







doi 10.18699/vjgb-25-131

## Транспозоны льна: роль в генетическом разнообразии, окультуривании и детерминации хозяйственно ценных признаков

М.А. Дук <sup>2</sup>, В.А. Станин<sup>1</sup>, А.А. Канапин <sup>1</sup>, А.А. Самсонова <sup>1</sup>, Т.А. Рожмина<sup>3</sup>, М.Г. Самсонова <sup>1</sup> <sup>1</sup> Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия<sup>2</sup> Физико-технический институт им. А.Ф. Иоффе Российской академии наук, Санкт-Петербург, Россия<sup>3</sup> Федеральный научный центр лубяных культур, Торжок, Россия m.g.samsonova@gmail.com







**Аннотация.** Лен – важная сельскохозяйственная культура многофункционального использования. Разнонаправленная селекция на содержание масла в семенах и волокна в стеблях привела к возникновению двух морфотипов – льна-долгунца и льна масличного. Ранее, используя однонуклеотидные полиморфизмы, мы охарактеризовали генетическое разнообразие у 306 образцов льна из коллекции российского Федерального научного центра лубяных культур. Вместе с тем более крупные структурные вариации также играют существенную роль в формировании агрономически важных признаков растений и могут быть применены для дальнейшего улучшения льна. Здесь мы задействовали ту же коллекцию льна, чтобы предсказать сайты новых инсерций транспозонов и оценить роль таких инсерций в формировании агрономически важных признаков, а также в процессе окультуривания льна. Мы обнаружили 588480 новых сайтов инсерций транспозонов, не содержащихся в референсном геноме льна (сборка NCBI ASM22429v2), из которых большая часть приходилась на ретротранспозоны суперсемейств *Copia* и *Gypsy*, а среди ДНК транспозонов чаще всего встречались сайты инсерции суперсемейств *MULE-MuDR*, *hAT* и *CMC-EnSpm*. В отличие от однонуклеотидных полиморфизмов, которых было значительно больше у льна масличного, чем у долгунцов, мы не обнаружили столь существенной разницы в числе инсерций разных семейств транспозонов на одно растение у образцов разного морфотипа. Анализ геномных областей, затронутых недавними селекционными усилиями, выявил в общей сложности 61 район-кандидат, из которых 18 районов пересекались с QTL, ассоциированными с важными агрономическими признаками. Интересно, что пять участков уменьшения генетического разнообразия у культурных сортов и кражей при сравнении их со староместными сортами также были идентифицированы как участки уменьшения разнообразия при использовании в качестве маркеров однонуклеотидных полиморфизмов. При полногеномном поиске ассоциаций найдено 50 инсерций TE, ассоциированных с разными фенотипическими признаками, причем многие ассоциации подтверждаются несколькими моделями или обнаруживаются в данных по нескольким годам. Таким образом, сайты инсерции транспозонов – важный источник генетического разнообразия у льна наряду с однонуклеотидными полиморфизмами, что позволяет использовать их для дальнейшего улучшения культуры при селекции.

**Ключевые слова:** лен; *Linum usitatissimum*; транспозоны; GWAS; генетическое разнообразие; селекция

**Для цитирования:** Дук М.А., Станин В.А., Канапин А.А., Самсонова А.А., Рожмина Т.А., Самсонова М.Г. Транспозоны льна: роль в генетическом разнообразии, окультуривании и детерминации хозяйственно ценных признаков. *Вавиловский журнал генетики и селекции*. 2025;29(8):1267-1276. doi 10.18699/vjgb-25-131

**Финансирование.** Работа выполнена при поддержке гранта РНФ № 23-16-00037.

## Flax transposons: unraveling their impact on domestication and agronomic trait variation

М.А. Duk <sup>2</sup>, V.A. Stanin<sup>1</sup>, A.A. Kanapin <sup>1</sup>, A.A. Samsonova <sup>1</sup>, T.A. Rozhmina<sup>3</sup>, M.G. Samsonova <sup>1</sup> <sup>1</sup> Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia<sup>2</sup> Ioffe Institute of the Russian Academy of Sciences, St. Petersburg, Russia<sup>3</sup> Federal Research Center for Bast Fiber Crops, Torzhok, Russia m.g.samsonova@gmail.com

**Abstract.** Flax is an important agricultural crop with multifunctional uses. Diversified breeding for oil content in seeds and fiber in stems has led to the emergence of two morphotypes – fiber flax and oilseed flax. Previously, using single nucleotide polymorphisms (SNPs), we characterized the genetic diversity of 306 flax samples from the collection of the Russian Federal Research Center for Bast Crops. However, larger structural variations, such as mobile genetic elements, also play a significant role in shaping agronomically important plant traits and can be used for

further flax improvement. Here, we used the same flax collection to predict sites of new transposon insertions and to assess the role of such insertions in the formation of agronomically important traits, as well as in the process of flax domestication. We discovered 588,480 new transposon insertion sites not present in the reference flax genome (NCBI assembly ASM22429v2), the majority of which were attributed to retrotransposons of the *Copia* and *Gypsy* superfamilies, while among DNA transposons, insertion sites of the *MULE-MuDR*, *hAT*, and *CMC-EnSpm* superfamilies were most common. Unlike SNPs, which were significantly more numerous in oilseed flax than in fiber flax, we did not find such a substantial difference in the number of insertions of different transposon families per plant among samples of different morphotypes. Analysis of genomic regions affected by recent breeding efforts revealed a total of 61 candidate regions, of which 18 regions overlapped with QTLs associated with important agronomic traits. Interestingly, 5 regions of reduced genetic diversity in kryazhs and cultivars compared to landraces were also identified as regions of reduced diversity when using single nucleotide polymorphisms as markers. A genome-wide association study (GWAS) identified 50 TE insertions associated with different phenotypic traits, with many associations confirmed by multiple models or detected in data from multiple years. Thus, transposon insertion sites are an important source of genetic diversity in flax, alongside single nucleotide polymorphisms, making them suitable for further crop improvement in breeding.

**Key words:** flax; *Linum usitatissimum*; transposons; GWAS; genetic diversity; breeding

**For citation:** Duk M.A., Stanin V.A., Kanapin A.A., Samsonova A.A., Rozhmina T.A., Samsonova M.G. Flax transposons: unraveling their impact on domestication and agronomic trait variation. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov J Genet Breed*. 2025;29(8):1267-1276. doi 10.18699/vjgb-25-131

## Введение

Лен – важная сельскохозяйственная культура, выращиваемая для получения как волокна, так и масла, используемая во многих сферах, таких как производство лаков и красок, линолеума, в производстве композитов, текстильной и пищевой промышленности (Goudenhooft et al., 2019). Длительная селекция льна на содержание масла в семенах и волокна в стеблях привела к появлению двух морфотипов – лена-долгунца и льна масличного. Для льна-долгунца характерна меньшая ветвистость, большие длина стебля и высота растения, тогда как лен масличный характеризуется большим количеством семян, а значит, большим числом соцветий при меньшей длине основного стебля. В конце XIX–начале XX века Россия была основным поставщиком качественного льняного волокна, получаемого из кражей – староместных сортов, имеющих смешанное происхождение из Индо-Афганского центра разнообразия и Плодородного полумесьяца и полученных в результате народной селекции. Кражи и российские местные сорта внесли решающий вклад в генофонд современных сортов льна (Helbaek, 1959; Duk et al., 2021).

