

DOI: 10.25205/978-5-4437-1691-6-14

НА ШАГ БЛИЖЕ К ПАНГЕНОМУ: ПОИСК КРУПНОМАСШТАБНЫХ ИНТРОГРЕССИЙ В ГЕНОМАХ РОССИЙСКИХ СОРТОВ МЯГКОЙ ПШЕНИЦЫ И ИДЕНТИФИКАЦИЯ ПОТЕНЦИАЛЬНЫХ ДОНОРОВ

TOWARD A PANGENOME: SEARCHING FOR LARGE-SCALE INTROGRESSIONS IN THE GENOMES OF RUSSIAN BREAD WHEAT VARIETIES, AND IDENTIFICATION OF PUTATIVE DONORS

А. В. Васильев¹, А. С. Ермолаев¹, Д. С. Ульянов¹, В. С. Воронежская^{1,2},
Н. М. Съедина¹, С. В. Тошчаков⁴, И. Ф. Лапочкина⁵, М. Г. Дивашук^{1,5}

¹Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии, Москва

²Всероссийский научно-исследовательский институт радиологии и агроэкологии,
НИЦ «Курчатовский институт», Обнинск

³Федеральный исследовательский центр «Немчиновка», Москва

⁴Курчатовский геномный центр, НИЦ «Курчатовский институт», Москва

A. V. Vasiliev¹, A. S. Ermolaev¹, D. S. Ulyanov¹, V. S. Voronezhskaya^{1,2},
N. M. Syedina¹, S. V. Toshchakov⁴, I. F. Lapochkina⁵, M. G. Divashuk^{1,5}

¹All-Russia Research Institute of Agricultural Biotechnology, Moscow

²All-Russia Research Institute of Radiology and Agroecology, National Research Centre “Kurchatov Institute”, Moscow

³Federal Research Center “Nemchinovka”, Moscow

⁴Kurchatov Genomic Center, National Research Center “Kurchatov Institute”, Moscow

✉ Va01007@mail.ru

Аннотация

В работе был проведен биоинформатический анализ коллекции образцов мягкой пшеницы (*Triticum aestivum*) Арсенал для поиска крупных геномных перестроек и идентификации предполагаемого донора интрогрессий. Предполагаемый вид — источник интрогрессий был идентифицирован с помощью соревновательного картирования (competitive mapping) на синтетические аллополиплоидные сборки (САС), собранные из сборки генома мягкой пшеницы и предполагаемых доноров.

Abstract

A bioinformatic analysis of the bread wheat (*Triticum aestivum*) sample collection Arsenal was conducted to search for large genomic rearrangements and to identify the putative donor of the introgression. The presumed source species of the introgressions was identified using competitive mapping on synthetic allopolyploid assemblies (SAA) compiled from the bread wheat genome assembly and the presumed donors.

Поиск полезных интрогрессий в геноме мягкой пшеницы, полученных из дикой природы, — один из путей повышения генетического разнообразия для последующей селекции. Так как интрогрессии и инсерции приносят генетический материал из видов вторичного или третичного пула, они могут нести гены, критически важные для сельского хозяйства. Гибридизация мягкой пшеницы имеет ряд ограничений и часто невозможна со многими представителями своего семейства. Этот фактор делает поиск интрогрессий из других родов и видов ключевым способом внедрения новых ценных генов, наряду со скрещиванием через посредника. Гены, найденные в интрогрессиях, могут быть включены в пангеном мягкой пшеницы, так как были получены путем опыления с близкородственным организмом и последующей рекомбинации. В работе мы используем метод конкурентного картирования, который хорошо себя зарекомендовал и уже использовался в исследованиях, посвященных человеку, животным и метагеномике, но не применялся еще к растениям.

Материалы и методы. В работе были использован генетический материал, полученный генотипированием секвенированием (Genotyping-by-sequencing, GBS), выделенный из линий коллекции Арсенал. Коллекция была составлена из растений, которые скрещивали с представителями рода *Aegilops* (*Ae. speltoides*, *Ae. triuncialis*), а также с *T. kiharae* в качестве опылителей. В рамках работы нами был проведен углубленный поиск интрогрессий с использованием всех доступных в GenBank сборок геномов первичного, вторичного и третичного пулов генов

мягкой пшеницы, собранных до хромосом или выше [1]. Для того чтобы не только найти позиции интрогрессии, но и идентифицировать возможного донора, был использован подход соревновательного картирования (competitive mapping) [2]. Подход заключается в выравнивании генетического материала на САС, составленную из геномов реципиента и предполагаемого донора. Таким образом, картирование ридов будет происходить на две сборки одновременно. Всего было использовано 19 САС, состоящих из геномной сборки мягкой пшеницы (Chinese Spring v2.1) с видами — донорами H, U, G, S, A, V, E, R, D и C геномов. Результаты картирования позволили идентифицировать не только зону интрогрессии, но и предполагаемого донора (рис. 1). Преимуществом соревновательного картирования является более четкое определение границ коротких интрогрессий за счет картирования ридов с этих участков в более «подходящие» места на геноме-доноре (рис. 2). Так как программы

