БИОФИЗИКА, 2023, том 68, № 2, с. 268–274

МОЛЕКУЛЯРНАЯ БИОФИЗИКА

УДК 519.876.5

ПОЛНОГЕНОМНЫЙ АНАЛИЗ АССОЦИАЦИЙ ВАРИАЦИЙ КОПИЙНОСТИ В КОНТЕКСТЕ СТАБИЛЬНОСТИ ГЕНОМА

© 2023 г. М.А. Дук*, А.А. Канапин*, Т.А. Рожмина, А.А. Самсонова*, #**

**Санкт-Петербургскии? государственный университет, Университетская набережная, 7/9, Санкт-Петербург, 199034, Россия*

***Институт льна – обособленное подразделение Федерального научного центра лубяных культур, ул. Луначарского, 35, Торжок, Тверской области, 172002, Россия*

#E-mail: a.samsonova@spbu.ru

Поступила в редакцию 28.11.2022 г.
После доработки 05.12.2022 г.
Принята к публикации 21.12.2022 г.

Классические методы поиска генетических вариантов, ассоциированных с теми или иными фенотипическими признаками, обычно сводятся к анализу однонуклеотидных замен. Вариации копийности и, шире, структурные варианты могут предоставить гораздо больший объем информации в силу масштабности вносимых ими изменений. Однако их использование в анализе полногеномных ассоциаций затруднено недостаточной точностью их локализации в геноме. Тем не менее, в отдельных случаях такой анализ возможен и может дать достоверные результаты. Ранее нами был проведен полногеномный поиск ассоциаций однонуклеотидных замен в геноме льна по отношению к фенотипическим признакам, определяющим качество получаемого волокна. В данной работе мы используем новый набор данных, полученный с большим значением покрытия секвенирования, что позволяет предсказать координаты вариаций копийности с большей точностью. В результате анализа получен список из 41 гена-кандидата, ассоциированных с пятью количественными фенотипическими признаками. Разработанная ранее метрика стабильности генома позволила также классифицировать регионы, содержащие вариации копийности на более и менее стабильные. Результаты анализа позволяют предположить, что менее стабильные, и, как следствие, более пластичные области генома более подвержены изменениям, связанным с изученными фенотипическими признаками.

Ключевые слова: вариации копийности, стабильность генома, Linum usitasissimum L., волокно льна, полногеномные ассоциации.

DOI: 10.31857/S0006302923020084, **EDN:** CARJRB

По мере роста популярности применения методов полногеномного анализа ассоциаций (GWAS) для нахождения однонуклеотидных полиморфизмов (SNPs), отвечающих за сложные признаки у растений, со временем, стала очевидна их полигенная архитектура [1, 2]. Тем не менее, SNPs не объясняют всю наблюдаемую фенотипическую изменчивость признаков. Этот феномен известен как «недостающая наследственность». Одно из предложенных объяснений его возникновения - вклад дополнительных типов генетических вариантов, таких как вариации числа копий (CNV) [3, 4].

Характеризующиеся как делеции или дупликации фрагментов ДНК размером более 50 пар оснований [5], CNV представляют собой весьма разнообразный класс мутаций, которые, благодаря своему подчас большому размеру, являются мощными фенотипическими модификаторами, действующими, например, посредством изменения дозы генов, т.е. делетирования и слияния генов. Путем отмены маскировки рецессивных аллелей или за счет повреждения *цис*-регуляторных элементов генов [6]. Геномные данные, полученные у различных видов растений в рамках крупномасштабных проектов секвенирования, указывают на CNV как на один из факторов, обеспечивающих природное разнообразие на геномном уровне [7, 8]. CNV у растений вовлечены в контроль времени цветения, устойчивости к насекомым, РНК-интерференции, ответов на солевой и обезвоживающий стрессы [7].

Лен является важной сельскохозяйственной культурой двойного назначения [9]. Семена масличного льна – ценный источник высококаче-

Сокращения: SNP – однонуклеотидный полиморфизм, CNV – вариации числа копий.