Ранее с использованием однонуклеотидных полиморфизмов мы охарактеризовали генетическое разнообразие у 306 образцов льна из коллекции Федерального научного центра лубяных культур (ФНЦ ЛК, Россия). Мы обнаружили значительную дифференциацию между популяциями льна масличного и льна-долгунца, а также выявили регионы генома, затронутые современной селекцией (Kanapin et al., 2022; Duk и др., 2025).

Вместе с тем более крупные структурные вариации, такие как инсерции транспозонов (ТЕ), также играют существенную роль в формировании агрономически важных признаков растений и могут быть использованы для дальнейшего улучшения культуры льна. Известно, что ТЕ составляют большую часть генома растений (Quesneville, 2020) и их встройки могут приводить к перестройке генома, сайленсингу генов и изменению генных сетей (Bourque et al., 2018), а также быть материалом для возникновения новых генов и некодирующих РНК (Pulido, Casacuberta, 2023).

Транспозоны условно разделяют на два класса. К классу I относятся ретротранспозоны, которые увеличивают число своих копий в геноме, встраивая их через РНК-посредник (Mhiri et al., 2022), в результате чего длинные терминальные повторы могут составлять до 90 % генома растения. К классу II относят ДНК-транспозоны, работающие по принципу «вырезать-вставить»; они перемещаются по геному и обычно не увеличивают количество своих копий. Наибольшая активность ТЕ наблюдается в периоды стресса (Schrader, Schmitz, 2019). Чаще всего новые инсерции оказывают отрицательное воздействие и удаляются из популяции, однако иногда они могут способствовать адаптации растений к неблагоприятным условиям внешней среды (Niu et al., 2019) и благодаря этому сохраниться в популяции как результат действия положительного отбора.

В настоящей работе мы использовали ту же коллекцию из 306 образцов льна ФНЦ ЛК (Duk et al., 2021; Kanapin et al., 2022), чтобы предсказать сайты новых инсерций ТЕ в группах образцов разного морфотипа и селекционного статуса и сравнить характер распределения сайтов инсерций ТЕ и полиморфных сайтов по геному. Мы также оценили роль инсерций ТЕ в формировании агрономически важных признаков и в процессе окультуривания льна.

## Материалы и методы

Материалом исследования были 306 образцов льна из коллекции ФНЦ ЛК (г. Торжок, Россия), из которых 182 образца по морфотипу относятся к льну-долгунцу и 120 – к льну масличному: 99 образцов – к «межеумкам», 16 – к «кудряшам» и 5 – к крупносеменному льну; морфотип остальных четырех образцов не определен. По селекционному статусу образцы разделяются на староместные сорта – 40 образцов, кражи – 31 образец, и 230 культурных сортов – 89 селекционных линий и 141 селекционный сорт.

ДНК была выделена из листьев с помощью набора DNeasy Plant Mini Kit (Qiagen, Нидерланды). Секвенирование ДНК проводили в BGI с помощью протокола Illu-

mina, генерирующего парные прочтения длиной 150 п. о. Получено 1143.850625 Гб и 7626 млн сырых прочтений со средним покрытием генома  $9.3\times$  или 3.7 Гб на образец.

Предсказание координат инсерций ТЕ выполнено с применением пакета программ TEMP2 (Yu T. et al., 2021), использован режим поиска insertion2, при котором осуществляется идентификация ТЕ, не содержащихся в референсном геноме (сборка NCBI ASM22429v2). Консенсусные последовательности ТЕ, применявшиеся при поиске, были предсказаны *de novo* с помощью модуля TEdenovo из пакета REPET (Flutre et al., 2011). Поскольку существующие методы анализа не обеспечивают определение координат вставки с точностью до одного нуклеотида, дополнительно проведена операция унификации для инсерций одного и того же ТЕ, имеющих разницу координат, не превышающую удвоенную длину фрагмента секвенирования (в данном случае 200 п. о.). Всем вставкам данного ТЕ, разность между координатами начала вставки которых не превышала указанную длину, была присвоена одна и та же координата, соответствующая центральной позиции региона, содержащего эти инсерции. Для дальнейшего анализа данные были конвертированы в формат VCF. Валидация ключевых инсерций выполнена с помощью программы Integrative Genomics Viewer, которая делает возможной визуализацию выравнивания ридов в месте инсерции.

Для популяционного анализа и поиска ассоциаций данные были преобразованы таким образом, что инсерции ТЕ у разных образцов, расположенные на расстоянии менее 50 п. н., считались одним событием встройки, а позицией инсерции была середина этого окна. Такой подход позволил уменьшить вероятные неточности относительно предсказания сайтов инсерций ТЕ и увеличить количество встроек, найденных более чем у 5 % образцов, что уменьшает вероятность случайного ложноположительного определения ассоциации варианта с признаком. Выбор размера окна 50 п. н. был обусловлен тем, что при таком размере окна количество инсерций, обнаруживаемых более чем у 5 % образцов, было максимальным и уменьшалось при увеличении или уменьшении размера окна.

Для анализа популяционной структуры применяли метод главных компонент, а также метод ADMIXTURE (Alexander et al., 2009) с количеством популяций  $K = 2, 3, 4, 5$ . Для филогенетического анализа использовали функцию построения дерева по методу ближайших соседей из пакета *ape* и функцию визуализации из пакета *ggtree* для R (Yu G., 2020). Для оценки генетического разнообразия, обусловленного инсерциями ТЕ, и расчета индекса фиксации (*Fst*) был взят инструмент VCFtools (Danecsek et al., 2011) с размером окна 200 kb. Размер окна был выбран в соответствии со средним размером участков, найденных в качестве «горячих точек» в пакете PrimatR для R (<https://github.com/daewo000/primatR>).

Чтобы выявить возможные районы генома, испытывавшие действие отбора при селекции льна масличного и льна-долгунца, мы сравнили сорта с соответствующими староместными сортами и кряжами. В каждом сравнении мы рассчитывали статистики *Fst* и  $ROD = 1 - \pi_1/\pi_2$ , где  $\pi$  – генетическое разнообразие соответствующей группы образцов. Области генома с высокими популяционными

различиями между двумя группами (наивысшие значения *Fst*, верхние 5 % всего генома и верхние 2.5 % значений *ROD*) рассматривались в качестве возможных областей со следами селекции.

