

АНАЛИЗ ГЕНОМНЫХ АССОЦИАЦИЙ МЕСТНЫХ И СЕЛЕКЦИОННЫХ СОРТОВ КУЛЬТУРНОГО НУТА

© 2023 г. М.А. Дук*, **, #, А.А. Канапин*, М.П. Банкин*, М.А. Вишнякова***, С.В. Булынецв***, М.Г. Самсонова*

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

#E-mail: duk@mail.ioffe.ru

**Физико-Технический институт им. А.Ф. Иоффе, Политехническая ул., 26, Санкт-Петербург, 194021, Россия

***Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова, Большая Морская ул., 42-44, Санкт-Петербург, 190000, Россия

Поступила в редакцию 26.09.2023 г.

После доработки 0264.09.2023 г.

Принята к публикации 04.10.2023 г.

Нут (*Cicer arietinum*) является важной зернобобовой культурой, особо распространенной на Ближнем Востоке. Из-за сильной подверженности нута при влажной погоде таким грибковым заболеваниям, как аскохитоз и фузариоз, важна селекция устойчивых, а также скороспелых сортов. В данной работе мы исследуем геномные ассоциации для 171 образца нута, выращенных на двух опытных станциях — в Краснодарском крае (Кубанская опытная станция) и Астрахани (Астраханская опытная станция), их связь с двенадцатью фенотипическими признаками, а также с тремя признаками устойчивости к трем патогенам: фузариозу, аскохитозу и совке. При исследовании методами полногеномного анализа ассоциаций были выявлены варианты, ассоциированные с различными фенотипическими признаками.

Ключевые слова: геномные ассоциации, нут, устойчивость к патогенам, GWAS.

DOI: 10.31857/S0006302923060066, **EDN:** RNBACN

Нут бараний (*Cicer arietinum*) является важной пищевой зернобобовой культурой, применяемой во множестве традиционных блюд, особенно распространенных на Ближнем Востоке и в Индии [1]. Предок культурного нута (*Cicer reticulatum*) произрастает на территории современной Турции и Сирии [2]. Это однолетнее самоопыляющееся теплолюбивое растение, являющееся культурой длинного дня.

Сорта нута сильно отличаются по времени созревания и семенной продуктивности и массе бобов. Существует межсортная и внутрисортная изменчивость по числу семян в бобе (от одного до четырех) и внутривидовая изменчивость по типу семян, определяющая дифференциацию генофонда вида на два типа. Образцы со светлыми крупными округлыми семенами преимущественно продовольственного назначения относят к типу кабули, с темноокрашенными угловатыми мелкими семенами — к типу дези, используемому для кормов [3]. Также нут очень сильно подвер-

жен грибковым заболеваниям — аскохитозу и фузариозу [4], особенно при влажной погоде.

В связи с этими особенностями при селекции обращают внимание на скороспелость, массу всего растения и семян с одного растения, число ветвей и листьев, длину периода цветения [5] а также на устойчивость к абиотическим [6] и биотическим стрессам: грибковым заболеваниям и вредителям.

При доместикации значительно уменьшилось генетическое разнообразие вида, что затрудняет создание новых сортов [7], тем не менее, поскольку выращивание нута имеет огромное значение для экономики стран Азии, ведутся обширные исследования по селекционному улучшению культуры [8].

МАТЕРИАЛЫ И МЕТОДЫ

Генотипические и фенотипические данные. Использовали генетические данные, фильтрованные по MAF 0.05, для 171 образца нута, из которых 147 местных сортов, 23 селекционных сорта и

Сокращение: SNP — однонуклеотидные замены.

1 образец неопределенного типа. Образцы относились к разным группам по окраске семени: 89 образцов — к типу дези, 81 — к типу кабули, 1 образец неопределенного типа. Фенотипические данные были получены за 2022 г. с двух опытных станций ВИР, расположенных в Краснодарском крае (Кубанская ОС) и Астрахани (Астраханская ОС). Данные Кубанской ОС станции также содержали оценки устойчивости к грибковым заболеваниям (фузариозу и аскохитозу) и вредителям (совке и минирующей мухе).

Подготовка данных. Фенотипические данные с двух опытных станций были квантильно нормированы. Для перевода в формат HarMap использовали программу TASSEL5 [9], для перевода в BED-формат — программу PLINK [10].

Анализ принципиальных компонент и популяционной структуры. Для анализа принципиальных компонент был использован пакет SNPRelate для R [11], графики строились в R. Для анализа популяционной структуры использовали программу ADMIXTURE [12] для $N = 2, 3, 4, 5$.

Анализ геномных ассоциаций. Для анализа геномных ассоциаций использовали пакет GAPIT3 для R [13], модели: FarmCPU, Blink, MLM, GLM, SUPER. В качестве ковариатов была взята популяционная структура (Q-матрица) для пяти популяций из ADMIXTURE. Порог MAF был взят равным 0.05.

Подтверждение аллельного эффекта найденных ассоциаций. Для подтверждения аллельного эффекта рассматривали фенотипические признаки в группах растений, содержащих (генотип 0/1 или 1/1) и не содержащих альтернативный аллель (генотип 0/0); в случаях, когда все растения содержали альтернативный аллель, рассматривались фенотипические признаки в группах гетерозигот (генотип 0/1) и гомозигот (генотип 1/1) по альтернативному аллелю; для варианта, предполагающего два альтернативных аллеля, растения, содержащие альтернативные аллели объединяли в одну группу и сравнивали с группой растений, содержащих только референсный аллель; p -value критерия Манна–Уитни и графики получены с помощью соответствующих функций в R.

Нуклеотидное разнообразие, индекс фиксации, D-статистика Таджимы. Для вычисления π -разнообразия, F_{st} , $TajimaD$ была использована программа vcftools [14]. Данные величины вычислялись в окнах, равных 10000 п.н.

Убывание неравновесия по сцеплению. Для расчета убывания неравновесия по сцеплению (LD) была использована программа PopLDDecay [15], для построения рисунка использовали скрипт в R.

Пересечения с генами. Корреляцию признаков, графики и попадания в гены исследовали с помощью программы, написанной в R, данные о координатах генов были взяты из GFF-файла версии 1 Cicer_arietinum_GA_v1.0.gene.gff, функциональное описание генов нута было получено из базы данных (<https://www.pulsedb.org/Analysis/1869759>).

РЕЗУЛЬТАТЫ

В данном исследовании мы рассмотрели геномные ассоциации для 171 образца нута (147 местных сортов и 23 селекционных сорта) относящихся к разным группам по окраске семени (89 образцов относятся к типу дези, 81 к типу кабули, 1 образец неопределенного типа), выращенных на двух опытных станциях в Краснодарском крае (Кубанская ОС ВИР) и в Астрахани (Астраханская ОС ВИР).