ственных ненасыщенных кислот, лигнинов, легко усваиваемых протеинов, диетической клетчатки, витаминов и минеральных элементов. Лендолгунец в современных условиях рассматривается как стратегическая культура России, позволяющая заменить хлопок-сырец, который перешел в разряд импортного сырья. Поэтому получение высокопродуктивных сортов льна-долгунца с высоким качеством волокна является одной из актуальных и приоритетных селекционных задач [10].

Элементарное волокно у льна – это отдельная клетка склеренхимы с исключительно толстой клеточной стенкой и экстремальным отношением длины к диаметру (более 1000 раз) [11] Клетки происходят из апикальной меристемы и развиваются в две стадии: удлинение и утолщение стенки. Вначале удлинение льняных волокон происходит синхронно с соседними клетками, однако через несколько часов начинается интрузивный рост, который приводит к образованию волокон, достигающих нескольких сантиметров в длину. Утолщение клеточной стенки включает отложение нескольких новых слоев изнутри первичной клеточной стенки и начинается, когда удлинение прекращается. Первый дополнительный слой (так называемый S-слой) очень тонкий, а последующие слои характеризуются высоким содержанием целлюлозы, осевой ориентацией целлюлозных микрофибрилл, почти полным отсутствием ксилана и лигнина и присутствием рамногалактуронана-I [11–13]. Волокна собираются в пучки, расположенные по периферии стебля, обеспечивая его укрепление [9, 14]. Каждый пучок состоит из 15–50 элементарных волокон, заключенных в гелеобразную матрицу и соединенных между собой.

Техническое качество волокна — это комплексный признак, определяемый генотипом растений, окружающей средой и послеуборочной обработкой. Признаки волокна проявляют значительную степень вариабельности между сортами [9, 10, 14]. Наиболее важными морфологическими признаками, связанными с качеством волокна, являются следующие: техническая длина стебля (от основания стебля до первого цветочного разветвления), количество междоузлий и диаметр стебля. Сорта льна с более длинным техническим стеблем имеют компактные, плотные пучки луба, состоящие из длинных элементарных волокон. Растения с длинными междоузлиями и небольшим количеством листьев имеют более высокое качество волокна, так как пучки волокон разрываются в месте прикрепления листовых пластинок. У растений с более толстыми стеблями пучки волокон редкие и увеличенные, с хорошо развитыми одревеснев шими сердцевинами, поэтому они дают грубое и менее гибкое волокно с низким качеством прядения. Цилиндрическая форма стебля указывает на равномерное распределение элементарных волокон по его длине, что обеспечивает однородное созревание волокна и равномерное распределение цвета [14].

Ранее, используя однонуклеотидные полиморфизмы, мы идентифицировали ряд районов генома, контролирующие качество волокна у льна. Однако, у льна также обнаружена значительная структурная вариабельность в форме вариаций числа копий [15]. Мобилизация этого дополнительного источника геномной изменчивости поможет уточнить ранее идентифицированные районы, а также обнаружить новые, не маркированные однонуклеотидными полиморфизмами в изученной выборке сортов. В этой работе, используя вариацию числа копий в качестве маркеров, замещающих однонуклеотидные полиморфизмы, мы применили различные модели полногеномного поиска ассоциаций для идентификации районов, ассоциированных с признаком волокна.

МАТЕРИАЛЫ И МЕТОДЫ

В работе были использованы данные полногеномного секвенирования 100 образцов льна из коллекции Федерального научного центра лубяных культур (Институт льна, Торжок Тверской области, Россия). Секвенирование было выполнено с использованием технологии коротких парных прочтений по технологии Illumina Inc.; средняя глубина покрытия в образцах составила около 20×. Поиск вариаций копийности осуществлялся с помощью пакета программ CNVnator [16], версия 0.4.1. Общее число обнаруженных вариантов составило 8275 (7379 делеций и 896 дупликаций).

