

ПОИСК АГРООСТРОВКОВ В ГЕНОМЕ НУТА

© 2021 г. А.Б. Соколкова*, С.В. Булынецв**, П.Л. Чанг***, Н. Карраскила-Гарсия***, Д.Р. Кук***, Э. Веттберг****, М.А. Вишнякова*****, С.В. Нуждин*, *****, М.Г. Самсонова*

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

**Кубанская опытная станция Федерального исследовательского центра «Всероссийский научно-исследовательский институт растениеводства имени Н.И. Вавилова», 352183, Краснодарский край, п. Ботаника, Центральная ул., 2

***Калифорнийский университет в Дэвисе, 95616, Дэвис, США

****Университет Вермонта, 05405, Берлингтон, США

*****Федеральный исследовательский центр «Всероссийский научно-исследовательский институт растениеводства имени Н.И. Вавилова», 190000, Санкт-Петербург, Большая Морская ул., 42–44

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

E-mail: m.samsonova@spbstu.ru

Поступила в редакцию 07.12.2020 г.

После доработки 07.12.2020 г.

Принята к публикации 25.12.2020 г.

Концентрация генов, контролирующих признаки доместикизации в так называемых «островках одомашнивания, или агроостровках» затрудняет селекцию, поэтому информация о том, в каком месте генома локализованы такие островки и с какими признаками доместикизации они ассоциированы, очень важна для дальнейших исследований, направленных на улучшение современных сортов. Для получения информации о таких районах в геномах староместных сортов нута использована программа BayPass, которая оценивает ассоциацию однонуклеотидных полиморфизмов со специфичными для популяций ковариатами, в качестве которых выбраны комплексы признаков синдрома доместикизации одинакового целевого назначения. Районы генома, с которыми эти ковариаты ассоциированы, локализируются на первой, пятой, шестой, седьмой и восьмой хромосомах. Эти области являются потенциальными агроостровками, что косвенно подтверждается ранее полученными данными.

Ключевые слова: нут (*Cicer arietinum* L.), генотипирование путем секвенирования, однонуклеотидные полиморфизмы, геномный анализ ассоциаций, адаптация.

DOI: 10.31857/S0006302921030066

Окультуривание растений началось примерно 10000 лет назад и впоследствии независимо повторялось во многих местах по всему миру. Этот процесс включал выращивание и отбор растений с более крупными съедобными частями и более легким сбором урожая — признаками, которые позволяли производить излишки пищи. Такой комплекс признаков, который принято называть синдромом одомашнивания, часто перекрывается у культур одинакового целевого назначения. Так, например, у зернобобовых синдром одомашнивания включает в себя более крупные семена, уменьшение их растрескиваемости и периода покоя, а также изменение фенологии, т.е. количе-

ственные признаки, контролируемые большим числом генов.

Таким образом, уже на начальных этапах селекции, которые, по-видимому, были непредумышленными, сильный отбор строго определенных фенотипов должен был приводить к эффекту бутылочного горлышка, т.е. к уменьшению генетического разнообразия в окультуриваемой популяции за счет выбора только части аллелей генов из генофонда диких видов [1]. Согласно современным представлениям, окультуривание растений происходило постепенно и обмен генами между доместизируемыми и дикими формами все же был возможен в течение довольно длительного времени [2], что приводило к гомогенизации и препятствовало дивергенции окультуриваемых и диких форм, в то время как селекция усиливала такую дивергенцию [3, 4].

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

Наиболее вероятным результатом этих двух разнонаправленных процессов может быть концентрация генов, контролирующих признаки доместикации на так называемых «островках одомашнивания, или агроостровках», подобных «островкам видообразования» комаров [5]. Эта гипотеза неявно исходит из соображения, что гены синдрома одомашнивания с большей вероятностью сохранятся в популяции, несмотря на обмен генами и рекомбинацию, если они находятся в неравновесном сцеплении и в одном или нескольких местах в геноме. Сильная совместная локализация QTL, влияющих на несколько признаков в нескольких регионах генома, обнаруженная у многих культурных растений, убедительно подтверждает эту точку зрения [6].

