которых образец будет считаться вирус-отрицательным (Belkina et al., 2023).

Второй подход к идентификации вирусов с помощью HTS основан на сборке *de novo* последовательностей из перекрывающихся чтений – контигов, и сравнении их с аннотированными базами данных. Сборка контигов позволяет обнаружить новые виды вирусов, новые растения-хозяева известных вирусов, а также описать различные

варианты их геномов. Кроме того, этот подход исключает риск ложноотрицательных результатов, связанных с возможной неполнотой набора референсных геномов вирусов целевой культуры.

Для сборки контигов могут использоваться различные алгоритмы, каждый из которых имеет свои сильные и слабые стороны (White et al., 2017). При достаточной глубине секвенирования контиги могут соответствовать почти полным геномам вирусов, без коротких 5'- и 3'-концевых участков. Эффективность сборки одним и тем же алгоритмом различается на разных наборах данных, поэтому невозможно выбрать один универсальный (Sutton et al., 2019; Shvets et al., 2022). На качество сборки могут положительно влиять предварительное удаление чтений растения-хозяина и выбор подходящих настроек (Kutnjak et al., 2021).

Наиболее популярными алгоритмами сборки для коротких чтений являются SPAdes (Bankevich et al., 2012), Velvet (Zerbino, Birney, 2008), Geneious (<https://www.geneious.com>), CLC (<https://digitalinsights.qiagen.com>), MIRA (Chevreux et al., 1999), ABySS (Simpson et al., 2009), IDBA-UI (Peng et al., 2012), SOAPdenovo2 (Luo et al., 2012); для длинных чтений – Canu (Koren et al., 2017), Falcon (Chin et al., 2016) и Pomoxis (<https://github.com/nanoporetech/pomoxis>). Некоторые алгоритмы, например SPAdes и Unicycler (Wick et al., 2017), показывают хорошие результаты при сборке геномов из совокупности коротких и длинных чтений (Pappas et al., 2021).

Для получения более надежных результатов рекомендуется использовать несколько алгоритмов сборки. При этом нужно учитывать, что, в зависимости от размера библиотеки и выбранного алгоритма, процесс сборки может занимать продолжительное время и требовать больших объемов оперативной памяти.

После завершения сборки контиги сравнивают с базами данных аннотированных последовательностей. Достоверность результатов во многом зависит от полноты и качества используемых баз данных. Многие из них доступны онлайн, но зачастую бывает удобнее развернуть базу данных на облачном сервере или локально, если имеется достаточное дисковое пространство.

Наиболее удобными, обновляемыми и часто используемыми являются базы данных от NCBI, которые к началу 2025 г. в совокупности содержат 6 млн нуклеотидных и 1 млн аминокислотных последовательностей (Sayers et al., 2025). NCBI предоставляет как курируемую базу данных, содержащую по одному геному каждого организма, – RefSeq (<https://www.ncbi.nlm.nih.gov/refseq>), так и ежедневно пополняемую исследователями со всего мира базу данных GenBank (<https://www.ncbi.nlm.nih.gov/genbank>). База данных GenBank содержит преобладающее большинство общедоступных нуклеотидных и аминокислотных последовательностей организмов, но качество последовательностей и правильность аннотации в ней почти не проверяются. Поэтому сравнение контигов с этой базой данных может привести к ошибкам в идентификации. Использование базы данных референсных геномов вирусов (Viral RefSeq) потенциально даст наиболее надежный

результат, однако она медленно пополняется новыми записями и может не содержать значительное количество недавно открытых вирусов.

Самым широко используемым инструментом для сравнения контигов с базами данных является BLAST, который объединяет несколько алгоритмов анализа нуклеотидных и аминокислотных последовательностей. Определение того, какому вирусу из базы данных соответствует тот или иной контиг, можно выполнить с помощью самого быстрого алгоритма – megablast. Для поиска новых вирусов лучше использовать blastx или tblastx, которые улавливают отдаленные сходства между последовательностями. Главный недостаток blastx и особенно tblastx – низкая скорость анализа, поэтому при ограниченном времени можно задействовать более быстрый алгоритм DIAMOND (Buchfink et al., 2015). Однако некоторые исследователи считают, что он менее эффективен для идентификации вирусов (Kutnjak et al., 2021).

