———— ОБЗОРЫ ———

УДК 577.21

# СЕКВЕНИРОВАНИЕ ГЕНОМОВ РАСТЕНИЙ: СОВРЕМЕННЫЕ ТЕХНОЛОГИИ И НОВЫЕ ВОЗМОЖНОСТИ ДЛЯ СЕЛЕКЦИИ

#### © 2022 г. А. А. Дмитриев<sup>а</sup>, Е. Н. Пушкова<sup>а</sup>, Н. В. Мельникова<sup>а, \*</sup>

<sup>а</sup>Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, 119991 Россия \*e-mail: mnv-4529264@yandex.ru

Поступила в редакцию 17.07.2021 г. После доработки 24.08.2021 г. Принята к публикации 24.08.2021 г.

Изучение геномов растений имеет огромное значение для фундаментальных изысканий и практической селекции. В 1977 году Ф. Сэнгером предложен метод секвенирования ДНК, позволивший установить полные нуклеотидные последовательности ряда геномов. Затем появились высокопроизводительные и экономически эффективные методы секвенирования нового/второго поколения, генерирующие до миллиардов коротких прочтений, что сделало возможным секвенирование геномов значительного числа видов и обеспечило прорыв в генетических исследованиях растений. Наконец, были разработаны технологии секвенирования третьего поколения, определяющие последовательности единичных молекул длиной до миллиона нуклеотидов, что имеет ключевое значение для получения высококачественных сборок геномов. Актуальной задачей является создание пангенома, включающего всю совокупность нуклеотидных последовательностей, представленных в различных генотипах одного вида. Секвенирование геномов растений позволило оценить внутривидовой полиморфизм, идентифицировать ключевые гены, влияющие на формирование значимых свойств, разработать молекулярные маркеры хозяйственно ценных признаков и стало основой для развития маркер-ориентированной и геномной селекции. В данном обзоре представлена информация о последних достижениях в области технологий секвенирования и сборки геномов растений, а также возможностей, которые они открывают для фундаментальных и прикладных работ.

**Ключевые слова:** растения, геном, пангеном, секвенирование, селекция **DOI:** 10.31857/S0026898422040048

#### **ВВЕДЕНИЕ**

Растения - основные производители органических соединений и кислорода, они обеспечивают существование жизни на нашей планете и используются человеком для получения пищи, одежды, стройматериалов, топлива, лекарств и промышленного сырья [1]. За миллионы лет сформировалось огромное разнообразие видов растений, лишь немногие из которых активно возделываются человеком, однако значительно большее число видов играют важную роль в природных экосистемах и могут быть полезны человеку в будущем [2]. Направленный отбор ценных генотипов, а затем и научно обоснованная селекция позволили создать высокопродуктивные сорта растений, без которых сложно представить существование человечества. Изменения климата и рост населения Земли требуют интенсификации сельскохозяйственного производства, однако при этом необходимо сохранить биологическое разнообразие, что невозможно без разработки новых подходов к обеспечению баланса между удовлетворением потребностей человече-

531

ства и поддержанием сложившихся экосистем [3-6].

Изучение геномов растений является первоочередной задачей для фундаментальных исследований и практической селекции. История секвенирования полных геномов растений насчитывает более 20 лет, она началась в 2000 году с момента получения генома Arabidopsis thaliana с использованием секвенирования по Сэнгеру [7]. Этот метод активно применялся в 2000-е годы, он позволил определить последовательности полных геномов ряда важных сельскохозяйственных растений, в том числе риса [8], кукурузы [9] и сои [10]. Секвенирование по Сэнгеру характеризуется высокой стоимостью, трудозатратностью и низкой продуктивностью. На смену ему пришли высокопроизводительные и экономически выгодные методы секвенирования нового поколения (next-generation sequencing, NGS), чаще называемые методами секвенирования второго поколения (second-generation sequencing). Эти методы позволяют генерировать миллионы и миллиарды коротких высокоточных прочтений за один за-

пуск прибора (платформы 454, SOLiD, Ion Torrent/Proton, Illumina, BGI), что сделало возможным секвенирование и ресеквенирование геномов многих видов [11, 12]. Однако большинство геномов, секвенированных только с использованием коротких прочтений, собраны лишь до уровня контигов и скаффолдов, имеют недостаточное качество сборки (в первую очередь малую длину контигов) и содержат значительное число ошибок. Появление технологий секвенирования третьего поколения (third-generation sequencing), определяющих последовательности единичных молекул длиной до 1 млн. нуклеотидов (платформы Pacific Biosciences (PacBio) и Oxford Nanopore Technologies (ONT)), позволило частично преодолеть трудности сборки геномов из коротких прочтений [13, 14]. Однако платформы третьего поколения все еще уступают в точности секвенаторам второго поколения (отдельную проблему представляет секвенирование гомополимеров), а стоимость секвенирования 1 млрд.н. (gigabases, Gb) остается достаточно высокой [12, 15, 16]. В то же время, прогресс в технологиях третьего поколения стремителен, а точность прочтений значительно повышается: у РасВіо появился вариант HiFi (highfidelity), когда одна последовательность ДНК секвенируется многократно, благодаря чему средняя точность превышает 99% [17], а ONT в 2021 году анонсировала новые проточные ячейки R10.4 и химию Q20+, использование которых должно повысить среднюю точность до более чем 99%, в том числе благодаря решению проблемы с ошибками в секвенировании гомополимеров (по крайней мере длиной до 10 н.) [18]. Совершенствование технологий секвенирования позволило определить нуклеотидные последовательности геномов сотен видов растений, собрать референсные геномы и провести ресеквенирование геномов для десятков и сотен генотипов одного вида [19-21]. В последнее время появляются сборки высокого качества (вплоть до уровня хромосом) геномов все большего числа видов растений, при этом залогом успеха проекта является выбор оптимальных подходов к секвенированию и сборке генома с учетом особенностей изучаемого организма [16]. Стоит отметить, что понятие "высокое качество" геномной сборки весьма расплывчато и до сих пор отсутствует единая номенклатура, позволяющая однозначно оценивать качество сборки [22]. Сборка генома человека от теломеры до теломеры получена лишь в 2021 году [23], а сборки геномов растений уровня хромосом представляют собой преимущественно контиги, расположение которых друг относительно друга определяется посредством генетических и/или физических карт или даже выравнивания на референсный геном [24], однако точность и протяженность геномных сборок растений растут год от года [25].

## СЛОЖНОСТИ СЕКВЕНИРОВАНИЯ И СБОРКИ ГЕНОМОВ РАСТЕНИЙ

Размеры геномов наземных растений варьируют от небольших (0.06 млрд.н.) до огромных значений (150 млрд.н.) [26, 27]. Многие растения, особенно культурные, являются алло- и автополиплоидами [28]. Кроме того, растительные геномы часто содержат большое количество повторяющихся элементов [21]. Секвенирование полных геномов большого размера требует серьезных финансовых вложений, а наличие повторов и полиплоидность могут существенно осложнять сборку и требовать увеличения глубины покрытия генома прочтениями на 50-100% [29]. Еще одна проблема – гетерозиготность, требующая увеличения покрытия при секвенировании по меньшей мере на 30% [16]. Кроме того, при секвенировании геномов растений около 5-10% данных могут составлять прочтения, полученные с ДНК органелл (пластид и митохондрий), что также осложняет сборку [29].

Известно, что генотипы одного вида растений могут значительно различаться не только однонуклеотидными полиморфизмами, но и копийностью генов и повторяющихся некодирующих участков, а также иметь последовательности, специфичные для определенного генотипа, включающие, в том числе, гены, кодирующие ценные признаки [15, 30–33], которые не будут идентифицированы при сборке геномов из коротких прочтений с использованием референсного генома. В связи с этим более предпочтительно получение высококачественных *de novo* сборок геномов с применением длинных прочтений.

#### ТЕХНОЛОГИИ СЕКВЕНИРОВАНИЯ ТРЕТЬЕГО ПОКОЛЕНИЯ ДЛЯ ПОЛУЧЕНИЯ СБОРОК ГЕНОМОВ РАСТЕНИЙ

Две основные платформы для секвенирования третьего поколения - РасВіо и ОNТ - позволяют секвенировать длинные молекулы ДНК без предварительной амплификации, что важно для последующей сборки генома. При секвенировании на платформе ONT одноцепочечная ДНК проходит через белковые нанопоры, находящиеся в полимерной мембране, вызывая характерные изменения силы тока, детектируемые сенсором, что при последующей обработке данных (basecalling) позволяет получать информацию о последовательности нуклеотидов [34]. Технология PacBio SMRT (single molecule real-time) представляет секвенирование синтезом и дает возможность определять включение флуоресцентно меченных нуклеотидов в кольцевую цепь ДНК в реальном времени — детекция происходит на чипе SMRT cell в лунках zero-mode waveguide (ZMW) диаметром в десятки нанометров, на дне которых закреплена полимераза. При этом секвенирование одной и той же молекулы ДНК может происходить многократно в случае точных HiFi прочтений, длина которых составляет 15–20 т.п.н. [17, 35, 36]. Платформа ONT позволяет получать прочтения длиной до 1 млн.н. [37] и имеет бо́льший выход данных при меньшей стоимости [38], однако до недавнего времени эта платформа уступала в точности РасВіо (появление новых проточных ячеек R10.4 и химии Q20+ может изменить ситуацию [18]).

Секвенирование третьего поколения дает возможность получения информации не только о последовательности нуклеотидов в молекуле ДНК, но и об их модификациях, включая метилирование ДНК [39]. Модификации нуклеотидов влияют на кинетику полимеразы на платформе РасВіо [40], а на платформе ONT вызывают отклонения в изменениях силы тока при прохождении определенного нуклеотида через пору [41, 42]. Важную роль в регуляции экспрессии генов играет метилирование ДНК, которое управляет множеством процессов, происходящих в клетках растений, в том числе ростом, развитием, ответом на стрессовые воздействия [43-45], поэтому данные с платформ третьего поколения могут быть использованы для расширения нашего представления о роли этой эпигенетической модификации в жизни растений.

