

Прогресс в секвенировании геномов растений – направления исследований

М.К. Брагина¹, Д.А. Афонников^{1, 2}, Е.А. Салина¹

¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ e-mail: koltunova@bionet.nsc.ru

Геномные технологии претерпели значительные изменения с момента публикации первой последовательности генома растения *Arabidopsis thaliana*. Исследователи приняли на вооружение новые алгоритмы и технологии секвенирования и биоинформационные подходы для получения геномной последовательности, транскриптома и экзома для модельных и культурных видов растений, что позволило сделать глубокие выводы о биологии растений. В результате снижения затрат на секвенирование благодаря улучшению методов сборки и анализа геномов, количество и качество секвенированных геномов растений постоянно растет. В течение последних двадцати лет опубликовано более 300 геномных последовательностей растений. Хотя многие из опубликованных геномов считаются неполными, они, тем не менее, оказались ценным инструментом для идентификации генов-кандидатов, участвующих в формировании хозяйственно ценных признаков растений, для проведения работ по маркер-ориентированной и геномной селекции и сравнительного анализа геномов растений с целью установления основных закономерностей происхождения различных видов растений. В связи с тем, что высокого уровня покрытия и разрешения полногеномного секвенирования не хватает для обнаружения всех изменений в сложных образцах, стало развиваться целевое (таргетное) секвенирование, которое заключается в выделении и секвенировании определенной области генома. Основным преимуществом целевого секвенирования являются его высокая мощность обнаружения (способность идентифицировать новые варианты) и более высокое разрешение. Кроме того, активно развивается экзомное секвенирование (метод секвенирования только белок-кодирующих участков генов), позволяющее секвенировать участки генома, которые обогащены функциональными вариантами и демонстрируют низкий уровень содержания повторяющихся областей. В настоящем обзоре проведен анализ развития работ по секвенированию и построению «референсных» геномов растений. Сравняются методы целевого секвенирования, базирующиеся на использовании референсной последовательности ДНК. Ключевые слова: растения; подходы к секвенированию; геном; экзомное секвенирование; целевое секвенирование.

Для цитирования: Брагина М.К., Афонников Д.А., Салина Е.А. Прогресс в секвенировании геномов растений – направления исследований. Вавиловский журнал генетики и селекции. 2019;23(1):38-48. DOI 10.18699/VJ19.459

Progress in plant genome sequencing: research directions

M.K. Bragina¹, D.A. Afonnikov^{1, 2}, E.A. Salina¹

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

✉ e-mail: koltunova@bionet.nsc.ru

Since the first plant genome of *Arabidopsis thaliana* has been sequenced and published, genome sequencing technologies have undergone significant changes. New algorithms, sequencing technologies and bioinformatic approaches were adopted to obtain genome, transcriptome and exome sequences for model and crop species, which have permitted deep inferences into plant biology. As a result of an improved genome assembly and analysis methods, genome sequencing costs plummeted and the number of high-quality plant genome sequences is constantly growing. Consequently, more than 300 plant genome sequences have been published over the past twenty years. Although many of the published genomes are considered incomplete, they proved to be a valuable tool for identifying genes involved in the formation of economically valuable plant traits, for marker-assisted and genomic selection and for comparative analysis of plant genomes in order to determine the basic patterns of origin of various plant species. Since a high coverage and resolution of a genome sequence is not enough to detect all changes in complex samples, targeted sequencing, which consists in the isolation and sequencing of a specific region of the genome, has begun to develop. Targeted sequencing has a higher detection power (the ability to identify new differences/variants) and resolution (up to one basis). In addition, exome sequencing (the method of sequencing only protein-coding genes regions) is actively developed, which allows for the sequencing of non-expressed alleles and genes that cannot be found with RNA-seq. In this review, an analysis of sequencing technologies development and

the construction of “reference” genomes of plants is performed. A comparison of the methods of targeted sequencing based on the use of the reference DNA sequence is accomplished.

Key words: plants; sequencing approaches; genome; exome sequencing; targeted sequencing.

For citation: Bragina M.K., Afonnikov D.A., Salina E.A. Progress in plant genome sequencing: research directions. *Vavilovskii Zhurnal Genetiki i Selektzii*=Vavilov Journal of Genetics and Breeding. 2019;23(1):38-48. DOI 10.18699/VJ19.459 (in Russian)

Введение

В настоящее время эффективные генетические исследования растений уже невозможно представить без этапа расшифровки генома (Varabaschi et al., 2016). Геномные последовательности являются фундаментальной основой для решения таких важных задач, как идентификация генов и генных сетей, обеспечивающих развитие, адаптацию растений к условиям стресса. Определение набора генов – отправная точка для детальной характеристики функций генов, биохимических и регуляторных путей или количественных локусов признаков. Последовательности генов предоставляют ценную информацию о белковых комплексах, регуляторных взаимодействиях и метаболических процессах, которые определяют физиологические и биохимические свойства клетки, органа или организма (Bassel et al., 2012).

Новые технологии селекции растений опираются на информацию о маркерах и ассоциированных с ними количественных признаках, при этом наличие данных о геноме позволяет использовать самый широкий спектр маркеров, включая как повторы (Bai et al., 2017), так и SNP (Strain et al., 2018; Li et al., 2018). Структурно-функциональная разметка геномов помогает существенно облегчить определение молекулярных механизмов, обеспечивающих формирование целевых признаков растений. Информация о полной последовательности генома является ключевой и в случае подбора участков для геномного редактирования и поиска сайтов, имеющих наиболее высокую степень эффективности разрезания ДНК и специфичности по отношению к остальной геномной ДНК (Scheben et al., 2017; Zhu et al., 2017). Еще одной задачей, решение которой зависит от знания геномных последовательностей, является сравнительный анализ геномов растений с целью установить основные закономерности происхождения различных видов растений и сельскохозяйственных культур (Avni et al., 2017).

Со времени определения первого генома растения, *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000), благодаря развитию технологий секвенирования и улучшению алгоритмов сборки и анализа геномов количество и качество секвенированных геномов растений постоянно растет. В идеале геномная сборка представляет собой набор последовательностей молекул ДНК всех его хромосом с суммарной длиной, равной размеру гаплоидного генома. Такие модели геномов служат основой для решения огромного числа задач, как связанных с поиском генов, идентификацией маркеров, так и опирающихся на знание детальной структуры генома – сравнительной геномики, идентификации синтенных групп хромосом. Нуклеотидные последовательности, собранные на уровне хромосом, включают, помимо кодирующих, участки, содержащие повторенные последовательности и мобильные

элементы. Информация о них позволяет охарактеризовать различные классы повторяющихся элементов, идентифицировать крупномасштабную коллинеарность генов у родственных видов и восстановить организацию и эволюцию мобильных элементов (Bennetzen, Wang, 2014).

Модели геномов на уровне хромосом содержат полный набор генов, необходимый для проверки того, может ли дупликация или потеря биохимических или сигнальных путей у конкретных видов растений объяснить структурные и физиологические адаптации, необходимые для выживания в экстремальных условиях. Полные хромосомные последовательности служат основой («референсным» геномом) для изучения других геномов того же вида с использованием гораздо меньших ресурсов, чем было затрачено на получение референсной последовательности. Однако такие модели геномов в настоящее время доступны лишь для небольшого числа растений, таких как *A. thaliana* (The Arabidopsis Genome Initiative, 2000), *Brachypodium distachyon* (The International Brachypodium Initiative, 2010), рис (Goff et al., 2002; Yu et al., 2002), ячмень (The International Barley Genome Sequencing Consortium, 2012), кукуруза (Schnable et al., 2009), картофель (The Potato Genome Sequencing Consortium, 2011), мох *Physcomitrella patens* (Rensing et al., 2008), пшеница *T. dicoccoides* (Anvi et al., 2017; Akpinar et al., 2018), *T. aestivum* (International Wheat Genome Sequencing Consortium, 2018), и некоторых других (<http://plants.ensembl.org>).

Несмотря на то что последовательности большинства из опубликованных геномов представлены пока лишь фрагментами ДНК, а информация об их порядке в хромосомах отсутствует, даже такая модель геномной последовательности содержит в себе достаточно информации для идентификации участков, кодирующих гены, подбора маркеров для задач геномной селекции, сравнительного анализа генов, кодирующих белки с гомологичными последовательностями из других видов.