Все признаки показали значительную вариабельность среди 171 образца нута (см. табл. 1), что подтверждает возможность применения метода полногеномного анализа ассоциаций (Genome-wide association analysis, GWAS) для обнаружения значимых вариаций.

Применение методов уменьшения размерности — метода принципиальных компонент (рис. 1а), tSNE, UMAP (здесь не приведены) — выявило, что образцы генетически не показывают значительного разделения на группы ни по типу окраски (дези/кабули), ни по селекционному статусу (местные и селекционные), ни по странам происхождения (результаты не показаны). На графиках принципиальных компонент можно увидеть слаборазличимое разделение на группы по первой, второй и четвертой компоненте, не связанное, однако, ни с одним из признаков. На рис. 1б показаны результаты применения программы ADMIXTURE к данным. Можно видеть, что большая часть сортов трактуется как смесь нескольких популяций, и среди образцов невозможно выделить достаточно замкнутые группы, обладающие одним происхождением. Все это указывает на активное генетическое смешение и дрейф генов даже между местными сортами, несмотря на то что растение является самоопыляемым. Это, вероятно, связано с активными попытками селекции на протяжении истории культивации нута.

Для поиска геномных ассоциаций были отдельно использованы фенотипические данные, полученные на разных опытных станциях. Среди признаков, для которых проводился поиск ассоциаций, были следующие: продолжительность периодов всходы-созревание, всходы-цветение, цветения, а также высота растения, размер листа,

Таблица 1. Статистические характеристики рассматриваемых фенотипических признаков

Признак	Среднее		Стандартное отклонение		Коэффициент вариации, %	
	К	А	К	А	К	А
Опытная станция*	К	А	К	А	К	А
Высота растения, см	44.33	51.57	8.4	12.2	18.95	23.66
Высота прикрепления первого боба, см	19.5	19.92	4.95	4.26	25.38	21.39
Число ветвей 1-го порядка	1.26	3.4	0.37	0.56	29.37	16.47
Число ветвей 2-го порядка	10.87	10.37	2.85	2.88	26.22	27.77
Масса растения с бобами, г	67.86	60.83	34.25	26.43	50.47	43.45
Масса бобов, г	24.4	29.22	10.98	11.89	45	40.69
Масса семян с растения, г	14.78	20.84	6.76	8.32	45.74	39.92
Размер листа, баллы	4.99	4.25	1.33	1.83	26.65	43.06
Период: всходы-созревание, дни	96.4	72.67	12.32	6.78	12.78	9.33
Период: всходы-цветение, дни	30.94	31.74	7.58	4.08	24.5	12.85
Период цветения, дни	26.8	33.6	7.31	6.57	27.28	19.55
Устойчивость к аскохитозу, баллы	8.12	–	1.3	–	16.01	–
Устойчивость к фузариозу, баллы	7.32	–	1.51	–	20.63	–
Устойчивость к совке, баллы	4.25	–	1.34	–	31.53	–

Примечание. * – К – Кубанская опытная станция, А – Астраханская опытная станция.

число ветвей первого и второго порядка, высота прикрепления первого боба, вес растения с бобами, масса бобов с одного растения, число бобов, масса семян с одного растения, число семян, а также устойчивость к аскохитозу, фузариозу, совке и минирующей мухе.

Данные для всех признаков, кроме различной устойчивости, были получены на обеих опытных станциях, данные по устойчивости получены на Кубанской ОС. Из всех признаков не были найдены только ассоциации, связанные с числом семян в бобе, числом бобов и с устойчивостью к минирующей мухе.

Большинство одинаковых признаков для данных с разных опытных станций показывают относительную корреляцию друг с другом, например, длина периода «всходы–цветение» для данных Кубанской и Астраханской ОС, а также высота растения, размер листа, высота прикрепления первого боба (см. рис. 2), что говорит о том, что данные признаки с большей вероятностью

связаны с генотипом растения, а не с условиями выращивания.

Лишь немногие различные признаки показывают некоторую корреляцию друг с другом, например, высота растения и высота прикрепления первого боба, масса семян и масса бобов, что кажется очевидным. Однако некоторые одинаковые признаки, рассчитанные на разных опытных станциях, не показывают корреляции друг с другом, например, период цветения, число ветвей первого и второго порядка, вес семян и число бобов, что указывает на то, что эти признаки значительно связаны с условиями выращивания, имеющимися на разных опытных станциях. В таком случае геномные ассоциации, полученные на основе фенотипических данных с разных опытных станций, тоже могут быть различны, так как будут связаны с некоторыми внутренними биологическими процессами, запустившимися при разных условиях выращивания.