Секвенирование третьего поколения предъявляет крайне высокие требования к чистоте ДНК; кроме того, для получения высококачественных сборок геномов необходима ДНК большой длины [16]. Выделение чистой высокомолекулярной ДНК из растений представляет сложную задачу, обусловленную присутствием полисахаридов клеточной стенки и различных метаболитов, включая фенолы, терпены, алкалоиды и флавоноиды, затрудняющих очистку [46]. При этом каждый вид растений содержит свой специфический спектр соединений, препятствующих выделению высококачественной ДНК, что затрудняет разработку универсальных протоколов экстракции ДНК. На сайте сообщества ONT можно найти ряд протоколов, хорошо зарекомендовавших себя при использовании на отдельных видах растений (https://community.nanoporetech.com/extraction methods). Эти протоколы могут быть взяты за основу при оптимизации методики выделения ДНК из интересующего исследователя вида растений. На данный момент предложены различные подходы к экстракции ДНК, пригодной для секвенирования на платформах третьего поколения как с применением коммерческих наборов, так и буферов собственного приготовления. Часто используются модификации с ЦТАБ (цетилтриметиламмонийбромид) в составе лизирующего буфера, в том числе лизирующий буфер Carlson, рекомендованный в нескольких протоколах выделения ДНК из листьев растений (https://community.nanoporetech.com/extraction methods#plant&modal=plant), с последующей дополнительной очисткой с помощью колонок Genomic-tips ("Qiagen", США), магнитных частиц SPRI (solid-phase reversible immobilization), экстракция ДНК определенной длины из агарозного геля с использованием коммерческих наборов или системы BluePippin ("Sage Science", США) [46-51]. Хорошо зарекомендовала себя экстракция ДНК из предварительно изолированных ядер [52, 53]. Для удаления коротких фрагментов ДНК перед секвенированием целесообразно использовать наборы Short Read Eliminator Kit ("Circulomics", США). Чистоту ДНК обычно контролируют спектрофотометрически, например, с помощью NanoDrop ("Thermo Fisher Scientific", США; параметры 260/280 и 260/230 нм должны составлять 1.8-2.0 и 2.0-2.2 соответственно), сравнивают также значения концентраций, измеренные на флуориметрах (часто применяют Qubit, "Thermo Fisher Scientific") и спектрофотометрах – близость значений свидетельствует о высокой чистоте ДНК [16, 48, 49, 51].

Выбор оптимального подхода к секвенированию на платформах третьего поколения с целью *de novo* сборки генома зависит от размера генома и его сложности. При небольших и средних размерах эффективным может быть применение только секвенирования третьего поколения, однако, если размер генома превышает 1 млрд.п.н., то более целесообразным с финансовой точки зрения может быть использование гибридного подхода, когда получают как длинные, так и короткие NGS-прочтения [16]. Кроме того, в случае сложных геномов может потребоваться привлечение дополнительных методов, таких как Hi-C (chromosome conformation capture sequencing) [54], оптическое картирование (optical mapping) [55] или 10× Genomics [56], позволяющих более точно определить относительное расположение участков генома [13, 14].

#### СБОРКА ГЕНОМОВ РАСТЕНИЙ

Помимо прогресса в технологиях секвенирования, важную роль в получении высококачественных сборок геномов растений сыграло развитие методов биоинформатического анализа [57-61]. Выбор стратегии секвенирования и последующей сборки генома зависит от многих параметров, включая размер генома, плоидность, гетерозиготность, сложность выделения высококачественной ДНК и прочее. Возможна de novo сборка генома только из длинных прочтений с платформ ONT или PacBio, однако стоимость такого проекта может быть весьма значительной, особенно в случае геномов большого размера, так как для сборки высокого качества рекомендуется более чем 100-кратное покрытие. Еще один вариант *de novo* сборки геномов – гибридный, когда используются как длинные, так и короткие прочтения, полученные на разных платформах, что позволяет снизить покрытие длинными прочтениями (минимум до 60-кратного в случае геномов длиной до 3 млрд.п.н.) и избавиться от ряда ошибок в сборке благодаря коротким высокоточным прочтениям [29, 62].

Как правило, первичную сборку генома проводят на основе длинных прочтений с платформ третьего поколения. К настоящему времени разработано значительное количество сборщиков геномов, в том числе Canu [63], Flye [64], Shasta [65], wtdbg2 [66], MARVEL [67], MaSuRCA [59], FALCON [68], Minimap/miniasm [69], Raven [70], которые могут использоваться в исследованиях растений, однако отсутствует "идеальный" сборщик – в зависимости от размера и сложности генома и полученного объема и качества данных секвенирования лучшие результаты могут давать разные приложения [29, 48, 70–73]. Стоит также учитывать, что требования сборщиков к вычислительным ресурсам существенно различаются, например, одному из первых сборщиков, Canu, хорошо зарекомендовавшему себя в том числе при сборке геномов растений, необходимы значительные мощности и количество часов для обработки данных, в то время как вышедшие позже приложения (Flye, Shasta, wtdbg2, Raven и др.) требуют в разы меньше ресурсов [48, 72, 74]. Качество геномной сборки оценивают, используя различные параметры, из которых особенно важны размер полученной сборки, число и длина контигов, в первую очередь метрики N50 (максимальная длина, при которой суммарная длина контигов этой длины и более составляет не менее 50% длины сборки) и L50 (число контигов с длиной N50 и более, т.е. минимальное число контигов, суммарная длина которых составляет не менее 50% длины сборки), часто определяемые приложением QUAST [75]), а также метрики BUSCO (benchmarking universal single-copy orthologs), показывающие процент консервативных однокопийных генов-ортологов, характерных для подавляющего большинства представителей той или иной группы видов, и отражающие полноту сборки [76]. Повысить точность полученной сборки может "полировка" с использованием как длинных, так и коротких прочтений – программы Racon [77], Pilon [78], POLCA [59], Nanopolish [79], Medaka (https:// github.com/nanoporetech/medaka) и др. Дополнительным эффективным подходом к сравнительному анализу сборок генома, полученных с применением различных комбинаций сборшиков и полировщиков, и выбора наилучшей является их сопоставление с высококачественным геномом представителя того же или близкородственного вида (обычно с помощью QUAST) [49, 73, 80]. Кроме того, актуален вопрос определения гаплотипов и идентификации аллелей в сборке генома одного растения, что имеет особое значение в

эволюционных исследованиях, а также при изучении гетерозиса, в связи с чем в последнее время предложены различные подходы к фазированию гаплотипов (haplotype phasing) [81]. Существуют также приложения, которые, наоборот, позволяют слить разделенные гаплотипы, такие как Purge Haplotigs, что может быть необходимо, например, при использовании собранного генома в качестве референсного при поиске полиморфизмов или оценке экспрессии генов [82].

#### ВАЖНОСТЬ ПОЛУЧЕНИЯ КАЧЕСТВЕННЫХ СБОРОК ГЕНОМОВ РАСТЕНИЙ

Получение высококачественных сборок геномов растений имеет ключевое значение для последующего анализа с целью идентификации интересующих исследователя генов и регуляторных областей, а также обнаружения гомологов [5].

Секвенирование третьего поколения сделало возможным определение нуклеотидных последовательностей участков генома, сборку и анализ которых затрудняет присутствие нескольких копий генов, повторов, а также транслокаций, в том числе для полового локуса двудомных растений и локуса самонесовместимости растений (self-incompatibility locus, S-locus) [83]. В качестве примера можно привести изучение детерминации пола у видов рода Populus, подавляющее большинство которых являются двудомными — на одних растениях развиваются только женские цветки, на других – только мужские. Секвенирование на платформе Illumina позволило выявить ассоциированные с полом полиморфизмы. Однако в качестве референсного в тот момент использовали геном только женского растения Populus trichocarpa, поэтому эти полиморфизмы оказались ошибочно локализованы на разных хромосомах, и собрать половой локус на основе данных секвенирования второго поколения не удавалось [84, 85]. В 2020-2021 годах сразу несколько научных групп опубликовали результаты изучения видов тополя и осины, полученные с применением секвенирования третьего поколения, в которых был идентифицирован половой локус и показано, что ключевую роль в детерминации пола играет ген ARR17. Частичные повторы гена ARR17 выявили в этом локусе только у мужских растений *P. trichocarpa*, *P. trem*ula, P. deltoides, P. davidiana, P. × sibirica, P. euphratiса и показали, что с двух из этих повторов синтезируются малые РНК, которые опосредованно метилируют ген ARR17 или расщепляют его транскрипт, что приводит к подавлению экспрессии этого гена (локализованного не в половом локусе). Когда ген *ARR17* включен (активен), развиваются женские цветки, а когда он выключен — мужские. Гетерогаметным у перечисленных видов является мужской пол – система XY детерминации пола, а половой локус локализован в теломерной (*P. trichocarpa*, *P. deltoides*, *P. × sibirica*) или перицентромерной (*P. tremula*, *P. davidiana*) области хромосомы 19 или на конце хромосомы 14 (*P. euphratica*) [50, 80, 86–88]. В то же время показано, что *P. alba* имеет ZW систему детерминации пола, ген *ARR17* отсутствует у мужских растений, однако женские растения содержат три его копии [50]. Таким образом, использование для секвенирования геномов длинных прочтений сыграло решающую роль в установлении механизмов детерминации пола у видов *Рориlus* и выявлении разнообразия в локализации полового локуса.

Важное направление, существенно продвинуться на котором помогло знание полных геномов изучение эволюции растений. Серия геномных исследований водорослей, мхов и папоротников позволила пролить свет на эволюцию стрептофитов [89-94]. Получение высококачественных сборок геномов роголистников и водяных лилий привело к выявлению уникальных биологических особенностей, которых нет у других растений, что значительно расширило представление об эволюции наземных растений [95, 96]. Получение генома Ginkgo biloba высокого качества и его анализ привнесли новые данные, касающиеся формирования жгутика сперматофора, необходимые для понимания эволюции голосеменных растений [97]. а изучение геномов водяных лилий оказалось важным для установления ранних этапов эволюции цветковых растений [98, 99]. Наличие значительного числа высококачественных сборок геномов растений позволяет проводить сравнительный анализ как полногеномный, так и отдельных генов или их семейств, играющих ключевую роль в формировании признаков, интересующих исследователей. Полученные за последние годы геномы водорослей, мхов, ликофитов и папоротников не только расширили представление об эволюции наземных растений, но и позволили идентифицировать специфичные для бессемянных растений биологические особенности [100]. Использование геномных данных открыло новые возможности для понимания роли определенных генов в эволюции основных функциональных различий между голосеменными и покрытосеменными растениями [101], а полногеномный филогенетический анализ покрытосеменных растений, выполненный на основе микросинтении, дал дополнительные сведения о спорных вопросах в установлении филогенетических отношений [102].

Еще одно направление, прогресс на котором достигнут благодаря развитию технологий секвенирования и биоинформатических подходов к анализу данных – изучение полных геномов сельскохозяйственных растений и их дикорастущих сородичей. Так, например, удалось значительно продвинуться в понимании эволюции геномов видов *Brassica* и процессов, происходящих при полиплоидизации, что также важно для более эффективной селекции растений семейства Капустные [103, 104].