Стратегии секвенирования геномов

Стратегии секвенирования растительных геномов развивались в зависимости от методов секвенирования и методов фракционирования геномной ДНК. К основным методам секвенирования в порядке их разработки относятся секвенирование по Сэнгеру, высокопроизводительное пиросеквенирование, секвенирование на молекулярных кластерах с использованием флуоресцентно меченых нуклеотидов, циклическое лигазное секвенирование, полупроводниковое секвенирование, одномолекулярное секвенирование, секвенирование через нанопоры. Каждый из этих методов имел или имеет свои преимущества и ограничения, заключающиеся в длине, качестве и стоимости прочтений, которые неоднократно были изложены в различных обзорах (Krasnov et al., 2014; Mardis, 2017;

Shendure et al., 2017). К настоящему времени часть подходов (секвенирование по Сэнгеру, пиросеквенирование) уже утратила свою актуальность в связи с появлением новых, более совершенных и дешевых технологий. К методу фракционирования геномной ДНК можно отнести предварительное клонирование крупных геномов в ВАС-векторах (bacterial artificial chromosome) с последующим секвенированием ВАС-клонов, в том числе собранных в более протяженные контиги методами физического картирования, – так называемое ВАС-by-ВАС секвенирование (Bolger et al., 2014). Стратегией, заключающейся в отказе от предварительного деления геномной ДНК на фракции (отдельные хромосомы или ВАС-клоны) и секвенировании суммарной геномной ДНК, является метод дробовика (whole genome shotgun – WGS). В данном случае с помощью сайт-неспецифичных нуклеаз проводится случайная фрагментация геномной ДНК с созданием библиотек для секвенирования. Стоит отметить, что секвенирование первых растений давало частичную информацию о геномах и не позволяло получить полную референсную последовательность.

Важным этапом не только для исследования растений, но и для геномного секвенирования, стало секвенирование генома модельного растения *A. thaliana* в 2000 г. (The Arabidopsis Genome Initiative, 2000). Это был один из первых полученных геномов многоклеточных организмов. Выбранный подход ВАС-by-ВАС основывался на секвенировании по Сэнгеру, сборке и выравнивании последовательностей ВАС-клонов ((100–150) · 10³ п. н.) в соответствии с физической картой так, чтобы идентичные последовательности перекрывались, а смежные (контиги) собирались в геномную последовательность высокого качества (Feuillet et al., 2011). Однако трудоемкость этого подхода ограничивала его применение для других геномов, и только два года спустя с его помощью был секвенирован геном риса (Goff et al., 2002; Yu et al., 2002), размер которого примерно в три раза превышает геном *A. thaliana*. Кроме того, в качестве альтернативного подхода к секвенированию больших и полиплоидных геномов, таких как гексаплоидная пшеница, было разработано выделение и секвенирование хромосом и плеч (Doležel et al., 2007; Paux et al., 2008; Hernandez et al., 2012; Mayer et al., 2014). До недавнего времени секвенирование минимального перекрывающегося пути выполнялось с использованием автоматизированного секвенирования по Сэнгеру (Groenen et al., 2012; Howe et al., 2013), хотя платформы секвенирования следующего поколения (NGS) были применены для крупномасштабного секвенирования ВАС-клонов (Zhang et al., 2012; Choulet et al., 2014; Li et al., 2015).

Следующим секвенированным культурным растением стал тополь (Tuskan et al., 2006). С помощью секвенирования методом дробовика (WGS) его ДНК была амплифицирована, случайным образом фрагментирована сайт-неспецифичными нуклеазами для получения перекрывающихся фрагментов, которые затем были секвенированы и собраны. Хотя эта стратегия значительно снижает временные затраты, как правило, в результате получается более фрагментированная последовательность генома. Полногеномное секвенирование методом дробовика так-

же было успешно применено для других растений, в том числе для винограда *Vitis vinifera* (Jaillon et al., 2007), папайи *Carica papaya* (Ming et al., 2008), огурца *Cucumis sativus* (Huang et al., 2009), яблони *Malus × domestica* Borkh. (Velasco et al., 2010), *B. distachyon* (The International Brachypodium Initiative, 2010), сои *Glycine max* (Schmutz et al., 2010) и картофеля *Solanum tuberosum* (The Potato Genome Sequencing Consortium, 2011).

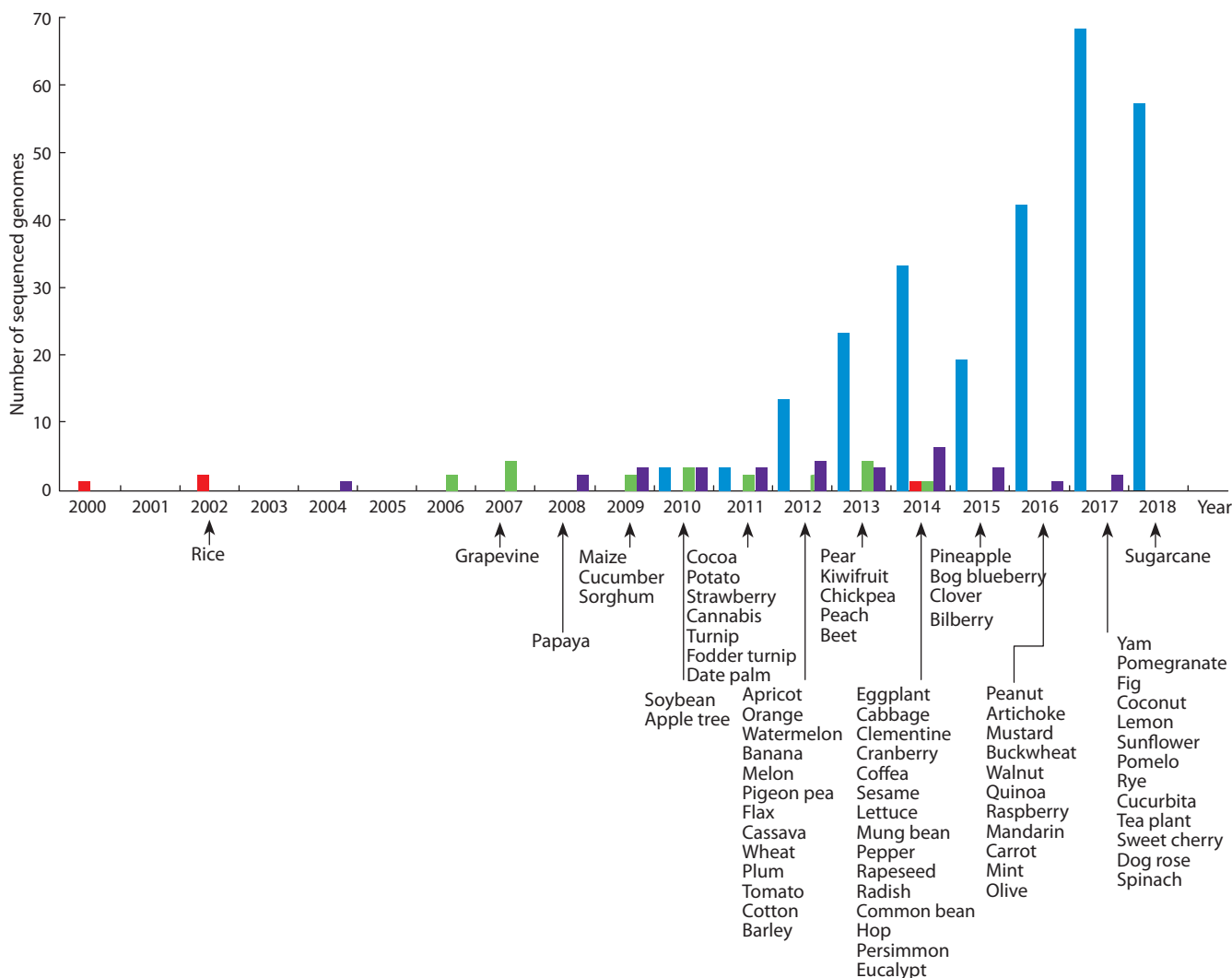
Известно, что большинство геномов растений характеризуются повышенным содержанием повторяющейся ДНК и наличием дупликаций генома из-за событий полиплоидизации. Важно отметить, что все геномные проекты, основанные только на секвенировании по Сэнгеру, были реализованы для растений, геномы которых значительно меньше 5.8 · 10⁹ п. н. Хотя подходы ВАС-by-ВАС могут снизить сложность генома более чем в 10 тысяч раз, сборка больших геномов оставалась трудоемкой и дорогостоящей. Секвенирование методом дробовика сокращает необходимое время и усилия, однако получение последовательности генома было затратным из-за высокой стоимости секвенирования по Сэнгеру и требовало много времени. Кроме того, оставались несеквенированные области, что обусловлено низким покрытием при секвенировании по Сэнгеру (×4–7) и техническими проблемами, связанными со вторичной структурой ДНК и гомополимерами.