Полногеномный анализ ассоциаций проводили для следующих фенотипических признаков, связанных с качеством производимого волокна: техническая длина волокна, высота растения, число междоузлий, технический вес растения, расстояние между узлами (детальное описание фиенотипов приводится в работе [14]). Количественные фенотипические данные были нормализованы с помощью метода квантильной нормализации. Полногеномный анализ ассоциаций проводили с применением программы GAPIT3 [17], были использованы следующие статистические модели: GLM, MLM, SUPER, MLMM, FarmCPU и Blink. Для анализа использовали вариации копийности со значением MAF (minor allele frequency, частота минорного аллеля), не ниже 0.05, таким образом, общее число вариантов, использованных в анализе, составило 3805. Для определения наиболее значимых вариаций копийности результаты полногеномного поиска ассоциаций были отфильтрованы по критерию FDR (false discovery rate, уровень ложноположительных результатов) составлявшим 0.01.

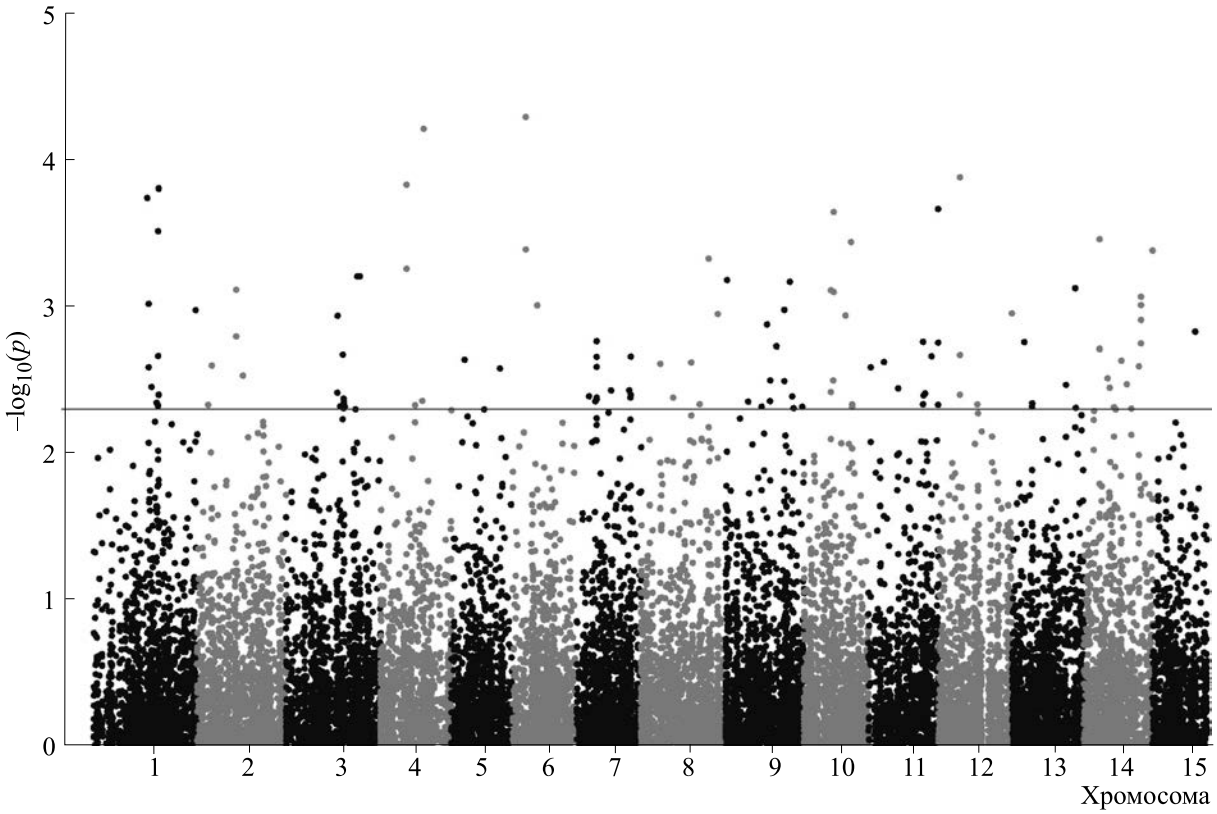

Рис. 1. Диаграмма Manhattan plot вариаций копийности. По оси абсцисс отложены номера хромосом, по оси ординат – значения логарифма *P*-value. Горизонтальная линия соответствует порогу по *P*-value в 0.005.