При поиске ассоциаций использованы значения 13 фенотипических признаков, измеренных у растений, выращенных на экспериментальных полях ФНЦ ЛК в 2019 (одно выращивание), 2020 (два выращивания со сдвигом в две недели) и 2021 (одно выращивание) годах (Kanašin et al., 2022): *DSI* – индекс тяжести фузариозного увядания; *EFL* – длина элементарного волокна; *FC* – содержание волокна; *FW* – вес волокна; *IL* – длина соцветия; *NI* – число междоузлий; *Oil* – содержание масла в семенах; *PH* – высота растения; *Nsed* – количество семян с растения; *STI* – сбежистость; *TL* – техническая длина стебля; *TW* – вес технической части растения; *Tswgt* – вес 1000 семян. Для анализа взяты модели Blink, FarmCPU, SUPER, MLM, GLM в пакете GAPIT3 (Wang, Zhang, 2021) с порогом  $FDR.rate = 0.9$ . Для связи ассоциаций с генами использовали аннотацию генома, предоставленную группой S. Cloutier (You, Cloutier, 2020). Найденные ассоциации были также подтверждены прямым *t*-тестом между значениями признака у образцов, содержащих вставку и не содержащих ее для исключения ложноположительных ассоциаций. Для расчета эффекта вставки на признак сравнивали значения признака у образцов, содержащих и не содержащих вставку, достоверность подтверждалась статистическим тестом Манна–Уитни (Mann, Whitney, 1947).

## Результаты

### Состав мобилома льна

Всего в коллекции выявлено 588480 новых инсерций ТЕ в 306 образцах, из них 172984 инсерции не были классифицированы (рис. 1). Из классифицированных больше всего инсерций (41 % от общего числа или 58 % от классифицированных инсерций) приходится на ТЕ суперсемейства *Copia* и порядка 15 % всех инсерций или 20 % классифицированных инсерций, – на ТЕ суперсемейства *Gypsy*. Ретротранспозоны порядка *LINE* составляют 1 % инсерций в коллекцию. Среди ТЕ класса II часто встречаются элементы суперсемейства *Mutator* (*MULE-MuDR*) – около 6 % всех вставок. На ДНК-транспозоны суперсемейства *hAT* приходится 4.3 % всех вставок, *CMC-EnSpm* составляют лишь 1.3 % классифицированных инсерций.

Визуализация данных о вставках ТЕ методом главных компонент показала, что первая главная компонента дифференцирует образцы разного морфотипа (льна-долгунца от льна масличного) только в случае ТЕ суперсемейств *Copia* и *Gypsy*, а также ТЕ семейства *hAT-Ac* (Приложение, рис. S1)<sup>1</sup>.

### Ландшафт мобилома льна

Расположение сайтов инсерций ТЕ относительно генов льна различалось для разных суперсемейств. Внутри известных генов или на расстоянии менее 2000 п. н. от генов располагалось 22 % всех вставок (рис. 2, табл. S1).

<sup>1</sup> Рис. S1–S4 и табл. S1–S5 Приложения см. по адресу: <https://vavilovj-icg.ru/download/pict-2025-29/appx49.pdf>

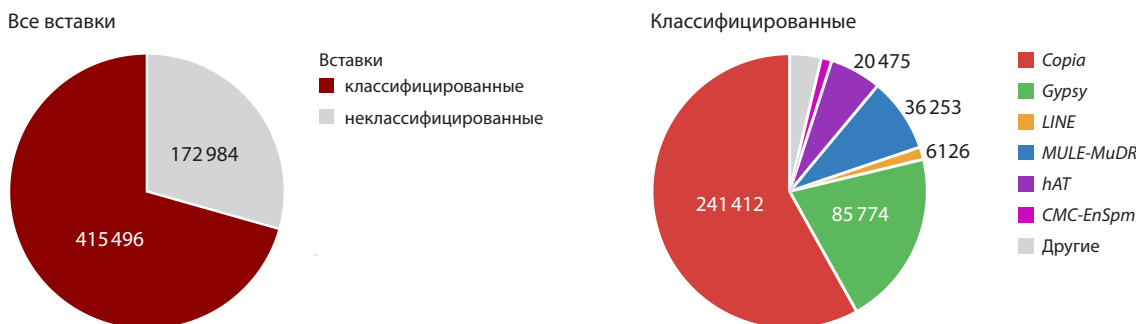


Рис. 1. Общее количество новых вставок TE и количество вставок отдельных суперсемейств TE в образцах коллекции.

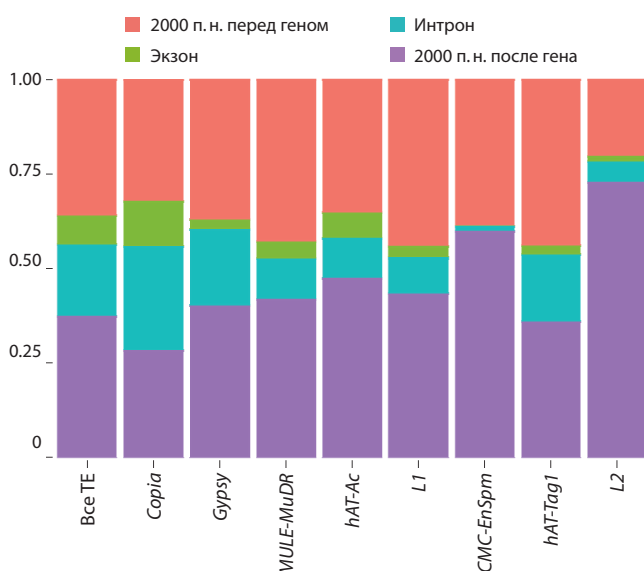


Рис. 2. Расположение сайтов инсерций TE относительно генов.

А 85 % инсерций TE суперсемейств *Gypsy* и 74 % инсерций TE суперсемейства *CMC-EnSpm* располагались далеко (>2000 п. н.) от известных генов, в отличие от TE *LINE* и *hAT*, для которых около половины инсерций находится вблизи генов (<2000 п. н.). TE *Copia* в 1.4 раза чаще встраивались в гены, как в интроны, так и в экзоны, чем в среднем у вставок всех типов. TE *Gypsy*, попавшие в ген, в 1.3 раза чаще встречались в интронах, чем в целом по всем вставкам TE, и в 3 раза реже в экзонах. Также инсерции в экзоны практически не наблюдались у транспозонов *CMC-EnSpm* и *L2*. В целом 29 % инсерций TE в ген приходилось на экзон и 71 % – на интрон.