Стратегия секвенирования методом дробовика была основана на секвенировании по Сэнгеру и применяется до сих пор, однако начиная с 2005 г. все большие обороты стало набирать так называемое секвенирование следующего поколения (next-generation sequencing – NGS). Его применение улучшило соотношение между получаемыми данными и стоимостью секвенирования генома. Технология NGS охватывает широкий круг подходов, в которых идет параллельное секвенирование последовательностей нескольких фрагментов ДНК с получением значительно большего числа секвенированных последовательностей, но, как правило, более короткой длины (25–500 п. н.) и более низкого качества, чем в случае секвенирования по Сэнгеру. Эти подходы реализуются на секвенаторах нового поколения производства коммерческих компаний Illumina, Thermo Fisher Scientific, Pacific Biosciences и Oxford Nanopore Technologies. Часть разработанных платформ уже ушли с рынка (например, GS FLX, 454/Roche, HeliScope/Helicos Bioscience).

NGS применяют как для ресеквенирования (анализа геномов организмов с доступным референсным геномом), так и для секвенирования *de novo*. Для упрощения сборки в случае больших по размеру и сложно устроенных геномов используют комбинированный подход, в котором сочетают платформы, генерирующие как короткие, так и длинные последовательности (The Potato Genome Sequencing Consortium, 2011; Brenchley et al., 2012; The Tomato Genome Consortium, 2012), или сочетание нескольких стратегий секвенирования (см. рисунок).

Геномные проекты

В результате резкого снижения затрат на секвенирование генома в течение последних двадцати лет было опубликовано 322 геномных последовательности растений



Transition from the formerly applied Sanger method to genome sequencing only by NGS (blue)
Purple: combined sequencing by BAC and WGS; green: shotgun sequencing (WGS); red: Sanger sequencing of BAC clones. Below: examples of sequenced crops.

(по состоянию на июль 2018 г., www.plabipd.de), из них геномов двудольных растений – 207, однодольных – 68, водорослей – 35, непокрытосеменных растений – 12 (Приложение 1)¹. До недавнего времени выполнением этих проектов занимались крупные консорциумы по секвенированию генома, объединяя опыт во многих областях. Благодаря эволюции секвенирования следующего поколения постепенно стало возможным проводить целые проекты сборки генома растений силами отдельных лабораторий или небольших консорциумов (Pucker et al., 2016; Jiao et al., 2017). Дальнейшее усовершенствование технологий с длинным прочтением позволит охватывать повторяющиеся районы, что ранее было серьезным препятствием для завершения секвенирования генома (Jiao et al., 2017).

В 2008 г. стартовал проект «1001 Genomes Project» для получения точной геномной последовательности *A. thaliana* на основе секвенирования 1001 образца (<http://1001genomes.org>). Первый этап проекта недавно был завершен

¹ Приложения 1 и 2 см. по адресу: <http://www.bionet.nsc.ru/vogis/download/pict-2019-23/appx2.pdf>

публикацией подробного анализа геномов 1135 инбредных линий, который дал информацию об истории видов и распределении генетического разнообразия (The 1001 Genomes Consortium, 2016).

В 2012 г. была создана международная инициатива для секвенирования транскриптома 1000 различных видов растений – «1KP» (<https://sites.google.com/a/ualberta.ca/onekp/>). Секвенирование проводится на платформе Illumina (GA2 или HiSeq). Растения отбирали таким образом, чтобы предоставить образцы для всех основных линий зеленых растений, включая хвойные деревья, папоротники, мхи, зеленые водоросли и цветковые растения. Следует отметить, что большинство этих видов никогда не подвергались крупномасштабному секвенированию. Например, были секвенированы представители почти всех 415 известных семейств покрытосеменных, а примерно пятую часть секвенированных видов представляют водоросли. Полученные данные являются транскриптомными сборками, предполагаемыми кодирующими последовательностями, ортогруппами, а также филогенетическими деревьями (Matasci et al., 2014).

В 2014 г. опубликована работа о ресеквенировании 3000 образцов риса из 89 стран (<http://iric.igri.org/resources/3000-genomes-project>). Все геномы имели среднюю глубину секвенирования $\times 14$, со средними коэффициентами покрытия референсного генома и картирования 94.0 и 92.5 % соответственно. После выравнивания полученных данных с референсным сортом *O. japonica* Nipponbare было обнаружено приблизительно 18.9 миллионов SNP. Филогенетические анализы, основанные на данных SNP, подтвердили дифференциацию генофонда *O. sativa* на пять сортовых групп – *indica*, *aus/boro*, *basmati/sadri*, тропическую *japonica* и умеренную *japonica* (The 3,000 rice genomes project, 2014).

Общая цель Проекта генома паразитических растений (PPGP) (Westwood et al., 2012) заключается в проведении сравнительного функционального геномного анализа паразитических растений для выявления изменений генома, которые привели к установлению паразитического образа жизни. В этом проекте были секвенированы *Lindenbergia philippensis*, *Orobanchaeae* и *Mimulus* – транскрипты паразитических родов от *Orobanchaceae* (*Triphysaria*, *Striga* и *Orobanche*) и двух тесно связанных непаразитов. В настоящее время они секвенированы в независимом проекте (<http://www.mimulusevolution.org>). Эти виды растений эволюционно связаны между собой, но охватывают диапазон паразитарных способностей от свободного существования до полностью гетеротрофного паразита.

В июле 2017 г. китайский гигант секвенирования BGI (Beijing Genomics Institute) и Национальный банк Китая (CNGB) объявили о планах по секвенированию не менее 10000 геномов, представляющих каждую основную ветвь (группу, кладу) растений и эукариотических микробов. План 10КР станет ключевой частью проекта по получению биогенома Земли (Earth BioGenome Project – EBP), амбициозной и все еще развивающейся схемы, позволяющей получить данные о геномных последовательностях 1.5 миллионов эукариотических видов, включая детальные последовательности одного члена из каждого из 9000 эукариотических семейств. План 10КР основывается на предыдущем проекте 1КР; предполагается, что цели исследования будут включать роль дубликации генома, корреляцию между геномными и морфологическими изменениями и изменения темпов эволюции с течением времени.

Проблемы полногеномного секвенирования

Наиболее широкое применение новых подходов к секвенированию растений заключается в полногеномном секвенировании для получения референсной последовательности и генетической структуры геномов. Большой размер генома, дубликация и повторяющийся контент являются факторами, которые затрудняют получение полноценной информации о структуре генома растений и построении «золотого стандарта» референсной последовательности. Эти проблемы активно преодолеваются как разработкой новых биоинформатических ресурсов и использованием при секвенировании различных платформ, так и привлечением данных РНК-Seq и по экзомному секвенированию, а также привлечением данных по физическому картирова-

нию хромосом (The International Wheat Genome Sequencing Consortium, 2018).

В 2013 г. была опубликована референсная последовательность генома сахарной свеклы (*Beta vulgaris*). В этом исследовании использовалось сочетание 454, Illumina и Sanger платформ для секвенирования. В общей сложности был идентифицирован 27421 белок-кодирующий ген, что подтверждается данными РНК-Seq. На основе внутривидового геномного анализа пяти различных видов сахарной свеклы идентифицировано 7 миллионов геномных вариантов, а также большие консервативные области. Полученный геном сахарной свеклы позволяет обнаружить агрономически важные признаки, которые могут повысить качество и продуктивность растения. Последовательности геномов внесли также вклад в сравнительные исследования с видами Caryophyllales и другими цветковыми растениями (Dohm et al., 2014).