Функциональную аннотацию вариаций копийности, имеющих значимые ассоциации с использованными фенотипическими признаками, проводили путем поиска общих участков с белок-кодирующими генами льна, версия генома 2.0, аннотация из работы [18]

Ранее нами была описана метрика локальной стабильности генома, успешно примененная в анализе грибкового патогена льна *Fusarium oxysporum* f.sp. *lini* [19]. В данной работе мы также использовали этот подход. Метрика стабильности была рассчитана на всех 15 хромосомах льна, размер региона составлял 16384 п.о.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В результате полногеномного поиска ассоциаций нами было обнаружено 113 вариаций копийности, связанных с хозяйственными признаками, определяющими качество волокна. Результаты представлены в виде диаграммы Manhattan plot на рис. 1. Табл. 1 включает гены, общим числом 41, затронутые геномными регионами, содержащими варианты числа копий.

Следует отметить, что обнаруженные гены не содержатся в списке тех, которые были определены ранее с использованием однонуклеотидных замен [20]. Основной причиной этого является низкая частота встречаемости SNP в данных генах среди 306 образцов, проанализированных в предыдущей работе. Данный факт еще раз указывает на важность разработки альтернативных методов поиска полногеномных ассоциаций, не ограничивающихся только SNP.

Для всех участков генома, содержащих варианты числа копий, ассоциированные с признаками, определяющими качество волокна, нами было рассчитано среднее значение метрики локальной стабильности генома. Результаты анализа приведены на рис. 2. Тепловая карта демонстрирует наличие по крайней мере двух выраженных кластеров, отличающихся по значению метрики стабильности генома. Больший кластер соответствует регионам с низким значением метрики (среднее значения для регионов, входящих в кластер, составляет 4.6), меньший кластер содержит менее стабильные регионы со средним значением метрики 14.1 В табл. 1 гены, попадающие в регионы кластера 2, выделены курсивом. Среди них преобладают белки, вовлеченные в синтез и мо-

Таблица 1. Список генов, затронутых вариациями числа копий, ассоциированных с фенотипическими признаками качества волокна