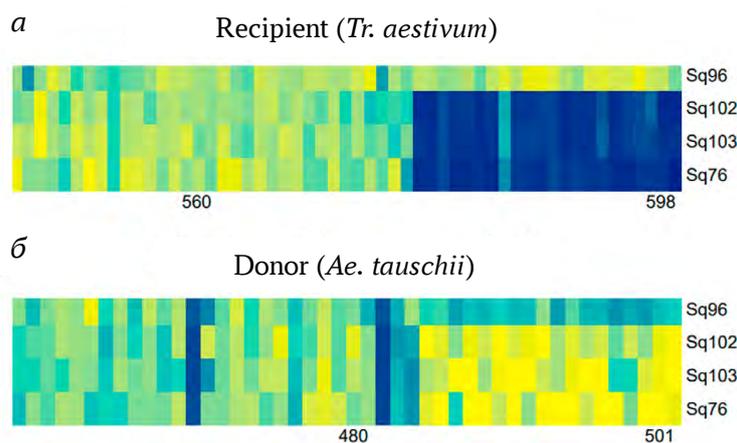


Рис. 1. Участок тепловой карты результатов картирования генетического материала на гибридную сборку *Triticum aestivum* и *Aegilops tauschii*:
а — участок на реципиенте, б — участок на доноре

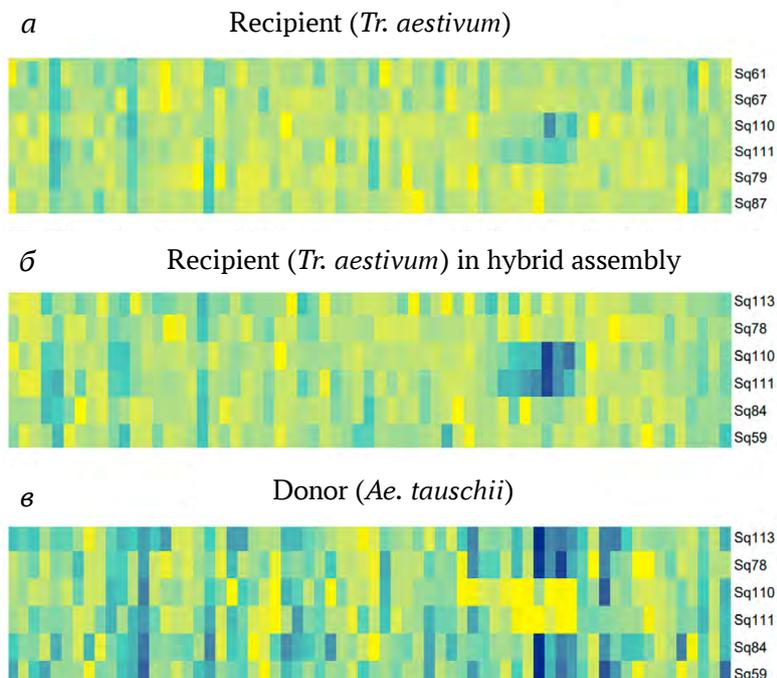


Рис. 2. Участок тепловой карты результатов картирования генетического материала на гибридную сборку *Triticum aestivum* и *Aegilops tauschii* и обычную сборку *Triticum aestivum*: а — участок мягкой пшеницы на обычной сборке, б — участок мягкой пшеницы на гибридной сборке, в — участок на доноре в составе гибридной сборки

для картирования коротких ридов на геномный референс настроены на выравнивании генетического материала с максимальным соответствием, то возможно определить участки, которые более специфично картируются на геном донора, чем на геном мягкой пшеницы. Такие участки будут характеризоваться повышенным покрытием на геноме-доноре и пониженным покрытием на геноме-реципиенте (см. рис. 2).

Для определения зон интрогрессий количество картированных ридов было посчитано в неперекрывающихся окнах генома размером 1 Mbp и нормализовано на общее количество картированных ридов с помощью скрипта на Python. Для визуализации результаты картирования были визуализированы в виде хитмапов, где была проведена дополнительная нормализация среди одинаковых окон (совпадающие хромосома и координаты) у разных образцов. Анализ тепловых карт с целью поиска интрогрессий, инсерций, дубликации и предполагаемых делеций проводился визуально.

Результаты. Анализ тепловых карт картирования генетического материала коллекции Арсенал на основе 19 САС позволил обнаружить 5 инсерций, 13 дубликаций, 19 предполагаемых делеций и 126 интрогрессий. Размеры обнаруженных интрогрессий составляли от 3 до 802 Mbp (почти целая хромосома). На основании существующих аннотаций регион хромосомных перестроек был функционально аннотирован. По итогам работы гены, представленные в интрогрессиях и инсерциях, могут быть использованы в создании пангенома пшеницы мягкой пшеницы. Несмотря на простоту реализации, существенным недостатком данного метода является низкое разрешение, что позволяет находить только большие интрогрессии, а также для близкородственных геномов (например, для семейства S-геномов) невозможна однозначная идентификация вида-донора.

Литература

1. Keilwagen J., Lehnert, H., Berner, T. et al. Detecting major introgressions in wheat and their putative origins using coverage analysis // *Sci Rep.* 2022. Vol. 12. P. 1908.
2. Feuerborn T. R., Palkopoulou E., van der Valk T. et al. Competitive mapping allows for the identification and exclusion of human DNA contamination in ancient faunal genomic datasets // *BMC Genomics.* 2020. Vol. 21. P. 844.