Для поиска геномных ассоциаций (GWAS) были использованы различные модели: GLM,

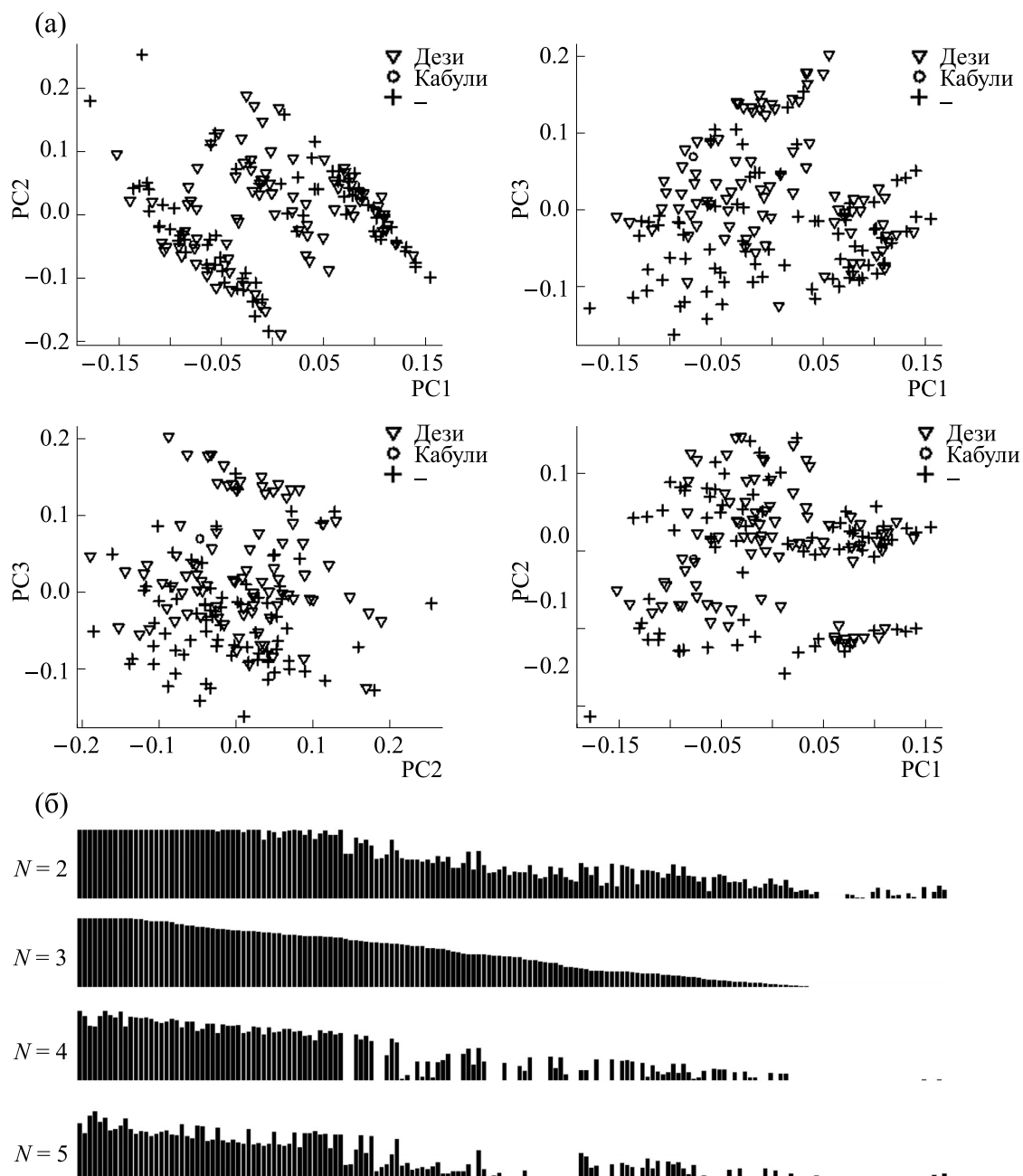


Рис. 1. (а) — Результаты анализа для 1-4 принципиальных компонент, образцы изображены в соответствии с принадлежностью к подтипу по окраске семян: desi/kabuli; (б) — результат работы ADMIXTURE для количества популяций $N = 2-5$.

MLM, SUPER, FarmCPU, Blink, представленные в пакете GAPIT3 для R. В результате было найдено 160 однонуклеотидных замен (single nucleotide polymorphism, SNP), связанных с 15 различными признаками, из них 123 имеют статистически значимый аллельный эффект. 23 варианта были найдены несколькими моделями, 4 варианта из них были ассоциированы с несколькими фенотипи-

ческими признаками. В табл. 2 показаны 22 SNP, выявленные несколькими моделями, аллельный эффект которых был подтвержден. Для подтверждения аллельного эффекта каждого варианта был рассмотрен критерий Манна–Уитни между значениями фенотипических признаков у групп образцов содержащих и не содержащих альтернативный аллель (в случаях, когда альтернативный

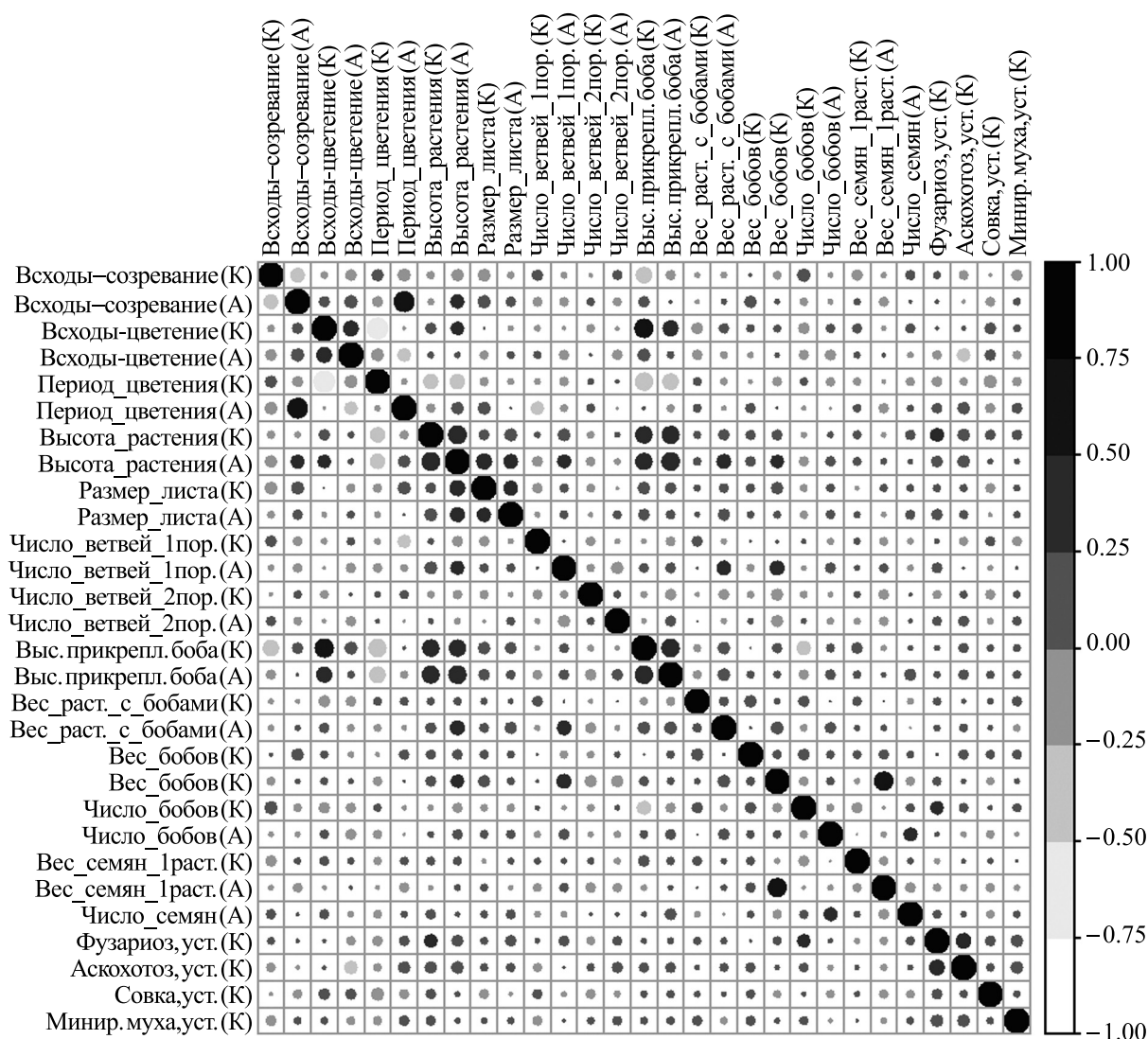


Рис. 2. Корреляция между фенотипическими признаками. (К) – Данные Кубанской опытной станции, (А) – данные Астраханской опытной станции.