Для запуска BLAST можно использовать веб-интерфейс на сайте NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Он особенно удобен, когда нужно проверить небольшое число последовательностей. Через сайт осуществляется доступ ко всем базам данных нуклеотидных и белковых последовательностей NCBI, при этом на поиск не расходуются вычислительные ресурсы пользователя. Кроме того, BLAST встроен в программы с графическим интерфейсом, такие как Geneious Prime и CLC Genomics Workbench. При работе с командной строкой Linux, Mac и Windows применяется инструмент BLAST+ (Camacho et al., 2009).

В результате сравнения с базой данных последовательностей вирусов контиги с той или иной достоверностью относят к конкретным видам. Обычно уровень достоверности указывают с помощью E-value – показателя вероятности случайного совпадения между контигом и базой данных (VanderWeele, Ding, 2017). Чем ближе E-value к 0, тем надежнее выполнена идентификация.

Подходы, основанные на таксономической классификации чтений без предварительной сборки или картирования, менее распространены в исследованиях растительных виромов. Однако при наличии необходимых вычислительных ресурсов они позволяют получать результаты намного быстрее рассмотренных выше подходов. Среди таксономических классификаторов чтений популярны такие, как Kraken2 (Wood et al., 2019), Kaiju (Menzel et al., 2016), CLARK-S (Ounit, Lonardi, 2016) и Centrifuge (Kim et al., 2016). Как правило, для работы им требуются обширные базы данных, большие вычислительные мощности и умение пользователя работать в командной строке Linux.

## Валидация и интерпретация результатов анализа виroma

После того как любым методом биоинформатического анализа в библиотеке обнаружены чтения того или иного вируса, важно правильно интерпретировать эту информацию. Поскольку качество секвенирования зависит от многих факторов, не существует универсальных критериев того, какое растение следует считать зараженным вирусом: сколько чтений и контигов достоверно свидетельствует

о наличии инфекции (Villamor et al., 2019). Более того, анализ одного набора данных с использованием разных программ или с разными настройками одних и тех же программ может приводить к различающимся результатам.

Поэтому в каждом конкретном случае необходимо всесторонне и критически подходить к анализу полученных данных, а на завершающих этапах анализа вирома большое значение приобретает экспертная роль исследователя, его «биологический взгляд» на биоинформатические данные. В частности, исследователь должен понимать, что не все обнаруженные в образцах вирусы являются патогенами исследуемой культуры. В виромных данных могут присутствовать вирусы насекомых, грибов и бактерий, ассоциированных с растением, а также вирусы человека и животных (Cobbin et al., 2021). Это особенно важно учитывать, когда речь идет об обнаружении ранее не описанных вирусов.

Для достоверной идентификации вирусов растений в метагеномных данных следует подтверждать результаты биоинформатического анализа с помощью молекулярно-генетических методов, например ПЦР или ПЦР с обратной транскрипцией (ОТ-ПЦР), в зависимости от типа генома вируса. Если вирус будет обнаружен в образце в результате секвенирования вирома, так и молекулярно-биологической диагностики, можно с большой долей вероятности считать идентификацию точной.

## Заключение

Анализ вирома с помощью высокопроизводительного секвенирования – это мощный метод, позволяющий комплексно подойти к решению фундаментальных и прикладных задач в области сельского хозяйства. С каждым годом высокопроизводительное секвенирование становится более доступным: расширяется как приборно-техническая, так и аналитическая база. В будущем с появлением правовой основы этот метод станет основанием для принятия решений в сфере государственной продовольственной безопасности.