#### Использование данных об инсерциях TE для анализа популяционной структуры и родства образцов

Популяционную структуру образцов коллекции, а также родство между ними исследовали, используя данные об инсерциях TE как замещающих однонуклеотидные полиморфизмы. После снижения размерности методом главных компонент данные были визуализированы для выявления разделения образцов на группы (рис. 3, а, б). Дифференциация образцов по морфотипу наблюдалась только по второй и третьей принципиальным компонентам, хотя

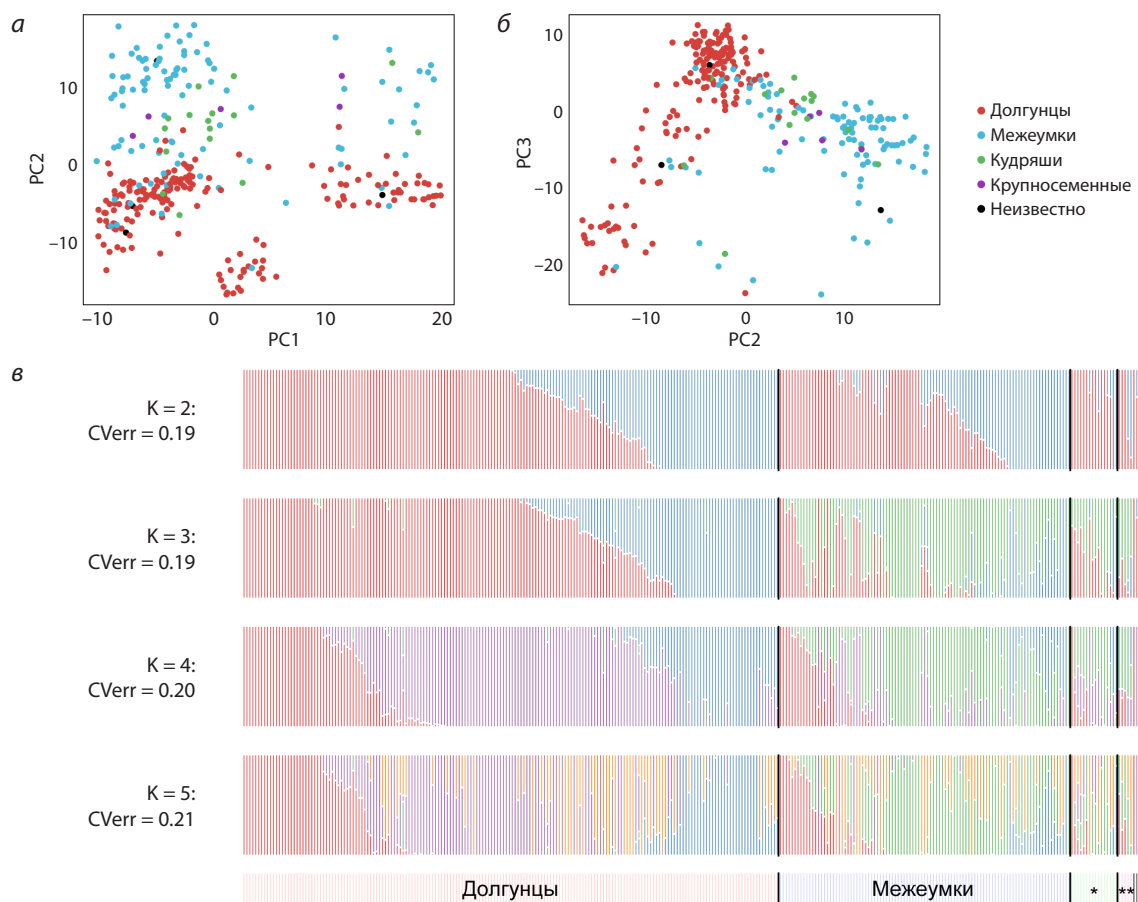
и оставалась нечеткой. Группировка образцов по селекционному статусу отсутствовала (графики не приведены). При запуске программы ADMIXTURE лучшее количество популяций по оценке кросс-валидационной ошибки было определено как  $K = 2$ , однако для  $K = 3$  ошибка была лишь немногим больше (см. рис. 3, в), при дальнейшем увеличении числа популяций ошибка растет. При количестве популяций больше 2 можно видеть наличие у образцов льна масличного примеси популяции, отмеченной на рис. 3 зеленым цветом. У групп с разным селекционным статусом заметных различий в картине примесей не наблюдалось.

На филогенетическом древе (рис. 4) образцы образуют три клады, две из которых, клада I и клада II, практически полностью состоят из образцов льна-долгунца, а клада III содержит практически все образцы льна масличного.

Большая часть кряжей попадает в кладу I, что отражает их близкое родство и общее происхождение (Duk et al., 2021). Как оказалось, каждая из клад обогащена инсерциями некоторых TE. Всего обнаружено 10 последовательностей TE (табл. S2), характерных для клады I, из которых пять принадлежат суперсемейству *Copia*, две – *Gypsy* и по одной – *RC/Helitron* и *hAT-Ac*. Клады II и III оказались обогащенными соответственно двумя и тремя последовательностями TE, принадлежащими к суперсемейству *Copia*.

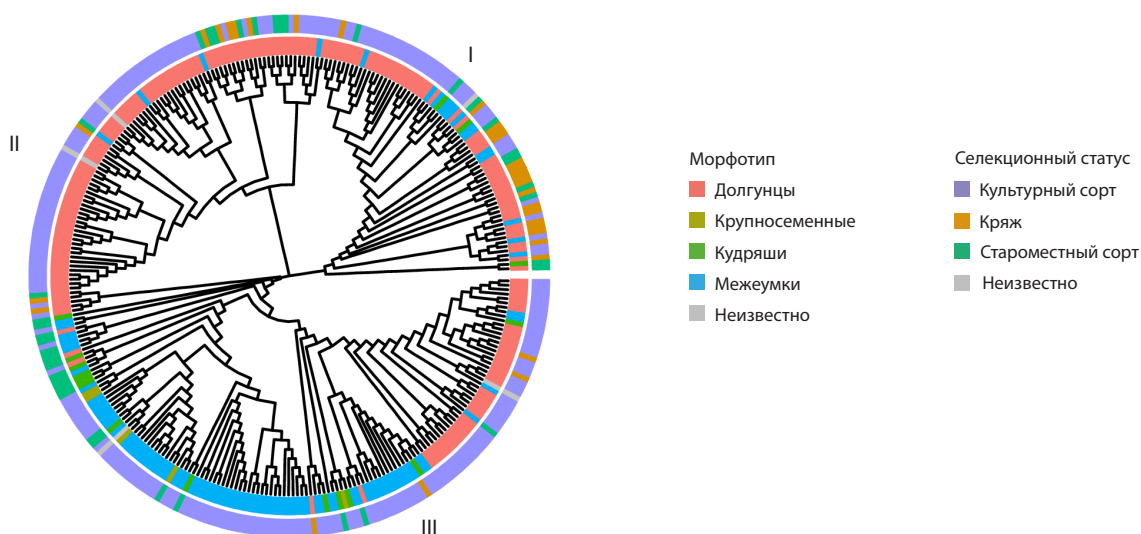
#### Роль инсерций транспозонов в формировании агрономически важных признаков растений

Для выявления следов селекции в геномах образцов проведен поиск районов, которые демонстрируют существенное снижение разнообразия: а) в сортах льна-долгунца по сравнению с местными сортами льна-долгунца или кряжами; б) в кряжах по сравнению с местными сортами льна-долгунца и в) в масличных сортах по сравнению с местными масличными сортами. Также сравнивалось генетическое разнообразие между сортами льна-долгунца и льна масличного для выявления следов селекции по агрономическим признакам, дифференцирующим эти морфотипы. Сопоставляя упомянутые выше группы, мы идентифицировали в общей сложности 61 район-кандидат (рис. 5), из которых 9 были найдены для двух сравнений. Из этих районов 18 пересекаются с QTL (quantitative trait locus – локусы количественных признаков), ассоциированными с важными агрономическими признаками (You, Cloutier, 2020) (табл. S3).