В недавней работе (Nystedt et al., 2013) проведено *de novo* секвенирование норвежской ели (*Picea abies*) методом дробовика с использованием технологии Illumina. Иерархическая стратегия сборки генома была разработана для объединения гаплоидных и диплоидных геномных данных и данных РНК-Seq. Размер генома *P. abies* оценивается в $19.6 \cdot 10^9$ п. н. Однако только 28354 последовательности, кодирующие белок с высокой степенью достоверности, были предсказаны по данным EST и транскриптомам. Кроме того, предложена модель эволюции генома хвойных пород, которая предполагает, что удаление мобильных элементов у них шло менее активно, чем у большинства других видов растений (Bennetzen et al., 2005), причем вставки мобильных элементов в гены приводят к появлению больших интронов и псевдогенов. Дополнительное секвенирование генома хвойных пород позволит провести сравнительный анализ и предоставит дополнительные ресурсы для понимания эволюции важных признаков у семенных растений.

В 2014 г. был собран референсный геном *Eucalyptus grandis* (Myburg et al., 2014), для чего проведено полногеномное секвенирование по Сэнгеру и парное секвенирование ВАС-концов. Это первый референсный геном, опубликованный для эвдикотов миртоцветных, обеспечивающий ресурс для получения информации о генетической природе больших древесных многолетников.

С использованием подхода Illumina WGS были секвенированы три инбредных сорта *Nicotiana tabacum* (Zhang et al., 2011). Размеры полученных геномов были оценены в $4.41 \cdot 10^9$ п. н. для *N. tabacum* TN90, $4.60 \cdot 10^9$ п. н. для *N. tabacum* K326 и $4.57 \cdot 10^9$ п. н. для *N. tabacum* BX (с покрытием $\times 49$, $\times 38$ и $\times 29$ соответственно). Авторы предполагали, что данный проект внесет значительный вклад в функциональные геномные исследования модельного организма *N. tabacum* (Sierra et al., 2014).

Помимо полногеномного секвенирования, проводятся исследования по ресеквенированию, которое используется в основном для профилирования транскриптому и обнаружения SNP для разработки маркеров (Deschamps et al., 2012). Таким образом был получен высококачественный референсный геном картофеля и идентифицировано 3.67 миллиона SNP при сравнении гомозиготной двойной гаплоидной линии с гетерозиготной диплоидной (Xu et

al., 2011). Кроме того, ресеквенированы и сопоставлены друг с другом несколько образцов арбуза. В общей сложности идентифицировано 6 784 860 SNP, представляющих генетическое разнообразие видов сельскохозяйственных культур (Guo et al., 2013).

В 2016 г. была создана компания NRGene, разрабатывающая новейшее программное обеспечение и алгоритмы для секвенирования сложных больших геномов сельскохозяйственных культур, животных и водных организмов. NRGene позволила секвенировать *de novo* $10.5 \cdot 10^9$ п. н. (98.4 %) геном тетраплоидной пшеницы (Avni et al., 2017), $4.09 \cdot 10^9$ п. н. (98.6 %) геном диплоидной пшеницы (Luo et al., 2017), $2.18 \cdot 10^9$ п. н. (98.2 %) референсный геном кукурузы (Springer et al., 2018) и $0.82 \cdot 10^9$ п. н. (98.4 %) геном томата (Kol, 2016) с использованием платформы Illumina и программного обеспечения DeNovoMAGIC. Также была секвенирована мягкая пшеница, и данные по сборке будут объединены с данными физического картирования для получения высококачественной последовательности каждой хромосомы с точно локализованными генами, регуляторными элементами и маркерами (<https://www.wheatgenome.org>).

Целевое секвенирование

Переход от секвенирования по Сэнгеру к NGS привел к снижению затрат и увеличению числа секвенированных геномов. Однако полногеномное секвенирование не подходит для большинства исследований. Высокого уровня покрытия и разрешения полногеномного секвенирования не хватает для обнаружения всех изменений в сложных образцах. В связи с этим стало развиваться и продолжает набирать обороты целевое (таргетное) секвенирование, которое заключается в выделении и секвенировании определенной области генома или подмножества генов.

Для растений, которые обладают большим или полиплоидным геномом, существует стратегия уменьшения сложности генома с использованием стратегий целевого обогащения. Целевое секвенирование включает в себя три этапа: отбор интересующей фракции ДНК (обогащение образцов), секвенирование отобранного материала и анализ полученных результатов. Целевое обогащение заключается в выделении специфических локусов генома (например, генов, молекулярных маркеров, больших геномных областей, геномов органелл) в сочетании с секвенированием второго поколения. Обогащение может проводиться с помощью ПЦР, метода молекулярной инверсии, гибридационного обогащения и его улучшенной версии – обогащения в растворе (табл. 1, Приложение 2).

В настоящее время для растений используют гибридационные методы обогащения – на твердой основе (т. е. покровное стекло, микрочип) или в растворе. Наиболее распространенные и надежные в исследованиях растений методы предоставлены компаниями Agilent Technologies (SureSelect), Roche NimbleGen (SeqCap EZ), MYcroarray (MYbaits) и Ion Torrent (TargetSeq) (Terracciano et al., 2016). Плоидность, размер генома и особенности ДНК (GC-состав) изучаемых видов вместе с особенностями зондов (длина зонда, температура гибридизации) могут повлиять на эффективность обогащения. Было показано, что уровень обогащения значительно снижается в про-

торной области, 5'UTR области и в первом экзоне генов из-за высокого содержания GC в этих областях (Dohm et al., 2008). Высокое или низкое содержание GC снижает эффективность ПЦР амплификации и гибридизации зонда (Aird et al., 2011). Эффективность обогащения зависит также от выбранного протокола и от используемой технологии секвенирования (табл. 2).

За последние несколько лет была доказана эффективность целевого обогащения для исследования нуклеотидного разнообразия полиплоидных видов с большим, повторяющимся и гетерозиготным геномом и диплоидных видов с геномной гетерогенностью внутри группы (см. Приложение 2). Так, например, был ресеквенирован участок экзона $3.5 \cdot 10^6$ п. н. аллотетраплоидной пшеницы для характеристики SNP, вариации числа копий генов и изучения дивергенции гомологичных последовательностей в кодирующих участках (Saintenac et al., 2011). В 2012 г. разработан протокол ресеквенирования и обогащения для изучения $56.5 \cdot 10^6$ п. н. геномной ДНК восьми сортов мягкой пшеницы, в результате чего идентифицировано около 500 000 новых SNP (Winfield et al., 2012).

Целевое секвенирование позволило уменьшить сложность генома ячменя более чем в 50 раз, для чего было проведено обогащение кодирующих областей, которые включали в себя предсказанные гены из геномной сборки сорта Mogex, находящуюся в открытом доступе полноразмерную кДНК и *de novo* собранную консенсусную последовательность RNA-Seq (Mascher et al., 2013).

Разработаны методики анализа целевых геномных областей, связанных с агрономически важными признаками, для выявления разнообразия ДНК-последовательностей кукурузы (Muraya et al., 2015), рапса (Clarke et al., 2013; Schiessl et al., 2014), хлопка (Salmon et al., 2012) и маниоки (Pootakham et al., 2014). Кроме того, целевое обогащение последовательностей и повторное секвенирование проведены для нескольких видов деревьев, а именно ладанной сосны, черного тополя и эвкалипта, с целью выявления полиморфизмов для построения генетической карты (Neves et al., 2013, 2014), генотипирования (Zhou et al., 2012), а также разработки маркеров, ассоциированных с ксилогенезом (Dasgupta et al., 2015).

Картирование с помощью секвенирования, которое сочетает в себе генетическое картирование с целевым ресеквенированием, было использовано для определения полиморфизмов потенциальных генов ячменя (Pankin et al., 2014), земляники (Tennessen et al., 2013) и *T. monocosmum* (Gardiner et al., 2014).

Даже геномы хлоропластов были подвергнуты целевому обогащению и массовому параллельному секвенированию (Stull et al., 2013). Был сконструирован набор РНК-зондов на основе полных последовательностей ДНК 22 ранее секвенированных хлоропластов эвдикотов. С помощью этого набора зондов проведено обогащение 24 покрытосеменных растений (22 эвдикота, 2 однодольных) с их последующим секвенированием для получения полных геномов пластид с исключительно высоким уровнем покрытия (717 в среднем).