## Список литературы / References

- Atallah S.S., Gómez M.I., Fuchs M.F., Martinson T.E. Economic impact of grapevine leafroll disease on *Vitis vinifera* cv. Cabernet Franc in Finger Lakes vineyards of New York. *Am J Enol Vitic.* 2012;63(1): 73-79. doi 10.5344/ajev.2011.11055
- Baizan-Edge A., Cock P., MacFarlane S., McGavin W., Torrance L., Jones S. *Kodoja*: a workflow for virus detection in plants using *k-mer* analysis of RNA-sequencing data. *J Gen Virol.* 2019;100(3):533-542. doi 10.1099/jgv.0.001210
- Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., ... Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-477. doi 10.1089/cmb.2012.0021
- Belkina D., Karpova D., Porotikova E., Lifanov I., Vinogradova S. Grapevine virome of the Don ampelographic collection in Russia has concealed five novel viruses. *Viruses.* 2023;15(12):2429. doi 10.3390/v15122429
- Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120. doi 10.1093/bioinformatics/btu170
- Buchfink B., Xie C., Huson D.H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12(1):59-60. doi 10.1038/nmeth.3176
- Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10(1):421. doi 10.1186/1471-2105-10-421
- Chevreaux B., Wetter T., Suhai S. Genome sequence assembly using trace signals and additional sequence information. In: Proceedings of the German Conference on Bioinformatics (GCB) 99. Hannover, 1999;45-56
- Chin C.-S., Peluso P., Sedlazeck F.J., Nattestad M., Concepcion G.T., Clum A., Dunn C., ... Luo C., Ecker J.R., Cantu D., Rank D.R., Schatz M.C. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods.* 2016;13(12):1050-1054. doi 10.1038/nmeth.4035
- Cholet F., Ijaz U.Z., Smith C.J. Reverse transcriptase enzyme and priming strategy affect quantification and diversity of environmental transcripts. *Environ Microbiol.* 2020;22(6):2383-2402. doi 10.1111/1462-2920.15017
- Cobbin J.C., Charon J., Harvey E., Holmes E.C., Mahar J.E. Current challenges to virus discovery by meta-transcriptomics. *Curr Opin Virol.* 2021;51:48-55. doi 10.1016/j.coviro.2021.09.007
- Crossley B.M., Bai J., Glaser A., Maes R., Porter E., Killian M.L., Clement T., Toohey-Kurth K. Guidelines for Sanger sequencing and molecular assay monitoring. *J Vet Diagn Invest.* 2020;32(6):767-775. doi 10.1177/1040638720905833
- Dobin A., Gingeras T.R. Mapping RNA-seq reads with STAR. *Curr Protoc Bioinformatics.* 2015;51(1):11.14.1-11.14.19. doi 10.1002/0471250953.bi1114s1
- Fall M.L., Xu D., Lemoyne P., Clément G., Moffett P., Ritzenthaler C. An innovative binding-protein-based dsRNA extraction method: comparison of cost-effectiveness of virus detection methods using high-throughput sequencing. *Mol Ecol Resour.* 2025;25(7):e14111. doi 10.1111/1755-0998.14111
- Filloux D., Dallot S., Delaunay A., Galzi S., Jacquot E., Roumagnac P. Metagenomics approaches based on virion-associated nucleic acids (VANA): an innovative tool for assessing without a priori viral diversity of plants. In: Lacomme C. (Ed.) Plant Pathology. Methods in Molecular Biology. Vol. 1302. New York: Humana Press, 2015; 249-257. doi 10.1007/978-1-4939-2620-6\_18
- Fitzpatrick A.H., Rupnik A., O'Shea H., Crispie F., Keaveney S., Cotter P. High throughput sequencing for the detection and characterization of RNA viruses. *Front Microbiol.* 2021;12:621719. doi 10.3389/fmicb.2021.621719
- Gaafar Y.Z.A., Ziebell H. Comparative study on three viral enrichment approaches based on RNA extraction for plant virus/viroid detection using high-throughput sequencing. *PLoS One.* 2020;15(8): e0237951. doi 10.1371/journal.pone.0237951
- Gallo Y., Marin M., Gutiérrez P. Detection of RNA viruses in *Solanum quitoense* by high-throughput sequencing (HTS) using total and double stranded RNA inputs. *Physiol Mol Plant Pathol.* 2021;113: 101570. doi 10.1016/j.pmpp.2020.101570
- Han Y., He J., Li M., Peng Y., Jiang H., Zhao J., Li Y., Deng F. Unlocking the potential of metagenomics with the PacBio high-fidelity sequencing technology. *Microorganisms.* 2024;12(12):2482. doi 10.3390/microorganisms12122482
- Hu X., Hurtado-Gonzales O.P., Adhikari B.N., French-Monar R.D., Malapi M., Foster J.A., McFarland C.D. PhytoPipe: a phytosanitary pipeline for plant pathogen detection and diagnosis using RNA-seq data. *BMC Bioinformatics.* 2023;24(1):470. doi 10.1186/s12859-023-05589-2
- Javaran V.J., Moffett P., Lemoyne P., Xu D., Adkar-Purushothama C.R., Fall M.L. Grapevine virology in the third-generation sequencing era: from virus detection to viral epitranscriptomics. *Plants.* 2021; 10(11):2355. doi 10.3390/plants10112355
- Jones R.A.C. Global plant virus disease pandemics and epidemics. *Plants.* 2021;10(2):233. doi 10.3390/plants10020233