аллель присутствовал у всех образцов – сравнивались фенотипические признаки между группами гетерозигот и гомозигот по альтернативному аллелю), значения p -value показаны на рис. 3.

Для рассматриваемого набора образцов наблюдается очень быстрое убывание неравновесия по сцеплению (рис. 4). Для имеющегося набора данных размер LD получается около 50 т.н., тогда как в литературе указываются величины 330 т.н. и 340 т.н. для кабули и дези соответственно. Такая разница может объясняться особенностью и качеством имеющихся у нас данных.

Нами были рассмотрены функциональные особенности генов в LD-окрестности найденных 123 ассоциаций, всего в окрестности найденных

SNP было отмечено 290 генов с различными функциями. Из 123 геномных ассоциаций 18 попали в тело известных генов, 3 из которых попали в тело одного гена, список генов приведен в табл. 3. SNP были ассоциированы с периодом «всходы–созревание», периодом цветения, высотой растения, числом ветвей 1-го и 2-го порядка, размером листа, весом растения и весом семян, а также с устойчивостью к фузариозному увяданию.

Три SNP, ассоциированные с весом семян, попали в ген, связанный с обменом жирных кислот. Варианты, ассоциированные с периодом «всходы–созревание», пересекаются с геном, кодирующим метилтрансферазу (*Ca_23694*), варианты,

Таблица 2. SNP, выявленные как значимые геномные ассоциации несколькими моделями

SNP ID	Хромо- сома	REF/ ALT	Положение в геноме	Модель	MAF	p-value	Эффект	Признак	Происхожде- ние данных
SCM001764.1_9199713	Ca1	A/G	9199713	FarmCPU, Blink	0.41	5.3·e-07	−0.48	Вес растения с бобами	Краснодар
SCM001764.1_18866042	Ca1	G/T	18866042	FarmCPU, Blink	0.31	7.1·e-11	−0.61	Высота растения	Краснодар
SCM001764.1_22963404	Ca1	G/A	22963404	FarmCPU, Blink	0.23	1.6·e-9	−0.49	Размер листа	Астрахань
SCM001764.1_24255799	Ca1	C/T	24255799	FarmCPU, Blink	0.26	5.1·e-07	0.61	Устойчи- вость к фузариозу	Краснодар
SCM001765.1_11031394	Ca2	T/C	11031394	FarmCPU, Blink	0.45	1.6·e-08	−0.52	Вес семян	Краснодар
SCM001765.1_11143297	Ca2	A/G	11143297	FarmCPU, Blink	0.49	1.5·e-07	2.16	Период: всходы- цветение	Краснодар
SCM001766.1_3964840	Ca3	A/G	3964840	FarmCPU, MLM, GLM, Blink	0.08	2.6·e-07	1.08	Высота прикрепле- ния боба	Краснодар
SCM001766.1_4378135	Ca3	G/C	4378135	FarmCPU, Blink	0.46	1.9·e-08	−0.76	Высота растения	Астрахань
SCM001766.1_5689101	Ca3	T/G	5689101	FarmCPU, Blink	0.06	3.8·e-08· 2.0·e-10	0.42, NA	Высота растения, высота прикрепле- ния боба	Астрахань/ Краснодар
SCM001767.1_7480193	Ca4	A/G	7480193	FarmCPU, Blink, SUPER	0.08	3.5·e-09· 5.5·e-09	−0.86, NA	Высота растения, высота прикреп- ления боба	Астрахань/ Краснодар
SCM001767.1_19608475	Ca4	A/G	19608475	FarmCPU, Blink	0.34	6.9·e-08	−0.75	Вес растения с бобами	Астрахань
SCM001767.1_28678319	Ca4	G/A	28678319	FarmCPU, Blink	0.29	3.5·e-09· 6.3·e-09	0.44, 0.69	Высота растения, высота прикрепле- ния боба	Астрахань/ Краснодар
SCM001767.1_30555335	Ca4	A/G	30555335	FarmCPU, Blink	0.32	2.5·e-09	−1.62	Число ветвей 2 порядка	Краснодар
SCM001768.1_24460834	Ca5	G/A	24460834	FarmCPU, Blink	0.31	9.4·e-08	−0.77	Вес растения с бобами	Астрахань
SCM001768.1_44202762	Ca5	C/T	44202762	FarmCPU, SUPER, Blink	0.21	6.8·e-08	0.71	Число ветвей 2 порядка	Астрахань
SCM001769.1_35831438	Ca6	A/T	35831438	FarmCPU, Blink	0.37	1.4·e-07	−0.74	Высота растения	Астрахань

Таблица 2. Окончание

SNP ID	Хромо- сома	REF/ ALT	Положение в геноме	Модель	MAF	<i>p</i> -value	Эффект	Признак	Происхоже- ние данных
SCM001770.1_25146185	Ca7	G/A	25146185	FarmCPU, Blink	0.15	1.9·e-08	0.41	Число ветвей 1 порядка	Краснодар
SCM001770.1_28975339	Ca7	G/A	28975339	FarmCPU, Blink	0.24	8.3·e-08	0.65	Число ветвей 1 порядка	Краснодар
SCM001769.1_141242	Ca6	A/T	141242	FarmCPU, Blink	0.45	9.7·e-08	−0.98	Устойчи- вость к фузариозу	Краснодар
SCM001769.1_39266797	Ca6	C/T	39266797	FarmCPU, Blink	0.49	1.0·e-07	1.31	Вес семян	Астрахань
SCM001769.1_46997240	Ca6	A/C	46997240	FarmCPU, Blink	0.48	1.0·e-07	−0.62	Вес растения с бобами	Краснодар
SCM001769.1_57993704	Ca6	G/A,C	57993704	FarmCPU, Blink	0.50	1.6·e-08 2.4·e-13	−0.02, NA	Размер листа, устойчи- вость к аскохитозу	Астрахань/ Краснодар

Примечание. Показаны средние по моделям значения *p*-value и величина эффекта (если модель его дает).

