

УДК 519.876.5

СЛЕДЫ ЭКОЛОГИЧЕСКОЙ АДАПТАЦИИ В ГЕНОМАХ СТАРОМЕСТНЫХ СОРТОВ НУТА

© 2020 г. А.Б. Соколкова*, П.Л. Чанг**, Н. Карраскила-Гарсия**, Н.В. Нуждина***, Д.Р. Кук**, С.В. Нуждин*, ****, М.Г. Самсонова*

*Санкт-Петербургский политехнический университет Петра Великого, 195251, Санкт-Петербург, Политехническая ул., 29

**Факультет патологии растений Калифорнийского университета в Дэвисе, Дэвис, Калифорния 95616 США

***Факультет географии Калифорнийского университета в Лос-Анджелесе, Лос-Анджелес, Калифорния 90095 США

****Университет Южной Калифорнии, СА 90089, Лос-Анджелес, США

E-mail: m.samsonova@spbstu.ru

Поступила в редакцию 04.12.2019 г.

После доработки 04.12.2019 г.

Принята к публикации 22.01.2020 г.

Банк семян Всероссийского института генетических ресурсов растений имени Н.И. Вавилова (ВИР) содержит многочисленные местные сорта, собранные почти сто лет назад, которые могут нести следы естественного отбора, действовавшего в ходе диверсификации сельскохозяйственных культур. В работе проведен анализ 407 местных сортов, собранных из основных исторических центров выращивания нута и вторичной диверсификации. Предложена гипотеза о том, что некоторые однонуклеотидные полиморфизмы связаны с сильным отбором к условиям окружающей среды, которые испытывал нут при одомашнивании и последующем географическом распространении. С помощью пакета BayPass определены 13 полиморфизмов, которые группируются по условиям среды и являются сильными кандидатами как маркеры локальной адаптации.

Ключевые слова: нут, биоклиматический анализ, генотипирование путем секвенирования, однонуклеотидные полиморфизмы, неравновесное сцепление.

DOI: 10.31857/S0006302920020088

В начале XX века (1911–1940) под руководством Н.И. Вавилова были предприняты систематические усилия по сбору и сохранению разнообразия культур, хранящихся в настоящее время в коллекции Всероссийского института генетических ресурсов растений имени Н.И. Вавилова (ВИР) в Санкт-Петербурге [1]. Географический разброс и генетическое разнообразие большинства культур, собранных в тот период, вероятно, должны отражать исторические условия их выращивания, сложившиеся в предшествующие тысячелетия. Особенно интересной представляется возможность идентифицировать следы в геноме исторического отбора на адаптацию к различным условиям окружающей среды [2].

В работе исследована такая возможность на данных для нута (*Cicer arietinum*), выращивание

которого в России в настоящее время быстро растет. Изначально нут был одомашнен примерно 10 тыс. лет назад в регионе Плодородного полумесяца (Ближний Восток) и затем распространился в Индию (~6000 лет назад), а также в Эфиопию и Северную Африку (~3000 лет назад) [3]. В ходе вторичной диверсификации нут стали культивировать в ряде новых сред, климатических условий и методов выращивания, что, по всей видимости, стало возможным благодаря селекции сегрегирующих в популяции мутаций или возникновению новых мутаций. Для проверки данной гипотезы были использованы методы биогеографии, геномики и вычислительной биологии, с помощью которых были идентифицированы конкретные гаплотипы, потенциально реагирующие на отбор. Полученные результаты могут быть использованы в будущем для улучшения сельскохозяйственных культур.

Сокращения: ОНП – однонуклеотидные полиморфизмы, SMS – оценка синтетической морфологии (Synthetic Morphology Score), BF – фактор Байеса (Bayes factor).

Таблица 1. Список биоклиматических ковариат и их сокращенные обозначения

Биоклиматическая ковариата	Сокращение
Среднегодовая температура, °С * 10	BIO ₁
Средняя разница между максимальной и минимальной температурами/день, °С * 10	BIO ₂
Изотермальность, %	BIO ₃
Температурная сезонность, стандартное отклонение * 100	BIO ₄
Макс. температура самого теплого месяца, °С * 10	BIO ₅
Мин. температура самого холодного месяца, °С * 10	BIO ₆
Среднегодовой диапазон температуры, °С * 10	BIO ₇
Средняя температура самого влажного квартала, °С * 10	BIO ₈
Средняя температура самого теплого квартала, °С * 10	BIO ₁₀
Средняя температура самого холодного влажного квартала, °С * 10	BIO ₁₁
Осадки за год, мм	BIO ₁₂
Осадки самого влажного месяца, мм	BIO ₁₃
Осадки самого сухого месяца, мм	BIO ₁₄
Сезонность осадков, мм	BIO ₁₅
Осадки самого влажного квартала, мм	BIO ₁₆
Осадки самого сухого квартала, мм	BIO ₁₇
Осадки самого теплого квартала, мм	BIO ₁₈
Осадки самого холодного квартала, мм	BIO ₁₉
Цифровая модель рельефа, м	DEM

МАТЕРИАЛЫ И МЕТОДЫ

Генетический материал и геномные данные.