- Kim D., Song L., Breitwieser F.P., Salzberg S.L. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 2016;26(12):1721-1729. doi 10.1101/gr.210641.116
- Kim D., Paggi J.M., Park C., Bennett C., Salzberg S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907-915. doi 10.1038/s41587-019-0201-4
- Koren S., Walenz B.P., Berlin K., Miller J.R., Bergman N.H., Phillippy A.M. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* 2017; 27(5):722-736. doi 10.1101/gr.215087.116
- Kurochkin V.E., Alekseev Y.I., Petrov D.G., Evstrapov A.A. Domestic devices for molecular genetic analysis: developments of the IAP RAS and SINTOL LLC. *Russ Mil Med Acad Rep.* 2021;40(3): 69-74. doi 10.17816/rmmar76918 (in Russian)
- Kutnjak D., Tamisier L., Adams I., Boonham N., Candresse T., Chiumenti M., De Jonghe K., ... Rollin J., Rott M., Schumpp O., Massart S., Haegeman A. A primer on the analysis of high-throughput sequencing data for detection of plant viruses. *Microorganisms.* 2021;9(4):841. doi 10.3390/microorganisms9040841
- Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357-359. doi 10.1038/nmeth.1923
- Lebas B., Adams I., Al Rwahnih M., Baeyen S., Bilodeau G.J., Blouin A.G., Boonham N., ... Vicente C.S.L., Vossenber B.T.L.H., Wetzel T., Ziebell H., Massart S. Facilitating the adoption of high-throughput sequencing technologies as a plant pest diagnostic test in laboratories: a step-by-step description. *EPPO Bull.* 2022;52(2): 394-418. doi 10.1111/epp.12863
- Lee H.-K., Kim S.-Y., Yang H.-J., Lee D.-S., Kwon B., Lee D.-Y., Oh J., Lee S.-H. The detection of plant viruses in Korean ginseng (*Panax ginseng*) through RNA sequencing. *Plant Pathol J.* 2020;36(6): 643-650. doi 10.5423/PPJ.NT.07.2020.0137
- Lefevre P., Martin D.P., Elena S.F., Shepherd D.N., Roumagnac P., Varsani A. Evolution and ecology of plant viruses. *Nat Rev Microbiol.* 2019;17(10):632-644. doi 10.1038/s41579-019-0232-3
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094-3100. doi 10.1093/bioinformatics/bty191
- Li H., Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760. doi 10.1093/bioinformatics/btp324
- Liu S., Rodriguez J.S., Munteanu V., Ronkowski C., Sharma N.K., Alser M., Andrae F., ... Ganda E., Davenport E.R., Pop M., Koslicki D., Mangul S. Analysis of metagenomic data. *Nat Rev Methods Primers.* 2025;5(1):5. doi 10.1038/s43586-024-00376-6
- Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., ... Li Y., Yang H., Wang J., Lam T.-W., Wang J. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience.* 2012;1(1):18. doi 10.1186/2047-217X-1-18
- Maina S., Donovan N.J., Plett K., Bogema D., Rodoni B.C. High-throughput sequencing for plant virology diagnostics and its potential in plant health certification. *Front Hortic.* 2024;3:1388028. doi 10.3389/fhort.2024.1388028
- Maliogka V.I., Minafra A., Saldarelli P., Ruiz-Garcia A.B., Glasa M., Katis N., Olmos A. Recent advances on detection and characterization of fruit tree viruses using high-throughput sequencing technologies. *Viruses.* 2018;10(8):436. doi 10.3390/v10080436
- Massart S., Olmos A., Jijakli H., Candresse T. Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Res.* 2014;188:90-96. doi 10.1016/j.virusres.2014.03.029
- Massart S., Adams I., Al Rwahnih M., Baeyen S., Bilodeau G.J., Blouin A.G., Boonham N., ... van de Vossenber B.T.L.H., Westenberg M., Wetzel T., Ziebell H., Lebas B.S.M. Guidelines for the reliable use of high throughput sequencing technologies to detect plant pathogens and pests. *Peer Commun J.* 2022;2:e62. doi 10.24072/pcjournal.181
- Menzel P., Ng K.L., Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun.* 2016;7(1):11257. doi 10.1038/ncomms11257