**Рис. 3.** Популяционная структура образцов коллекции.

а – график для первой и второй главных компонент с раскраской по морфотипу; б – график для второй и третьей главных компонент с раскраской по морфотипу; в – результаты ADMIXTURE для числа вероятных популяций K = 2–5. CVer – значение кросс-валидационной ошибки при анализе; \* кудряши; \*\* крупносеменные образцы.



**Рис. 4.** NJ-дерево, раскрашенное по морфотипу и селекционному статусу.

При сравнении образцов льна масличного и льнов-долгунцов только один выявленный сигнал селекционного выметания (selective sweeps) у долгунцов пересекается с QTL QOIL-Lu6.4 на хромосоме 6, отвечающим за содержание масла (см. табл. S3). При сравнении по селекцион-

ному статусу у культурных сортов обоих морфотипов наблюдалось уменьшение разнообразия в участках хромосомы 12, пересекающихся с QTL QIOD-Lu12.3, QLIN-Lu12.3, QLIO-Lu12.3 и ассоциированных с содержанием йода, линолевой и линоленовой кислот соответственно.



**Рис. 5.** Районы генома льна, демонстрирующие существенную разницу в разнообразии при сравнении групп образцов разного селекционного статуса и морфотипа. Сопоставлены: сорта льна-долгунца и льна масличного (Д/М), сорта льна-долгунца и кражи (Д/К), культурные и староместные сорта льна-долгунца (Д/С), кражи и староместные сорта льна-долгунца (К/С), сорта льна масличного и староместные сорта (М/С). QTL – QTL, опубликованные в (You, Cloutier, 2020).

У культурных сортов льна-долгунца сигналы селекционного выметания отмечены в районах этой же хромосомы, пересекающихся соответственно с QTL uq.C12–1 и uq.C12–1, в районе хромосомы 3, пересекающемся с Lu3-25559600, в районе хромосомы 8, пересекающемся с QOLE-Lu8.1, в районе хромосомы 9, пересекающемся с QSTE-Lu9.2, и в районах хромосомы 6, пересекающихся Lu2564 и QOIL-Lu6. Эти QTL ассоциированы с высотой растения и длиной стебля, содержанием семенной слизи, содержанием олеиновой и стеариновой кислот и содержанием масла соответственно.

У кражей сигналы селекционного выметания обнаружены в: 1) районах хромосомы 7, пересекающихся с QPM-crc-LG7 и QPAL-Lu7.3, ассоциированных соответственно с заболеваемостью мучнистой росой и содержанием пальмитиновой кислоты; 2) районах хромосомы 2, пересекающихся с scaffold43-1111162 и QOIL-Lu2.1, ассоциированных соответственно с весом 1000 семян и содержанием масла, и 3) в районах хромосомы 3, пересекающихся с QLio-LG3.1, QLin-LG3.1, Marker4371 и scaffold156-76129, для которых показана ассоциация с содержанием линоленовой и линолевой кислот, высотой растения и количеством коробочек соответственно.

У культурных сортов льна масличного также наблюдается уменьшение разнообразия в участках хромосомы 8, пересекающихся с scaffold635-43971 и QOLE-Lu8.1, связанных с количеством ветвей и содержанием олеиновой кислоты соответственно, и, наоборот, увеличение разнообразия, в сравнении с староместными сортами, в

участках хромосомы 7, пересекающихся с QLIN-Lu7.2, QLIO-Lu7.2, QPAL-Lu7.3 и QIOD-Lu7.2, ассоциированных с содержанием линоленовой, линолевой, пальмитиновой кислот и йода соответственно.

Далее мы использовали inserции транспозонов как маркеры, замещающие однонуклеотидные полиморфизмы при полногеномном поиске ассоциаций с признаками устойчивости к фузариозному увяданию, продуктивности и содержания волокна. Было найдено 50 inserций TE, ассоциированных с разными фенотипическими признаками, причем многие ассоциации подтверждаются несколькими моделями или обнаруживаются в данных по нескольким годам. Порядка 20 % из найденных ассоциированных inserций полиморфны, т.е. оказались ассоциированными сразу с несколькими признаками (см. таблицу и табл. S4). Четыре inserции попали в известные QTL и две – в участки уменьшения разнообразия между группами образцов с разным морфотипом и селекционным статусом, указанным на рис. 5.

Из представленных ассоциированных с признаками inserций TE 15 найдены несколькими моделями и 12 – для разных признаков или для измерения одного признака разных лет. Шесть таких inserций, широко представленных в образцах (более 50 образцов), были валидированы, результат валидации см. на рис. S2–S4.

## Обсуждение

Детальная характеристика генетического разнообразия льна имеет первостепенное значение для его долгосрочного и устойчивого производства и диверсификации, а также для общего успеха программ его селекции. Ранее, используя однонуклеотидные полиморфизмы, мы охарактеризовали генетическое разнообразие коровой коллекции льна (306 образцов) ФНЦ ЛК (Duk et al., 2021; Kanapin et al., 2022). Эта коллекция, одна из лучших в мире, включает сорта льна из Евразии со значительной долей местных сортов. В последнее время, благодаря успехам биоинформатики и совершенствованию технологий секвенирования, доступными для анализа стали другие источники геномного разнообразия, включая TE и структурную вариабельность, которые могут играть существенную роль в формировании агрономически важных признаков растений и могут быть использованы для дальнейшего улучшения сельскохозяйственных культур.

Анализ новых сайтов inserции TE, выполненный в этой работе, показал, что наряду с однонуклеотидными полиморфизмами TE являются важным источником генетического разнообразия у льна. Преобладающие сайты inserции, как и у многих других сельскохозяйственных растений (Станин и др., 2025), – это сайты inserции ретротранспозонов, а среди ДНК-транспозонов чаще всего встречаются сайты inserции TE суперсемейств MULE-MuDR, hAT и CMC-EnSpm (см. рис. 1).