Нут — зернобобовое растение, используемое в качестве источника белка особенно широко в странах Азии — не является исключением из этого правила. Как показали наши исследования, некоторые гаплоглобы генома этого растения обогащены однонуклеотидными полиморфизмами (ОНП), ассоциированными с признаками синдрома одомашнивания [7]. Хотя концентрация генов окультуривания в агроостровках является результатом селекции, она часто препятствует дальнейшему улучшению растений, особенно в случае, когда признаки отрицательно скоррелированы. Поэтому информация о том, в каком месте генома локализованы такие островки и с каким комплексом признаков доместикации они ассоциированы, очень важна для дальнейших исследований, направленных на улучшение современных сортов.

Получить информацию о таких районах можно с помощью программы BayPass [8], которая оценивает ассоциацию однонуклеотидных полиморфизмов со специфичными для популяций ковариатами, в качестве которых выбраны комплексы признаков синдрома доместикации одинакового целевого назначения. В этой работе мы применили BayPass к анализу шести популяций нута.

МАТЕРИАЛЫ И МЕТОДЫ

Генетический материал и геномные данные. Генетический материал был получен из староместных образцов нута коллекции ВИР, собранных из основных исторических центров выращивания нута и его вторичной диверсификации: Эфиопии, Ливана, Марокко, Турции, Индии и в более широких регионах Центральной Азии и Средиземноморья. 407 образцов, использованных при анализе, были разделены на шесть отдельных групп, отражающих географическое положение мест сбора образцов: Эфиопия («ETH»), Индия («IND»), Ливан («LEB»), Марокко («MOR»), Турция («TUR») и Центральная Азия («C_ASIA»).

Описание генотипирования путем секвенирования и отбора набора ОНП, использованных в данном исследовании, приведено в работе [7]. Данные секвенирования доступны в базе данных Национального центра биотехнологий, код BioProject PRJNA388691. Поиск ОНП был произведен с помощью программы Genome Analysis Tool Kit (GATK) [9], дальнейшая фильтрация ОНП была выполнена программой VCFtools [10]. Для анализа ассоциаций между фенотипическими признаками и ОНП были отобраны 2579 ОНП, идентифицированных в 407 староместных образцах нута.

Анализ ассоциаций между фенотипическими признаками и однонуклеотидными полиморфизмами. Для анализа ассоциаций между фенотипическими признаками и ОНП был применен пакет BayPass [8]. Данный пакет может идентифицировать ОНП, связанные с фенотипическими признаками на уровне популяций. Для анализа были использованы 16 количественных фенотипических признаков, оценка которых была произведена на Кубанской опытной станции ВИР в 2016 г. (табл. 1). Описание фенотипирования, проведенного в 2016 г., приведено в работе [7]. Для каждого фенотипического признака вычисляли средние значения отдельно для каждой из шести географически различных групп. По рекомендации авторов пакета BayPass [8] предварительно был проведен анализ корреляций между фенотипическими признаками. Коэффициенты корреляции Спирмена были рассчитаны с использованием функции «scor» из библиотеки «Hmisc» в R [11]. Согласно результатам анализа корреляций, одиннадцать признаков были отнесены к одной из четырех групп сильно коррелированных переменных. Средние значения каждой из этих четырех групп заменялись оценкой синтетической морфологии (Synthetic Morphology Score, SMS), которая является первой главной компонентой данных (табл. 2). Остальные пять фенотипических переменных не были сильно коррелированы с любыми другими переменными, поэтому их средние значения использовали в анализе без преобразований. Таким образом, пакетом BayPass [8] был проведен анализ ассоциаций между ОНП и девятью ковариатами фенотипических признаков. Анализ проводили в предположении о независимости переменных. Каждый ОНП оценивали с помощью алгоритма выборки по значимости [8], который вычисляет значение фактора Байеса (Bayes Factor, BF), эмпирическое значение P -value и соответствующие регрессионные коэффициенты. Значимость ассоциации определяли по эмпирическому значению P -value (> 4), а силу ассоциации ОНП с переменными определяли по шкале оценки фактора в соответствии с правилом Джеффриса [12]: «очень убедительное доказательство» при $15 < BF < 20$ и «решающее доказа-