Ген	Функциональная аннотация	Гомолог *A. thaliana*	Имя гена	Категория GO	Фенотип	CNV
Lus10006089	DEAD/DEAH box RNA helicase family protein	AT1G35530.2	–	GO:DNA primase activity GO:DNA replication, synthesis of RNA primer GO:ATP binding GO:DNA binding GO:hydrolase activity	Pht	DUP
Lus10018956	Major facilitator superfamily protein	AT3G05165.4	–	GO:integral component of membrane GO:transmembrane transporter activity GO:transmembrane transport	TeLen	DEL
Lus10033225	with no lysine (K) kinase 8	AT5G41990.1	ATWNK8, WNK8	GO:ATP binding GO:protein kinase activity GO:protein phosphorylation GO:protein serine/threonine kinase activity	Pht	DUP
Lus10033226	Yippee family putative zinc-binding protein	AT4G27745.1	–	–	Pht	DUP
Lus10033227	early nodulin-like protein 8	AT1G64640.1	AtENODL8, ENODL8	GO:electron transfer activity	Pht	DUP
Lus10033228	Major facilitator superfamily protein	AT1G64650.1	–	GO:integral component of membrane GO:molybdate ion transmembrane transporter activity GO:molybdate ion transport GO:transmembrane transporter activity GO:transmembrane transport	Pht	DUP
Lus10023481	*binding;RNA binding*	*AT2G39260.1*	–	*GO:RNA binding* *GO:protein binding* *GO:spore wall* *GO:sporulation resulting in formation of a cellular spore* *GO:ATP binding*	*TeLen*	*DEL*
Lus10038498	Arabidopsis thaliana gibberellin 2-oxidase 1	AT1G78440.1	ATGA2OX1, GA2OX1	GO:oxidation-reduction process GO:oxidoreductase activity	TeLen	DEL
Lus10008854	*defective in meristem silencing 3*	*AT3G49250.1*	*DMS3, IDN1*	–	*Ufilen*	*DEL*
Lus10008855	*defective in meristem silencing 3*	*AT3G49250.1*	*DMS3, IDN1*	–	*Ufilen*	*DEL*
Lus10022351	*NB-ARC domain-containing disease resistance protein*	*AT3G14470.1*	–	*GO:protein binding* *GO:ADP binding*	*Ninodes*	*DEL*
Lus10010513	RING-H2 group F1A	AT4G14220.1	RHF1A	GO:metal ion binding GO:zinc ion binding GO:anaphase-promoting complex-dependent catabolic process GO:anaphase-promoting complex GO:cullin family protein binding GO:ubiquitin protein ligase activity	Tewgt	DUP
Lus10010514	P-loop containing nucleoside triphosphate hydrolases superfamily protein	AT5G19210.2	–	GO:ATPase activity GO:ATP binding GO:nucleic acid binding GO:DNA binding GO:hydrolase activity	Tewgt	DUP
Lus10010515	ADP glucose pyrophosphorylase large subunit 1	AT5G19220.1	ADG2, APL1	GO:biosynthetic process GO:nucleotidyltransferase activity GO:transferase activity, transferring phosphorus-containing groups	Tewgt	DUP
Lus10010516	Glycoprotein membrane precursor GPI-anchored	AT3G06035.1	–	GO:integral component of membrane	Tewgt	DUP
Lus10010517	peroxisomal adenine nucleotide carrier 1	AT3G05290.1	PNC1	–	Tewgt	DUP
Lus10016577	syntaxin of plants 52	AT1G79590.1	ATSYP52,SYP52	GO:cell adhesion GO:Golgi vesicle transport GO:membrane	Ufilen	DUP
Lus10016578	TATA BOX ASSOCIATED FACTOR II 59	AT1G04950.1	ATTAF6, TAF6, TAFII59	GO:DNA binding GO:DNA-templated transcription, initiation GO:transcription initiation from RNA polymerase II promoter	Ufilen	DUP
Lus10016579	homolog of bacterial cytokinesis Z-ring protein FTSZ 1-1	AT5G55280.1	ATFTSZ1-1, CPFTSZ, FTSZ1-1	GO:GTPase activity	Ufilen	DUP