Вместе с тем, в отличие от однонуклеотидных полиморфизмов, которых было значительно больше у льна масличного, чем у долгунцов, мы не обнаружили столь существенной разницы в числе inserций разных семейств TE на одно растение у образцов разного морфотипа. Тем не менее, как показывают филогенетический анализ и метод главных компонент (для TE суперсемейств *Copia*,

Инсерции TE, ассоциированные с фенотипическими признаками и расположенные на расстоянии менее чем 3500 п. н. от генов с известной функцией

Координаты инсерции TE (суперсемейство)	Признак (эффект) ***	Положение инсерции относительно ближайшего гена	Ближайший ген (ортолог)	Аннотация
Chr1_3002494 ( <i>Copia</i> )	Oil (+)	Интрон	<i>Lus10042426</i> ( <i>AT5G10770.1</i> )	Белок семейства эукариотических аспартилпротеаз
Chr1_10250548 (-)	EFL (-) TL (-)		<i>Lus10022657</i> ( <i>AT3G04380.1</i> )	SET-домен, содержащий белок семейства лизинметил-трансфераз
Chr1_11870656 ( <i>hAT-Tag1</i> )	TL (-) NI (-) PH (-)		<i>Lus10009386</i> ( <i>AT1G08230.2</i> )	Трансмембранный белок семейства переносчиков аминокислот
Chr1_13735350 ( <i>Copia</i> ; <i>MULE-MuDR</i> )	TL (-)		<i>Lus10008435</i> ( <i>AT5G37290.1</i> )	Суперсемейство белков с повторами ARM
Chr1_20792291 (-)	TW (-)		<i>Lus10015841</i> ( <i>AT3G62870.1</i> )	Рибосомальный белок семейства L7Ae/L30e/S12e/Gadd45
Chr1_28316405 ( <i>hAT-Ac</i> )	Tswgt (+)	1309 перед геном	<i>Lus10018915</i> ( <i>AT1G71890.1</i> )	Белок семейства главных посредников
Chr2_2089015 ( <i>Copia</i> )	Tswgt (+)	Экзон	<i>Lus10016394</i> ( <i>AT3G08030.1</i> )	Белок неизвестной функции, DUF642
Chr3_522289 ( <i>Copia</i> )	TW (+)	1318 перед геном	<i>Lus10013432</i> ( <i>AT5G23130.1</i> )	Белок, содержащий домен LysM, связывающий пептидогликан
Chr3_19617078 ( <i>Gypsy</i> )	DSI (-)	Интрон	<i>Lus10008232</i> ( <i>AT5G46050.1</i> )	Пептидный транспортер 3
Chr3_20336100 ( <i>Copia</i> )	STI (-)	1882 после гена	<i>Lus10007194</i> ( <i>AT4G01690.1</i> )	Семейство флавин-содержащих аминоксидо-редуктаз
Chr3_24121740 ( <i>Copia</i> )	FC (+)	Экзон	<i>Lus10017112</i> ( <i>AT1G48130.1</i> )	1-цистеин пероксиредоксин 1
Chr3_24426254 ( <i>MULE-MuDR</i> ), *M/C	FC (-)	190 перед геном	<i>Lus10017063</i> ( <i>AT5G49480.1</i> )	Ca <sup>2+</sup> -связывающий белок 1
Chr4_11952218 ( <i>Copia</i> )	DSI (-)	853 после гена	<i>Lus10036789</i> ( <i>AT1G13980.1</i> )	Белок, содержащий домен sec7
Chr4_14756320 ( <i>hAT-Ac</i> )	EFL (-), FW (-), IL (+), NI (-), PH (-), TL (-), FC (-), TW (-)	2059 перед геном	<i>Lus10041548</i> ( <i>AT3G61120.1</i> )	AGAMOUS-подобный 13, **QYLD-Lu4.1 (YLD), QPLH-Lu4.3 (PLH), QDTM-Lu4.1 (DTM)
Chr4_14994752 ( <i>Copia</i> )	TL (-), NI (-)	Интрон	<i>Lus10041602</i> ( <i>AT3G52150.1</i> )	РНК-связывающий белок (мотивы RRM/RBD/RNP) семейства, **QPLH-Lu4.3 (PLH), QDTM-Lu4.1 (DTM)
Chr6_1356894 ( <i>Copia</i> )	TL (-)		<i>Lus10033725</i> ( <i>AT1G12470.1</i> )	Связывание ионов цинка
Chr6_11075470 ( <i>Copia</i> )	TL (+)	2456 перед геном	<i>Lus10002212</i> ( <i>AT5G01300.1</i> )	Белок семейства PEBP (фосфатидил-этанолламин-связывающий белок)
Chr6_11780716 ( <i>hAT-Ac</i> )	Tswgt (+)	1006 перед геном	<i>Lus10017263</i> ( <i>AT4G36040.1</i> )	Шаперон DnaJ-домен суперсемейства белков
Chr6_17837550 ( <i>Copia</i> )	TW (+)	400 перед геном	<i>Lus10024282</i> ( <i>AT1G58250.1</i> )	Домен белка локализации Гольджи; промотор РНК-полимеразы II; домен белка Fmp27
Chr7_15829004 ( <i>Copia</i> )	FW (-)	709 перед геном	<i>Lus10038370</i> ( <i>AT4G20050.1</i> )	Суперсемейство белков, подобных пектинлазе **QIOD-Lu7.2 (IOD), QLIN-Lu7.2 (LIN), QLIO-Lu7.2 (LIO)
Chr8_4775776 (-)	EFL (-), TW (+), PH (-), TL (-)	34 перед геном	<i>Lus10023944</i> ( <i>AT4G38380.1</i> )	MATE эффлюкс семейства белков
Chr8_12565858 ( <i>hAT-Ac</i> )	NI (+), PH (+)	1199 после гена	<i>Lus10025285</i> ( <i>AT2G37370.1</i> )	Центросомальный белок

