

# Идентификация микросателлитных локусов по данным секвенирования ВАС-клонов и их физическое картирование на хромосому 5В мягкой пшеницы

М.А. Нестеров<sup>1</sup>, Д.А. Афонников<sup>1</sup>, Е.М. Сергеева<sup>1</sup>, Л.А. Мирошниченко<sup>2</sup>, М.К. Брагина<sup>1</sup>, А.О. Брагин<sup>1</sup>, Г.В. Васильев<sup>1</sup>, Е.А. Салина<sup>1</sup>

<sup>1</sup> Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия

<sup>2</sup> Федеральное государственное бюджетное учреждение науки «Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук», Новосибирск, Россия

Необходимость изучения микросателлитных локусов пшеницы, в первую очередь, обусловлена актуальностью работ по выявлению полиморфных маркеров для участков хромосом, определяющих хозяйственно ценные признаки. В настоящей работе проведено насыщение отдельных районов короткого плеча хромосомы 5В (5BS) мягкой пшеницы SSR-маркерами, разработанными по данным секвенирования ВАС-клонов. 130 клонов, отобранных случайным образом из ВАС-библиотеки 5BS, были секвенированы на платформе IonTorrent и собраны в контиги с использованием программы MIRA. Характеристики сборки (N50 = 4 136 п.н.) сравнимы с таковыми для сборок генома пшеницы и родственных видов, полученными в последнее время, и приемлемы для решения задачи идентификации микросателлитных локусов. Для выявления последовательностей ДНК с повторяющейся единицей 2–4 п.н. использовался алгоритм, основанный на свойствах сложностных разложений, формирующихся в режиме скользящего окна. По данным анализа 17 770 контигов общей протяженностью 25 879 921 п.н., разработано 113, 79 и 67 маркеров микросателлитных (SSR) локусов с повторяющейся единицей 2, 3 и 4 п.н. соответственно. SSR-маркеры с мотивом 3 п.н. были проверены на нулли-тетрасомных линиях пятой гомеологической группы хромосом сорта пшеницы Чайниз Спринг (CS). Выявлен 21 маркер, специфичный для хромосомы 5В. Были локализованы 8 маркеров в дистальном районе хромосомы (бин 5BS6) с использованием серии делеционных линий CS по 5BS. Для 8 и 4 маркеров определена локализация в интерстициальном районе в бинах 5BS5 и 5BS4 соответственно, один маркер был локализован в прицентромерном бине. Сравнительный анализ распределения тринуклеотидных микросателлитов по хромосоме 5В пшеницы и у различных видов злаков указывает на пролиферацию и поддержание количественного содержания повтора (AAG)<sub>n</sub> в процессе эволюции злаков.

Ключевые слова: мягкая пшеница; *Triticum aestivum*; ВАС-клоны; IonTorrent; MIRA; микросателлиты; SSR-маркеры; хромосома 5В; (AAG)<sub>n</sub>.

## Identification of microsatellite loci according to BAC sequencing data and their physical mapping to the bread wheat 5B chromosome

M.A. Nesterov<sup>1</sup>, D.A. Afonnikov<sup>1</sup>, E.M. Sergeeva<sup>1</sup>, L.A. Miroshnichenko<sup>2</sup>, M.K. Bragina<sup>1</sup>, A.O. Bragin<sup>1</sup>, G.V. Vasiliev<sup>1</sup>, E.A. Salina<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia  
<sup>2</sup> Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia

The shortage of polymorphic markers for the regions of wheat chromosomes that encode commercially valuable traits determined the need for studying wheat microsatellite loci. In this work, SSR markers for individual regions in the short arm of bread wheat chromosome 5B (5BS) were designed based on sequencing data for BAC clones, and the regions of the corresponding chromosome were saturated with these markers. Totally, 130 randomly selected BAC clones from the 5BS library were sequenced on the Ion Torrent platform and assembled in contigs using MIRA software. The assembly characteristics (N50 = 4 136 bp) are comparable to the recently obtained data for wheat and relative species and acceptable for identification of microsatellite loci. An algorithm utilizing the properties of complexity decompositions in the sliding-window mode was used to detect DNA sequences with a repeat unit of 2–4 bp. Analysis of 17 770 contigs with the total length of 25 879 921 bp allowed for designing 113, 79, and 67 microsatellite (SSR) loci with a repeat unit of 2, 3, and 4 bp, respectively. The SSR markers with a motif of 3 bp were tested using nullitetrasonic lines of Chinese Spring wheat homoeologous group 5. Thus, 21 markers specific for chromosome 5B were detected. Eight of these markers were mapped to the distal region of this chromosome (bin 5BS6) using a set of Chinese Spring deletion lines for 5BS. Eight and four markers were mapped to the interstitial region (bins 5BS5 and 5BS4, respectively). One marker was mapped to a pericentromeric bin. A comparative analysis of the distribution of trinucleotide microsatellites over

wheat chromosome 5B and in different cereal species suggests that the (AAG)<sub>n</sub> repeat has proliferated and has been maintained during the evolution of cereals.

Key words: bread wheat; *Triticum aestivum*; BAC clones; Ion Torrent; MIRA; microsatellites; chromosome 5B; SSR markers; (AAG)<sub>n</sub>.

#### КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ?

Нестеров М.А., Афонников Д.А., Сергеева Е.М., Мирошниченко Л.А., Брагина М.К., Брагин А.О., Васильев Г.В., Салина Е.А. Идентификация микросателлитных локусов по данным секвенирования ВАС-клонов и их физическое картирование на хромосому 5В мягкой пшеницы. Вавиловский журнал генетики и селекции. 2015;19(6):707-714. DOI 10.18699/VJ15.086

#### HOW TO CITE THIS ARTICLE?

Nesterov M.A., Afonnikov D.A., Sergeeva E.M., Miroshnichenko L.A., Bragina M.K., Bragin A.O., Vasiliev G.V., Salina E.A. Identification of microsatellite loci according to BAC sequencing data and their physical mapping to the bread wheat 5B chromosome. Vavilovskii Zhurnal Genetiki i Selekcii = Vavilov Journal of Genetics and Breeding. 2015;19(6):707-714. DOI 10.18699/VJ15.086

Физическое картирование хромосом, а именно прямая локализация последовательностей ДНК на хромосомы, впервые получило свое развитие с появлением методов гибридизации *in situ*. Физическое картирование можно успешно проводить с использованием различных генетических ресурсов растений, включающих анеуплоидные, замещенные, интрогрессивные и делеционные линии, в связи с чем его иногда называют цитогенетическим картированием (Sourdille et al., 2004). В век геномного секвенирования термин «физическое картирование» все чаще используется при локализации протяженных контигов ДНК или ВАС-клонов в определенном участке хромосомы, при этом методы локализации могут быть различны, включая в том числе и биоинформатические подходы (Paux et al., 2008).

Мягкая пшеница *Triticum aestivum* L. ( $2n = 42$ ) является естественным аллополиплоидом с геномной формулой ВВААДД (Feldman, 2001), в образовании которого принимали участие диплоидные виды *Triticum* и *Aegilops*. Э. Сирс (Sears, 1966) показал, что хромосомы трех геномов мягкой пшеницы можно разбить на 7 гомеологических групп, внутри которых одна хромосома в экстра- (тетра-) дозе компенсирует нарушения, обусловленные отсутствием другой. Способность гомеологичных хромосом служить буфером при потере хромосом или их фрагментов привела к созданию коллекций анеуплоидных и делеционных линий, которые явились основой для нового этапа развития генетических и молекулярных работ по анализу генома пшеницы.

Первый этап таких работ был связан с использованием серии анеуплоидных линий мягкой пшеницы для изучения генетического вклада каждой хромосомы в наследование различных признаков пшеницы, локализации и распределения маркеров и генов по группам сцепления, анализа взаимодействия отдельных генов при формировании признаков (Plaschke et al., 1996).

Разработка новых коллекций, основанных на стабильности генома полиплоидных форм при потере отдельных участков хромосом, связана с созданием серии делеционных линий. Коллекции этих линий по каждому хромосомному плечу были тщательно охарактеризованы цитологическими методами (http://www.ksu.edu/wgrc/Germplasm/Deletions/). Создание перекрывающихся делеционных линий явилось необходимым этапом для анализа генома растений. В последние годы данная коллекция интенсивно используется Консорциумом по секвениро-

ванию генома пшеницы при построении физических карт индивидуальных хромосом (<http://www.wheatgenome.org/Projects/IWGSC-Bread-Wheat-Projects/Physicalmapping>).

Несмотря на то что число маркеров для генома пшеницы существенно увеличилось в последние пять лет, особенно с появлением первых результатов его секвенирования и SNP-чипов для анализа генома пшеницы, задача по разработке новых маркеров все еще остается актуальной (Akhunov et al., 2009; Brenchley et al., 2012). Сопоставление физических и генетических карт хромосом мягкой пшеницы указывает на неравномерность распределения молекулярных маркеров по хромосоме (Sourdille et al., 2004; Тимонова и др., 2013). Микросателлитные локусы, или тандемные повторы с повторяющей единицей от 2 до 6 п. н. (Simple Sequence Repeat, SSR) успешно используются для разработки SSR-маркеров. Преимущество этих маркеров заключается в высоком уровне полиморфизма и стабильности в получении результатов, их недостаток – более низкая частота встречаемости относительно SNP- и ISBP-маркеров (Li et al., 2002).

Короткое плечо хромосомы 5В имеет протяженность 290 млн п. н. Ряд генов, влияющих на устойчивость к грибным патогенам, скрещиваемость с рожью, мягкозерность зерна, устойчивость к гессенской мухе, локализованы на данном плече хромосомы ([http://www.shigen.nig.ac.jp/wheat/komugi/genes/symbol\\_ClassList.jsp](http://www.shigen.nig.ac.jp/wheat/komugi/genes/symbol_ClassList.jsp)). В настоящий момент количество маркеров на физической и генетической картах хромосомы 5В явно недостаточно для маркирования генов и их дальнейшего использования для позиционного клонирования генов или их применения в маркер-ориентированной селекции.

Разработка маркеров на основе ВАС-клонов преследует две цели, одна из которых – расширение пула маркеров, вторая – разработка маркеров непосредственно к ВАС-клонам для их последующей локализации на хромосоме. Концевое секвенирование ВАС-клонов активно проводилось в последние годы для решения этих задач (Paux et al., 2008). Однако при использовании данного метода эффективность разработки маркеров крайне низка из-за ограничений по протяженности и количеству секвенируемых последовательностей (не более двух) для индивидуального ВАС-клона.

Настоящая работа направлена на разработку новых SSR-маркеров по данным высокопроизводительного секвенирования пула ВАС-клонов короткого плеча хромосомы 5В и их локализацию на физической карте хромосомы.

## Материалы и методы

### Растительный материал

В работе использовались сорт мягкой пшеницы Чайниз Спринг; нуллитетрасомные линии N5BT5A (хромосома 5B замещена на 5A) и N5BT5D (хромосома 5B замещена на 5D); дителосомная линия Dt5BL (отсутствует короткое плечо 5BS) и набор из 10 делеционных линий мягкой пшеницы сорта Чайниз Спринг по хромосоме 5B, полученных и описанных ранее (Endo, Gill, 1996). Делеционные линии определяют 7 делеционных участков (бинов) 5BS: C-5BS3-0.41; 5BS3-0.41-0.42; 5BS2-0.42-0.43; 5BS4-0.43-0.56; 5BS8-0.56-0.71; 5BS5-0.71-0.81; 5BS6-0.81-1.00. Номенклатура делеционных линий и бинов подробно описана Л. Ки с коллегами (Qi et al., 2003). Делеционные линии были любезно предоставлены центром генетических ресурсов Канзасского университета ([http://www.k-state.edu/wgrc/Germplasm/Deletions/del\\_index.html](http://www.k-state.edu/wgrc/Germplasm/Deletions/del_index.html)). Линии были размножены и проверены на наличие делеций (Тимонова и др., 2013).