Lus10003208	Dof-type zinc finger DNA-binding family protein	AT4G24060.1	–	GO:DNA binding GO:regulation of transcription, DNA-templated	TeLen	DEL
Lus10042690	*Tudor/PWWP/MBT domain-containing protein*	*AT2G48160.1*	–	*GO:mitochondrial large ribosomal subunit* *GO:structural constituent of ribosome*	*Ninodes*	*DEL*
Lus10032760	Primosome PriB/single-strand DNA-binding	AT1G47720.1	OSB1	GO:single-stranded DNA binding GO:nucleic acid binding GO:regulation of transcription, DNA-templated	Pht	DEL
Lus10036335	Heat shock protein DnaJ with tetratricopeptide repeat	AT4G02100.1	–	GO:extracellular region GO:hormone activity GO:protein binding	Tewgt	DEL
Lus10008989	P-loop containing nucleoside triphosphate hydrolases superfamily protein	AT3G18600.1	–	GO:ATP binding GO:nucleic acid binding GO:ATP:ADP antiporter activity GO:integral component of membrane GO:nucleotide transport GO:DNA binding GO:hydrolase activity GO:cell adhesion GO:extracellular region GO:ossification	Ufilen	DEL
Lus10026328	polyol/monosaccharide transporter 5	AT3G18830.1	ATPLT5, ATPMT5, PMT5	GO:integral component of membrane GO:transmembrane transporter activity GO:transmembrane transport	Ninodes	DEL
Lus10031392	*xyloglucan endotransglucosylase/hydrolase 16*	*AT3G23730.1*	*XTH16*	*GO:carbohydrate metabolic process* *GO:hydrolase activity, hydrolyzing O-glycosyl compounds* *GO:apoplast* *GO:cellular glucan metabolic process* *GO:cell wall* *GO:xyloglucan:xyloglucosyl transferase activity*	*Pht*	*DEL*
Lus10031393	*xyloglucan endotransglucosylase/hydrolase 16*	*AT3G23730.1*	*XTH16*	*GO:carbohydrate metabolic process* *GO:hydrolase activity, hydrolyzing O-glycosyl compounds* *GO:apoplast* *GO:cellular glucan metabolic process* *GO:cell wall* *GO:xyloglucan:xyloglucosyl transferase activity*	*Pht*	*DEL*
Lus10026103	*nudix hydrolase homolog 3*	*AT1G79690.1*	*atnudt3, NUDT3*	*GO:hydrolase activity*	*Tewgt*	*DUP*
Lus10026104	*RING/U-box superfamily protein*	*AT3G14250.1*	–	*GO:metal ion binding*	*Tewgt*	*DUP*
Lus10026105	*rhodanese-like domain-containing protein / PPIC-type PPIASE domain-containing protein*	*AT5G19370.1*	–	*GO:peptidyl-prolyl cis-trans isomerase activity*	*Tewgt*	*DUP*
Lus10026106	*Galactose mutarotase-like superfamily protein*	*AT5G57330.1*	–	*GO:carbohydrate metabolic process* *GO:isomerase activity*	*Tewgt*	*DUP*
Lus10026107	*FASCICLIN-like arabinogalactan protein 14 precursor*	*AT3G12660.1*	*FLA14*	–	*Tewgt*	*DUP*
Lus10026108	*Protein kinase superfamily protein*	*AT1G79640.1*	–	*GO:ATP binding* *GO:protein kinase activity* *GO:protein phosphorylation* *GO:protein dimerization activity*	*Tewgt*	*DUP*
Lus10031928	indeterminate(ID)-domain 4	AT2G02080.1	AtIDD4,IDD4	-	Ufilen	DEL
Lus10006824	receptor like protein 37	AT3G23110.1	AtRLP37,RLP37	GO:protein binding	Ninodes	DEL
Lus10006825	receptor like protein 7	AT1G47890.1	AtRLP7,RLP7	GO:protein binding	Ninodes	DEL
Lus10035647	xylem NAC domain 1	AT5G64530.1	ANAC104,XND1	GO:DNA binding GO:regulation of transcription, DNA-templated	Tewgt	DEL
Lus10035648	xylem NAC domain 1	AT5G64530.1	ANAC104,XND1	GO:DNA binding GO:regulation of transcription, DNA-templated	Tewgt	DEL
Lus10008323	heat shock protein 70 (Hsp 70) family protein	AT5G28540.1	BIP1	GO:ATPase activity GO:ATP binding	Tewgt	DEL

Lus10008324	Heat shock protein 70 (Hsp 70) family protein	AT1G09080.1	BIP3	GO:ATPase activity GO:ATP binding	Tewgt	DEL
Lus10018096	*Protein kinase superfamily protein*	*AT3G19300.1*	–	*GO:ATP binding* *GO:ATP binding* *GO:protein kinase activity* *GO:protein kinase activity* *GO:protein phosphorylation* *GO:protein phosphorylation*	*TeLen* *Ufilen*	*DEL*

Примечание. Курсивом выделены гены, находящиеся в областях с высоким уровнем нестабильности генома. Акронимы фенотипов расшифровываются следующим образом: TeLen – техническая длина волокна, Ufilen – длина междоузлий, Tewgt – технический вес волокна, Ninodes – число междоузлий, Pht – высота растений (для расшифровки фенотипов см. работу [14]).