ассоциированные с периодом цветения — с геном, кодирующим гуанилилтрансферазу (*Ca_09426*), варианты, связанные с высотой растения, пересекаются с геном, кодирующим рибосомный белок (*Ca_01938*) и ДНК-праймазу (*Ca_15072*), работа этих ферментов связана с делением клеток и внутриклеточным транспортом. Тем не менее, большинство вариантов не пересекались с генами, а находились в непосредственной близости от них. В табл. 4 показаны 13 генов, находящихся вблизи четырех вариантов, выявленных несколькими моделями для нескольких фенотипических признаков в данных двух опытных станций.

В большинстве случаев функции белка, производимого с гена, оказались не охарактеризованы. Известные функции оказались связаны с обменом пектина, аденилированием, и энергетическим обменом и транспортом полисахаридов в клетках растений.

Несмотря на то, что устойчивость к грибковым заболеваниям и вредителям коррелирует друг с другом больше, чем с другими признаками (см. рис. 2), значимые для устойчивости к разным за-

болеваниям варианты расположены далеко друг от друга в геноме (см. рис. 5).

Варианты, связанные с устойчивостью к аскохитозу, оказываются вблизи генов, связанных с поражением растительными вирусами и иммунитетом (*Ca_22054*, *Ca_21959*) и фосфорным обменом (*Ca_24946*, *Ca_15816*). Варианты, ассоциированные с устойчивостью к совке, оказались вблизи генов, связанных с регуляцией транскрипции (*Ca_08447*, *Ca_08450*), а также генов, связанных с реакцией на освещенность (*Ca_14991*) и температуру (*Ca_08449*).

Нами было рассмотрено нуклеотидное разнообразие (π -разнообразие) в нашем наборе данных, а также Fst и ROD (уменьшение разнообразия) между группами местных и селекционных сортов. Среднее по окнам в 10000 п.н. нуклеотидное разнообразие для каждой хромосомы показано на рис. 6. Можно видеть, что для нашего набора данных нуклеотидное разнообразие очень мало и слабо отличается по хромосомам. Относительно меньшее разнообразие наблюдается у четвертой и восьмой хромосом.

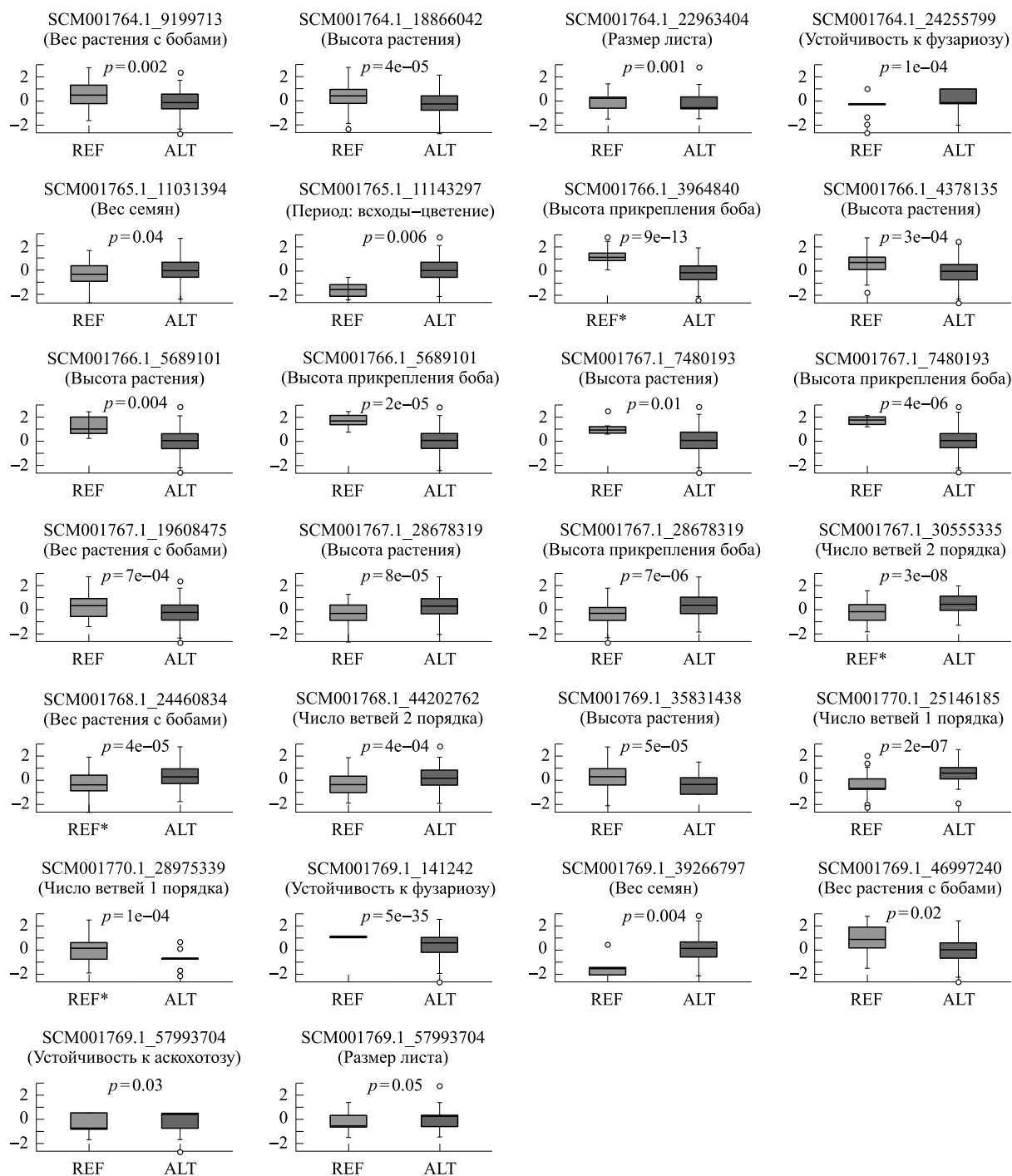


Рис. 3. Сравнение значений квантильно нормированных фенотипических данных между образцами, содержащими и не содержащими альтернативный аллель в соответствующих вариантах. REF/ALT — соответственно референсный и альтернативный аллель, REF*/ALT — генотипы 0/1 и 1/1 по альтернативному аллелю соответственно (в случаях, когда альтернативный аллель присутствует у всех образцов, дикий тип гетерозиготен).