Данные были получены из растений, произрастающих в таких странах, как Эфиопия, Ливан, Марокко, Турция, Индия и в более широких регионах Центральной Азии и Средиземноморья. Данные генотипирования путем секвенирования и однонуклеотидные полиморфизмы (ОНП) были взяты из работы [4]. Все данные Illumina доступны в базе данных Национального центра биотехнологий, код BioProject PRJNA388691. ОНП были выделены с использованием GATK pipeline [5] и далее отфильтрованы с помощью VCFtools [6]. 2579 ОНП прошли все фильтры, 407 образцов остались для дальнейшего анализа.

Генетический и биоклиматический анализ. Пакет BayPass [7] был применен для идентификации

генетических маркеров, связанных с биоклиматическими переменными, специфичными для конкретной местности. Для каждой биоклиматической переменной (см. табл. 1) вычислялись средние значения отдельно для каждой из шести географически различных групп. Поскольку многие биоклиматические переменные сильно коррелированы, средние значения каждой из пяти групп сильно коррелированных переменных заменяли оценкой синтетической морфологии (Synthetic Morphology Score, SMS), которая является первой главной компонентой данных (табл. 2). Три биоклиматические переменные не были сильно коррелированы с любыми другими переменными, поэтому их средние значения использовались в анализе без преобразований. Таким образом, были построены восемь перемен-

Таблица 2. Оценки синтетической морфологии, вычисленные на основании биоклиматических ковариат

Оценки синтетической морфологии	Биоклиматические переменные
SMS1	BIO ₁ , BIO ₃ , BIO ₆ , BIO ₈ , BIO ₁₁ , BIO ₁₅
SMS2	BIO ₁₂ , BIO ₁₃ , BIO ₁₆ , BIO ₁₈
SMS3	BIO ₅ , BIO ₁₀



Анализ ассоциаций с помощью BayPass. ОНП со значением BF > 15 показаны треугольниками. Когда одна позиция ассоциируется с рядом биоклиматических ковариат с разным BF, представлен только наиболее значимый ОНП. Внизу манхэттенского графика показана плотность распределения ОНП вдоль хромосом.

ных для исследования ассоциации с генетическими признаками с помощью BayPass. Анализ проводили в предположении о независимости переменных. Каждый ОНП оценивали с помощью алгоритма выборки по значимости [7], который вычисляет значение фактора Байеса (Bayes Factor, BF), эмпирическое значение P -value и соответствующие регрессионные коэффициенты. Ассоциации считались значимыми при эмпирическом значении P -value больше четырех. Чтобы количественно оценить силу ассоциации ОНП с переменными, была использована шкала оценки фактора в соответствии с правилом Джеффриса [8]: «очень убедительное доказательство» при $15 < BF < 20$ и «решающее доказательство» при $BF > 20$. Манхэттенские графики были выполнены в R с использованием библиотеки CMplot [9].

РЕЗУЛЬТАТЫ

407 образцов были разделены на шесть отдельных групп, отражающих географическое положение: Эфиопия («ETH»), Индия («IND»), Ливан («LEB»), Марокко («MOR»), Турция («TUR») и Центральная Азия («UZB»). Тест Манна–Уитни–Уилкоксона [10] использовали для выявления различий между группами по каждой биоклиматической ковариате.

Был проведен анализ ассоциаций найденных ранее 2579 ОНП с градиентами окружающей среды, с использованием пакета BayPass [7]. Этот пакет может идентифицировать генетические маркеры, связанные с популяционными переменными, такими как биоклиматические переменные в местах сбора. В соответствии с рекомендациями авторов пакета BayPass для каждой из шести географически различных групп были построены восемь ковариат: пять оценок синтетической мор-

фологии (табл. 1 и 2) и три усредненные некоррелированные ковариаты (BIO_{2_av}, BIO_{19_av}, DEM_av). Каждая SMS представляет собой линейную комбинацию усредненных значений коррелированных биоклиматических ковариат, которая имеет максимально возможную дисперсию. SMS1, SMS3 и SMS5 соответствуют биоклиматическим ковариатам температуры, тогда как SMS2 и SMS4 соответствуют биоклиматическим переменным осадков. Для оценки значимости ассоциации были использованы следующие критерии: эмпирическое байесовское P -value больше четырех и (согласно правилу Джеффриса [8]) $BF > 15$.