### ВАС-клоны

ВАС-клоны 5BS получены в Институте экспериментальной ботаники (Оломоук, Чехия) и любезно предоставлены профессором Я. Долежелем.

### Выделение ДНК

Выделение и очистку ДНК из растений проводили согласно ранее описанной методике (Plaschke et al., 1995).

ДНК из ВАС-клонов выделяли с использованием набора NucleoSpin 96 Flash (Macherey-Nagel), в основе которого лежит SDS-щелочной лизис. Методика была адаптирована под имеющуюся аппаратуру с включением дополнительной стадии очистки методом центрифугирования.

### Анализ ПЦР

Анализ проводили в реакционной смеси объемом 20 мкл, содержащей 50 нг ДНК-матрицы, 65 мМ трис-HCl (рН 8,9), 1,5 мМ MgCl<sub>2</sub>, 16 мМ (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 0,05 % Tween 20, по 0,2 мМ каждого дНТФ, 1 ед. ДНК-полимеразы Taq, 0,25 мкМ прямого и обратного специфических праймеров (Дополнительные материалы 1<sup>1</sup>). Реакцию проводили при денатурации – 30 с при 94 °С; отжиге – 30 с при 55 °С; полимеризации – 30 с при 72 °С; числе циклов – 35. ПЦР осуществляли в амплификаторе «BioRad T100» (США). ПЦР-фрагменты разделяли электрофорезом в 2 %-м агарозном геле или 10 % ПААГ, в зависимости от уровня полиморфизма между фрагментами.

### Секвенирование на платформе IonTorrent

ДНК 130 индивидуально выделенных ВАС-клонов была пулирована в приблизительно эквимольных количествах. Для приготовления библиотек IonTorrent 5 мкг ДНК фрагментировано ультразвуком на приборе Covaris, фрагменты длиной 330–450 п. н. выделены на приборе LabchipXT, после репарации концов проведено лигирование адаптеров с помощью набора Ion Plus Fragment Library Kit. Библиотека мультиплицирована с 8 циклами

амплификации, очищена набором AmPureXP и проверена на приборе Bioanalyzer2100. Далее проведены эмульсионная ПЦР с набором Ion PGM Template OT2 400 Kit и секвенирование на приборе IonTorrent на расчетную длину 400 п. н. с использованием чипа максимальной емкости – 318.v2.

Для сборки последовательностей ДНК была использована программа MIRA (Chevreux et al., 1999). Оценку глубины покрытия сборки генома и доли ПЦР-дубликатов проводили на основе картирования исходных секвенированных фрагментов длиной более 30 п. н. на последовательности контигов программой Bowtie2 (Langmead, Salzberg, 2012) и дальнейшей обработки выравниваний программой Samtools (Li et al., 2009).

### Идентификация SSR-локусов, разработка маркеров

Для выявления последовательностей с повторяющимся мотивом 2–4 п. н. использовался алгоритм, основанный на свойствах сложных разложений, формирующихся в режиме скользящего окна (Gusev et al., 2009).

Разработка праймеров к повторяющимся мотивам 2–4 п. н. (SSR-локусам) осуществлялась с помощью программы PrimerQuest (<http://eu.idtdna.com/PrimerQuest/Home/Index>) при установке следующих параметров: длина ампликона (150–450 символов); длина праймера (17–25 символов); (C+G) состав (40–60 %); Tm (59–65 °С); C или G на 3'-конце праймера. Остальные параметры PrimerQuest использовались по умолчанию.

Из нескольких (до 5) пар праймеров, предложенных программой PrimerQuest, для каждого SSR-локуса выбиралась одна пара. Преимуществом пользовались праймеры с высокой сложностью: лингвистической или ДНК-ориентированной мерой LZ (Gusev et al., 1999). Исключались праймеры с неравновесным составом нуклеотидов, с наличием длинных (G+C)-, (A+G)-, (A+C)-, (A+T)-, (C+T)-, (G+T)-трактов, а также праймеры, в которых присутствовали повторяющиеся мотивы.

## Результаты

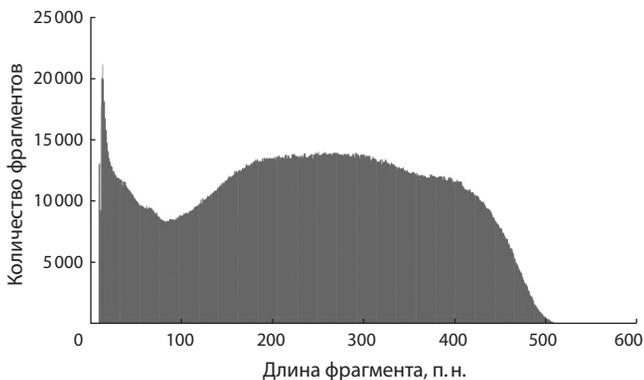
### Секвенирование ВАС-клонов и сборка контигов

Количество клонов, отобранных для секвенирования, было рассчитано исходя из не менее чем стократного прочтения изучаемых последовательностей ДНК, средней длины ВАС-клона около 100 т. п. н. и размера чипа IonTorrent максимальной емкости. Таким образом, расчетное число ВАС-клонов составило 130.

В ходе секвенирования пула ВАС-клонов были получены результаты по плотности нанесения и качеству секвенирования выше средних, наблюдаемых при использовании данного метода. Распределение длин прочитанных фрагментов (рис. 1) показало, что большинство из них имеет длину 150–400 п. н.

Общий объем полученных данных составил более 5,5 млн индивидуальных чтений со средней длиной 239, в сумме более 1,3 млрд нуклеотидов, доля последовательностей с Q > 20 составила 74,6 %. В результате было получено расчетное стократное среднее прочтение взятых в анализ ВАС-клонов.