Таблица 1. Фенотипические признаки, измеренные в 2016 г. на Кубанской опытной станции, отобранные для анализа ассоциаций

Фенотипический признак	Единица измерения
Длительность цветения	сутки
Высота прикрепления нижнего боба	см
Время от посева до начала всходов	сутки
Длительность созревания	сутки
Время от начала всходов до начала цветения	сутки
Время от конца цветения до начала созревания	сутки
Масса 1000 семян	г
Ширина боба	мм
Длина боба	мм
Число семян с одного растения	шт
Число бобов с одного растения	шт
Вес семян с одного растения	г
Вес бобов	г
Высота растений	см
Вес растения без бобов	г
Вес всего растения (с бобами)	г

тельство» при $BF > 20$. Далее с помощью базы данных Legume information system [13] была проведена аннотация значимо ассоциированных полиморфизмов. Манхэттенские графики были выполнены в R с использованием библиотеки *CM-plot* [14].

РЕЗУЛЬТАТЫ

Для анализа ассоциаций между ОНП и фенотипическими ковариатами с учетом популяционной структуры были использованы 16 количественных фенотипических признаков, оценка

которых была произведена на Кубанской опытной станции ВИР в 2016 г. (табл. 1). Для оценки ассоциаций на уровне популяций 407 образцов, использованных в анализе, были разделены на шесть отдельных групп, отражающих географическое положение мест сбора образцов: Эфиопия («ETH»), Индия («IND»), Ливан («LEV»), Марокко («MOR»), Турция («TUR») и Центральная Азия («C_ASIA») (рис. 1а).

Для определения групп сильно коррелированных фенотипических признаков для 16 фенотипических признаков был проведен анализ корреляций. Результаты анализа приведены на рис. 1б.

Таблица 2. Оценки синтетической морфологии фенотипических признаков

Оценка синтетической морфологии (SMS)	Фенотипические признаки и коэффициенты корреляции между ними
SMS1	Количество дней цветения – количество дней от конца цветения до начала созревания ($r = -0.76^*$)
SMS2	Масса 1000 семян – ширина боба ($r = 0.74^*$); Масса 1000 семян – длина боба ($r = 0.76^*$); Ширина боба – длина боба ($r = 0.86^*$)
SMS3	Число семян с одного растения – число бобов с одного растения ($r = 0.85^*$); Число семян с одного растения – вес семян с одного растения ($r = 0.74^*$); Число семян с одного растения – вес бобов ($r = 0.60^*$); Число бобов с одного растения – вес семян с одного растения ($r = 0.82^*$); Число бобов с одного растения – вес бобов ($r = 0.77^*$); Вес семян с одного растения – вес бобов ($r = 0.88^*$)
SMS4	Вес растения без бобов – вес всего растения (с бобами) ($r = 0.92^*$)

Примечание. * – $p < 0.0001$.