Разнообразие по хромосомам между местными и селекционными сортами оказывается практически одинаковым, уменьшение разнообразия у селекционных сортов наблюдается лишь в неко-

торых участках, приведенных в табл. 5, пересекающихся с генами, связанными с энергетическим обменом (протеаза, полигалактуроназа, белок хлоропластов). Эти гены, вероятно, обеспечива-

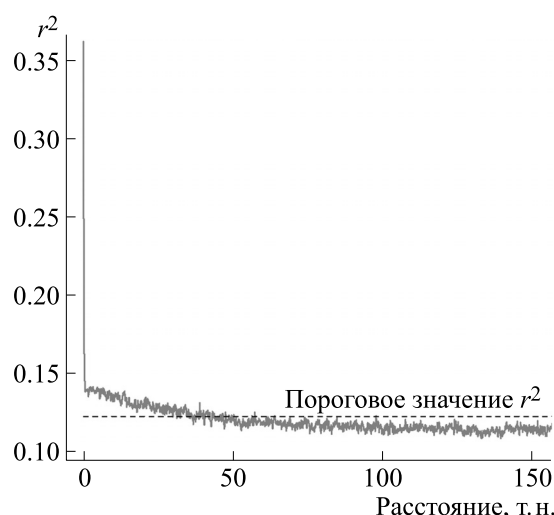


Рис. 4. Убывание неравновесия по сцеплению (linkage disequilibrium) для имеющегося набора данных нута.

ют большую продуктивность селекционных сортов.

Участки, показывающие уменьшение разнообразия среди сортов, устойчивых к грибковым заболеваниям (фузариозу и аскохитозу), также пересекаются с различными важными функциональными генами в первой-седьмой хромосомах. Для устойчивых к фузариозу образцов был выявлен 21 ген, среди которых присутствуют гены ферментов, связанных с белковым и энергетическим обменом — протеинкиназы (*Ca_14253*, *Ca_01018*), изомеразы (*Ca_07236*), гликозидазы (*Ca_04950*). Для устойчивых к аскохитозу образцов были выявлены 75 генов, среди которых гены ферментов, связанных с энергетическим и белковым обменом: регулятор роста (*Ca_06980*), протеинкиназы (*Ca_11633*, *Ca_2074*, *Ca_19214*), изомеразы (*Ca_07236*, *Ca_03117*), протеазы (*Ca_17088*), а также гены белков, связанных с повышенной устойчивостью к заболеваниям (*Ca_08433*, *Ca_05008*, *Ca_15790*).

ОБСУЖДЕНИЕ

Несмотря на то, что выборка сортов оказалась не очень генетически разнообразна, что не позволило выделить среди образцов четко прослеживаемые родственные группы и изучить нуклеотидное разнообразие между ними, методы FarmCPU и Blink для поиска геномных ассоциаций показали на этих данных хорошие результаты. Для 15 различных признаков для данных с двух опытных станций было найдено 123 значимых варианта, аллельный эффект которых был статистически подтвержден, на всех восьми хромосомах при фильтрации параметров $MAF = 0.05$. В данном

случае фильтрация была необходима, так как количество образцов было сравнительно небольшим. При фильтрации соответственно исключались редкие варианты, которые могут быть ассоциированы с признаками и найдены на большей выборке.

Из 123 найденных вариантов 22 варианта были найдены несколькими моделями для одного и того же признака, и 4 варианта — найдены несколькими моделями для разных признаков (по данным обеих опытных станций). Обнаружение одних и тех же вариантов несколькими моделями увеличивает надежность полученных данных.

18 выявленных вариантов располагается внутри известных генов, 3 из которых расположены в одном гене. Эти гены связаны с различными признаками, некоторые из них имеют важное значение для скорости роста и деления клеток и для обмена жирных кислот, тем не менее, большая часть найденных вариантов либо располагалась вне известных генов, либо в генах с неизвестными функциями. Многие варианты располагаются в непосредственной близости к важным генам, а варианты, ассоциированные с устойчивостью к грибковым заболеваниям — аскохитозу и фузариозу — располагались вблизи генов, связанных с энергетическим и белковым обменом и устойчивостью к болезням.

Уменьшение нуклеотидного разнообразия у селекционных сортов по сравнению с местными сортами наблюдалось лишь в некоторых участках хромосом, связанных, вероятно, с большей продуктивностью, однако в целом нуклеотидное разнообразие в рассматриваемой выборке было мало. Отдельное рассмотрение участков уменьше-

Таблица 3. Значимые SNP, попавшие в тело известных генов

Хромосома	Положение в геноме	ID гена	Начало гена	Конец гена	Фенотипические признаки	Происхождение данных	Ортолог <i>A. thaliana</i>	Описание функций
Ca4	43512220	ID=Ca_23694; evid_id=GAR_10004833	43511982	43530939	Период: всходы-созревание	Краснодар	AT2G16910	Вероятная S-аденозилметионин-зависимая метилтрансфераза At5g38100
Ca5	24177498	ID=Ca_17652; evid_id=GAR_10010965	24157761	24181955	Период: всходы-созревание	Краснодар	AT3G13330	Гипотетический белок DVH24_038707 (ортолог – активатор протеосомы)
Ca3	8285781	ID=Ca_24378; evid_id=GAR_10006335	8282802	8301564	Период: всходы-созревание	Астрахань	—	Гомеобокс-лейциновый белок zipper PROTODERMAL FACTOR 2
Ca3	20257421	ID=Ca_09426; evid_id=GAR_10018848	20257070	20257745	Период цветения	Краснодар	—	тРНК (His) гуанилтрансфераза 1
Ca6	52822315	ID=Ca_27699; evid_id=GAR_10001048	52822150	52822566	Период цветения	Краснодар	—	Белок LYK2
Ca1	16671522	ID=Ca_06900; evid_id=GAR_10022047	16670592	16673062	Период цветения	Астрахань	—	Укороченный фактор транскрипции CAULIFLOWER A-подобный
Ca8	1108528	ID=Ca_15072; evid_id=GAR_10007089	1107812	1110239	Высота растения	Астрахань	AT3G03150	ДНК-примаза, малая субъединица
Ca8	6085438	ID=Ca_01938; evid_id=GAR_10025825	6083232	6086628	Высота растения	Астрахань	—	40S рибосомный белок S21-2
Ca2	30921855	ID=Ca_12470; evid_id=GAR_10015200	30902774	30933225	Число ветвей 1 порядка	Краснодар	AT5G05560	Белок 6, содержащий повторы armadillo (ортолог - E3 убиквитинлигаза)
Ca3	11951992	ID=Ca_22252; evid_id=GAR_10005852	11951640	11952416	Число ветвей 1 порядка	Краснодар	—	Кальциевый унипортерный белок 2, митохондриальный
Ca6	29780621	ID=Ca_25982; evid_id=GAR_10003992	29753481	29792628	Вес семян с 1 растения	Краснодар	—	Цинковый палец, содержащий домен CCHH
Ca4	48755155; 48760624; 48767615	ID=Ca_23023; evid_id=GAR_10003504	48740176	48771978	Вес семян с 1 растения	Краснодар	AT1G75890	Муцин-3А, частичный (ортолог - GDSL-подобный белок суперсемейства липазы/ацилгидролазы)
Ca1	9199713	ID=Ca_02875; evid_id=GAR_10025740	9199030	9199943	Вес растения с бобами	Краснодар	—	Неописанный белок
Ca5	24460834	ID=Ca_15581; evid_id=GAR_10012684	24452880	24463063	Вес растения с бобами	Астрахань	AT1G07380	Синтаза слезоточивого фактора (ортолог - Нейтральная/щелочная нелизосомальная церамидаза)
Ca1	22963404	ID=Ca_22347; evid_id=GAR_10007799	22963280	22964164	Размер листа	Астрахань	—	Неописанный белок
Ca6	11454891	ID=Ca_23429; evid_id=GAR_10002517	11454538	11454915	Устойчивость к фузариозу	Краснодар	—	Альфа-подобная субъединица фактора инициации трансляции eIF-2B