В последнее время появляется все больше геномов растений, собранных с высоким качеством отдельными научными группами, а также реализуются крупномасштабные проекты по получению референсных геномов для видов эукариот [105], что открывает новые возможности исследования растений, однако необходимы подходы, позволяющие максимально эффективно использовать имеющиеся данные [106].

#### ПАНГЕНОМЫ РАСТЕНИЙ

Высококачественные сборки геномов отдельных представителей видов используются в широком спектре исследований растений. Однако не менее важна оценка внутривидового полиморфизма на геномном уровне. В последнее время анализ так называемого пангенома (pan-genome, от греческого  $\pi \alpha v$  – "все"), включающего в себя всю совокупность нуклеотидных последовательностей, найденных в различных представителях одного вида или популяции, привлекает все больше внимания. Пангеномые исследования направлены не на изучение геномов индивидуальных растений, а на установление разнообразия геномных последовательностей на внутривидовом уровне. Первые подобные работы, проведенные на бактериях, показали, что штаммы существенно различаются по наличию/отсутствию значительного числа генов [107]. Изучение пангеномов эукариот, включая растения, расширило представление о пангеноме, в который помимо генов включили и некодирующие области, представленные, в том числе, промоторами и энхансерами. Пангеном разделяют на основной геном (соге genome, эти последовательности найдены у всех представителей определенного вида), необходимый для выживания организма, и необязательный (вариабельный) геном (dispensable genome, эти последовательности встречаются у одних представителей вида, но отсутствуют у других), причем доля основного генома у разных видов растений существенно различается [108].

Выделяют следующие типы полиморфизмов: однонуклеотидные (single-nucleotide polymorphisms, SNP), непротяженные инсерции/делеции (менее 50 н.) и структурные вариации (structural variants, SV) длиной от 50 до миллионов нуклеотидов [109]. Изучение структурных вариаций, включая различия в наличии/отсутствии определенных последовательностей или их копийности, а также присутствие хромосомных перестроек, представляет важнейший аспект пангеномных исследований [110–112]. Благодаря развитию технологий секвенирования и подходов к сборке геномов стало возможным проводить пангеномный анализ при сравнении сборок геномов высокого качества.

Многие пангеномные работы сфокусированы на определении вариаций в наличии/отсутствии генов, но ценность представляют не только гены, но и регуляторные области, в том числе *цис*-регуляторные элементы (*cis*-regulatory elements) [113, 114], а также повторяющиеся последовательности. Изучение основного генома (core genome) необходимо для выявления генов, без которых организм не может существовать, однако нужно учитывать, что значимость гена для конкретного организма зависит от условий внешней среды и генного окружения: так, один ген может компенсировать отсутствие другого [115].

Разработано несколько подходов к получению пангеномов, в том числе: 1) *de novo* сборка и последующее сравнение полных геномов множества представителей вида; 2) выравнивание прочтений для множества представителей на референсный геном с последующей сборкой невыравненных последовательностей и добавлением их в пангеном; 3) подход, основанный на графах (graph-based pangenome assembly). Третий подход, вероятно, наиболее перспективен для пангеномных исследований растений, однако требует значительных вычислительных мощностей и имеет ограничения, связанные с размером генома и числом анализируемых генотипов [116].

К настоящему времени собраны пангеномы риса [117–119], пшеницы [120, 121], ячменя [122], кукурузы [123], рапса [124], сои и родственных ей видов [110, 125], *Brachypodium* [126] и других видов растений [108, 116, 127], что позволило получить новые данные о генах, влияющих на биомассу растения, вес и состав плодов и семян, цветение и созревание, устойчивость к патогенам и абиотическим стрессорам, другие хозяйственно ценные признаки, а также о генах, на которые оказали влияние процессы доместикации и селекции, идентифицировать новые гены, отсутствующие в референсных геномах, а также выявить важную роль транспозонов в формировании генетического разнообразия растений.

Стоит также отметить получение значительного числа сборок геномов высокого качества грибных патогенов растений (База NCBI, https://www.ncbi. nlm.nih.gov/genome/?term=Fungi). Использование этих данных, в том числе для сравнительного анализа образцов одного вида, характеризующихся различной вирулентностью, может помочь в понимании механизмов взаимодействия растений и патогенов, что будет способствовать созданию устойчивых к биотическим стрессорам сортов. Выполнены пангеномные исследования нескольких фитопатогенов с особым вниманием к генам-эффекторам, связанным с вирулентностью [128–130].

Известно, что применение в геномном анализе только одного, пусть даже хорошо собранного генома, может приводить к искажению результатов. Так, выбор референсного генома оказывает значительное влияние на результаты оценки экспрессии генов и полногеномного поиска ассоциаций (GWAS, genome-wide association studies) [73, 131]. Использование пангенома вместо генома индивидуального генотипа в качестве референсного позволяет с большей точностью идентифицировать полиморфизмы при ресеквенировании геномов и оценивать экспрессию при транскриптомных исследованиях представителей того же вида [116]. Таким образом, изучение пангеномов растений – это перспективное активно развивающееся направление.

#### ИСПОЛЬЗОВАНИЕ ДАННЫХ СЕКВЕНИРОВАНИЯ ГЕНОМОВ В СЕЛЕКЦИИ РАСТЕНИЙ

Успех селекционной работы во многом определяется ценностью и степенью изученности исходного материала, эффективный подбор которого позволит существенно ускорить создание сортов с заданными свойствами [132-135]. В условиях изменяющегося климата и роста населения нашей планеты необходимо совершенствование методов селекционной работы для получения достаточного объема качественной сельскохозяйственной продукции. Использование классических подходов, основанных на отборе наиболее ценных растений по их фенотипическим данным, требует значительных затрат времени, однако привлечение генетических маркеров может сделать процесс создания сортов более эффективным [136, 137]. Благодаря технологическому прогрессу, в том числе совершенствованию методов секвенирования, получены огромные массивы последовательностей геномов растений и собраны референсные геномы наиболее важных сельскохозяйственных культур, что является основой для идентификации генов и маркеров, ассоциированных с хозяйственно значимыми признаками [138]. Развитие маркер-ориентированной селекции, при которой отбор ценных генотипов проводят с использованием отдельных ДНК-маркеров, а затем и геномной селекции, когда селекционную значимость генотипа определяют на основе данных о тысячах ДНКполиморфизмов, позволило существенно повысить эффективность подбора пар для скрещивания и последующего отбора ценных генотипов с целью создания улучшенных сортов [139–143].

Поиск ассоциаций между хозяйственно важными характеристиками и геномными последовательностями проводят с использованием QTL-анализа (quantitative trait loci, локусы количественных признаков), основанного на изучении картирующих популяций, полученных от скрещиваний родителей с контрастными проявлениями признаков, и полногеномного поиска ассоциаций (GWAS), представляющего собой мощный инструмент для изучения генетики комплексных характеристик в популяциях растений [140, 141, 144]. Эти подходы позволяют идентифицировать различные типы ДНК-маркеров, ассоциированных с интересующими исследователя признаками, такие как полиморфизм длин рестрикционных фрагментов (RFLP. restriction fragment length polymorphism), полиморфизм длин амплифицируемых фрагментов (AFLP, amplified fragment length polymorphism), микросателлитные маркеры (SSR, simple sequence repeats), полиморфизм рестрикционных фрагментов амплифицированной ДНК (CAPS, cleaved amplified polymorphic sequences), вставки транспозонов (SSAP, sequence specific amplification polymorphism), однонуклеотидные полиморфизмы (SNP), инсерции/делеции (InDel) [145]. Технологии высокопроизводительного секвенирования позволили получать достаточно плотное покрытие генома маркерами, а полные геномы сделали возможным идентификацию не только ассоциированных с признаком ДНК-маркеров, но и определение вероятных генов, детерминирующих этот признак. В качестве примера использования полногеномной сборки для картирования ранее найденных QTL и предсказания генов-кандидатов, определяющих ценные характеристики, можно привести работу You и Cloutier, выполненную на растениях льна [146]. В настоящее время полногеномный поиск ассоциаций, направленный на изучение признаков, связанных с урожайностью, размером и архитектоникой растений, временем цветения, биохимическим составом, устойчивостью к биотическим и абиотическим стрессорам, проведен для большинства основных сельскохозяйственных культур, в том числе пшеницы, риса, кукурузы, сои, ячменя, сорго и хлопка, с применением различных методов анализа [141, 145].

Необходимо отметить, что использование ресеквенирования и только одного генома в качестве референсного ограничивает возможность идентификации структурных вариаций (SV), которые часто встречаются в сельскохозяйственных культурах и играют важную роль в детерминации ценных признаков [112]. Проблему выявления структурных вариаций может решить пангеномный анализ. Необязательный (dispensable) геном содержит значительное число генов, ассоциированных с хозяйственно важными характеристиками. включая устойчивость к биотическим и абиотическим стрессорам, а его изучение поможет более точно и эффективно отобрать селекционный материал для создания сортов и линий с заданными свойствами [112]. Кроме того, структурные вариации могут вносить вклад в гетерозис, их выявление необходимо для классификации гетеротических групп [147, 148], что будет способствовать более обоснованному выбору родительских линий и получению гибридов с улучшенными признаками. Более того, пангеномный анализ с привлечением староместных сортов и диких предков культивируемых растений перспективен для идентификации генов, важных для доместикации и адаптации, знания о которых представляют ценность для направленной селекционной работы по улучшению сельскохозяйственных растений с эффективным привлечением существующего генофонда [6, 112]. Таким образом, прогресс в технологиях секвенирования и анализа геномов растений предоставляет новые инструменты для более эффективного ведения селекции и создания высокопродуктивных сортов с комплексом хозяйственно ценных характеристик и улучшенными свойствами, необходимыми для удовлетворения потребностей растущего населения Земли.

#### ЗАКЛЮЧЕНИЕ

Современные технологии секвенирования открыли новые возможности проведения генетических исследований растений. Разработаны и развиваются подходы к получению высококачественных сборок геномов растений, а также пангеномов, служащих основой для функциональных исследований, изучения генетического разнообразия и селекционной работы. Сложно представить дальнейшее эффективное ведение сельского хозяйства без привлечения достижений современной науки, при этом нельзя забывать о балансе между удовлетворением потребностей человечества и сохранением нашей планеты.

Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований в рамках научного проекта № 20-116-50144.

Настоящая статья не содержит каких-либо исследований с участием людей или животных в качестве объектов исследований.