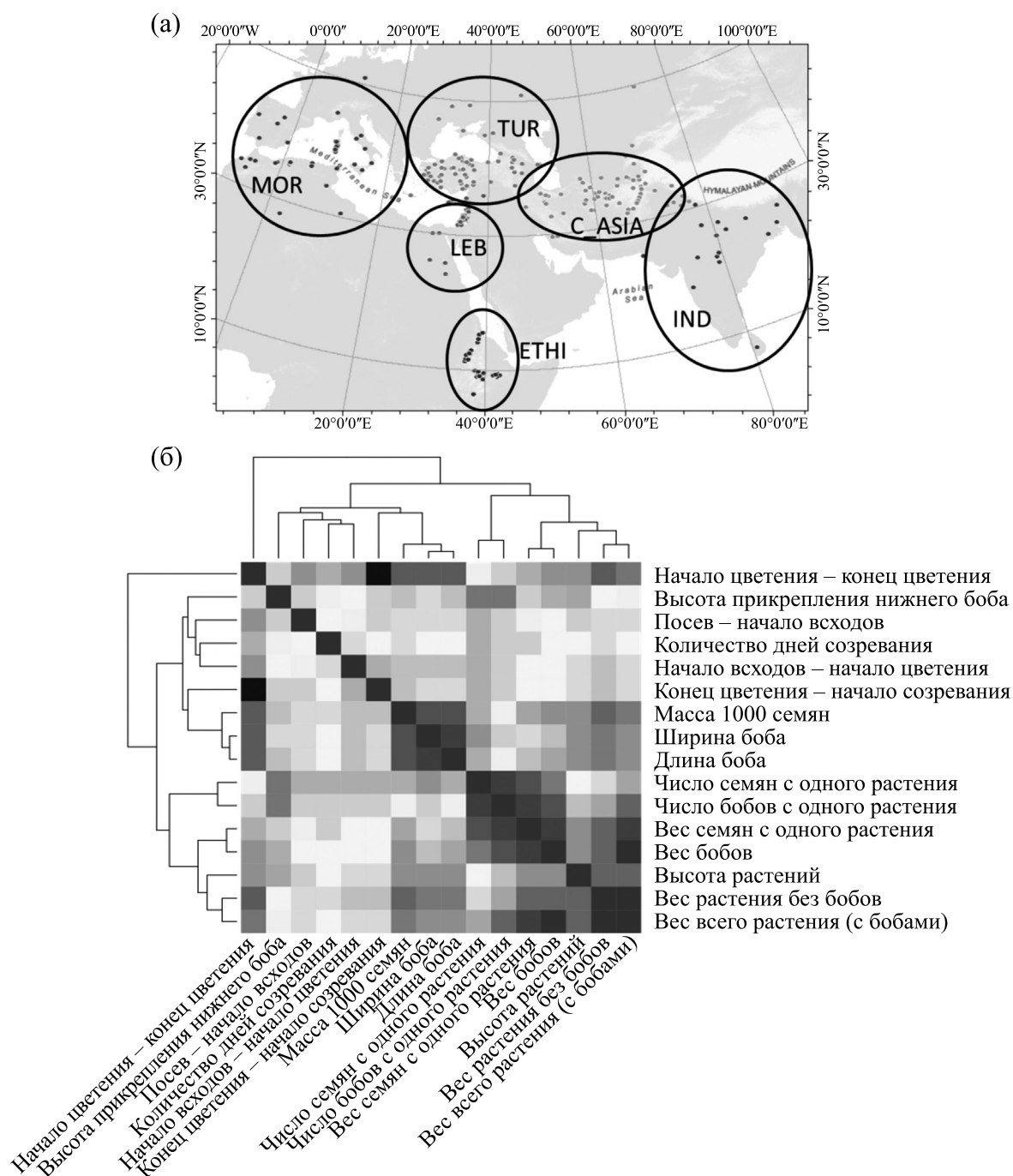


Рис. 1. (а) — Географическое происхождение образцов и их отнесение к одной из шести географических групп; (б) — результат анализа корреляций фенотипических признаков.

Одиннадцать признаков были отнесены к одной из четырех групп сильно коррелированных переменных. В соответствии с рекомендациями авторов пакета *VayPass* для каждой из шести географически различных групп были построены девять ковариат: четыре оценок синтетической морфологии (SMS) (табл. 1 и 2) и пять усредненных некоррелированных ковариат.