- Moubset O., François S., Maclot F., Palanga E., Julian C., Claude L., Fernandez E., ... Massart S., Ogliastro M., Martin D.P., Filloux D., Roumagnac P. Virion-associated nucleic acid-based metagenomics: a decade of advances in molecular characterization of plant viruses. *Phytopathology.* 2022;112(11):2253-2272. doi 10.1094/PHYTO-03-22-0096-RVW
- Nabeshima T., Abe J. High-throughput sequencing indicates novel *Varicosavirus*, *Emaravirus*, and *Deltapartitivirus* infections in *Vitis coignetiae*. *Viruses.* 2021;13(5):827. doi 10.3390/v13050827
- Ounit R., Lonardi S. Higher classification sensitivity of short metagenomic reads with CLARK-S. *Bioinformatics.* 2016;32(24):3823-3825. doi 10.1093/bioinformatics/btw542
- Panno S., Davino S., Caruso A.G., Bertacca S., Crnogorac A., Mandić A., Noris E., Matić S. A review of the most common and economically important diseases that undermine the cultivation of tomato crop in the mediterranean basin. *Agronomy.* 2021;11(11):2188. doi 10.3390/agronomy11112188
- Pappas N., Roux S., Hölzer M., Lamkiewicz K., Mock F., Marz M., Dutilh B.E. Virus bioinformatics. In: Encyclopedia of Virology. Elsevier, 2021;124-132. doi 10.1016/B978-0-12-814515-9.00034-5
- Pecman A., Kutnjak D., Gutiérrez-Aguirre I., Adams I., Fox A., Boonham N., Ravnikar M. Next generation sequencing for detection and discovery of plant viruses and viroids: comparison of two approaches. *Front Microbiol.* 2017;8:1998. doi 10.3389/fmicb.2017.01998
- Peng Y., Leung H.C.M., Yiu S.M., Chin F.Y.L. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012;28(11):1420-1428. doi 10.1093/bioinformatics/bts174
- Piombo E., Abdelfattah A., Droby S., Wisniewski M., Spadaro D., Schena L. Metagenomics approaches for the detection and surveillance of emerging and recurrent plant pathogens. *Microorganisms.* 2021;9(1):188. doi 10.3390/microorganisms9010188
- Rang F.J., Kloosterman W.P., de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* 2018;19(1):90. doi 10.1186/s13059-018-1462-9
- Roossinck M.J., Martin D.P., Roumagnac P. Plant virus metagenomics: advances in virus discovery. *Phytopathology.* 2015;105(6):716-727. doi 10.1094/PHYTO-12-14-0356-RVW
- Rose R., Constantinides B., Tapinos A., Robertson D.L., Prosperi M. Challenges in the analysis of viral metagenomes. *Virus Evol.* 2016; 2(2):vew022. doi 10.1093/ve/vew022
- Rott M., Xiang Y., Boyes I., Belton M., Saeed H., Kesnakurti P., Hayes S., Lawrence T., Birch C., Bhagwat B., Rast H. Application of next generation sequencing for diagnostic testing of tree fruit viruses and viroids. *Plant Dis.* 2017;101(8):1489-1499. doi 10.1094/PDIS-03-17-0306-RE
- Roy S., Coldren C., Karunamurthy A., Kip N.S., Klee E.W., Lincoln S.E., Leon A., Pullambhatla M., Temple-Smolkin R.L., Voelkerding K.V., Wang C., Carter A.B. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines. *J Mol Diagn.* 2018;20(1):4-27. doi 10.1016/j.jmoldx.2017.11.003
- Rubino L., Abrahamian P., An W., Aranda M.A., Ascencio-Ibañez J.T., Bejerman N., Blouin A.G., ... Whitfield A.E., Wylie S.J., Yang C., Zerbini F.M., Zhang S. Summary of taxonomy changes ratified by the International Committee on Taxonomy of Viruses from the Plant Viruses Subcommittee, 2025. *J Gen Virol.* 2025;106(7):002114. doi 10.1099/jgv.0.002114
- Sayers E.W., Beck J., Bolton E.E., Brister J.R., Chan J., Connor R., Feldgarden M., ... Wang J., Ye J., Zellers E., Schneider V.A., Pruitt K.D. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res.* 2025;53(D1):D20-D29. doi 10.1093/nar/gkae979
- Shvets D., Sandomirsky K., Porotikova E., Vinogradova S. Metagenomic analysis of ampelographic collections of Dagestan revealed the presence of two novel grapevine viruses. *Viruses.* 2022;14(12): 2623. doi 10.3390/v14122623