В 2013 г. вышла работа по идентификации вариантов геномных последовательностей у 84 сортов картофеля (Uitdewilligen et al., 2013). И для томата, и для картофеля

Table 1. The most important features of the commercially available target enrichment kits (according to Terracciano et al., 2016)

Parameter	Enrichment by hybridization					
	on a solid support (microarray)		in-solution			
	NimbleGen Sequence Capture Array	Agilent Microarray	NimbleGen SeqCap EZ	Agilent SureSelect	MYcroarray MYbait	IonTorrent Target Seq
Bait type	DNA	DNA	DNA	RNA	RNA	DNA
Bait length, bp	60	60	55–105	114–126	80–120	50–120
Target size	До 30 Mb	?	Up to 200 Mb	1 kb to 24 Mb	Up to 200000 baits	100 kb to 10 Mb

Table 2. Comparison of target-enrichment methods

Parameter	PCR	MIP	Enrichment by hybridization	
			on a solid support (microarray)	in-solution
Cost	High	High (<10 samples) Low (>100 samples)	Medium	Medium (<10 samples) Low (>10 samples)
Ease of use	Low	High	Medium	High
DNA mass	~8 µg for 1 Mb	200 ngg	10–15 µg for up to 30 Mb	3 µg for up to 30 Mb
Sensitivity ¹	>99.5 %	>98 %, with stringent design constraints	98.6 % of CTR ⁵	>99.5 % of CTR ⁵
Specificity ²	93 % for HapMap DNA samples, 72% for whole genome–amplified samples	>98 %	Up to 70 % mapping to CTR for exons; higher for contiguous regions	Up to 80 % mapping to CTR for exons; higher for contiguous regions
Uniformity ³	80 % of bases within twofold range of median	58 % of CTR within tenfold coverage range; 88 % within 100-fold coverage range	60 % of CTR within 0.5–1.5-fold of mean coverage (mapping quality 30) ⁶	61 % of CTR within 0.5–1.5-fold of mean coverage (mapping quality 30) ⁶
Reproducibility ⁴	Up to 100 %	Up to 92 %	>95 %	>96 %

¹ Sensitivity: the percentage of the target bases that are represented by one or more sequence reads; ² Specificity: the percentage of sequences that map to the intended targets; ³ Uniformity: the variability in sequence coverage across target regions; ⁴ Reproducibility: how closely the results obtained from replicate experiments correlate; ⁵ CTR, capture target region; ⁶ Mapping qualities were calculated with MAQ mapping software (<http://maq.sourceforge.net>). They indicate the probability that the mapping location is correct. A score of 30 or greater indicates that the quality of a read was good, and that it mapped unambiguously to that location with few mismatches.

использовалось обогащение и секвенирование генов устойчивости NB-LRR, направленное на обнаружение и аннотирование новых генов (Jupe et al., 2013; Andolfo et al., 2014).

Преимущество целевого секвенирования заключается в возможности определения последовательности в конкретной области с гораздо более высоким уровнем покрытия по сравнению с полногеномным секвенированием или секвенированием по Сэнгеру. Полногеномное секвенирование обычно дает покрытие $\times 30$ – 75 раз, в то время как целевое позволяет секвенировать с покрытием $\times 5000$ или выше. Целевое секвенирование – это быстрый и экономичный способ обнаружения известных и новых вариантов в выбранных наборах генов или геномных областях.

Экзомное секвенирование

Экзомное секвенирование – это метод секвенирования только белок-кодирующих участков генов с помощью одной из платформ следующего поколения – Illumina,

SOLID, IonTorrent, PacBio (см. Приложение 2). Экзом составляет 1–2 % генома в зависимости от вида и может быть расширен до целевых функциональных некодирующих элементов (например, микроРНК, длинной некодирующей РНК и др.) и определенных локусов-кандидатов (Warr et al., 2015). Секвенирование экзона включает в себя идентификацию генов (целый экзом, определенный класс генов), дизайн праймеров или чипа для обогащения образцов и секвенирование полученной библиотеки. Экзомное секвенирование позволяет секвенировать неэкспрессирующиеся аллели и гены, которые не могут быть найдены с помощью РНК-seq. Однако подходы экзомного секвенирования основаны на существовании высококачественных референсных геномных последовательностей с точной аннотацией. Некачественная аннотация геномов может привести к потере данных.

Экзомное секвенирование обычно применяется для идентификации мутаций в белок-кодирующих генах (Schneeberger, 2014). В недавнем исследовании (Henry et al., 2014) для выявления мутаций в популяциях мутантов

риса (*O. sativa*) и пшеницы (*T. aestivum*) было использовано секвенирование экзома ($42 \cdot 10^6$ и $107 \cdot 10^6$ п. н. соответственно). Этот метод обеспечил обнаружение мутаций, позволил быстро и недорого создать базу данных полиморфизмов. Обнаружение SNP и SNV с помощью упомянутого метода было проведено и в другой работе, посвященной просу *P. virgatum* (Evans et al., 2014), где выявлено в общей сложности 1 395 501 SNP и 8 173 предполагаемых SNV. В аналогичном эксперименте на гексаплоидной пшенице был разработан 10251 SNP-маркер с применением целевого ресеквенирования пшеничного экзома для получения большого количества геномных данных для восьми сортов. Полученные SNP-маркеры обеспечили значимый источник информации, особенно для селекционных исследований (Allen et al., 2013). Секвенирование экзома ячменя использовалось для идентификации маркеров *H. vulgare* L. и *H. bulbosum* L. – дикого вида, который обладает лучшей устойчивостью к патогенам по сравнению с домашними видами (Wendler et al., 2014). Таким образом, экзомное секвенирование может применяться для изучения геномных вариаций у полиплоидных видов с большими, повторяющимися и гетерозиготными геномами.

Экзомное секвенирование сои позволило обнаружить делеции в генах и провести генетическое скринирование на основе фенома на полногеномном уровне (Bolon et al., 2011).

Секвенирование экзома оказалось важным инструментом для оценки естественной эволюции растений, изучения взаимодействия патогенов хозяина и улучшения производства сельскохозяйственных культур, поскольку экзомы помогают интерпретировать аллельные вариации генов и их влияние на фенотип. Метод можно использовать для разработки стратегий борьбы с патогеном, что было реализовано для идентификации генов, участвующих в растительно-грибных взаимодействиях (O'Brien et al., 2011; Venu et al., 2011).

Кроме того, секвенирование экзома может быть использовано для восстановления таксономических связей и идентификации генов, участвующих в эволюции сельскохозяйственных культур. Так, в 2016 г. были секвенированы экзомы 267 сортов ячменя с целью изучения изменений генов адаптации растения к разным географическим условиям (Russell et al., 2016).

Заключение

В последние годы были секвенированы многие из основных культурных растений, хотя и с разным качеством и полнотой прочтения. Эти геномные последовательности обеспечивают беспрецедентный ресурс, который может быть использован различными способами. Одно из основных направлений применения секвенированных геномов культурных растений заключается в насыщении всех участков генома молекулярными маркерами с последующим их вовлечением в картирование, идентификацию и анализ генов-кандидатов, определяющих агрономически важные признаки, в проведение работ по маркер-ориентированной и геномной селекции.

Секвенирование последовательности генома, даже на текущих уровнях покрытия, способствовало существен-

ному увеличению работ, направленных на идентификацию генов-кандидатов, участвующих в формировании морфологических и хозяйственно ценных признаков растений; значительно расширило пул маркеров, вовлеченных в геномную и маркер-ориентированную селекцию. Подходы по целевому и экзомному секвенированию существенно расширили возможности идентификации генов, оценки их полиморфизма и вклада в формирование фенотипа. Трудности, с которыми столкнулись исследователи при сборке больших геномов растений, полиплоидных геномов, районов с высоким содержанием GC и гомополимеров, районов, насыщенных транспозиционными элементами и другими фракциями повторяющейся ДНК, успешно преодолевались благодаря активному развитию биоинформационных подходов и методов секвенирования.