Окончание таблицы

Координаты инсерции TE (суперсемейство)	Признак (эффект) ***	Положение инсерции относительно ближайшего гена	Ближайший ген (ортолог)	Аннотация
Chr8_16467884 (MULE-MuDR)	TW (-), FW (-)	972 после гена	Lus10004540 (AT1G29850.2)	Белок семейства двуцепочечных ДНК-связывающих белков
Chr8_16923232 (Copia), *Д/М	FC (+)	Интрон	Lus10000673 (AT1G32450.1)	Транспортер нитрата 1.5
Chr8_18366764 (Copia)	PH (+), NI (+)	Экзон	Lus10007810 (AT5G36930.2)	Семейство белков устойчивости к болезням (класс TIR-NBS-LRR)
Chr8_22233522 (Copia)	PH (+), NI (+)	2260 перед геном	Lus10039488 (AT5G43700.1)	Белок семейства регуляторов транскрипции AUX/IAA **QOLE-Lu8.1 (OLE)
Chr11_14960430 (Copia)	TW (-), TL (-), NI (-)	3421 после гена	Lus10008566 (AT1G55200.1)	Протеинкиназный белок с доменом, подобным адениннуклеотид-альфа-гидролазе
Chr12_11155523 (MULE-MuDR)	TW (-)	373 перед геном	Lus10030109 (AT4G02530.1)	Хлоропласт тилакоид люмен белок
Chr13_1437052 (Copia)	EFL (-)	Экзон	Lus10010648 (AT3G03770.1)	Белок с богатым лейцином повтором, белок семейства киназ
Chr13_2213745 (Copia)	Oil (+)	271 перед геном	Lus10026978 (AT2G26110.1)	Белок неизвестной функции (DUF761)
Chr13_3391752 (Copia)	PH (-), NI (-), TW (-)	1071 после гена	Lus10026032 (AT4G36670.1)	Белок семейства главных посредников
Chr13_6346942 (hAT-Ac)	FW (-)	253 перед геном	Lus10002084 (AT5G64040.1)	Субъединица реакционного центра фотосистемы I PSI-N, хлоропласт, предполагаемый/PSI-N, предполагаемый (PSAN)
Chr13_11641862 (Copia)	EFL (-)	972 после гена	Lus10009758 (AT4G17650.1)	Поликетидциклаза/дегидраза и белок транспорта липидов
Chr13_14372739 (Copia)	PH (-)	780 перед геном	Lus10034660 (AT4G29680.1)	Белок семейства щелочно-фосфатазоподобных белков
Chr14_9124491 (Copia)	DSI (-)	153 перед геном	Lus10005526 (AT2G37690.1)	Фосфорибозил-аминоимидазол-карбоксилаза, предполагаемая/AIR-карбоксилаза, предполагаемая
Chr14_9853868 (hAT-Ac)	Nsed (+)	984 после гена	Lus10000815 (AT2G27410.1)	Домен неизвестной функции (DUF313)
Chr14_16743646 (Copia)	FW (-)	466 после гена	Lus10017981 (AT1G14590.1)	Белок семейства нуклеотид-дифосфо-сахарных трансфераз

\* Вставка попадает в регион различия разнообразия между группами: Д/М – долгунов и масличных сортов, М/С – культурных и староместных сортов масличного льна. \*\* Вставка также попадает в QTL, опубликованный в (You, Cloutier, 2020). Координата инсерции TE – середина окна длиной в 50 п.н.  
\*\*\* Эффект инсерции TE на признак: (+) – положительный, (-) – отрицательный.

*Gypsy* и *hAT-Ac*), картины инсерции TE у льна масличного и у льна-долгунца четко дифференцируют эти морфотипы и подтверждают близкое родство и общее происхождение кряжей – русских староместных сортов (см. рис. S1).  
Расположение TE относительно генов существенно влияет на их экспрессию и может привести к различным фенотипическим изменениям. У льна большинство инсерций TE *Gypsy* и *CMC-EnSpm* попадает в межгенные области, а среди инсерций TE в ген только 29 % приходится на экзоны, что отражает предпочтительное выметание таких вставок под действием естественного отбора (см. рис. 2).  
Анализ геномных областей, затронутых недавними селекционными усилиями (см. табл. S3), показал неожиданную разницу в количестве районов со следами селекции у сортов льна масличного (9 районов) и льна-долгунца (32 района). Хромосомы 2, 5, 6, 9–11 и 15 вовсе

не содержат сигналов селекционного улучшения льна масличного (см. рис. 5), тем не менее 2 из 9 таких сигналов пересекаются с известными QTL, ассоциированными с синтезом жирных кислот. Десять геномных областей содержат следы разнонаправленной селекции по агрономически важным признакам льна масличного и льна-долгунца. Сравнение сортов льна-долгунца и кряжей со староместными сортами выявило много районов селекционного улучшения, которые, однако, практически не пересекаются, что, по всей видимости, отражает различные требования к улучшению льна-долгунца в современную эпоху селекции в свете изменения климата и появления новых областей применения. Эта гипотеза подтверждается обнаружением 13 районов уменьшения разнообразия в геномах сортов льна-долгунца при сравнении с кряжами.

Интересно, что пять участков уменьшения генетического разнообразия у культурных сортов и кряжей при сравнении их со староместными сортами также выделяются как участки уменьшения разнообразия при использовании в качестве маркеров однонуклеотидных полиморфизмов (табл. S5). Из них можно отметить участок уменьшения разнообразия у культурных сортов льна масличного в районах Chr4:12400001–12600000, в который попадает ген *Lus10036915*, связанный с защитой от патогенов, а также район Chr8:22400001–22600000, пересекающийся с QTL scaffold635-43971 и QOLE-Lu8.1, ассоциированными с количеством ветвей и содержанием олеиновой кислоты соответственно.

При полногеномном поиске ассоциаций наиболее обширно была представлена ассоциация инсерции Chr4\_14756320, которая была обнаружена 22 раза для разных признаков и данных разных лет всеми моделями (см. таблицу). Это инсерция TE семейства *hAT-Ac* на расстоянии ~2000 п.о. перед геном *Lus10041548*. Данный ген относится к генам типа AGAMOUS, которые определяют прекращение роста меристемы и формирование цветка (Yanofsky et al., 1990). Инсерция TE перед этим геном положительно повлияла на длину соцветия и отрицательно – на различные характеристики волокна, что, вероятно, указывает на негативное влияние данной инсерции на инициацию клеток волокна в апикальной меристеме. Также эта инсерция попадала в три QTL: QYLD-Lu4.1, QPLH-Lu4.3 и QDTM-Lu4.1, ассоциированных с урожайностью семян, высотой растения и скороспелостью.

Интересно, что в некоторых сайтах у разных образцов обнаруживаются инсерции TE разных семейств, например, в Chr1\_13735350 встречаются TE суперсемейств *Copia* и *MULE-MuDR*, а в Chr13\_2207778 – типов *MULE-MuDR* и *hAT-Tag1*. Однако вставки разных элементов оказывают одинаковое влияние на соответствующий признак.

Несколько инсерций TE в экзоны генов, тем не менее, положительно влияют на ассоциированный признак, как, например, инсерции TE *Copia* Chr2\_2089015 в ген *Lus10016394*, Chr3\_24121740 в ген *Lus10017112* и Chr8\_18366764 в ген *Lus10007810*, что позволяет сделать заключение о негативной роли этих генов в формировании соответствующих признаков (см. таблицу).