Авторы заявляют об отсутствии конфликта интересов.

#### СПИСОК ЛИТЕРАТУРЫ

- 1. Schaal B. (2019) Plants and people: our shared history and future. *Plants, People, Planet.* **1**, 14–19.
- Corlett R.T. (2016) Plant diversity in a changing world: status, trends, and conservation needs. *Plant Diversity*. 38, 10–16.
- 3. Jimenez-Alfaro B., Frischie S., Stolz J., Galvez-Ramirez C. (2020) Native plants for greening Mediterranean agroecosystems. *Nat. Plants.* **6**, 209–214.
- 4. Zenda T., Liu S., Dong A., Duan H. (2021) Advances in cereal crop genomics for resilience under climate change. *Life*. **11**, 502.

- 5. Pourkheirandish M., Golicz A.A., Bhalla P.L., Singh M.B. (2020) Global role of crop genomics in the face of climate change. *Front. Plant Sci.* **11**, 922.
- Razzaq A., Kaur P., Akhter N., Wani S.H., Saleem F. (2021) Next-generation breeding strategies for climate-ready crops. *Front. Plant Sci.* 12, 620420.
- 7. The *Arabidopsis* Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. **408**, 796–815.
- 8. International Rice Genome Sequencing Project & Takuji Sasaki. (2005) The map-based sequence of the rice genome. *Nature*. **436**, 793–800.
- 9. Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F., Pasternak S., Liang C., Zhang J., Fulton L., Graves T.A., Minx P., Reily A.D., Courtney L., Kruchowski S.S., Tomlinson C., Strong C., Delehaunty K., Fronick C., Courtney B., Rock S.M., Belter E., Du F., Kim K., Abbott R.M., Cotton M., Levy A., Marchetto P., Ochoa K., Jackson S.M., Gillam B., Chen W., Yan L., Higginbotham J., Cardenas M., Waligorski J., Applebaum E., Phelps L., Falcone J., Kanchi K., Thane T., Scimone A., Thane N., Henke J., Wang T., Ruppert J., Shah N., Rotter K., Hodges J., Ingenthron E., Cordes M., Kohlberg S., Sgro J., Delgado B., Mead K., Chinwalla A., Leonard S., Crouse K., Collura K., Kudrna D., Currie J., He R., Angelova A., Rajasekar S., Mueller T., Lomeli R., Scara G., Ko A., Delaney K., Wissotski M., Lopez G., Campos D., Braidotti M., Ashlev E., Golser W., Kim H., Lee S., Lin J., Duimic Z., Kim W., Talag J., Zuccolo A., Fan C., Sebastian A., Kramer M., Spiegel L., Nascimento L., Zutavern T., Miller B., Ambroise C., Muller S., Spooner W., Narechania A., Ren L., Wei S., Kumari S., Faga B., Levy M.J., McMahan L., Van Buren P., Vaughn M.W., Ying K., Yeh C.T., Emrich S.J., Jia Y., Kalyanaraman A., Hsia A.P., Barbazuk W.B., Baucom R.S., Brutnell T.P., Carpita N.C., Chaparro C., Chia J.M., Deragon J.M., Estill J.C., Fu Y., Jeddeloh J.A., Han Y., Lee H., Li P., Lisch D.R., Liu S., Liu Z., Nagel D.H., McCann M.C., SanMiguel P., Myers A.M., Nettleton D., Nguyen J., Penning B.W., Ponnala L., Schneider K.L., Schwartz D.C., Sharma A., Soderlund C., Springer N.M., Sun Q., Wang H., Waterman M., Westerman R., Wolfgruber T.K., Yang L., Yu Y., Zhang L., Zhou S., Zhu Q., Bennetzen J.L., Dawe R.K., Jiang J., Jiang N., Presting G.G., Wessler S.R., Aluru S., Martienssen R.A., Clifton S.W., McCombie W.R., Wing R.A., Wilson R.K. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science. 326, 1112-1115.
- Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X.C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C., Jackson S.A. (2010) Genome sequence of the palaeopolyploid soybean. *Nature.* 463, 178–183.
- Reuter J.A., Spacek D.V., Snyder M.P. (2015) Highthroughput sequencing technologies. *Mol. Cell.* 58, 586–597.

- Goodwin S., McPherson J.D., McCombie W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351.
- 13. Li C., Lin F., An D., Wang W., Huang R. (2017) Genome sequencing and assembly by long reads in plants. *Genes.* **9**, 6.
- Jiao W.B., Schneeberger K. (2017) The impact of third generation genomic technologies on plant genome assembly. *Curr. Opin. Plant Biol.* 36, 64–70.
- Jiao Y., Peluso P., Shi J., Liang T., Stitzer M.C., Wang B., Campbell M.S., Stein J.C., Wei X., Chin C.S., Guill K., Regulski M., Kumari S., Olson A., Gent J., Schneider K.L., Wolfgruber T.K., May M.R., Springer N.M., Antoniou E., McCombie W.R., Presting G.G., Mc-Mullen M., Ross-Ibarra J., Dawe R.K., Hastie A., Rank D.R., Ware D. (2017) Improved maize reference genome with single-molecule technologies. *Nature*. 546, 524–527.
- 16. Li F.W., Harkess A. (2018) A guide to sequence your favorite plant genomes. *Appl. Plant Sci.* **6**, e1030.
- Wenger A.M., Peluso P., Rowell W.J., Chang P.C., Hall R.J., Concepcion G.T., Ebler J., Fungtammasan A., Kolesnikov A., Olson N.D., Topfer A., Alonge M., Mahmoud M., Qian Y., Chin C.S., Phillippy A.M., Schatz M.C., Myers G., DePristo M.A., Ruan J., Marschall T., Sedlazeck F.J., Zook J.M., Li H., Koren S., Carroll A., Rank D.R., Hunkapiller M.W. (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 1155–1162.
- Sereika M., Kirkegaard R.H., Karst S.M., Michaelsen T.Y., Sørensen E.A., Wollenberg R.D., Albertsen M. (2021) Oxford Nanopore R10.4 long-read sequencing enables near-perfect bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *bioRxiv*. 2021.10.27.466057.
- Брагина М.К., Афонников Д.А., Салина Е.А. (2019) Прогресс в секвенировании геномов растений – направления исследований. Вавиловский журн. генет. селекции. 23, 38–48.
- Chen F., Dong W., Zhang J., Guo X., Chen J., Wang Z., Lin Z., Tang H., Zhang L. (2018) The sequenced angiosperm genomes and genome databases. *Front. Plant Sci.* 9, 418.
- 21. Kersey P.J. (2019) Plant genome sequences: past, present, future. *Curr. Opin. Plant Biology.* **48**, 1–8.
- Lewin H.A., Graves J.A.M., Ryder O.A., Graphodatsky A.S., O'Brien S.J. (2019) Precision nomenclature for the new genomics. *GigaScience*. 8, giz086.
- 23. Nurk S., Koren S., Rhie A., Rautiainen M., Bzikadze A.V., Mikheenko A., Vollger M.R., Altemose N., Uralsky L., Gershman A., Aganezov S., Hoyt S.J., Diekhans M., Logsdon G.A., Alonge M., Antonarakis S.E., Borchers M., Bouffard G.G., Brooks S.Y., Caldas G.V., Cheng H., Chin C.-S., Chow W., de Lima L.G., Dishuck P.C., Durbin R., Dvorkina T., Fiddes I.T., Formenti G., Fulton R.S., Fungtammasan A., Garrison E., Grady P.G.S., Graves-Lindsay T.A., Hall I.M., Hansen N.F., Hartley G.A., Haukness M., Howe K., Hunkapiller M.W., Jain C., Jain M., Jarvis E.D., Kerpedjiev P., Kirsche M., Kolmogorov M., Korlach J., Kremitzki M., Li H., Maduro V.V., Marschall T., McCartney A.M., McDaniel J., Miller D.E.,

Mullikin J.C., Myers E.W., Olson N.D., Paten B., Peluso P., Pevzner P.A., Porubsky D., Potapova T., Rogaev E.I., Rosenfeld J.A., Salzberg S.L., Schneider V.A., Sedlazeck F.J., Shafin K., Shew C.J., Shumate A., Sims Y., Smit A.F.A., Soto D.C., Sović I., Storer J.M., Streets A., Sullivan B.A., Thibaud-Nissen F., Torrance J., Wagner J., Walenz B.P., Wenger A., Wood J.M.D., Xiao C., Yan S.M., Young A.C., Zarate S., Surti U., McCoy R.C., Dennis M.Y., Alexandrov I.A., Gerton J.L., O'Neill R.J., Timp W., Zook J.M., Schatz M.C., Eichler E.E., Miga K.H., Phillippy A.M. (2021) The complete sequence of a human genome. *bioRxiv*. 2021.05.26.445798.

- 24. Zwyrtkova J., Simkova H., Dolezel J. (2021) Chromosome genomics uncovers plant genome organization and function. *Biotechnol. Adv.* **46**, 107659.
- Sun Y., Shang L., Zhu Q.H., Fan L., Guo L. (2021) Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci.* S1360-1385(21)00281-8.
- Fleischmann A., Michael T.P., Rivadavia F., Sousa A., Wang W., Temsch E.M., Greilhuber J., Muller K.F., Heubl G. (2014) Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Ann. Botany.* 114, 1651–1663.
- Hidalgo O., Pellicer J., Christenhusz M., Schneider H., Leitch A.R., Leitch I.J. (2017) Is there an upper limit to genome size? *Trends Plant Sci.* 22, 567–573.
- Adams K.L., Wendel J.F. (2005) Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* 8, 135–141.
- 29. Jung H., Winefield C., Bombarely A., Prentis P., Waterhouse P. (2019) Tools and strategies for long-read sequencing and *de novo* assembly of plant genomes. *Trends Plant Sci.* 24, 700–724.
- Thudi M., Khan A.W., Kumar V., Gaur P.M., Katta K., Garg V., Roorkiwal M., Samineni S., Varshney R.K. (2016) Whole genome re-sequencing reveals genomewide variations among parental lines of 16 mapping populations in chickpea (*Cicer arietinum L.*). *BMC Plant Biol.* 16 Suppl. 1, 10.
- Bayer P.E., Hurgobin B., Golicz A.A., Chan C.K., Yuan Y., Lee H., Renton M., Meng J., Li R., Long Y., Zou J., Bancroft I., Chalhoub B., King G.J., Batley J., Edwards D. (2017) Assembly and comparison of two closely related *Brassica napus* genomes. *Plant Biotechnol. J.* 15, 1602–1610.
- 32. Schatz M.C., Maron L.G., Stein J.C., Hernandez Wences A., Gurtowski J., Biggers E., Lee H., Kramer M., Antoniou E., Ghiban E., Wright M.H., Chia J.M., Ware D., McCouch S.R., McCombie W.R. (2014) Whole genome *de novo* assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biol.* **15**, 506.
- 33. Minio A., Lin J., Gaut B.S., Cantu D. (2017) How single molecule real-time sequencing and haplotype phasing have enabled reference-grade diploid genome assembly of wine grapes. *Front. Plant Sci.* **8**, 826.
- Deamer D., Akeson M., Branton D. (2016) Three decades of nanopore sequencing. *Nat. Biotechnol.* 34, 518–524.
  - МОЛЕКУЛЯРНАЯ БИОЛОГИЯ том 56 № 4 2022