Оценка синтетической морфологии SMS1 получена на основе двух последовательных вегетационных периодов, связанных сильной отрицательной корреляционной связью: количества дней цветения и количества дней от конца цветения до начала созревания (табл. 2). Оценки синтетической морфологии SMS2, SMS3 и SMS4 соответствуют фенотипическим признакам, описы-

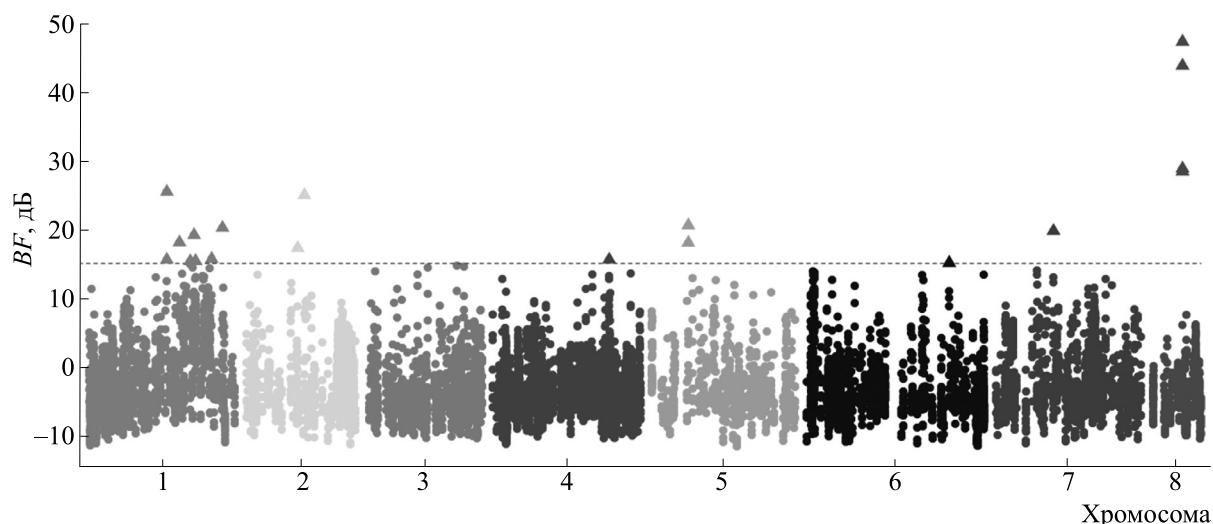


Рис. 2. Анализ ассоциаций с помощью BayPass. ОНП со значением фактора Байеса (BF) > 15 обозначены треугольниками. Когда одна позиция ассоциируется с рядом фенотипических ковариат с разным BF , представлены все ассоциации.

вающим весовые характеристики растений, бобов и семян, размеры бобов и количество бобов и семян с одного растения (табл. 2).

Таким образом, с помощью программного пакета BayPass [8] был проведен анализ ассоциаций между ОНП и девятью ковариатами фенотипических признаков. Для оценки значимости ассоциации были использованы следующие критерии: эмпирическое байесовское P -value > 4 и (согласно правилу Джеффриса [12]) фактор Байеса > 15 .

Анализ ассоциаций с помощью BayPass [8] выявил 14 ОНП, достоверно связанных с фенотипическими ковариатами (рис. 2, табл. 3): семь ОНП на первой хромосоме, два — на второй и по одному ОНП на четвертой, пятой, шестой, седьмой и восьмой хромосомах. Ранее нами в работе [7] был проведен поиск блоков неравновесного сцепления (гапблоков), включающих в себя 2579 ОНП, расположенных на хромосомах, с помощью программы Haploview [15]. Каждый гапблок рассматривался как набор ОНП, расположенных в данном гапблоке, и использовался для аннотации значимо ассоциированных маркеров. Также была проведена аннотация ОНП с помощью базы данных Legume information system [13]. В результате два идентифицированных в анализе ОНП попадают в последовательность известных генов и три ОНП входят в регионы неравновесного сцепления (табл. 3). Два ОНП (на первой и пятой хромосомах) значимо ассоциированы с двумя ковариатами — SMS3 и SMS4. ОНП на восьмой хромосоме (8: 10314452) значимо ассоциирован сразу с четырьмя фенотипическими ковариатами — SMS1, SMS3, SMS4 и ковариатом высоты растений (табл. 3). Данный ОНП был идентифицирован нами и в предыдущих GWAS-