<sup>1</sup> Дополнительные материалы 1–3 см. в Приложении 5 по адресу: <http://www.bionet.nsc.ru/vogis/download/pict-2015-12/appx5.pdf>



**Рис. 1.** Распределение секвенированных фрагментов библиотеки ВАС-клонов по длине.

Одним из первых этапов при сборке последовательностей ДНК после секвенирования ВАС-клонов является удаление последовательностей ДНК вектора и ДНК *E. coli*. В исходных данных было выявлено менее 2,2 % ДНК *E. coli* (*Escherichia coli* str. K12 substr. DH10B), что указывало на высокое качество выделяемой ДНК ВАС-клонов.

Результаты сборки последовательностей после удаления ДНК *E. coli* с использованием программы MIRA представлены в таблице. Отметим, что характеристики для контигов, имеющих длину более 500 п.н. (~5 тыс. контигов), демонстрируют почти двукратное увеличение параметров N50 и более чем двукратное для параметров N90 и N95.

Для дополнительной оценки качества секвенирования полученных библиотек мы оценили содержание ПЦР дубликатов в них и среднюю глубину покрытия последовательности ДНК ВАС-клонов прочтениями. Для этого на собранные контиги мы картировали исходные прочтения (длиной более 30 п.н.) и по данным выравнивания вычисляли указанные характеристики. Оказалось, что доля ПЦР дубликатов составила около 14 %, а средняя глубина покрытия – примерно 25 прочтений на позицию контига. Кроме того, программа MIRA выдает оценку средней глубины покрытия для контигов, имеющих длину 500 и среднее покрытие не менее 15. Значение среднего покрытия для таких качественных контигов составило ~40 прочтений на позицию. Полученные оценки позволяют заключить, что библиотеки и собранные на основе их секвенирования контиги являются пригодными для решения задачи идентификации SSR-маркеров в последовательностях ВАС-клонов.

### Идентификация и маркирование SSR-локусов

Исходный материал (секвенированные последовательности ДНК) содержит повторяющиеся фрагменты. Некоторые последовательности повторяются целиком, другие являются фрагментами более длинных последовательностей. Как следствие, часть последовательностей ДНК, входящих в состав SSR-локусов, тоже повторяются. Для выявления повторов была разработана программа, основанная на алгоритме поиска «образцов», которая позволяла удалять из анализа идентичные SSR-локусы.

На первом этапе, по данным анализа 17770 контигов общей протяженностью 25 879 921 п.н., выявлено 253, 156 и 27 последовательностей ДНК, содержащих SSR-локусы с периодом повторности 2, 3 и 4 п.н. соответственно и длиной не менее 20 нуклеотидов. На втором этапе было проведено удаление повторяющихся мотивов, расположенных на концах контигов, повторяющихся SSR-локусов и локусов, окружение которых не позволяет разработать праймеры. В результате было отобрано 113 праймеров для SSR-локусов с повторяющимся мотивом 2 п.н., 79 – для SSR-локусов с мотивом 3 п.н. и 23 – для SSR-локусов с мотивом 4 п.н.

Для 67 пар праймеров, подобранных к SSR-локусам с повторяющимся мотивом 3 п.н., была проведена ПЦР с ДНК нулли-тетрасомных и делеционных линий и изучена возможность их использования в качестве маркеров различных участков хромосомы 5В пшеницы (Доп. материалы 1).

### Оценка и физическое картирование новых SSR-маркеров

Результаты ПЦР ДНК сорта CS с праймерами, разработанными к SSR-локусам с повторяющимся мотивом 3 п.н., представлены в Доп. материалах 2. Продукты амплификации для большинства изученных маркеров были четкими и хорошо выявляемыми как в ПААГ, так и в агарозных гелях. Только четыре (Xicgc740, Xicgc1738, Xicgc1755, Xicgc1931) из 67 изученных SSR-маркеров в стандартных ПЦР условиях давали слабые фрагменты амплификации.

Физическое картирование маркеров на хромосоме 5BS проводили с использованием 7 делеционных линий (5BS6, 5BS5, 5BS1, 5BS8, 5BS4, 5BS2 и 5BS3) (рис. 2; Доп. материалы 3). В качестве контроля использовали сорт мягкой пшеницы CS и нулли-тетрасомная линия CSN5BT5A (хромосома 5В замещена на 5А).

На рис. 2 представлено расположение точек разрывов у 7 делеционных линий короткого плеча хромосомы 5В. Для примера, в делеционной линии 5BS6 отсутствует дистальный участок хромосомы до точки разрыва 5BS6 (делеционный бин 5BS6), а у делеционной линии 5BS5 делеция крупнее и распространяется до точки разрыва 5BS5 (делеционный бин 5BS5). Анализ серии делеционных линий пшеницы позволяет локализовать маркеры на определенном участке (бине) хромосомы, т.е. провести их физическое картирование. Примеры локализации маркеров в бинах 5BS6, 5BS5, 5BS4 и C-5BS3 приведены на рисунке в Доп. материалах 3.

Всего при сравнительном анализе ПЦР фрагментов CS и линии CSN5BT5A на хромосоме 5В локализован 21 SSR-маркер. Остальные 46 маркеров имели близкие по длине фрагменты амплификации между хромосомами 5-й гомеологической группы. Анализ 21 полиморфного SSR-маркера с серией делеционных линий CS по хромосоме 5В позволил локализовать данные маркеры на физическую карту 5BS. Так, 8 маркеров (Xicgc178, Xicgc178\_2, Xicgc284, Xicgc456, Xicgc14c009, Xicgc15c020, Xicgc16c004\_2, Xicgc16c041) были локализованы в дистальном районе хромосомы (бин 5BS6), 8 маркеров (Xicgc122, Xicgc131, Xicgc229, Xicgc299, Xicgc307, Xicgc342,

## Основные характеристики сборки фрагментов библиотеки BAC-клонов

Характеристика сборки	Все контиги	Контиги длиной более 500 п. н.
Число контигов	17 770	4 780
Общая протяженность, п. н.	25 879 921	18 560 927
Длина наиболее протяженного контига, п. н.	250 809	250 809
N50, п. н.	4 136	7 767
N90, п. н.	504	1 431
N95, п. н.	416	942

Xicgc686, Xicgc1572) – в бине 5BS5, четыре маркера (Xicgc495, Xicgc498, Xicgc69, Xicgc1699) – в бине 5BS4 и Xicgc1988 – в бине C-5BS3 (рис. 2).