- Rhoads A., Au K.F. (2015) PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*. 13, 278–289.
- van Dijk E.L., Jaszczyszyn Y., Naquin D., Thermes C. (2018) The third revolution in sequencing technology. *Trends Genet. TIG.* 34, 666–681.
- Payne A., Holmes N., Rakyan V., Loose M. (2019) BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*. 35, 2193–2198.
- De Coster W., Weissensteiner M.H., Sedlazeck F.J. (2021) Towards population-scale long-read sequencing. *Nat. Rev. Genetics.* 22, 572–587.
- 39. Gouil Q., Keniry A. (2019) Latest techniques to study DNA methylation. *Essays Biochem.* **63**, 639–648.
- Flusberg B.A., Webster D.R., Lee J.H., Travers K.J., Olivares E.C., Clark T.A., Korlach J., Turner S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Meth.* 7, 461–465.
- Simpson J.T., Workman R.E., Zuzarte P.C., David M., Dursi L.J., Timp W. (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Meth.* 14, 407–410.
- 42. Rand A.C., Jain M., Eizenga J.M., Musselman-Brown A., Olsen H.E., Akeson M., Paten B. (2017) Mapping DNA methylation with high-throughput nanopore sequencing. *Nat. Meth.* **14**, 411–413.
- Kumar S., Mohapatra T. (2021) Dynamics of DNA methylation and its functions in plant growth and development. *Front. Plant Sci.* 12, 596236.
- Ramos-Cruz D., Troyee A.N., Becker C. (2021) Epigenetics in plant organismic interactions. *Curr. Opin. Plant Biol.* 61, 102060.
- Akhter Z., Bi Z., Ali K., Sun C., Fiaz S., Haider F.U., Bai J. (2021) In response to abiotic stress, DNA methylation confers epigenetic changes in plants. *Plants*. 10, 1096.
- Vaillancourt B., Buell C.R. (2019) High molecular weight DNA isolation method from diverse plant species for use with Oxford nanopore sequencing. *bioRxiv*. 783159.
- 47. Siadjeu C., Pucker B., Viehover P., Albach D.C., Weisshaar B. (2020) High contiguity *de novo* genome sequence assembly of trifoliate yam (*Dioscorea dumetorum*) using long read sequencing. *Genes.* **11**, 274.
- Krasnov G.S., Pushkova E.N., Novakovskiy R.O., Kudryavtseva L.P., Povkhova T.A., Dvorianinova E.M., Povkhova L.V., Kudryavtseva A.V., Dmitriev A.A., Melnikova N.V. (2020) High-quality genome assembly of *Fusarium oxysporum* f. sp. lini. *Front. Genet.* 11, 959.
- 49. Dmitriev A.A., Pushkova E.N., Novakovskiy R.O., Beniaminov A.D., Rozhmina T.A., Zhuchenko A.A., Bolsheva N.L., Muravenko O.V., Povkhova L.V., Dvorianinova E.M., Kezimana P., Snezhkina A.V., Kudryavtseva A.V., Krasnov G.S., Melnikova N.V. (2020) Genome sequencing of fiber flax cultivar atlant using Oxford nanopore and illumina platforms. *Front Genet.* 11, 590282.
- Müller N.A., Kersten B., Leite Montalvão A.P., Mähler N., Bernhardsson C., Bräutigam K., Carracedo Lorenzo Z., Hoenicka H., Kumar V., Mader M.,

Pakull B., Robinson K.M., Sabatti M., Vettori C., Ingvarsson P.K., Cronk Q., Street N.R., Fladung M. (2020) A single gene underlies the dynamic evolution of poplar sex determination. *Nat. Plants.* **6**, 630–637.

- Schalamun M., Nagar R., Kainer D., Beavan E., Eccles D., Rathjen J.P., Lanfear R., Schwessinger B. (2019) Harnessing the MinION: an example of how to establish long-read sequencing in a laboratory using challenging plant tissue from *Eucalyptus pauciflora*. *Mol. Ecol. Resources.* **19**, 77–89.
- Zhang M., Zhang Y., Scheuring C.F., Wu C.C., Dong J.J., Zhang H.B. (2012) Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat. Protocols.* 7, 467–478.
- Li Z., Parris S., Saski C.A. (2020) A simple plant highmolecular-weight DNA extraction method suitable for single-molecule technologies. *Plant Meth.* 16, 38.
- Lieberman-Aiden E., van Berkum N.L., Williams L., Imakaev M., Ragoczy T., Telling A., Amit I., Lajoie B.R., Sabo P.J., Dorschner M.O., Sandstrom R., Bernstein B., Bender M.A., Groudine M., Gnirke A., Stamatoyannopoulos J., Mirny L.A., Lander E.S., Dekker J. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 326, 289–293.
- 55. Yuan Y., Chung C.Y., Chan T.F. (2020) Advances in optical mapping for genomic research. *Comput. Struct. Biotechnol. J.* **18**, 2051–2062.
- 56. Ma Z.S., Li L., Ye C., Peng M., Zhang Y.P. (2019) Hybrid assembly of ultra-long nanopore reads augmented with 10×-Genomics contigs: demonstrated with a human genome. *Genomics*. **111**, 1896–1901.
- 57. Schmidt M.H., Vogel A., Denton A.K., Istace B., Wormit A., van de Geest H., Bolger M.E., Alseekh S., Mass J., Pfaff C., Schurr U., Chetelat R., Maumus F., Aury J.M., Koren S., Fernie A.R., Zamir D., Bolger A.M., Usadel B. (2017) *De novo* assembly of a new *Solanum pennellii* accession using nanopore sequencing. *Plant Cell.* 29, 2336–2348.
- Michael T.P., Jupe F., Bemm F., Motley S.T., Sandoval J.P., Lanz C., Loudet O., Weigel D., Ecker J.R. (2018) High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* 9, 541.
- 59. Zimin A.V., Puiu D., Luo M.C., Zhu T., Koren S., Marcais G., Yorke J.A., Dvorak J., Salzberg S.L. (2017) Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* 27, 787–792.
- Zimin A.V., Stevens K.A., Crepeau M.W., Puiu D., Wegrzyn J.L., Yorke J.A., Langley C.H., Neale D.B., Salzberg S.L. (2017) An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *GigaScience*. 6, 1–4.
- 61. Mondal T.K., Rawal H.C., Gaikwad K., Sharma T.R., Singh N.K. (2017) First *de novo* draft genome sequence of *Oryza coarctata*, the only halophytic species in the genus *Oryza*. *F1000Research*. **6**, 1750.

- 62. Zhang H., Jain C., Aluru S. (2020) A comprehensive evaluation of long read error correction methods. *BMC genomics.* **21**, 889.
- 63. Koren S., Walenz B.P., Berlin K., Miller J.R., Bergman N.H., Phillippy A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736.
- Kolmogorov M., Yuan J., Lin Y., Pevzner P.A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546.
- Shafin K., Pesout T., Lorig-Roach R., Haukness M., Olsen H.E., Bosworth C., Armstrong J., Tigyi K., Maurer N., Koren S., Sedlazeck F.J., Marschall T., Mayes S., Costa V., Zook J.M., Liu K.J., Kilburn D., Sorensen M., Munson K.M., Vollger M.R., Monlong J., Garrison E., Eichler E.E., Salama S., Haussler D., Green R.E., Akeson M., Phillippy A., Miga K.H., Carnevali P., Jain M., Paten B. (2020) Nanopore sequencing and the Shasta toolkit enable efficient *de novo* assembly of eleven human genomes. *Nat. Biotechnol.* 38, 1044–1053.
- 66. Ruan J., Li H. (2020) Fast and accurate long-read assembly with wtdbg2. *Nat. Meth.* **17**, 155–158.
- Nowoshilow S., Schloissnig S., Fei J.F., Dahl A., Pang A.W.C., Pippel M., Winkler S., Hastie A.R., Young G., Roscito J.G., Falcon F., Knapp D., Powell S., Cruz A., Cao H., Habermann B., Hiller M., Tanaka E.M., Myers E.W. (2018) The axolotl genome and the evolution of key tissue formation regulators. *Nature*. 554, 50–55.
- Chin C.S., Peluso P., Sedlazeck F.J., Nattestad M., Concepcion G.T., Clum A., Dunn C., O'Malley R., Figueroa-Balderas R., Morales-Cruz A., Cramer G.R., Delledonne M., Luo C., Ecker J.R., Cantu D., Rank D.R., Schatz M.C. (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Meth.* 13, 1050–1054.
- Li H. (2016) Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*. 32, 2103–2110.
- Vaser R., Šikić M. (2021) Time- and memory-efficient genome assembly with Raven. *Nat. Comput. Sci.* 1, 332–336.
- Jayakumar V., Sakakibara Y. (2019) Comprehensive evaluation of non-hybrid genome assembly tools for third-generation PacBio long-read sequence data. *Brief. Bioinform.* 20, 866–876.
- Murigneux V., Rai S.K., Furtado A., Bruxner T.J.C., Tian W., Harliwong I., Wei H., Yang B., Ye Q., Anderson E., Mao Q., Drmanac R., Wang O., Peters B.A., Xu M., Wu P., Topp B., Coin L.J.M., Henry R.J. (2020) Comparison of long-read methods for sequencing and assembly of a plant genome. *Giga-Science*. 9, giaa146.
- Dvorianinova E.M., Pushkova E.N., Novakovskiy R.O., Povkhova L.V., Bolsheva N.L., Kudryavtseva L.P., Rozhmina T.A., Melnikova N.V., Dmitriev A.A. (2021) Nanopore and Illumina genome sequencing of *Fusarium oxysporum* f. sp. lini strains of different virulence. *Front. Genet.* 12, 662928.