исследованиях (исследованиях геномного анализа ассоциаций): ОНП значимо ассоциирован с весовыми характеристиками растений при фенотипировании на Кубани в 2016 г. [7], с периодом цветения на Кубани в 2017 г. [16]. Кроме того, в работе [17] при поиске программой BayPass [8] ОНП, значимо ассоциированных с биоклиматическими ковариатами в местах сбора образцов, две биоклиматические переменные, включающие температурные характеристики, были совместно ассоциированы с ОНП 8: 10314452. Также данный ОНП находится на расстоянии ~ 25 кб от ОНП, ассоциированного с весом растения в работе [18].

ОБСУЖДЕНИЕ

Концентрация генов, контролирующих признаки доместикиации в так называемых «островках одомашнивания, или агроостровках» затрудняет селекцию, поэтому информация о том, в каком месте генома локализованы такие островки и с какими признаками доместикиации они ассоциированы, очень важна для дальнейших исследований, направленных на улучшение современных сортов. Получить информацию о таких районах можно с помощью программы BayPass [2], которая оценивает ассоциацию ОНП со специфичными для популяций ковариатами, в качестве которых могут быть выбраны комплексы признаков синдрома доместикиации одинакового целевого назначения. В этой работе мы применили BayPass к анализу шести популяций староместных сортов нута из коллекции ВИР. Фенотипические признаки синдрома одомашнивания были сгруппированы в пять групп разного целевого назначения — описывающие фенологию, вес семян, число семян

Таблица 3. ОНП, выявленные по результатам анализа ассоциаций между ОНП и девятью ковариатами фенотипических признаков пакетом BayPass [4]

Позиция	Хромосома	Признак	Ген	Блок неравновесного сцепления (гапоблок)
25737022	1	SMS3	Ca_18581	—
		SMS4		—
29930356		Средняя высота прикрепления нижнего боба	—	—
33512355		Среднее время от начала всходов до начала цветения	—	—
34754596		SMS4	—	Ca1_Block_36
35261112		SMS4	—	Ca1_Block_37
40608784		SMS4	—	—
44150305		Средняя высота прикрепления нижнего боба	—	—
17432752	2	Средняя высота прикрепления нижнего боба	Ca_16006	—
19604640		Средняя высота прикрепления нижнего боба	—	—
38580554	4	Среднее время от начала всходов до начала цветения	—	Ca4_Block_61
12253018	5	SMS3	—	—
		SMS4	—	—
47657340	6	SMS4	—	—
19377547	7	SMS4	—	—
10314452	8	SMS1	—	—
		SMS3	—	—
		SMS4	—	—
		Средняя высота растений	—	—

и характеристики растения. Районы генома, с которыми эти ковариаты ассоциированы, локализируются на первой, пятой, шестой, седьмой и восьмой хромосомах. Эти области являются потенциальными агроостровками, что косвенно подтверждается перекрыванием некоторых из идентифицированных районов с гапблоками, обогащенными ОНП. Интересно, что некоторые из этих районов совпадают также с районами, ассоциированными с одиночными признаками при проведении полногеномного поиска ассоциаций в 2016 и 2017 гг. [7, 16], что является дополнительным указанием на правильность их классификации как агроостровков.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при финансовой поддержке Российского научного фонда (грант № 16-16-00007).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ НОРМ