Список литературы / References

- Aird D., Ross M.G., Chen W.-S., Danielsson M., Fennell T., Russ C., Gnirke A. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 2011;12(2):R18. DOI 10.1186/gb-2011-12-2-r18.
- Akpınar B.A., Biyiklioglu S., Alptekin B., Havránková M., Vrána J., Doležel J., Distelfeld A., Hernandez P., the IWGSC, Budak H. Chromosome-based survey sequencing reveals the genome organization of wild wheat progenitor *Triticum dicoccoides*. *Plant Biotechnol. J.* 2018;16:2077-2087. DOI 10.1111/pbi.12940.
- Allen A.M., Barker G.L., Wilkinson P., BurrIDGE A., Winfield M., Coghill J., Uauy C., Griffiths S., Jack P., Berry S., Werner P., Melichar J.P., McDougall J., Gwilliam R., Robinson P., Edwards K.J. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 2013;11:279-295. DOI 10.1111/pbi.12009.
- Andolfo G., Jupe F., Witek K., Etherington G.J., Ercolano M.R., Jones J.D.G. Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC Plant Biol.* 2014; 14(1):120. DOI 10.1186/1471-2229-14-120.
- Avni R., Nave M., Barad O., Baruch K., Twardziok S.O., Gundlach H., Distelfeld A. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science.* 2017;357(6346): 93-97. DOI 10.1126/science.aan0032.
- Bai B., Wang L., Lee M., Zhang Y., Rahmadsyah, Alfiko Y., Ye B.Q., Wan Z.Y., Lim C.H., Suwanto A., Chua N.-H., Yue G.H. Genome-wide identification of markers for selecting higher oil content in oil palm. *BMC Plant Biol.* 2017;17(1):93. DOI 10.1186/s12870-017-1045-z.
- Barabaschi D., Tondelli A., Desiderio F., Volante A., Vaccino P., Valè G., Cattivelli L. Next generation breeding. *Plant Sci.* 2016;242:3-13. DOI 10.1016/j.plantsci.2015.07.010.
- Bassel G.W., Gaudinier A., Brady S.M., Hennig L., Rhee S.Y., De Smet I. Systems analysis of plant functional, transcriptional, physical interaction, and metabolic networks. *Plant Cell.* 2012;24(10):3859-3875. DOI 10.1105/tpc.112.100776.
- Bennetzen J., Ma J., Devos K. Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 2005;95(1):127-132. DOI 10.1093/aob/mci008.
- Bennetzen J.L., Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* 2014;65(1):505-530. DOI 10.1146/annurev-arplant-050213-035811.
- Bolger M., Weisshaar B., Scholz U., Stein N., Usadel B., Mayer K. Plant genome sequencing – applications for crop improvement. *Curr. Opin. Biotechnol.* 2014;26:31-37. DOI 10.1016/j.copbio.2013.08.019.
- Bolon Y.-T., Haun W.J., Xu W.W., Grant D., Stacey M.G., Nelson R.T., Gerhardt D.J., Jeddloh J.A., Stacey G., Muehlbauer G.J., Orf J.H.,

- Naeve S.L., Stupar R.M., Vance C.P. Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiol.* 2011;156:240-253. DOI 10.1104/pp.110.170811.
- Brenchley R., Spannagl M., Pfeifer M., Barker G.L.A., D'Amore R., Allen A.M., McKenzie N., Kramer M., Kerhornou A., Bolser D., Kay S., Waite D., Trick M., Bancroft I., Gu Y., Huo N., Luo M.-C., Sehgal S., Gill B., Kianian S., Anderson O., Kersey P., Dvorak J., McCombie W.R., Hall A., Mayer K.F.X., Edwards K.J., Bevan M.W., Hall N. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature.* 2012;491:705-710. DOI 10.1038/nature11650.
- Choulet F., Alberti A., Theil S., Glover N., Barbe V., Daron J., Pingault L., Sourdille P., Couloux A., Paux E., Leroy P., Mangenot S., Guilhot N., Le Gouis J., Balfourier F., Alaux M., Jamilloux V., Poulain J., Durand C., Bellec A., Gaspin C., Safar J., Dolezel J., Rogers J., Vandepoele K., Aury J.-M., Mayer K., Berges H., Quesneville H., Wincker P., Feuillet C. Structural and functional partitioning of bread wheat chromosome 3B. *Science.* 2014;345(6194):1249721. DOI 10.1126/science.1249721.
- Clarke W.E., Parkin I.A., Gajardo H.A., Gerhardt D.J., Higgins E. Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L. *PLoS One.* 2013;8(12):e81992. DOI 10.1371/journal.pone.0081992.
- Crain J., Mondal S., Rutkoski J., Singh R.P., Poland J. Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *Plant Genome.* 2018;11(1). DOI 10.3835/plantgenome2017.05.0043.
- Dasgupta M.G., Dharanishanthi V., Agarwal I., Krutovsky K.V. Development of genetic markers in *Eucalyptus* species by target enrichment and exome sequencing. *PLoS One.* 2015;10(1):e0116528. DOI 10.1371/journal.pone.0116528.
- Deschamps S., Llaca V., May G.D. Genotyping-by-sequencing in plants. *Biology.* 2012;1(3):460-483. DOI 10.3390/biology1030460.
- Dohm J.C., Lottaz C., Borodina T., Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* 2008;36(16):e105. DOI 10.1093/nar/gkn425.
- Dohm J.C., Minoche A.E., Holtgräwe D., Capella-Gutiérrez S., Zakrzewski F., Tafer H., Rupp O., Sørensen T.R., Stracke R., Reinhardt R., Goesmann A., Kraft T., Schulz B., Stadler P.F., Schmidt T., Gabaldon T., Lehrach H., Weisshaar B., Himmelbauer H. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature.* 2014;505(7484):546-549. DOI 10.1038/nature12817.
- Doležel J., Kubaláková M., Paux E., Bartoš J., Feuillet C. Chromosome-based genomics in the cereals. *Chromosome Res.* 2007;15(1):51-66. DOI 10.1007/s10577-006-1106-x.
- Evans J., Kim J., Childs K.L., Vaillancourt B., Crisovan E., Nandety A., Gerhardt D.J., Richmond T.A., Jeddloh J.A., Kaeppler S.M., Casler M.D., Buell C.R. Nucleotide polymorphism and copy number variant detection using exome capture and next-generation sequencing in the polyploid grass *Panicum virgatum*. *Plant J.* 2014;79(6):993-1008. DOI 10.1111/tpj.12601.
- Feuillet C., Leach J.E., Rogers J., Schnable P.S., Eversole K. Crop genome sequencing: lessons and rationales. *Trends Plant Sci.* 2011;16(2):77-88. DOI 10.1016/j.tplants.2010.10.005.
- Gardiner L.J., Gawroński P., Olohan L., Schnurbusch T., Hall N., Hall A. Using genic sequence capture in combination with a syntenic pseudo genome to map a deletion mutant in a wheat species. *Plant J.* 2014;80(5):895-904. DOI 10.1111/tpj.12660.
- Goff S.A., Ricke D., Lan T.-H., Presting G., Wang R., Dunn M., Glazebrook J., ... Macalma T., Oliphant A., Briggs S. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science.* 2002;296(5565):92-100. DOI 10.1126/science.1068275.
- Groenen M., Archibald A.L., Uenishi H., Tuggle C.K., Takeuchi Y., ... Rogers J., Churcher C., Schook L.B. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature.* 2012;491:393-398. DOI 10.1038/nature11622.
- Guo S., Zhang J., Sun H., Salse J., Lucas W.J., Zhang H., ... Li Y., Fei Z., Xu Y. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 2013;45(1):51-58. DOI 10.1038/ng.2470.
- Henry I.M., Nagalakshmi U., Lieberman M.C., Ngo K.J., Krasileva K.V., Vasquez-Gross H., Akhunova A., Akhunov E., Dubcovsky J., Tai T.H., Comai L. Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *Plant Cell.* 2014;26:1382-1397. DOI 10.1105/tpc.113.121590.
- Hernandez P., Martis M., Dorado G., Pfeifer M., Gálvez S., Schaaf S., Jouve N., Šimková H., Valárik M., Doležel J., Mayer K.F. Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J.* 2012;69(3):377-386. DOI 10.1111/j.1365-313X.2011.04808.x.
- Howe K., Clark M.D., Torroja C.F., Torrance J., Berthelot C., ... Crollius H.R., Rogers J., Stemple D. The zebrafish reference genome sequence and its relationship to the human genome. *Nature.* 2013;496:498-503. DOI 10.1038/nature12111.
- Huang S., Li R., Zhang Z., Li L., Gu X., Fan W., ... Wang J., Du Y., Li S. The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* 2009;41:1275-1281. DOI 10.1038/ng.475.
- International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science.* 2018;361(6403):eaar7191. DOI 10.1126/science.aar7191.
- Jaillon O., Aury J.M., Noel B., Policriti A., Clepet C., Casagrande A., ... Weissenbach J., Quétier F., Wincker P. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature.* 2007;449(7161):463-467. DOI 10.1038/nature06148.
- Jiao Y., Peluso P., Shi J., Liang T., Stitzer M.C., Wang B., Campbell M.S., ... Hastie A., Rank D.R., Ware D. Improved maize reference genome with single-molecule technologies. *Nature.* 2017;546:524-527. DOI 10.1038/nature22971.
- Jupe F., Witek K., Verweij W., Sliwka J., Pritchard L., Etherington G.J., Maclean D., Cock P.J., Leggett R.M., Bryan G.J., Cardle L., Hein I., Jones J.D. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant J.* 2013;76(3):530-544. DOI 10.1111/tpj.12307.
- Kol G. The next revolution in genomics: exploring complex genome assembly to pan-genome interconnection. *Proc. of the Plant & Animal Genome Conference XXIV.* 2016;W3289.
- Krasnov Y.M., Guseva N.P., Sharapova N.A., Cherkasov A.V. Modern methods of DNA sequencing (scientific review). *Problemy Osobo Opasnykh Infektsiy = Problems of Particularly Dangerous Infections.* 2014;2:73-79. DOI 10.21055/0370-1069-2014-2-73-79. (in Russian)
- Li F., Fan G., Lu C., Xiao G., Zou C., Kohel R.J., Ma Z., ... Yu J.Z., Zhu Y., Yu S. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* 2015;33:524-530. DOI 10.1038/nbt.3208.
- Li Y., Ruperao P., Batley J., Edwards D., Khan T., Colmer T.D., Pang J., Siddique K.H.M., Sutton T. Investigating drought tolerance in chickpea using genome-wide association mapping and genomic selection based on whole-genome resequencing data. *Front. Plant Sci.* 2018;9:190. DOI 10.3389/fpls.2018.00190.
- Luo M.-C., Gu Y.Q., Puiu D., Wang H., Twardziok S.O., Deal K.R., Huo N., ... Salzberg S.L., Devos K.M., Dvořák J. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature.* 2017;551:498-502. DOI 10.1038/nature24486.
- Mardis E.R. DNA sequencing technologies: 2006–2016. *Nat. Protoc.* 2017;12(2):213-218. DOI 10.1038/nprot.2016.182.
- Mascher M., Richmond T.A., Gerhardt D.J., Himmelbach A., Clissold L., Sampath D., Ayling S., Steuernagel B., Pfeifer M., D'Ascenzo M., Akhunov E.D., Hedley P.E., Gonzales A.M., Mor-