Таблица 4. Гены, находящиеся вблизи SNP, выявленных несколькими моделями для данных с обеих опытных станций, ассоциированные с несколькими фенотипическими признаками

Хро-мо-сома	SNP ID	Положение	ID гена	Начало гена	Конец гена	Признаки	Ортолог <i>A. thaliana</i>	Функции гена
Ca3	SCM001766.1_5689101	5689101	ID=Ca_20649 evid_id=GAR_10006968	5671689	5676293	Высота растения, высота прикрепления боба	AT5G15800	Неописанный белок (ортолог – ТФ MADS-box, участвующий в развитии цветка)
			ID=Ca_20648 evid_id=GAR_10006967	5691039	5691242		—	Вероятная пектатлиаза 4
Ca4	SCM001767.1_7480193	7480193	ID=Ca_03488 evid_id=GAR_10024933	7454864	7460054	Высота растения, высота прикрепления боба	AT1G29220	sufE-подобный белок 1, хлоропластный/митохондриальный (ортолог – регулятор транскрипции)
			ID=Ca_03487 evid_id=GAR_10024932	7471601	7473607		AT4G16100	Неописанный белок (ортолог - белок теплового шока, предполагаемый)
			ID=Ca_03486 evid_id=GAR_10024931	7493983	7499371		AT2G47490	Полиаденилат-связывающий белок (ортолог – локализованный в хлоропластах транспортер NAD+)
Ca4	SCM001767.1_28678319	28678319	ID=Ca_26235 evid_id=GAR_10003030	28671220	28671933	Высота растения, высота прикрепления боба	—	Неописанный белок

Таблица 4. Окончание

Ca6	SCM001769.1_579 93704	57993704	ID=Ca_13640 evid_id=GAR_ 10011830	57959110	57971144	Размер листа, устойчи- вость к аскохитозу	AT3G18860	Неописанный белок (ортолог – Белок семейства трансдуцинов/белок семейства повторов WD-40)
			ID=Ca_13639 evid_id=GAR_ 10011829	57976183	57976758		AT3G18850	gag-полипротеин, частичный (орто- лог – лизофос- фатидил- ацилтрансфераза 5)
			ID=Ca_13638 evid_id=GAR_ 10011828	57986478	57987611		AT4G34320	Гипотетический белок L195_g053877, частичный (орто- лог – транс- мембранный белок, предполагаемый)
			ID=Ca_13637 evid_id=GAR_ 10011827	57991873	57992190		AT5G48490	Неописанный белок (ортолог - Бифункциональный ингибитор/белок- переносчик липидов/белок суперсемейства 2S- альбумина семенного запаса)
			ID=Ca_13636 evid_id=GAR_ 10011826	57994881	57999978		AT5G66680	Неописанный белок (ортолог – белковая субъединица олигосахарил- трансферазного комплекса)
			ID=Ca_13635 evid_id=GAR_ 10011825	58002601	58004211		AT4G36750	UDP- гликозилтрансфе- раза 1 (ортолог – белок семейства хинонредуктазы)
			ID=Ca_13634 evid_id=GAR_ 10011824	58006066	58008568		AT1G19360	Неописанный белок (ортолог – арабинозилтранс- фераза, модифицирующая белки экстенсина)

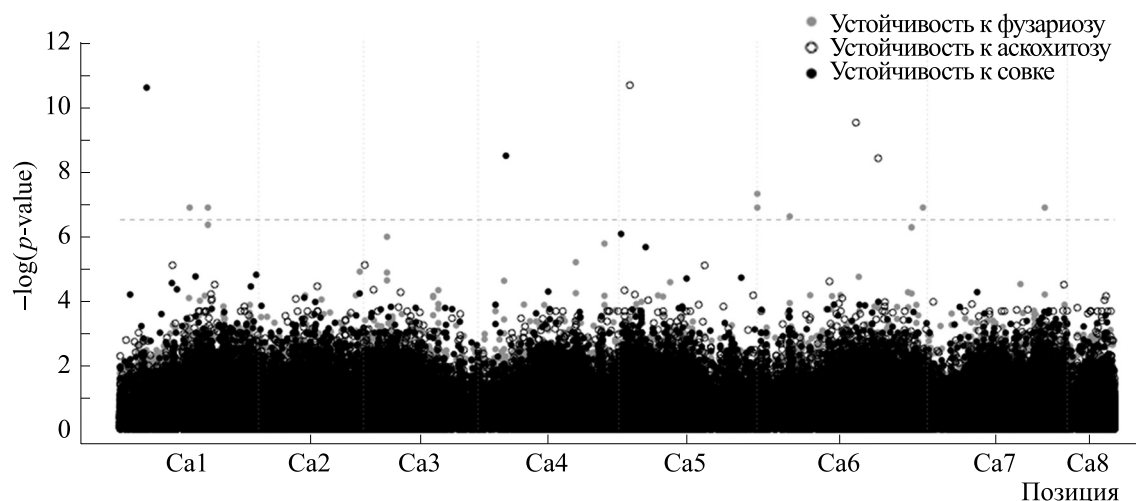


Рис. 5. Манхэттенские графики ассоциации вариантов с устойчивостью к фузариозу, аскохитозу и совке (все модели).