## Обсуждение

## Подходы и результативность сборки последовательностей ДНК у злаков

Для сборки генома использовали программу MIRA (Chevreux et al., 1999). С одной стороны, выбор был обусловлен тем, что эта программа достаточно хорошо зарекомендовала себя при сборке бактериальных геномов на основе библиотек, полученных по технологии Ion Torrent (Loman et al., 2012). С другой, эта программа была ранее использована при сборке геномных фрагментов ряда растений, в частности, для сборки BAC-клонов ячменя (Pasquariello et al., 2014), подсолнечника (*Helianthus annuus* L.) (Staton et al., 2012), для сборки фрагментов генома барбадосского ореха (*Jatropha curcas*) (Sato et al., 2011), а также генома пластид гнездовки настоящей (*Neottia nidus-avis*) (Logacheva et al., 2011) (во всех указанных случаях для секвенирования применялась технология Roche 454). В обзоре Штайна и Шторнагеля (Stein, Steuernagel, 2014), посвященном работам в области секвенирования генома ячменя, эта программа характеризуется как очень удачная для анализа относительно небольших объемов данных, полученных на разных платформах (Illumina GAIIx, HiSeq, Roche 454) и содержащих большое количество повторов.

В данной работе нам удалось получить достаточно качественные результаты сборки геномных фрагментов (таблица). Представляется интересным сравнить характеристики сборки с аналогичными данными по секвенированию геномных последовательностей злаковых растений, в особенности родственных пшенице. Анализ последних работ, посвященных секвенированию геномов мягкой пшеницы и родственных видов по технологии «дробовика» (shot-gun sequencing), показал, что характеристики полученной сборки BAC-клонов достаточно хорошо согласуются с результатами других групп. В частности, одной из наиболее информативных характеристик общего качества сборки является параметр N50 (значение, при котором контиги равной и большей длины составляют половину суммарной протяженности сборки). В данной работе она составила около 4 тыс. п. н., а при анализе контигов с длиной более 500 значение этого параметра стало 7 700 п. н. (таблица).

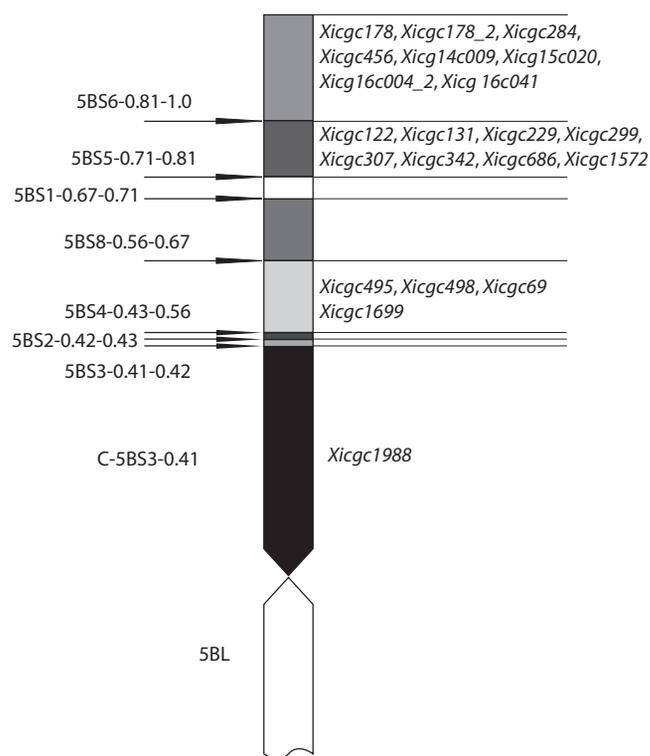


Рис. 2. Локализация SSR-маркеров в делеционных бинах хромосомы 5BS.

Сборка генома мягкой пшеницы на основе секвенирования с низким покрытием (~5×) по технологии 454 (Brenchley et al., 2012) характеризовалась величиной N50 для контигов около 900 п. н. Для сборки генома ячменя при секвенировании методом Illumina GAIIx и 50-кратном покрытии (International Barley Genome Sequencing Consortium, 2012) значение этого параметра для контигов составило 1 425 п. н. Для сборки генома *Aegilops tauschii* при секвенировании методом Roche 454 и 90-кратном покрытии значение N50 для контигов составило ~4 500 п. н. (Jia et al., 2013). Для сборки генома *Triticum urartu* значение N50 для контигов составило 3 400 п. н. при секвенировании методом Illumina HiSeq (2000) и 91-кратном покрытии (Ling et al., 2013). В работе, посвященной секвенированию и сборке последовательностей хромосом мягкой пшеницы с использованием технологии Illumina

и с покрытием между 30× и 241×, значение параметра N50 для разных хромосом находится в пределах 0,5~4,3 kb (International Wheat Genome Sequencing Consortium, 2014). Для недавней сборки генома пшеницы, полученной Чапманом с коллегами, методом Illumina по данным нескольких линий при общем ~175-кратном покрытии, N50 для контигов с длиной более 1 тыс. п. н. определяется значением ~8 300 (Chapman et al., 2015). Таким образом, характеристики сборки, полученные в нашей работе, в целом хорошо согласуются с таковыми для недавних вариантов сборок генома (Jia et al., 2013; Ling et al., 2013; Chapman et al., 2015) и достаточны для решения поставленных в работе задач.