- rell P.L., Kilian B., Blattner F.R., Scholz U., Mayer K.F.X., Flavell A.J., Muehlbauer G.J., Waugh R., Jeddelloh J.A., Stein N., Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* 2013;76(3):494-505. DOI 10.1111/tbj.12294.
- Matasci N., Hung L.-H., Yan Z., Carpenter E.J., Wickett N.J., Mirarab S., Nguyen N., ... Wang J., Leebens-Mack J., Wong G.K.-S. Data access for the 1,000 plants (1KP) project. *GigaScience.* 2014; 3(1):17. DOI 10.1186/2047-217X-3-17.
- Mayer K.F.X., Rogers J., Doležel J., Pozniak C., Eversole K., Feuillet C., Gill B., ... Loaec M., Keller B., Praud S. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Nature.* 2014;345(6194):1251788. DOI 10.1126/science.1251788.
- Ming R., Hou S., Feng Y., Yu Q., Dionne-Laporte A., Saw J.H., Senin P., ... Gonsalves D., Wang L., Alam M. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature.* 2008;452:991-996. DOI 10.1038/nature06856.
- Muraya M.M., Schmutz T., Ulpinnis C., Scholz U., Altmann T. Targeted sequencing reveals large-scale sequence polymorphism in maize candidate genes for biomass production and composition. *PLoS One.* 2015;10(7):e0132120. DOI 10.1371/journal.pone.0132120.
- Myburg A.A., Grattapaglia D., Tuskan G.A., Hellsten U., Hayes R.D., Grimwood J., Jenkins J., ... Van de Peer Y., Rokhsar D.S., Schmutz J. The genome of *Eucalyptus grandis*. *Nature.* 2014;510(7505):356-362. DOI 10.1038/nature13308.
- Neves L.G., Davis J.M., Barbazuk W.B., Kirst M. Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J.* 2013; 75:146-156. DOI 10.1111/tbj.12193.
- Neves L.G., Davis J.M., Barbazuk W.B., Kirst M. A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *G3: Genes Genomes Genetics.* 2014;4(1):29-37. DOI 10.1534/g3.113.008714.
- Nystedt B., Street N.R., Wetterbom A., Zuccolo A., Lin Y.-C., Scofield D.G., Vezzi F., ... Ingvarsson P.K., Lundeberg J., Jansson S. The Norway spruce genome sequence and conifer genome evolution. *Nature.* 2013;497(7451):579-584. DOI 10.1038/nature12211.
- O'Brien H.E., Thakur S., Guttman D.S. Evolution of plant pathogenesis in *Pseudomonas syringae*: A genomics perspective. *Annu. Rev. Phytopathol.* 2011;49(1):269-289. DOI 10.1146/annurev-phyto-072910-095242.
- Pankin A., Campoli C., Dong X., Kilian B., Sharma R., Himmelbach A., Saini R., Davis S.J., Stein N., Schneeberger K., von Korff M. Mapping-by-sequencing identifies *HvPHYTOCHROME C* as a candidate gene for the *early maturity 5* locus modulating the circadian clock and photoperiodic flowering in barley. *Genetics.* 2014;198(1):383-396. DOI 10.1534/genetics.114.165613.
- Paux E., Sourdille P., Salse J., Saintenac C., Choulet F., Leroy P., Korol A., Michalak M., Kianian S., Spielmeier W., Lagudah E., Somers D., Kilian A., Alaux M., Vautrin S., Bergès H., Eversole K., Appels R., Safar J., Simkova H., Dolezel J., Bernard M., Feuillet C. A physical map of the 1-gigabase bread wheat chromosome 3B. *Science.* 2008;322(5898):101-104. DOI 10.1126/science.1161847.
- Pootakham W., Shearman J.R., Ruang-areerate P., Sonthirod C., Sangsrakru D., Jomchai N., Yoocha T., Triwitayakorn K., Tragooonrun S., Tangphatsornruang S. Large-scale SNP discovery through RNA sequencing and SNP genotyping by targeted enrichment sequencing in cassava (*Manihot esculenta* Crantz). *PLoS One.* 2014; 9(12):e116028. DOI 10.1371/journal.pone.0116028.
- Pucker B., Holtgräwe D., Rosleff Sörensen T., Stracke R., Viehöver P., Weisshaar B. A *de novo* genome sequence assembly of the *Arabidopsis thaliana* accession Niederzenz-1 displays presence/absence variation and strong synteny. *PLoS One.* 2016;11(10):e0164321. DOI 10.1371/journal.pone.0164321.
- Rensing S.A., Lang D., Zimmer A.D., Terry A., Salamov A., Shapiro H., Nishiyama T., ... Grigoriev I.V., Quatrano R.S., Boore J.L. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science.* 2008;319(5859):64-69. DOI 10.1126/science.1150646.
- Russell J., Mascher M., Dawson I.K., Kyriakidis S., Calixto C., Freund F., Bayer M., ... Muehlbauer G.J., Stein N., Waugh R. Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nat. Genet.* 2016;48(9):1024-1030. DOI 10.1038/ng.3612.
- Saintenac C., Jiang D., Akhunov E.D. Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol.* 2011;12(9):R88. DOI 10.1186/gb-2011-12-9-r88.
- Salmon A., Udall J., Jeddelloh J., Wendel J. Targeted capture of homoeologous coding and noncoding sequence in polyploid cotton. *G3: Genes Genomes Genetics.* 2012;2(8):921-930. DOI 10.1534/g3.112.003392.
- Scheben A., Wolter F., Batley J., Puchta H., Edwards D. Towards CRISPR/Cas crops – bringing together genomics and genome editing. *New Phytologist.* 2017;216(3):682-698. DOI 10.1111/nph.14702.
- Schiessl S., Samans B., Hüttel B., Reinhard R., Snowdon R.J. Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* 2014;5:404. DOI 10.3389/fpls.2014.00404.
- Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., ... Stacey G., Shoemaker R.C., Jackson S.A. Genome sequence of the palaeopolyploid soybean. *Nature.* 2010;463:178-183. DOI 10.1038/nature08670.
- Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F., Pasternak S., Liang C., ... McCombie W.R., Wing R.A., Wilson R.K. The B73 maize genome: complexity, diversity, and dynamics. *Science.* 2009; 326(5956):1112-1115. DOI 10.1126/science.1178534.
- Schneeberger K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nat. Rev. Genet.* 2014; 15(10):662-676. DOI 10.1038/nrg3745.
- Shendure J., Balasubramanian S., Church G.M., Gilbert W., Rogers J., Schloss J.A., Waterston R.H. DNA sequencing at 40: Past, present and future. *Nature.* 2017;550(7676):345-353. DOI 10.1038/nature24286.
- Sierro N., Battey J.N.D., Ouadi S., Bakaher N., Bovet L., Willig A., Goepfert S., Peitsch M.C., Ivanov N.V. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* 2014;5(1):3833. DOI 10.1038/ncomms4833.
- Springer N.M., Anderson S.N., Andorf C.M., Ahern K.R., Bai F., Barad O., Barbazuk W.B., ... Woodhouse M.R., Xiong W., Brutnell T.P. The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* 2018;50(9):1282-1288. DOI 10.1038/s41588-018-0158-0.
- Stull G.W., Moore M.J., Mandala V.S., Douglas N.A., Kates H.-R., Qi X., Brockington S.F., Soltis P.S., Soltis D.E., Gitzendanner M.A. A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. *Appl. Plant Sci.* 2013;1(2):1200497. DOI 10.3732/apps.1200497.
- Tennessen J.A., Govindarajulu R., Liston A., Ashman T.-L. Targeted sequence capture provides insight into genome structure and genetics of male sterility in a gynodioecious diploid strawberry, *Fragaria vesca* ssp. *bracteata* (Rosaceae). *G3: Genes Genomes Genetics.* 2013;3:1341-1351. DOI 10.1534/g3.113.006288.
- Terracciano I., Cantarella C., D'Agostino N. Hybridization-Based Enrichment and Next Generation Sequencing to Explore Genetic Diversity in Plants. In: Rogato A., Zazzu V., Guarracino M. (Eds.). *Dynamics of Mathematical Models in Biology.* Springer, Cham., 2016. DOI 10.1007/978-3-319-45723-9_10.
- The 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell.* 2016;166(2):481-491. DOI 10.1016/j.cell.2016.05.063.
- The 3,000 rice genomes project. The 3,000 Rice Genomes Project. *GigaScience.* 2014;3(1):7. DOI 10.1186/2047-217X-3-7.