Таблица 5. Гены, пересекающиеся с участками, в которых у селекционных сортов наблюдается уменьшение нуклеотидного разнообразия

Хромосома	ID гена	Начало гена	Конец гена	Ортолог <i>A. thaliana</i>	Функциональное описание
Ca4	ID=Ca_13051 evid_id=GAR_10015638	39403291	39404805	—	deSI-подобный белок At4g17486
	ID=Ca_13050 evid_id=GAR_10015637	39405919	39407732	—	Этилен-чувствительный фактор транскрипции
	ID=Ca_10943 evid_id=GAR_10019818	41687794	41692786	AT4G32630	Вероятная полигалактуроназа (ортолог — ArfGap/RecO-подобный белок, содержащий домен цинкового пальца)
	ID=Ca_10942 evid_id=GAR_10019817	41698842	41704918	AT4G32620	Неописанный белок (ортолог — усилитель белка транскрипционного фактора)
Ca6	ID=Ca_25062 evid_id=GAR_10003009	11711556	11713555	—	Гомолог алкилированного белка репарации ДНК ALKBH8
Ca7	ID=Ca_09259 evid_id=GAR_10019148	12612996	12630579	AT3G50590	Цистеиновая протеаза (ортолог — трансдуктин / белок суперсемейства повторов WD40)
	ID=Ca_09225 evid_id=GAR_10019114	12956542	13028092	AT3G50380	nifU-подобный белок 1, хлоропласт (ортолог — белок, связанный с сортировкой вакуолярного белка, предполагаемый)
	ID=Ca_13788 evid_id=GAR_10013875	37043886	37046659	—	Неописанный белок
	ID=Ca_13789 evid_id=GAR_10013876	37048371	37049951	AT1G01780	Копия белка, частично (ортолог — белок семейства ТФ цинковых пальцев GATA-типа, регулирует организацию актинового цитоскелета)

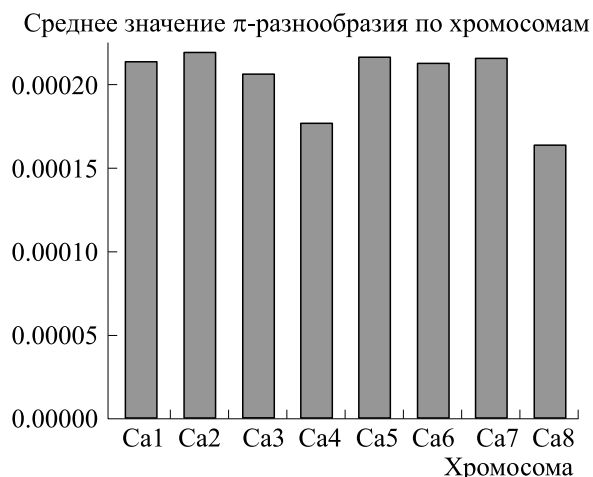


Рис. 6. Среднее по хромосомам нуклеотидное разнообразие, рассчитанное в окнах 10000 п.н..

ния нуклеотидного разнообразия для растений устойчивых и неустойчивых к грибковым заболеваниям позволило выявить в целом уменьшение нуклеотидного разнообразия в участках, пересекающихся с генами растительного иммунитета, а также с генами, связанными с белковым и фосфорным обменом. Тем не менее, применение D-статистики Таджимы не выявило участков, подтверждающих наличие направленной селекции, что указывает на слабую селекционную историю для рассматриваемых нами сортов, большая часть из которых относятся к местным формам.

БЛАГОДАРНОСТИ

Авторы выражают благодарность Суперкомпьютерному центру «Политехнический» (<https://research.spbstu.ru/skc/>) за предоставление вычислительных мощностей для проведения исследования.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при финансовой поддержке Российского научного фонда (грант № 22-46-02004).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит экспериментов с использованием людей и животных в качестве объектов исследований.

СПИСОК ЛИТЕРАТУРЫ

1. R. J. Redden and J. D., Berger, In *Chickpea Breeding & Management* (CABI: Wallingford, UK, 2007), pp. 1–13.
2. R. K. Varshney, M. Thudi, M. Roorkiwal, et al, *Nat. Genet.*, **51** (5), 857 (2019).
3. M. Thudi, et al., *Sci. Rep.*, **6**, 38636 (2016).
4. J. Kumar and S. Abbo, *Adv. Agronomy*, **72**, 107 (2001).
5. V. V. Gursky, K. N. Kozlov, S. V. Nuzhdin, and M. G. Samsonova, *Front. Genetics*, **9**, 547 (2018)
6. U. Ch. Jha, P. Ch. Kole, and N. P. Singh, *Legume Res.*, **44** (4), 382 (2019)
7. A. Sokolkova, S. V. Bulyntsev, P.L. Chang, et al., *Int. J. Mol. Sci.*, **21**, 3952 (2020)
8. R. K. Varshney, et al., *Nature*, **599**, 622 (2021)
9. P. J. Bradbury, Z. Zhang, D. E. Kroon, et al., *Bioinformatics* **23**, 2633 (2007)
10. Sh. Purcell, et al., *Am. J. Hum. Genet.*, **81** (3), 559 (2007)
11. X. Zheng, D. Levine, J. Shen, et al., *Bioinformatics*, **28** (24), 3326 (2012)
12. D. H. Alexander, J. Novembre, and K. Lange, *Genome Res.*, **19**, 1655 (2009)
13. J. Wang and Z. Zhang, *Genomics Proteomics Bioinformatics*, **19** (4), 629 (2021)
14. P. Danecek, A. Auton, G. Abecasis, et al., *Bioinformatics*, **27** (15), 2156 (2011)
15. Ch. Zhang, Sh.-Sh. Dong, J.-Y. Xu, et al., *Bioinformatics*, **35**, 1786 (2019)

Genome-Wide Association Analysis in Chickpea Landraces and Cultivars

M.A. Duk **, A.A. Kanapin*, M.P. Bankin*, M.A. Vishnyakova***,
S.V. Bulyntsev***, and M.G. Samsonova***

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

***Ioffe Institute, Polytekhnicheskaya ul. 26, St. Petersburg, 194021 Russia*

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

Chickpea (*Cicer arietinum*) is an important leguminous crop, which is widely grown especially in the Near East. In wet weather conditions, the susceptibility of chickpeas to fungal diseases such as Ascochyta blight and Fusarium blight increases. Thus, selection of disease-resistant and early-ripening varieties is critically needed. The present study was conducted to investigate genome associations in 171 samples of chickpea plants, grown in two experimental stations in Krasnodar (Kuban experimental station) and Astrakhan (Astrakhan experimental station), examine relationship between genes and 12 phenotypic traits as well as explore the association between genes and 3 hallmarks of resistance to pathogens: Fusarium blight, Ascochyta blight and Noctuidae. Variants associated with different phenotypic traits were identified using a genome-wide association study (GWAS).

Ключевые слова: genome associations, chickpea, cicer arietinum, pathogen resistance, GWAS