Анализ ассоциаций с помощью BayPass выявил тринадцать ОНП, достоверно связанных с ковариатами (рис. 1). Восемь ОНП на хромосоме 1 и один ОНП на хромосоме 2 были связаны с переменной BIO_{2_av}, представляющей среднюю разницу между максимальной и минимальной температурами в один день. Эти восемь ОНП на хромосоме 1 локализованы в области длиной 115 т.п.н. с сильным неравновесным сцеплением. Только один ОНП был связан с SMS5 (Ca4: 30608189). SMS4 ассоциирован с двумя ОНП на хромосоме 1 (Ca1: 23621481) и хромосоме 3 (Ca3: 36690022). Переменные SMS1 и SMS3, которые включают температурные характеристики, были совместно ассоциированы с одним ОНП на хромосоме 8 (Ca8: 10314452).

ЗАКЛЮЧЕНИЕ

В течение многих тысячелетий фермеры и селекционеры занимались отбором культур с желаемыми фенотипами [11]. Объединяя экологические и геномные данные, теперь можно иденти-

фицировать гаплотипы, отобранные местными фермерами. В этой работе мы идентифицировали 13 однонуклеотидных полиморфизмов, ассоциированных с биоклиматическими переменными в местах сбора образцов и маркирующих районы генома, которые по всей видимости испытали действие отбора, направленного на адаптацию к региональным условиям выращивания при вторичной диверсификации.

ФИНАНСИРОВАНИЕ РАБОТЫ

Биоклиматический и биогеографический анализы, а также анализ с помощью BayPass были выполнены при финансовой поддержке Российского научного фонда (грант № 16-16-00007 (АБС, СВН, МГС)). Работа была также поддержана соглашением о сотрудничестве от Агентства США по международному развитию в рамках программы «Накорми будущее» AID-OAA-A-14-00008 (ДРК, ПЛЧ, СВН), Фондом Зумберджа (СВН) и грантом IOS-1339346 от Программы генома растений Национального научного фонда США (ДРК).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит описания каких-либо исследований с использованием людей и животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. М. А. Вишнякова, М. О. Бурляева, С. В. Булынец и др., С.-х. биология **52** (5), 976 (2017).
2. E. Plekhanova, M. A. Vishnyakova, S. Bulyntsev, et al., *Sci. Rep.* **7**, 4816 (2017).
3. R. J. Redden and J. D. Berger, in *Chickpea Breeding & Management*, Ed. by S. S. Yadav, R. Redden, W. Chen, and B. Sharma (CABI, Wallingford, UK, 2007), pp. 1–13.
4. E. J. von Wettberg, P. L. Chang, A. Greenspan, et al., *Nature Comm.* **9**, Art. Nu. 649 (2018). doi: 10.1038/s41467-018-02867-z.
5. A. McKenna, M. Hanna, E. Banks, et al., *Genome Res.* **20**, 1297 (2010).
6. P. Danecek, A. Auton, G. Abecasis, et al., *Bioinformatics* **27**, 2156 (2011).
7. M. Gautier, *Genetics* **201**, 1555 (2015).
8. H. Jeffreys, *Theory of Probability*, 3rd ed. (Oxford Univ. Press, Oxford, 1961).
9. *CMplot: Circle Manhattan Plot*. URL: <https://github.com/YinLiLin/R-CMplot> [Accessed: 20th June 2018].
10. H. B. Mann and D. R. Whitney, *Ann. Math. Statistics* **18** (1), 50 (1947).
11. N. Maxted, M. E. Dooloo, and B. V. Ford-Lloyd, *Enhancing crop genepool use: capturing wild relative and landrace diversity for crop improvement* (CABI, Oxfordshire, UK, 2016).

Signatures of Ecological Adaptation in Genomes of Chickpea Landraces

A.B. Sokolkova*, P.L. Chang**, N. Carrasquilla-Garcia**, N.V. Nuzhdina***, D.R. Cook**, S.V. Nuzhdin*, ****, and M.G. Samsonova*

*Peter the Great St. Petersburg Polytechnic University, ul. Polytekhnicheskaya 29, St. Petersburg, 195251 Russia

**Faculty of Plant Pathology, University of California, Davis, CA 95616, USA

***Faculty of Geography, University of California, Los Angeles, CA 90095, USA

****University of Southern California, Los Angeles, CA 90089, USA

The seed bank of N.I. Vavilov All Russian Institute of Plant Genetic Resources (VIR) contains a wide range of landraces, collected almost one hundred years ago, in which natural selection might leave signatures through crop diversification. In this study, we analysed 407 landraces, sampled at centres of the chickpea origin and at sites of secondary diversity. We hypothesize that a fraction of single nucleotide polymorphisms might have exhibited strong selection to a range of environmental conditions that chickpea experienced during domestication and subsequent geographic distribution. Applying BayPass package we identified 13 polymorphisms; they assort by environmental conditions and are strong local adaptation candidates.

Keywords: chickpea, bioclimatic analysis, sequencing genotyping, single nucleotide polymorphisms, linkage disequilibrium