- The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000;408:796-815. DOI 10.1038/35048692.
- The International Barley Genome Sequencing Consortium. A physical, genetic and functional sequence assembly of the barley genome. *Nature*. 2012;491(7426):711-716. DOI 10.1038/nature11543.
- The International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*. 2010;463(7282):763-768. DOI 10.1038/nature08747.
- The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature*. 2011;475:189-195. DOI 10.1038/nature10158.
- The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*. 2012;485:635-641. DOI 10.1038/nature11119.
- Tuskan G.A., DiFazio S., Jansson S., Bohlmann J., Grigoriev I., Hellsten U., Putnam N., ... Sandberg G., Van de Peer Y., Rokhsar D. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 2006;313(5793):1596-1604. DOI 10.1126/science.1128691.
- Uitdewilligen J.G.A.M.L., Wolters A.-M.A., D'hoop B.B., Borm T.J.A., Visser R.G.F., van Eck H.J. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS One*. 2013;8(5):e62355. DOI 10.1371/journal.pone.0062355.
- Velasco R., Zharkikh A., Affourtit J., Dhingra A., Cestaro A., Kalyanaraman A., Fontana P., ... Van de Peer Y., Salamini F., Viola R. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat. Genet*. 2010;42:833-839. DOI 10.1038/ng.654.
- Venu R.C., Zhang Y., Weaver B., Carswell P., Mitchell T.K., Meyers B.C., Boehm M.J., Wang G.-L. Large Scale Identification of Genes Involved in Plant-Fungal Interactions Using Illumina's Sequencing-by-Synthesis Technology. In: Xu J.-R., Bluhm B.H. (Eds.). *Fungal Genomics: Methods and Protocols*. Springer, 2011;167-178. DOI 10.1007/978-1-61779-040-9_12.
- Warr A., Robert C., Hume D., Archibald A., Deeb N., Watson M. Exome sequencing: Current and future perspectives. *G3: Genes Genomes Genetics*. 2013;5(8):1543-1550. DOI 10.1534/g3.115.018564.
- Wendler N., Mascher M., Nöh C., Himmelbach A., Scholz U., Ruge-Wehling B., Stein N. Unlocking the secondary gene-pool of barley with next-generation sequencing. *Plant Biotechnol. J*. 2014;12:1122-1131. DOI 10.1111/pbi.12219.
- Westwood J.H., dePamphilis C.W., Das M., Fernández-Aparicio M., Honaas L.A., Timko M.P., Wafula E.K., Wickett N.J., Yoder J.I. The Parasitic Plant Genome Project: New tools for understanding the biology of *Orobanche* and *Striga*. *Weed Sci*. 2012;60(02):295-306. DOI 10.1614/WS-D-11-00113.1.
- Winfield M.O., Wilkinson P.A., Allen A.M., Barker G.L., Coghill J.A., Burrige A., Hall A., Brenchley R.C., D'Amore R., Hall N., Bevan M.W., Richmond T., Gerhardt D.J., Jeddloh J.A., Edwards K.J. *Plant Biotechnol. J*. 2012;10(6):733-742. DOI 10.1111/j.1467-7652.2012.00713.x.
- Xu X., Pan S., Cheng S., Zhang B., Mu D., Ni P., Zhang G., ... Gouvenet A., van Ham R.C.H.J., Visser R.G.F. (The Potato Genome Sequencing Consortium). Genome sequence and analysis of the tuber crop potato. *Nature*. 2011;475(7355):189-195. DOI 10.1038/nature10158.
- Yu J., Hu S., Wang J., Wong G.K.-S., Li S., Liu B., Deng Y., ... Zhu L., Yuan L., Yang H. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*. 2002;296(5565):79-92. DOI 10.1126/science.1068037.
- Zhang H., Scheuring C., Zhang M., Zhang Y., Wu C., Dong J., Li Y. Construction of BIBAC and BAC libraries from a variety of organisms for advanced genomics research. *Nat. Protoc*. 2012;7(3):479-499. DOI 10.1038/nprot.2011.456.
- Zhang J., Zhang Y., Du Y., Chen S., Tang H. Dynamic metabonomic responses of tobacco (*Nicotiana tabacum*) plants to salt stress. *J. Proteome Res*. 2011;10(4):1904-1914. DOI 10.1021/pr101140n.
- Zhou G., Chen Y., Yao W., Zhang C., Xie W., Hua J., Xing Y., Xiao J., Zhang Q. Genetic composition of yield heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA*. 2012; 109(39):15847-15852. DOI 10.1073/pnas.1214141109.
- Zhu C., Bortesi L., Baysal C., Twyman R. M., Fischer R., Capell T., Schillberg S., Christou P. Characteristics of genome editing mutations in cereal crops. *Trends Plant Sci*. 2017; 22(1):38-52. DOI 10.1016/j.tplants.2016.08.009.

ORCID ID

D.A. Afonnikov 0000-0001-9738-1409

Acknowledgements. This work was supported by State Budgeted Project № 0324-2019-0039.

Conflict of interest. The authors declare no conflict of interest.

Received September 6, 2018. Revised December 17, 2018. Accepted January 5, 2019.