Следует также отметить, что практически со всеми праймерами, разработанными по результатам сборки контигов, были получены хорошие продукты амплификации как на ВАС-клонах, используемых для секвенирования, так и на геномной ДНК пшеницы.

### Особенности организации микросателлитных локусов хромосомы 5В

Микросателлитные последовательности ДНК встречаются с высокой частотой во всех изученных геномах эукариот (Tautz, Renz, 1984; Li et al., 2002). Средняя частота встречаемости микросателлитов в геноме эукариот оценивается как 1 на 10 000 п. н. (Brown et al., 1996). Количество копий SSR может варьировать и достигать 100 и более повторностей в одном локусе (Tautz, Renz, 1984; Li et al., 2002). Первые оценки распределения по хромосомам микросателлитных локусов были сделаны по результатам картирования генома растений и на основании прямой *in situ* гибридизации. При изучении распределения микросателлитов для ряда объектов, например, у сахарной свеклы, методом *in situ* гибридизации было показано, что они группируются в определенных районах хромосом (Schmidt, Heslop-Harrison, 1996). При картировании микросателлитов томатов показано, что они локализируются в прицентромержной области хромосом (Areshchenkova, Ganai, 1999). Однако у большинства растительных видов микросателлитные локусы имеют более равномерное распределение по геному, что способствовало их успешному использованию в качестве маркеров как для генотипирования генома растений, так и в последующем их использовании в маркер-ориентированной селекции (Mason, 2015; <http://maswheat.ucdavis.edu/protocols/>).

Анализ результатов секвенирования выборки ВАС-клонов позволил дополнительно разработать 67 маркеров для короткого плеча хромосомы 5В, на котором расположен ряд генов, влияющих на формирование количественных признаков мягкой пшеницы и устойчивость к абиотическим и биотическим факторам внешней среды (<http://www.shigen.nig.ac.jp/wheat/komugi/genes/symbolClassList.jsp>). Примечательно, что 21 маркер выявил фрагменты амплификации, специфичные для короткого плеча хромосомы 5В, что позволило локализовать их на физической карте хромосомы 5В (в делеционных бинах). Наибольшее число маркеров из числа локализованных на хромосоме (76 %) присутствовало в дистальном участке хромосомы (в бинах 5BS6, 5BS5) и только один маркер – в прицентромержной бине хромосомы 5BS (рис. 2; Доп. материалы 2). Ана-

логичные результаты по распределению маркеров были получены нами ранее при локализации SSR-маркеров из базы данных по пшенице на хромосому 5В (Тимонова и др., 2013). Однако, в отличие от ранее полученных результатов, в настоящей работе 19 % маркеров было локализовано в интерстициальном бине 5BS4. Какой-либо зависимости от структуры микросателлитного локуса и его локализации на хромосоме по данным проведенного картирования выявлено не было (Доп. материалы 2). В то же время прямая гибридизация различных групп микросателлитов на хромосомы указывает на преимущественную их локализацию в прицентромержном районе (Cuadrado et al., 2000, 2008). Обратная закономерность, выявленная нами при физическом картировании В-специфичных хромосомных локусов, указывает на то, что, по-видимому, прицентромержные районы достаточно консервативны у гомеологичных хромосом, в том числе и у 5-й группы хромосом пшеницы, а отсутствие полиморфизма не позволило локализовать большинство (70 %) разработанных SSR-маркеров на хромосоме 5BS.

Распространенность различных микросателлитов в целом по геному в пределах одного вида зависит от структуры повторяющейся единицы и от ее протяженности. Так, анализ результатов секвенирования отдельных хромосом или всего генома ряда злаковых и кукурузы (Zhang et al., 2007; Qu, Liu, 2013; Sergeeva et al., 2014) указывает на то, что наиболее часто встречаются динуклеотидные повторы, затем частота встречаемости убывает при увеличении длины повторяющегося мотива. Полученные нами результаты по анализу микросателлитных локусов, по данным секвенирования выборки из 130 ВАС-клонов, также указывают на более высокую распространенность динуклеотидных последовательностей (253 последовательности), по сравнению с три- (156) и тетра-нуклеотидными (27) повторяющимися последовательностями ДНК. Наблюдаемая тенденция в распространенности повторяющихся мотивов, в зависимости от длины, соответствует ожидаемому распределению мотива, обратно пропорциональному его длине. Встречаемость ди-, три- и тетра-повторов варьирует также и в зависимости от участка генома. Оценка распределения ди- и тринуклеотидных повторов в транслируемых и нетранслируемых районах генных локусов показала, что встречающиеся в кодирующих районах генов микросателлиты более чем на 87 % состоят из тринуклеотидов, причем это характерно, например, как для риса, так и для кукурузы (Zhang et al., 2007; Qu, Liu, 2013). Это связано с тем, что увеличение копийности в тринуклеотидном повторе не приводит к сдвигу рамки считывания, и в ряде случаев может поддерживаться естественным отбором. Межвидовые различия у злаков заключаются в распространенности разных тринуклеотидных повторов. Так, по данным анализа последовательностей ДНК всего генома кукурузы, выявлено, что наиболее распространенный тринуклеотидный повтор имеет структуру (AGC)<sub>n</sub>, далее следуют повторы (ACG)<sub>n</sub>, (CCG)<sub>n</sub>, (ATC)<sub>n</sub>, (AAG)<sub>n</sub> с частотой встречаемости от 2,81 до 2,48 % от общего числа ди- и тринуклеотидных повторов (Qu, Liu, 2013). У риса распределение тринуклеотидных повторов по степени встречаемости несколько иное (Zhang et al., 2007). Наиболее часто встречаются

(CCG)<sub>n</sub>, затем (AGG)<sub>n</sub>, (AGC)<sub>n</sub>, (ACG)<sub>n</sub>, (AAG)<sub>n</sub>. Основные различия между видами наблюдаются по двум группам микросателлитов, а именно у кукурузы повторы (ATC)<sub>n</sub> встречаются часто, а (AGG)<sub>n</sub> – существенно реже, в то время как у риса наоборот. Общей характеристикой изученных геномов злаков (рис, кукуруза) является широкая представленность (AGC)<sub>n</sub>-, (ACG)<sub>n</sub>-, (CCG)<sub>n</sub>-, (AAG)<sub>n</sub>-повторов. Следует отметить, что в перечисленные выше группы повторов не входят тринуклеотидные мотивы, относящиеся к стоп-кодонам ядерного генома эукариот.