МОЛЕКУЛЯРНАЯ БИОЛОГИЯ том 56 № 4 2022

- 74. Jung H., Jeon M.S., Hodgett M., Waterhouse P., Eyun S.I. (2020) Comparative evaluation of genome assemblers from long-read sequencing for plants and crops. J. Agricult. Food Chem. 68, 7670–7677.
- Gurevich A., Saveliev V., Vyahhi N., Tesler G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 29, 1072–1075.
- Seppey M., Manni M., Zdobnov E.M. (2019) BUS-CO: assessing genome assembly and annotation completeness. *Meth. Mol. Biol.* **1962**, 227–245.
- Vaser R., Sovic I., Nagarajan N., Sikic M. (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746.
- Walker B.J., Abeel T., Shea T., Priest M., Abouelliel A., Sakthikumar S., Cuomo C.A., Zeng Q., Wortman J., Young S.K., Earl A.M. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One.* 9, e112963.
- Loman N.J., Quick J., Simpson J.T. (2015) A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat. Meth.* 12, 733–735.
- Melnikova N.V., Pushkova E.N., Dvorianinova E.M., Beniaminov A.D., Novakovskiy R.O., Povkhova L.V., Bolsheva N.L., Snezhkina A.V., Kudryavtseva A.V., Krasnov G.S., Dmitriev A.A. (2021) Genome assembly and sex-determining region of male and female *Populus × sibirica. Front. Plant Sci.* 12, 625416.
- Zhang X., Wu R., Wang Y., Yu J., Tang H. (2020) Unzipping haplotypes in diploid and polyploid genomes. *Comput. Struct. Biotechnol. J.* 18, 66–72.
- Roach M.J., Schmidt S.A., Borneman A.R. (2018) Purge Haplotigs: allelic contig reassignment for thirdgen diploid genome assemblies. *BMC Bioinformatics*. 19, 460.
- Vekemans X., Castric V., Hipperson H., Muller N.A., Westerdahl H., Cronk Q. (2021) Whole-genome sequencing and genome regions of special interest: lessons from major histocompatibility complex, sex determination, and plant self-incompatibility. *Mol. Ecol.* 30, 6072–6086.
- 84. Geraldes A., Hefer C.A., Capron A., Kolosova N., Martinez-Nunez F., Soolanayakanahally R.Y., Stanton B., Guy R.D., Mansfield S.D., Douglas C.J., Cronk Q.C. (2015) Recent Y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Mol. Ecol.* 24, 3243–3256.
- McKown A.D., Klapste J., Guy R.D., Soolanayakanahally R.Y., La Mantia J., Porth I., Skyba O., Unda F., Douglas C.J., El-Kassaby Y.A., Hamelin R.C., Mansfield S.D., Cronk Q.C.B. (2017) Sexual homomorphism in dioecious trees: extensive tests fail to detect sexual dimorphism in *Populus* (dagger). *Sci. Rep.* 7, 1831.
- Xue L., Wu H., Chen Y., Li X., Hou J., Lu J., Wei S., Dai X., Olson M.S., Liu J., Wang M., Charlesworth D., Yin T. (2020) Evidences for a role of two Y-specific genes in sex determination in *Populus deltoides*. *Nat. Commun.* 11, 5893.
- Zhou R., Macaya-Sanz D., Schmutz J., Jenkins J.W., Tuskan G.A., DiFazio S.P. (2020) Sequencing and analysis of the sex determination region of *Populus trichocarpa*. *Genes.* 11, 843.

МОЛЕКУЛЯРНАЯ БИОЛОГИЯ том 56 № 4 2022

- Yang W., Wang D., Li Y., Zhang Z., Tong S., Li M., Zhang X., Zhang L., Ren L., Ma X., Zhou R., Sanderson B.J., Keefover-Ring K., Yin T., Smart L.B., Liu J., DiFazio S.P., Olson M., Ma T. (2021) A general model to explain repeated turnovers of sex determination in the Salicaceae. *Mol. Biol. Evol.* 38, 968–980.
- 89. Nishiyama T., Sakayama H., de Vries J., Buschmann H., Saint-Marcoux D., Ullrich K.K., Haas F.B., Vanderstraeten L., Becker D., Lang D., Vosolsobe S., Rombauts S., Wilhelmsson P.K.I., Janitza P., Kern R., Heyl A., Rumpler F., Villalobos L., Clay J.M., Skokan R., Toyoda A., Suzuki Y., Kagoshima H., Schijlen E., Tajeshwar N., Catarino B., Hetherington A.J., Saltykova A., Bonnot C., Breuninger H., Symeonidi A., Radhakrishnan G.V., Van Nieuwerburgh F., Deforce D., Chang C., Karol K.G., Hedrich R., Ulvskov P., Glockner G., Delwiche C.F., Petrasek J., Van de Peer Y., Friml J., Beilby M., Dolan L., Kohara Y., Sugano S., Fujiyama A., Delaux P.M., Quint M., Theissen G., Hagemann M., Harholt J., Dunand C., Zachgo S., Langdale J., Maumus F., Van Der Straeten D., Gould S.B., Rensing S.A. (2018) The Chara genome: secondary complexity and implications for plant terrestrialization. Cell. 174, 448-464.
- 90. Hori K., Maruyama F., Fujisawa T., Togashi T., Yamamoto N., Seo M., Sato S., Yamada T., Mori H., Tajima N., Moriyama T., Ikeuchi M., Watanabe M., Wada H., Kobayashi K., Saito M., Masuda T., Sasaki-Sekimoto Y., Mashiguchi K., Awai K., Shimojima M., Masuda S., Iwai M., Nobusawa T., Narise T., Kondo S., Saito H., Sato R., Murakawa M., Ihara Y., Oshima-Yamada Y., Ohtaka K., Satoh M., Sonobe K., Ishii M., Ohtani R., Kanamori-Sato M., Honoki R., Miyazaki D., Mochizuki H., Umetsu J., Higashi K., Shibata D., Kamiya Y., Sato N., Nakamura Y., Tabata S., Ida S., Kurokawa K., Ohta H. (2014) *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5, 3978.
- 91. Banks J.A., Nishiyama T., Hasebe M., Bowman J.L., Gribskov M., dePamphilis C., Albert V.A., Aono N., Aoyama T., Ambrose B.A., Ashton N.W., Axtell M.J., Barker E., Barker M.S., Bennetzen J.L., Bonawitz N.D., Chapple C., Cheng C., Correa L.G., Dacre M., De-Barry J., Dreyer I., Elias M., Engstrom E.M., Estelle M., Feng L., Finet C., Floyd S.K., Frommer W.B., Fujita T., Gramzow L., Gutensohn M., Harholt J., Hattori M., Heyl A., Hirai T., Hiwatashi Y., Ishikawa M., Iwata M., Karol K.G., Koehler B., Kolukisaoglu U., Kubo M., Kurata T., Lalonde S., Li K., Li Y., Litt A., Lyons E., Manning G., Maruyama T., Michael T.P., Mikami K., Miyazaki S., Morinaga S., Murata T., Mueller-Roeber B., Nelson D.R., Obara M., Oguri Y., Olmstead R.G., Onodera N., Petersen B.L., Pils B., Prigge M., Rensing S.A., Riano-Pachon D.M., Roberts A.W., Sato Y., Scheller H.V., Schulz B., Schulz C., Shakirov E.V., Shibagaki N., Shinohara N., Shippen D.E., Sorensen I., Sotooka R., Sugimoto N., Sugita M., Sumikawa N., Tanurdzic M., Theisen G., Ulvskov P., Wakazuki S., Weng J.K., Willats W.W., Wipf D., Wolf P.G., Yang L., Zimmer A.D., Zhu Q., Mitros T., Hellsten U., Loque D., Otillar R., Salamov A., Schmutz J., Shapiro H., Lindquist E., Lucas S., Rokhsar D., Grigoriev I.V. (2011) The Selaginella ge-

nome identifies genetic changes associated with the evolution of vascular plants. *Science*. **332**, 960–963.

- 92. Lang D., Ullrich K.K., Murat F., Fuchs J., Jenkins J., Haas F.B., Piednoel M., Gundlach H., Van Bel M., Meyberg R., Vives C., Morata J., Symeonidi A., Hiss M., Muchero W., Kamisugi Y., Saleh O., Blanc G., Decker E.L., van Gessel N., Grimwood J., Hayes R.D., Graham S.W., Gunter L.E., McDaniel S.F., Hoernstein S.N.W., Larsson A., Li F.W., Perroud P.F., Phillips J., Ranjan P., Rokshar D.S., Rothfels C.J., Schneider L., Shu S., Stevenson D.W., Thummler F., Tillich M., Villarreal Aguilar J.C., Widiez T., Wong G.K., Wymore A., Zhang Y., Zimmer A.D., Quatrano R.S., Mayer K.F.X., Goodstein D., Casacuberta J.M., Vandepoele K., Reski R., Cuming A.C., Tuskan G.A., Maumus F., Salse J., Schmutz J., Rensing S.A. (2018) The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. Plant J.: Cell Mol. Biol. 93, 515-533.
- 93. Li F.W., Brouwer P., Carretero-Paulet L., Cheng S., de Vries J., Delaux P.M., Eily A., Koppers N., Kuo L.Y., Li Z., Simenc M., Small I., Wafula E., Angarita S., Barker M.S., Brautigam A., dePamphilis C., Gould S., Hosmani P.S., Huang Y.M., Huettel B., Kato Y., Liu X., Maere S., McDowell R., Mueller L.A., Nierop K.G.J., Rensing S.A., Robison T., Rothfels C.J., Sigel E.M., Song Y., Timilsena P.R., Van de Peer Y., Wang H., Wilhelmsson P.K.I., Wolf P.G., Xu X., Der J.P., Schluepmann H., Wong G.K., Pryer K.M. (2018) Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants.* 4, 460–472.
- 94. Wang S., Li L., Li H., Sahu S.K., Wang H., Xu Y., Xian W., Song B., Liang H., Cheng S., Chang Y., Song Y., Cebi Z., Wittek S., Reder T., Peterson M., Yang H., Wang J., Melkonian B., Van de Peer Y., Xu X., Wong G.K., Melkonian M., Liu H., Liu X. (2020) Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat. Plants.* 6, 95–106.
- 95. Li F.W., Nishiyama T., Waller M., Frangedakis E., Keller J., Li Z., Fernandez-Pozo N., Barker M.S., Bennett T., Blazquez M.A., Cheng S., Cuming A.C., de Vries J., de Vries S., Delaux P.M., Diop I.S., Harrison C.J., Hauser D., Hernandez-Garcia J., Kirbis A., Meeks J.C., Monte I., Mutte S.K., Neubauer A., Quandt D., Robison T., Shimamura M., Rensing S.A., Villarreal J.C., Weijers D., Wicke S., Wong G.K., Sakakibara K., Szovenyi P. (2020) Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants.* 6, 259–272.
- 96. Zhang J., Fu X.X., Li R.Q., Zhao X., Liu Y., Li M.H., Zwaenepoel A., Ma H., Goffinet B., Guan Y.L., Xue J.Y., Liao Y.Y., Wang Q.F., Wang Q.H., Wang J.Y., Zhang G.Q., Wang Z.W., Jia Y., Wang M.Z., Dong S.S., Yang J.F., Jiao Y.N., Guo Y.L., Kong H.Z., Lu A.M., Yang H.M., Zhang S.Z., Van de Peer Y., Liu Z.J., Chen Z.D. (2020) The hornwort genome and early land plant evolution. *Nat. Plants.* 6, 107–118.
- 97. Liu H., Wang X., Wang G., Cui P., Wu S., Ai C., Hu N., Li A., He B., Shao X., Wu Z., Feng H., Chang Y., Mu D., Hou J., Dai X., Yin T., Ruan J., Cao F. (2021) The nearly complete genome of *Ginkgo biloba* illuminates gymnosperm evolution. *Nat. Plants.* 7, 748–756.