дификацию компонентов клеточной стенки (Lus10026107, Lus10026106, Lus10031393, Lus10031392), ремоделинг хромосом/сайленсинг генов (Lus10042690, Lus10008854, Lus10008855), а также в ответ растения на стресс (Lus10018096, Lus10026108, Lus10022351), что позволяет предположить большую пластичность данных регионов с точки зрения изменчивости, и, таким образом, их потенциальную важность для создания новых сортов с заданными хозяйственными свойствами.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при финансовой поддержке Российского научного фонда (грант № 20-14-00072).

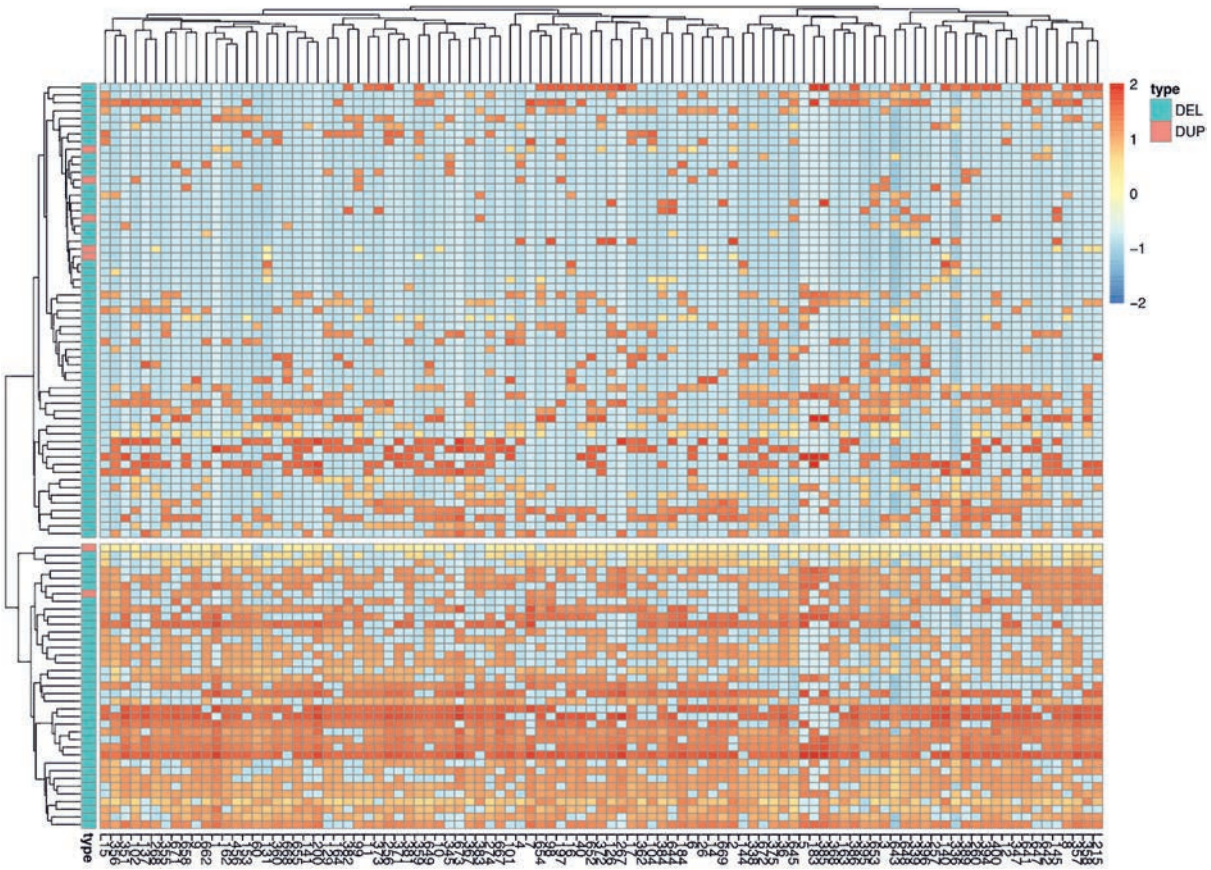

Рис. 2. Тепловая карта метрики локальной стабильности генома в участках, соответствующих вариациям копийности (строки) для исследованных образцов (столбцы). Дополнительная панель слева включает тип варианта числа копий (делеция – DEL, дупликация – DUP).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