У пшеницы и ее предшественников широкое распространение в геноме получил (AAG)<sub>n</sub>-повтор. Это подтверждается как полученными результатами (Доп. материалы 2), так и ранее проведенными исследованиями распределения тринуклеотидных микросателлитов методом прямой локализации на хромосомы олигонуклеотидов (Cuadrado et al., 2000, 2008; Adonina et al., 2015). Другой особенностью мягкой пшеницы, и преимущественно В-генома и его предшественника, является широкая представленность (AGG)<sub>n</sub>-повтора (Доп. материалы 2; Cuadrado et al., 2008). (AGG)<sub>n</sub> микросателлит широко встречается и у риса (Zhang et al., 2007). Такая мозаичная пролиферация (AGG)<sub>n</sub>-повтора в геномах отдельных видов злаков в процессе эволюции, возможно, происходила посредством мобильных элементов.

Таким образом, сравнительный анализ распределения тринуклеотидных микросателлитов по хромосоме 5В пшеницы и у различных видов злаков указывает на поддержание количественного содержания повтора (AAG)<sub>n</sub> в процессе эволюции злаков, в то время как содержание других тринуклеотидных повторов существенно варьирует у современных видов злаков и их предшественников.

### Благодарности

Авторы благодарят коллектив Центра структурной и функциональной геномики растений и его руководителя Я. Долежеля (Институт экспериментальной ботаники, Оломоуц, Чехия) за предоставление ВАС-клонов.

Работа выполнена при поддержке Минобрнауки России (соглашение № 14.604.21.0106 от 07.07.2014, идентификационный номер RFMEFI60414 X0106).

При выполнении работы были задействованы ЦКП «Биоинформатика» и ЦКП геномных исследований ИЦиГ СО РАН.

### Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

### Список литературы

Тимонова Е.М., Добровольская О.Б., Сергеева Е.М., Бильданова Л.Л., Сурдий П., Фойе К., Салина Е.А. Сравнительное генетическое и цитогенетическое картирование хромосомы 5В пшеницы с использованием интрогрессивных линий. *Генетика*. 2013;49(12):1200-1206.

Akhunov E., Nicolet C., Dvorak J. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor. Appl. Genet.* 2009;119:507-517. DOI 10.1007/s00122-009-1059-5

Adonina I.G., Goncharov N.P., Badaeva E.D., Sergeeva E.M., Petrash N.V., Salina E.A. (GAA)<sub>n</sub> microsatellite as an indicator of the A genome reorganization during wheat evolution and domestication.

*CompCytogen.* 2015;9(4):533-547. DOI 10.3897/CompCytogen.v9i4.5120

Areshchenkova T., Ganai M.W. Long tomato microsatellites are predominantly associated with centromeric regions. *Genome*. 1999;42:536-544.

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(7426):705-710. DOI 10.1038/nature11650

Brown S.M., Szewc-McFadden A.K., Kresovich S. Development and application of simple sequence repeat (SSR) loci for plant genome analysis. *Methods in Genome Analysis in Plants*. Boca Raton: CRC Press, 1996.

Chapman J.A., Mascher M., Buluç A., Barry K., Georganas E., Session A., Strnadova V., Jenkins J., Sehgal S., Olikier L., Schmutz J., Yelick K.A., Scholz U., Waugh R., Poland J.A., Muehlbauer G.J., Stein N., Rokhsar D.S. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology*. 2015;16(1):26. DOI 10.1186/s13059-015-0582-8

Chevreaux B., Wetter T., Suhai S. Genome sequence assembly using trace signals and additional sequence information. *Computer science and biology: Proc. of the German Conference on Bioinformatics*. 1999:45-56.

Cuadrado A., Schwarzacher T., Jouve N. Identification of different chromatin classes in wheat using *in situ* hybridization with simple sequence repeat oligonucleotides. *Theor. Appl. Genet.* 2000;101:711-717. DOI 10.1007/s001220051535

Cuadrado A., Cardoso M., Jouve N. Increasing the physical markers of wheat chromosomes using SSRs as FISH probes. *Genome*. 2008;51(10):809-815. DOI 10.1139/G08-065

Endo T.R., Gill B.S. The deletion stocks of common wheat. *J. Hered.* 1996;87(4):295-307.

Feldman M. The origin of cultivated wheat. *The World Wheat Book*. Paris: Lavoisier Publishing, 2001.

Gusev V.D., Miroshnichenko L.A., Chuzhanova N.A. The detection of fractal-like structures in DNA sequences. *Information science and computing*. Int. Book Series, No. 8: Classification, forecasting, data mining. Sofia: ITHEA, 2009.

Gusev V.D., Nemytikova L.A., Chuzhanova N.A. On the complexity measures of genetic sequences. *Bioinformatics*. 1999;15(12):994-999. DOI 10.1093/bioinformatics/15.12.994

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

International Wheat Genome Sequencing Consortium. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*. 2014;345(6194):1251788. DOI 10.1126/science.1251788

Jia J., Zhao S., Kong X., Li Y., Zhao G., He W., Appels R., Pfeifer M., Tao Y., Zhang X., Jing R., Zhang C., Ma Y., Gao L., Gao C., Spannagl M., Mayer K.F.X., Li D., Pan S., Zheng F., Hu Q., Xia X., Li J., Liang Q., Chen J., Wicker T., Gou C., Kuang H., He G., Luo Y., Keller B., Xia Q., Lu P., Wang J., Zou H., Zhang R., Xu J., Gao J., Middleton C., Quan Z., Liu G., Wang J., International Wheat Genome Sequencing Consortium; Yang H., Liu X., He Z., Mao L., Wang J. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*. 2013;496(7443):91-95. DOI 10.1038/nature12028

Langmead B., Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012;9:357-359.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R.; 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. 2009;25:2078-2079.