- Simpson J.T., Wong K., Jackman S.D., Schein J.E., Jones S.J.M., Birrol Í. ABySS: a parallel assembler for short read sequence data. *Genome Res.* 2009;19(6):1117-1123. doi 10.1101/gr.089532.108
- Sović I., Šikić M., Wilm A., Fenlon S.N., Chen S., Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. *Nat Commun.* 2016;7(1):11307. doi 10.1038/ncomms11307
- Sun K., Liu Y., Zhou X., Yin C., Zhang P., Yang Q., Mao L., Shentu X., Yu X. Nanopore sequencing technology and its application in plant virus diagnostics. *Front Microbiol.* 2022;13:939666. doi 10.3389/fmicb.2022.939666
- Sutton T.D.S., Clooney A.G., Ryan F.J., Ross R.P., Hill C. Choice of assembly software has a critical impact on virome characterisation. *Microbiome.* 2019;7(1):12. doi 10.1186/s40168-019-0626-5
- Turco S., Golyaev V., Seguin J., Gilli C., Farinelli L., Boller T., Schumpp O., Pooggin M.M. Small RNA-omics for virome reconstruction and antiviral defense characterization in mixed infections of cultivated *Solanum* plants. *Mol Plant Microbe Interact.* 2018; 31(7):707-723. doi 10.1094/MPMI-12-17-0301-R
- VanderWeele T.J., Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med.* 2017;167(4):268-274. doi 10.7326/M16-2607
- Villamor D.E.V., Ho T., Al Rwahnih M., Martin R.R., Tzanetakis I.E. High throughput sequencing for plant virus detection and discovery. *Phytopathology.* 2019;109(5):716-725. doi 10.1094/PHYTO-07-18-0257-RVW
- Vinogradova S., Porotikova E., Navrotskaya E., Galbacs Z.N., Mas-sart S., Varallyay E. The first virome of a Russian vineyard. *Plants.* 2023;12(18):3292. doi 10.3390/plants12183292
- Vivek A.T., Zahra S., Kumar S. From current knowledge to best practice: a primer on viral diagnostics using deep sequencing of virus-derived small interfering RNAs (vsiRNAs) in infected plants. *Methods.* 2020;183:30-37. doi 10.1016/j.ymeth.2019.10.009
- Wang Z., Qin L., Liu J., Jiang L., Zou X., Chen X., Song F., Dai H., Hou Y. Forensic nanopore sequencing of microhaplotype markers using QitanTech's QNome. *Forensic Sci Int Genet.* 2022;57:102657. doi 10.1016/j.fsigen.2021.102657
- White D.J., Wang J., Hall R.J. Assessing the impact of assemblers on virus detection in a de novo metagenomic analysis pipeline. *J Comput Biol.* 2017;24(9):874-881. doi 10.1089/cmb.2017.0008
- Wick R.R., Judd L.M., Gorrie C.L., Holt K.E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* 2017;13(6):e1005595. doi 10.1371/journal.pcbi.1005595
- Wood D.E., Lu J., Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20(1):257. doi 10.1186/s13059-019-1891-0
- Zerbino D.R., Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 2008;18(5):821-829. doi 10.1101/gr.074492.107
- Zheng Y., Gao S., Padmanabhan C., Li R., Galvez M., Gutierrez D., Fuentes S., Ling K.-S., Kreuze J., Fei Z. VirusDetect: an automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology.* 2017;500:130-138. doi 10.1016/j.virol.2016.10.017
- Zhuravlyov V.S., Dolgikh V.V., Timofeev S.A., Gannibal F.B. RNA interference method in plant protection against insect pests. *Plant Prot News.* 2022;105(1):28-39. doi 10.31993/2308-6459-2022-105-1-15219
- Zubov V.V., Chemeris D.A., Vasilov R.G., Kurochkin V.E., Alekseev Ya.I. Brief history of high-throughput nucleic acid sequencing methods. *Biomics.* 2021;13(1):27-46. doi 10.31301/2221-6197.bmcs.2021-4

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 03.09.2025. После доработки 31.10.2025. Принята к публикации 31.10.2025.