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

СПИСОК ЛИТЕРАТУРЫ

1. S. V. Nuzhdin, M. L. Friesen, and L. M. McIntyre, Trends Genet., **28**, 421 (2012).
2. P. K. Gupta, P. L. Kulwal, and V. Jaiswal, Adv. Genet., **104**, 75 (2019).
3. T. A. Manolio, et al., Nature, **461**, 747 (2009).
4. E. E. Eichler, et al., Nat. Rev. Genet., **11**, 446 (2010).
5. P. H. Sudmant, et al., Nature, **526**, 75 (2015).
6. T. H. Shaikh, Curr Genetic Med. Reports, **5**, 183 (2017).
7. A. Zmienko, et al., Plant Cell, **32**, 1797 (2020).
8. A. Dolatabadian, D. A. Patel, D. Edwards, and J. Batley, Theor. Appl. Genet., **130**, 2479 (2017).
9. C. Goudenhooft, A. Bourmaud, and C. Baley, Front. Plant Sci., **10**, 411 (2019).
10. C. Goudenhooft, A. Bourmaud, and C. Baley, Industrial Crops & Products, **97**, 56 (2017).
11. E. J. Mellerowicz and T. A. Gorshkova, J. Exp. Bot., **63**, 551 (2012).
12. M. J. Roach, et al., Plant Physiol., **156**, 1351 (2011).
13. T. Gorshkova, et al., Sci. Rep-uk, **8**, 14570 (2018).
14. T. Rozhmina, M. Bankin, A. Samsonova, et al., Data Brief, **37**, 107224 (2021).
15. М. А. Дук, А. А. Канапин, А. А. Самсонова и др., Биофизика, **67**, 234 (2022).
16. A. Abyzov, A. E. Urban, M. Snyder, and M. Gerstein, Genome Res., **21**, 974 (2011).
17. J. Wang and Z. Zhang, Genom Proteom Bioinform., **19**, 629–640 (2021).
18. F. M. You and S. Cloutier, Methods Protoc., **3** (2020). DOI:10.3390/mps3020028
19. A. Kanapin, et al., Mol. Plant Microbe Interact., **33**, 1112 (2020).
20. A. Kanapin, et al., Int. J. Mol. Sci., **23**, 14536 (2022).

Genome-Wide Association Study of Copy Number Variation in Flax Through the Lens of Genome Integrity

M.A. Duk*, A.A. Kanapin*, T.A. Rozhmina, and A.A. Samsonova***

**St. Petersburg State University, Univeritskaya nab. 7/9, St. Petersburg, 199034 Russia*

***Flax Institute – a separate subdivision of the Federal Scientific Center for Bast Crops, ul. Lunacharskogo 35, Thorzhok, Tver Region 172002 Russia*

Classical methods for identification of genetic variants associated with certain macroscopic phenotypic traits are, as a rule, limited to analyses of single nucleotide polymorphisms. Copy number variations, and more broadly structural variants may provide a plethora of useful information due to the magnitude of the changes they induce. However, their use in genome-wide association studies is seriously limited mostly due to the uncertainties in their discovery (i.e., failure to resolve an event with nucleotide resolution) by computational algorithms from genomic data. Nevertheless, in certain cases, such analyses are possible and may still yield valuable results. Our recent work has revealed genetic variants (single nucleotide polymorphisms) possibly related to phenotypic traits determining fibre quality. Here, we decided to extend the analyses to structural variants, namely copy number variations. Importantly, we use a novel high-coverage dataset allowing for accurate prediction of copy number variations. Overall, we compiled a list of 41 candidate genes associated with five quantitative phenotypic traits. Furthermore, the genome stability metric developed earlier facilitated stratification of copy number variant loci with regard to their stability. On the whole, our analyses suggest that the genomic regions less resilient to external and internal stresses are more susceptible to changes associated with the studied phenotypic traits.

Keywords: copy number variation, genome integrity, Linum usitasissimum L., flax fiber, genome-wide association study