- Li Y.-C., Korol A.B., Beiles A., Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.* 2002;11:2453-2465. DOI 10.1046/j.1365-294X.2002.01643.x
- Ling H.-Q., Zhao S., Liu D., Wang J., Sun H., Zhang C., Fan H., Li D., Dong L., Tao Y., Gao C., Wu H., Li Y., Cui Y., Guo X., Zheng S., Wang B., Yu K., Liang Q., Yang W., Lou X., Chen J., Feng M., Ji-an J., **Zhang X., Luo G., Jiang Y., Liu J., Wang Z., Sha Y., Zhang B.,** Wu H., Tang D., Shen Q., Xue P., Zou S., Wang X., Liu X., Wang F., Yang Y., An X., Dong Z., Zhang K., Zhang X., Luo M.-C., Dvorak J., Tong Y., Wang J., Yang H., Li Z., Wang D., Zhang A., Wang J. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*. 2013;496(7443):87-90. DOI 10.1038/nature11997
- Logacheva M.D., Schelkunov M.I., Penin A.A. Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. *Genome Biol. Evol.* 2011;3:1296-1303. DOI 10.1093/gbe/evr102
- Loman N.J., Misra R.V., Dallman T.J., Constantinidou C., Gharbia S.E., Wain J., Pallen M.J. Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* 2012;30(5):434-439. DOI 10.1038/nbt.2198
- Mason A.S. SSR genotyping. *Methods Mol. Biol.* 2015;1245:77-89. DOI 10.1007/978-1-4939-1966-6\_6
- Pasquariello M., Barabaschi D., Himmelmach A., Steuernagel B., Ariyadasa R., Stein N., Gandolfi F., Tenedini E., Bernardis I., Tagliafico E., **Pecchioni N., Francia E.** The barley Frost resistance-H2 locus. *Funct. Integr. Genomic.* 2014;14(1):85-100. DOI 10.1007/s10142-014-0360-9
- Paux E., Sourdille P., Salse J., Sautenac 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., Doležel J., Bernard M., Feuillet C.** Physical map of the 1-Gigabase bread wheat chromosome 3B. *Science*. 2008;322:101-104. DOI 10.1126/science.1161847
- Plaschke J., Ganai M.W., Röder M.S. Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor. Appl. Genet.* 1995;91:1001-1007. DOI 10.1007/BF00223912
- Plaschke J., Börner A., Wendehake K., Ganai M.W., Röder M.S. The use of wheat aneuploids for the assignment of microsatellite loci. *Euphytica*. 1996;89:33-40. DOI 10.1007/BF00015716
- Sato S., Hirakawa H., Isobe S., Fukai E., Watanabe A., Kato M., Kawashima K., Minami C., Muraki A., Nakazaki N., Takahashi C., Nakayama S., Kishida Y., Kohara M., Yamada M., Tsuruoka H., Sasamoto S., Tabata S., Aizu T., Toyoda A., Shin-i T., Minakuchi Y., Kohara Y., Fujiyama A., Tsuchimoto S., Kajiyama S., Makigano E., Ohmido N., Shibagaki N., Cartagena J.A., Wada N., Kohinata T., Atefeh A., Yuasa S., Matsunaga S., Fukui K. Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Res.* 2011;18(1):65-76. DOI 10.1093/dnares/dsq030
- Sears E.R. Nullisomic-tetrasomic combinations in hexaploid wheat. Chromosome manipulations and Plant Genetics. London: Oliver and Boyd, 1966. DOI 10.1007/978-1-4899-6561-5\_4
- Sergeeva E.M., Afonnikov D.A., Koltunova M.K., Gusev V.D., Miroshnichenko L.A., Vrána J., Kubaláková M., Poncet C., Sourdille P., Feuillet C., Doležel J., Salina E.A. Common wheat chromosome 5B composition analysis using low-coverage 454 sequencing. *Plant Genome*. 2014;7(2):1-16. DOI 10.3835/plantgenome2013.10.0031
- Schmidt T., Heslop-Harrison J.S. The physical and genomic organization of microsatellites in sugar beet. *Proc. Natl Acad. Sci. USA*. 1996;93:8761-8765.
- Staton S.E., Bakken B.H., Blackman B.K., Chapman M.A., Kane N.C., Tang S., Ungerer M.C., Knapp S.J., Rieseberg L.H., Burke J.M. The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* 2012;72(1):142-153. DOI 10.1111/j.1365-313X.2012.05072.x
- Stein N., Steuernagel B. Advances in sequencing the barley genome. Genomics of plant genetic resources. Springer Netherlands, 2014. DOI 10.1007/978-94-007-7572-5\_16
- Sourdille P., Singh S., Cadalen T., Brown-Guedira G.L., Gay G., Qi L., Gill B.S., Dufour P., Murigneux A., Bernard M. Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics*. 2004;4:12-25. DOI 10.1007/s10142-004-0106-1
- Tautz D., Renz M. Simple sequences are ubiquitous repetitive component of eukaryotic genomes. *Nucl. Acid. Res.* 1984;12:4127-4138. DOI: 10.1093/nar/12.10.4127
- Qi L., Echalié B., Friebe B., Gill B.S. Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. *Funct. Integr. Genomics*. 2003;3:39-55. DOI 10.1007/s10142-002-0063-5
- Qu J., Liu J. A genome-wide analysis of simple sequence repeats in maize and the development of polymorphism markers from next-generation sequence data. *BMC Res. Notes*. 2013;6:403. DOI 10.1186/1756-0500-6-403
- Zhang Z., Deng Y., Tan J., Hu S., Yu J., Xue Q. A genome-wide microsatellite polymorphism database for the indica and japonica rice. *DNA Res.* 2007;14:37-45. DOI 10.1093/dnares/dsm005