- Yang Y., Sun P., Lv L., Wang D., Ru D., Li Y., Ma T., Zhang L., Shen X., Meng F., Jiao B., Shan L., Liu M., Wang Q., Qin Z., Xi Z., Wang X., Davis C.C., Liu J. (2020) Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat. Plants.* 6, 215–222.
- 99. Zhang L., Chen F., Zhang X., Li Z., Zhao Y., Lohaus R., Chang X., Dong W., Ho S.Y.W., Liu X., Song A., Chen J., Guo W., Wang Z., Zhuang Y., Wang H., Chen X., Hu J., Liu Y., Qin Y., Wang K., Dong S., Liu Y., Zhang S., Yu X., Wu Q., Wang L., Yan X., Jiao Y., Kong H., Zhou X., Yu C., Chen Y., Li F., Wang J., Chen W., Chen X., Jia Q., Zhang C., Jiang Y., Zhang W., Liu G., Fu J., Chen F., Ma H., Van de Peer Y., Tang H. (2020) The water lily genome and the early evolution of flowering plants. *Nature*. **577**, 79–84.
- Szovenyi P., Gunadi A., Li F.W. (2021) Charting the genomic landscape of seed-free plants. *Nat. Plants.* 7, 554–565.
- De La Torre A.R., Piot A., Liu B., Wilhite B., Weiss M., Porth I. (2020) Functional and morphological evolution in gymnosperms: a portrait of implicated gene families. *Evol. Appl.* 13, 210–227.
- 102. Zhao T., Zwaenepoel A., Xue J.Y., Kao S.M., Li Z., Schranz M.E., Van de Peer Y. (2021) Whole-genome microsynteny-based phylogeny of angiosperms. *Nat. Commun.* 12, 3498.
- 103. He Z., Ji R., Havlickova L., Wang L., Li Y., Lee H.T., Song J., Koh C., Yang J., Zhang M., Parkin I.A.P., Wang X., Edwards D., King G.J., Zou J., Liu K., Snowdon R.J., Banga S.S., Machackova I., Bancroft I. (2021) Genome structural evolution in *Brassica crops. Nat. Plants.* 7, 757–765.
- 104. Perumal S., Koh C.S., Jin L., Buchwaldt M., Higgins E.E., Zheng C., Sankoff D., Robinson S.J., Kagale S., Navabi Z.K., Tang L., Horner K.N., He Z., Bancroft I., Chalhoub B., Sharpe A.G., Parkin I.A.P. (2020) A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral *Brassica* genome. *Nat. Plants.* 6, 929–941.
- Lewin H.A., Robinson G.E., Kress W.J., Baker W.J., Coddington J., Crandall K.A., Durbin R., Edwards S.V., Forest F., Gilbert M.T.P., Goldstein M.M., Grigoriev I.V., Hackett K.J., Haussler D., Jarvis E.D., Johnson W.E., Patrinos A., Richards S., Castilla-Rubio J.C., van Sluys M.A., Soltis P.S., Xu X., Yang H., Zhang G. (2018) Earth BioGenome Project: sequencing life for the future of life. *Proc. Natl. Acad. Sci. USA.* 115, 4325–4333.
- 106. Exposito-Alonso M., Drost H.G., Burbano H.A., Weigel D. (2020) The Earth BioGenome project: opportunities and challenges for plant genomics and conservation. *Plant J. : Cell Mol. Biol.* **102**, 222–229.
- 107. Tettelin H., Masignani V., Cieslewicz M.J., Donati C., Medini D., Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S., Deboy R.T., Davidsen T.M., Mora M., Scarselli M., Margarit y Ros I., Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R., Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A., Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N., Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J., Smith S., Utterback T.R., White O., Rubens C.E.,

МОЛЕКУЛЯРНАЯ БИОЛОГИЯ том 56 № 4 2022

Grandi G., Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R., Fraser C.M. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci. USA.* **102**, 13950–13955.

- Golicz A.A., Bayer P.E., Bhalla P.L., Batley J., Edwards D. (2020) Pangenomics comes of age: from bacteria to plant and animal applications. *Trends Genet.*: *TIG.* 36, 132–145.
- 109. Ho S.S., Urban A.E., Mills R.E. (2020) Structural variation in the sequencing era. *Nat. Rev. Genet.* **21**, 171–189.
- 110. Li Y.H., Zhou G., Ma J., Jiang W., Jin L.G., Zhang Z., Guo Y., Zhang J., Sui Y., Zheng L., Zhang S.S., Zuo Q., Shi X.H., Li Y.F., Zhang W.K., Hu Y., Kong G., Hong H.L., Tan B., Song J., Liu Z.X., Wang Y., Ruan H., Yeung C.K., Liu J., Wang H., Zhang L.J., Guan R.X., Wang K.J., Li W.B., Chen S.Y., Chang R.Z., Jiang Z., Jackson S.A., Li R., Qiu L.J. (2014) *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052.
- 111. Anderson J.E., Kantar M.B., Kono T.Y., Fu F., Stec A.O., Song Q., Cregan P.B., Specht J.E., Diers B.W., Cannon S.B., McHale L.K., Stupar R.M. (2014) A roadmap for functional structural variants in the soybean genome. *G3.* **4**, 1307–1318.
- Tao Y., Zhao X., Mace E., Henry R., Jordan D. (2019) Exploring and exploiting pan-genomics for crop improvement. *Mol. Plant.* 12, 156–169.
- Swinnen G., Goossens A., Pauwels L. (2016) Lessons from domestication: targeting *cis*-regulatory elements for crop improvement. *Trends Plant Sci.* 21, 506–515.
- 114. Gao L., Gonda I., Sun H., Ma Q., Bao K., Tieman D.M., Burzynski-Chang E.A., Fish T.L., Stromberg K.A., Sacks G.L., Thannhauser T.W., Foolad M.R., Diez M.J., Blanca J., Canizares J., Xu Y., van der Knaap E., Huang S., Klee H.J., Giovannoni J.J., Fei Z. (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051.
- Rancati G., Moffat J., Typas A., Pavelka N. (2018) Emerging and evolving concepts in gene essentiality. *Nat. Rev. Genet.* 19, 34–49.
- Bayer P.E., Golicz A.A., Scheben A., Batley J., Edwards D. (2020) Plant pan-genomes are the new reference. *Nat. Plants.* 6, 914–920.
- 117. Zhou Y., Chebotarov D., Kudrna D., Llaca V., Lee S., Rajasekar S., Mohammed N., Al-Bader N., Sobel-Sorenson C., Parakkal P., Arbelaez L.J., Franco N., Alexandrov N., Hamilton N.R.S., Leung H., Mauleon R., Lorieux M., Zuccolo A., McNally K., Zhang J., Wing R.A. (2020) A platinum standard pan-genome resource that represents the population structure of Asian rice. *Sci. Data.* 7, 113.
- 118. Wang W., Mauleon R., Hu Z., Chebotarov D., Tai S., Wu Z., Li M., Zheng T., Fuentes R.R., Zhang F., Mansueto L., Copetti D., Sanciangco M., Palis K.C., Xu J., Sun C., Fu B., Zhang H., Gao Y., Zhao X., Shen F., Cui X., Yu H., Li Z., Chen M., Detras J., Zhou Y., Zhang X., Zhao Y., Kudrna D., Wang C., Li R., Jia B., Lu J., He X., Dong Z., Xu J., Li Y., Wang M.,

МОЛЕКУЛЯРНАЯ БИОЛОГИЯ том 56 № 4 2022

Shi J., Li J., Zhang D., Lee S., Hu W., Poliakov A., Dubchak I., Ulat V.J., Borja F.N., Mendoza J.R., Ali J., Li J., Gao Q., Niu Y., Yue Z., Naredo M.E.B., Talag J., Wang X., Li J., Fang X., Yin Y., Glaszmann J.C., Zhang J., Li J., Hamilton R.S., Wing R.A., Ruan J., Zhang G., Wei C., Alexandrov N., McNally K.L., Li Z., Leung H. (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*. **557**, 43–49.