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

СПИСОК ЛИТЕРАТУРЫ

1. B. L. Gross and K. M. Olsen, Trends Plant Sci. **15**, 529 (2010).
2. T. Pyhäjärvi, M. B. Hufford, S. Mezouk, and J. Ross-Ibarra, Genome Biol. Evol. **5**, 1594 (2013).
3. L. A. F. Frantz, et al., Nat. Genet. **47**, 1141 (2015).
4. S. Abbo and A. Gopher, Trends Plant Sci. **22** (6), 491 (2017). DOI: 10.1016/j.tplants.2017.03.010
5. T. L. Turner, M. W. Hahn, and S. V. Nuzhdin, PLoS Biol. **3**, 1572 (2005).

6. R. Bohar, A. Chitkineni, and R. K. Varshney, *Biotechniques* **69** (3), 158 (2020). DOI: 10.2144/btn-2020-0066
7. A. Sokolkova, et al., *Int. J. Mol. Sci.* **21**, 3952 (2020).
8. M. Gautier, *Genetics* **201**, 1555 (2015).
9. A. McKenna, M. Hanna, E. Banks, et al., *Genome Res.* **20**, 1297 (2010).
10. P. Danecek, A. Auton, G. Abecasis, et al., *Bioinformatics* **27**, 2156 (2011).
11. F. E. Harrell Jr, *Hmisc: Harrell Miscellaneous. R package version 4.1-1* (2018). <https://CRAN.R-project.org/package=Hmisc> (Accessed 20th October, 2020).
12. H. Jeffreys, *Theory of Probability*, 3rd ed. (Oxford Univ. Press, Oxford, 1961).
13. S. Dash, J. D. Campbell, E. K. Cannon, et al., *Nucl. Acids Res.* **44**, D1181 (2016).
14. *CMplot: Circle Manhattan Plot*. URL <https://github.com/YinLiLin/RCMplot> (Accessed 25th November, 2020).
15. J. C. Barrett, B. Fry, J. Maller, and M. J. Daly, *Bioinformatics* **21**, 263 (2005).
16. Соколкова А.Б. и др., *Биофизика* **66** (1), 40 (2021).
17. Соколкова А.Б. и др., *Биофизика* **65** (2), 276 (2020).
18. R. K. Varshney, M. Thudi, M. Roorkiwal, et al., *Nat. Genet.* **51**, 857 (2019). DOI: 10.1038/s41588-019-0401-3

Search for Agroislands in Chickpea Genome

A.B. Sokolkova*, S.V. Bulyntsev, P.L. Chang***, N. Carrasquilla-Garcia***, D.R. Cook***, E. von Wettberg****, M.A. Vishnyakova*****, S.V. Nuzhdin*, *****, and M.G. Samsonova***

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

**Kuban Experimental Station, Federal Research Center "Vavilov All-Russian Institute of Plant Genetic Resources", ul. Tsentralnaya 2, pos. Botanika, Krasnodar Region, 352183 Russia

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

****University of Vermont, Burlington, Vermont, VT 05405, USA

*****Federal Research Center "N.I. Vavilov All-Russian Institute of Plant Genetic Resources", ul. Bolshaya Morskaya 42–44, St. Petersburg, 190000 Russia

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

The concentration of genes that control domestication traits in the so-called "agroislands" complicates selection and, therefore, information about where such islands are located in the genome and with what traits of domestication they are associated is very important for further research aimed at improving modern varieties. To obtain information about such regions in the genomes of chickpea landraces, the BayPass program was used to assess the association of single nucleotide polymorphisms with population-specific covariates, for which complexes of the domestication syndrome traits of the same target purpose were selected. The regions of the genome with which these covariates are associated are located on chromosomes 1, 5, 6, 7, and 8. These regions are potential agroislands, which is indirectly confirmed by previously obtained data.

Keywords: chickpea (Cicer arietinum L.), genotyping by sequencing, single nucleotide polymorphisms, genomic analysis of associations, adaptation