- Zhao Q., Feng Q., Lu H., Li Y., Wang A., Tian Q., Zhan Q., Lu Y., Zhang L., Huang T., Wang Y., Fan D., Zhao Y., Wang Z., Zhou C., Chen J., Zhu C., Li W., Weng Q., Xu Q., Wang Z.X., Wei X., Han B., Huang X. (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284.
- 120. Walkowiak S., Gao L., Monat C., Haberer G., Kassa M.T., Brinton J., Ramirez-Gonzalez R.H., Kolodziej M.C., Delorean E., Thambugala D., Klymiuk V., Byrns B., Gundlach H., Bandi V., Siri J.N., Nilsen K., Aquino C., Himmelbach A., Copetti D., Ban T., Venturini L., Bevan M., Clavijo B., Koo D.H., Ens J., Wiebe K., N'Diave A., Fritz A.K., Gutwin C., Fiebig A., Fosker C., Fu B.X., Accinelli G.G., Gardner K.A., Fradgley N., Gutierrez-Gonzalez J., Halstead-Nussloch G., Hatakeyama M., Koh C.S., Deek J., Costamagna A.C., Fobert P., Heavens D., Kanamori H., Kawaura K., Kobayashi F., Krasileva K., Kuo T., McKenzie N., Murata K., Nabeka Y., Paape T., Padmarasu S., Percival-Alwyn L., Kagale S., Scholz U., Sese J., Juliana P., Singh R., Shimizu-Inatsugi R., Swarbreck D., Cockram J., Budak H., Tameshige T., Tanaka T., Tsuji H., Wright J., Wu J., Steuernagel B., Small I., Cloutier S., Keeble-Gagnere G., Muehlbauer G., Tibbets J., Nasuda S., Melonek J., Hucl P.J., Sharpe A.G., Clark M., Legg E., Bharti A., Langridge P., Hall A., Uauy C., Mascher M., Krattinger S.G., Handa H., Shimizu K.K., Distelfeld A., Chalmers K., Keller B., Mayer K.F.X., Poland J., Stein N., McCartney C.A., Spannagl M., Wicker T., Pozniak C.J. (2020) Multiple wheat genomes reveal global variation in modern breeding. Nature. 588, 277-283.
- 121. Montenegro J.D., Golicz A.A., Bayer P.E., Hurgobin B., Lee H., Chan C.K., Visendi P., Lai K., Dolezel J., Batley J., Edwards D. (2017) The pangenome of hexaploid bread wheat. *Plant J.: Cell Mol. Biol.* **90**, 1007–1013.
- 122. Jayakodi M., Padmarasu S., Haberer G., Bonthala V.S., Gundlach H., Monat C., Lux T., Kamal N., Lang D., Himmelbach A., Ens J., Zhang X.Q., Angessa T.T., Zhou G., Tan C., Hill C., Wang P., Schreiber M., Boston L.B., Plott C., Jenkins J., Guo Y., Fiebig A., Budak H., Xu D., Zhang J., Wang C., Grimwood J., Schmutz J., Guo G., Zhang G., Mochida K., Hirayama T., Sato K., Chalmers K.J., Langridge P., Waugh R., Pozniak C.J., Scholz U., Mayer K.F.X., Spannagl M., Li C., Mascher M., Stein N. (2020) The barley pangenome reveals the hidden legacy of mutation breeding. *Nature*. 588, 284–289.
- 123. Hufford M.B., Seetharam A.S., Woodhouse M.R., Chougule K.M., Ou S., Liu J., Ricci W.A., Guo T., Olson A., Qiu Y., Della Coletta R., Tittes S., Hudson A.I., Marand A.P., Wei S., Lu Z., Wang B., Tello-Ruiz M.K., Piri R.D., Wang N., Kim D.W., Zeng Y., O'Connor C.H.,

Li X., Gilbert A.M., Baggs E., Krasileva K.V., Portwood J.L., 2nd, Cannon E.K.S., Andorf C.M., Manchanda N., Snodgrass S.J., Hufnagel D.E., Jiang Q., Pedersen S., Syring M.L., Kudrna D.A., Llaca V., Fengler K., Schmitz R.J., Ross-Ibarra J., Yu J., Gent J.I., Hirsch C.N., Ware D., Dawe R.K. (2021) *De novo* assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science.* **373**, 655–662.

- 124. Song J.M., Guan Z., Hu J., Guo C., Yang Z., Wang S., Liu D., Wang B., Lu S., Zhou R., Xie W.Z., Cheng Y., Zhang Y., Liu K., Yang Q.Y., Chen L.L., Guo L. (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus. Nat. Plants.* 6, 34–45.
- 125. Liu Y., Du H., Li P., Shen Y., Peng H., Liu S., Zhou G.A., Zhang H., Liu Z., Shi M., Huang X., Li Y., Zhang M., Wang Z., Zhu B., Han B., Liang C., Tian Z. (2020) Pan-genome of wild and cultivated soybeans. *Cell.* 182, 162–176.
- 126. Gordon S.P., Contreras-Moreira B., Woods D.P., Des Marais D.L., Burgess D., Shu S., Stritt C., Roulin A.C., Schackwitz W., Tyler L., Martin J., Lipzen A., Dochy N., Phillips J., Barry K., Geuten K., Budak H., Juenger T.E., Amasino R., Caicedo A.L., Goodstein D., Davidson P., Mur L.A.J., Figueroa M., Freeling M., Catalan P., Vogel J.P. (2017) Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nat. Commun.* **8**, 2184.
- 127. Della Coletta R., Qiu Y., Ou S., Hufford M.B., Hirsch C.N. (2021) How the pan-genome is changing crop genomics and improvement. *Genome Biol.* **22**, 3.
- 128. Syme R.A., Tan K.C., Rybak K., Friesen T.L., Mc-Donald B.A., Oliver R.P., Hane J.K. (2018) Pan-parastagonospora comparative genome analysis-effector prediction and genome evolution. *Genome Biol. Evol.* 10, 2443–2457.
- 129. Plissonneau C., Hartmann F.E., Croll D. (2018) Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biol.* **16**, 5.
- Badet T., Oggenfuss U., Abraham L., McDonald B.A., Croll D. (2020) A 19-isolate reference-quality global pangenome for the fungal wheat pathogen *Zymoseptoria tritici. BMC Biol.* 18, 12.
- Gage J.L., Vaillancourt B., Hamilton J.P., Manrique-Carpintero N.C., Gustafson T.J., Barry K., Lipzen A., Tracy W.F., Mikel M.A., Kaeppler S.M., Buell C.R., de Leon N. (2019) Multiple maize reference genomes impact the identification of variants by genome-wide association study in a diverse inbred panel. *Plant Genome*. 12, 1–12.
- 132. Понажев В.П., Рожмина Т.А., Тихомирова В.Я. (2011) Инновационные разработки – льноводству: Селекция, семеноводство, возделывание, первичная обработка, экономика. Тверь: Твер. гос.ун-т.
- 133. Jung S., Lee T., Gasic K., Campbell B.T., Yu J., Humann J., Ru S., Edge-Garza D., Hough H., Main D. (2021) The Breeding Information Management System (BIMS): an online resource for crop breeding. *Database: J. Biol Databases Curation.* 2021, baab054.
- 134. Liu J., Fernie A.R., Yan J. (2021) Crop breeding from experience-based selection to precision design. *J. Plant Physiol.* 256, 153313.

- 135. Furbank R.T., Jimenez-Berni J.A., George-Jaeggli B., Potgieter A.B., Deery D.M. (2019) Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops. *New Phytologist.* 223, 1714–1727.
- Winter P., Kahl G. (1995) Molecular marker technologies for plant improvement. W. J. Microbiol. Biotechnol. 11, 438–448.
- 137. Andersen J.R., Lubberstedt T. (2003) Functional markers in plants. *Trends Plant Sci.* **8**, 554–560.
- 138. Varshney R.K., Graner A., Sorrells M.E. (2005) Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* **10**, 621–630.
- 139. Thudi M., Palakurthi R., Schnable J.C., Chitikineni A., Dreisigacker S., Mace E., Srivastava R.K., Satyavathi C.T., Odeny D., Tiwari V.K., Lam H.M., Hong Y.B., Singh V.K., Li G., Xu Y., Chen X., Kaila S., Nguyen H., Sivasankar S., Jackson S.A., Close T.J., Shubo W., Varshney R.K. (2021) Genomic resources in plant breeding for sustainable agriculture. *J. Plant Physiol.* **257**, 153351.
- Salgotra R.K., Stewart C.N., Jr. (2020) Functional markers for precision plant breeding. *Int. J. Mol. Sci.* 21, 4792.
- 141. Tibbs Cortes L., Zhang Z., Yu J. (2021) Status and prospects of genome-wide association studies in plants. *Plant Genome*. **14**, e20077.
- 142. Desta Z.A., Ortiz R. (2014) Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* **19**, 592–601.
- 143. Crossa J., Perez-Rodriguez P., Cuevas J., Montesinos-Lopez O., Jarquin D., de Los Campos G., Burgueno J., Gonzalez-Camacho J.M., Perez-Elizalde S., Beyene Y., Dreisigacker S., Singh R., Zhang X., Gowda M., Roorkiwal M., Rutkoski J., Varshney R.K. (2017) Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975.
- 144. Розанова И.В., Хлесткина Е.К. (2020) NGS-секвенирование в селекционно-генетических исследованиях ячменя. Вавиловский журн. генет. селекции. 24, 348–355.
- 145. Gupta P.K., Kulwal P.L., Jaiswal V. (2019) Association mapping in plants in the post-GWAS genomics era. *Adv. Genetics.* **104**, 75–154.
- 146. You F.M., Cloutier S. (2020) Mapping quantitative trait loci onto chromosome-scale pseudomolecules in flax. *Meth. Protocols.* **3**, 28.
- 147. Lai J., Li R., Xu X., Jin W., Xu M., Zhao H., Xiang Z., Song W., Ying K., Zhang M., Jiao Y., Ni P., Zhang J., Li D., Guo X., Ye K., Jian M., Wang B., Zheng H., Liang H., Zhang X., Wang S., Chen S., Li J., Fu Y., Springer N.M., Yang H., Wang J., Dai J., Schnable P.S., Wang J. (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat. Genetics.* 42, 1027–1030.
- 148. Darracq A., Vitte C., Nicolas S., Duarte J., Pichon J.P., Mary-Huard T., Chevalier C., Berard A., Le Paslier M.C., Rogowsky P., Charcosset A., Joets J. (2018) Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. *BMC Genomics*. **19**, 119.

### PLANT GENOME SEQUENCING: MODERN TECHNOLOGIES AND NOVEL OPPORTUNITIES FOR BREEDING

#### A. A. Dmitriev<sup>1</sup>, E. N. Pushkova<sup>1</sup>, and N. V. Melnikova<sup>1, \*</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia \*e-mail: mnv-4529264@yandex.ru

The investigation of plant genomes is of great importance for basic research and practical breeding. In 1977, F. Sanger proposed a DNA sequencing method, which allowed the complete sequences of a number of genomes to be determined. Then high-throughput and cost-effective next-generation/second-generation sequencing methods, producing up to billions of short reads, made it possible to sequence genomes of a significant number of species and provided a breakthrough in plant genetic studies. Finally, third-generation sequencing technologies allowed the determination of single-molecule sequences up to a million nucleotides in length, which is key for high-quality genome assemblies. An important task is to obtain a pan-genome, which includes an entire set of nucleotide sequences presented in various genotypes of the same species. The sequencing the formation of significant features, and develop molecular markers of economically valuable traits and became the basis for the development of marker-assisted and genomic selection. This review provides information on the latest advances in sequencing technologies and assembly of plant genomes, as well as the opportunities that they open up for basic and applied works.

Keywords: plants, genome, pan-genome, sequencing, breeding