Некоторые инсерции были ассоциированы с несколькими признаками. Например, инсерция TE Chr1\_11870656 семейства *hAT-Tag1* расположена в интроне гена *Lus10009386* и оказывает негативное воздействие на техническую длину стебля, высоту растений и число междоузлий. Ортолог этого гена отвечает за транспорт аминокислот (Meyer et al., 2006). Также негативное влияние на техническую длину стебля и число междоузлий оказывала инсерция Chr4\_14994752 типа *Copia*, попадавшая в интрон гена *Lus10041602*, ортолог которого является компонентом фотосинтетического аппарата, участвует в реакции устойчивости к абиотическим стрессам во время прорастания семян (Xu et al., 2013). Эта инсерция также попадала в QTL, ассоциированные с высотой растения и скороспелостью (см. таблицу).

Инсерция TE Chr8\_22233522 суперсемейства *Copia*, наоборот, оказывала положительное воздействие на вы-

соту растений и число междоузлий и была расположена перед геном *Lus10039488*, ортолог которого модулирует экспрессию генов раннего ответа на ауксин (Liscum, Reed, 2002), также эта инсерция попадала в QTL, ассоциированный с олеиновой кислотой.

Инсерция TE Chr1\_10250548 попала в интрон гена *Lus10022657*, ортолог которого способствует транскрипционному подавлению псевдогенов и транспозонов (Veiseth et al., 2011), и оказала негативное влияние на техническую длину стебля и длину волокна.

Некоторые инсерции также могут оказывать различное воздействие на разные признаки, так, например, инсерция Chr8\_4775776, расположенная перед геном *Lus10023944*, негативно влияла на длину волокна и положительно – на вес технической части; такое влияние более предпочтительно, если речь идет о масличном льне, и инсерция, вероятно, была сохранена в процессе отбора. Ортолог этого гена относится к семейству белков МАТЕ, которые участвуют в защите от токсинов и синтезе полезных соединений (Takanashi et al., 2014).

Ряд ассоциаций обнаружен для одного и того же признака несколькими моделями, что повышает их значимость. Среди них можно отметить инсерцию TE Chr1\_3002494 суперсемейства *Copia*, которая оказывала положительное воздействие на содержание масла и попадала в интрон гена *Lus10042426*, ортолог которого участвует в растительном иммунитете (Breitenbach et al., 2014). Другие ассоциации инсерций суперсемейства *Copia*, таких как Chr6\_1356894 и Chr8\_16923232, попавших в интроны генов, также обнаруживались несколькими моделями, ортологи соответствующих генов участвовали в различных процессах обмена веществ при развитии корней (Takemoto et al., 2018) и переносе нитрата (Lin et al., 2008) соответственно. Кроме того, инсерция Chr8\_16923232 попадает в участок уменьшения разнообразия у льна масличного в сравнении с долгунцами (см. таблицу).

Сорта, обладающие инсерциями TE, оказывающими благоприятный эффект на адаптацию растений, можно использовать при селекции для улучшения культуры. Так, например, к уменьшению индекса заболеваемости DSI приводит наличие инсерций Chr3\_19617078 и Chr4\_11952218, расположенных соответственно рядом с генами *Lus10008232*, ортолог которого участвует в обеспечении устойчивости семян к стрессу во время прорастания и в защите от опасных бактериальных патогенов (Karim et al., 2005), и *Lus10036789*, ортолог которого участвует в транспорте ауксина и требуется для правильной организации клеточной стенки (Geldner et al., 2003).

## Закключение

Вставки транспозонов в геноме льна расположены неравномерно и отражают собой следы разнонаправленного отбора образцов с разным морфотипом, выращиваемых для различных целей. Новые вставки могут влиять на величину различных фенотипических признаков и адаптацию растений к условиям внешней среды. Участие транспозонов в формировании агрономических признаков позволяет использовать их наравне с однонуклеотидными полиморфизмами для получения желаемых характеристик при селекции.

## Список литературы / References

- Дук М.А., Кананин А.А., Самсонова А.А., Банкин М.П., Самсонова М.Г. Метод HiVmrMLM обнаруживает новые генетические варианты, связанные с устойчивостью к фузариозному увяданию у льна. *Вавиловский журнал генетики и селекции*. 2025; 29(3):380-391. doi 10.18699/vjgb-25-41
- [Duk M.A., Kanapin A.A., Samsonova A.A., Bankin M.P., Samsonova M.G. The HiVmrMLM method uncovers new genetic variants associated with resistance to Fusarium wilt in flax. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2025;29(3): 380-391. doi 10.18699/vjgb-25-41]
- Станин В.А., Дук М.А., Кананин А.А., Самсонова А.А., Суркова С.Ю., Самсонова М.Г. Разнообразие нута, обусловленное полиморфизмом вставок транспозонов. *Вавиловский журнал генетики и селекции*. 2025;29(1):61-71. doi 10.18699/vjgb-25-08
- [Stanin V.A., Duk M.A., Kanapin A.A., Samsonova A.A., Surkova S.Y., Samsonova M.G. Chickpea diversity driven by transposon insertion polymorphism. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed*. 2025;29(1):61-71. doi 10.18699/vjgb-25-08]
- Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664. doi 10.1101/gr.094052.109
- Bourque G., Burns K.H., Gehring M., Gorbunova V., Seluanov A., Hammell M., Imbeault M., Izsvák Z., Levin H.L., Macfarlan T.S., Mager D.L., Feschotte C. Ten things you should know about transposable elements. *Genome Biol*. 2018;19(1):199. doi 10.1186/s13059-018-1577-z
- Breitenbach H.H., Wenig M., Wittek F., Jordá L., Maldonado-Alconada A.M., Sarioglu H., Colby T., Knappe C., Bichlmeier M., Pabst E., Mackey D., Parker J.E., Vlot A.C. Contrasting roles of the apoplastic aspartyl protease APOPLASTIC, ENHANCED DISEASE SUSCEPTIBILITY1-DEPENDENT1 and LEGUME LECTIN-LIKE PROTEIN1 in Arabidopsis systemic acquired resistance. *Plant Physiol*. 2014;165(2):791-809. doi 10.1104/pp.114.239665
- Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R.; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-2158. doi 10.1093/bioinformatics/btr330
- Duk M., Kanapin A., Rozhmina T., Bankin M., Surkova S., Samsonova A., Samsonova M. The genetic landscape of fiber flax. *Front Plant Sci*. 2021;12:764612. doi 10.3389/fpls.2021.764612
- Flutre T., Duprat E., Feuillet C., Quesneville H. Considering transposable element diversification in *de novo* annotation approaches. *PLoS One*. 2011;6(1):e16526. doi 10.1371/journal.pone.0016526
- Geldner N., Anders N., Wolters H., Keicher J., Kornberger W., Müller P., Delbarre A., Ueda T., Nakano A., Jürgens G. The Arabidopsis GNOM ARF-GEF mediates endosomal recycling, auxin transport, and auxin-dependent plant growth. *Cell*. 2003;112(2):219-230. doi 10.1016/s0092-8674(03)00003-5
- Goudenhooff C., Bourmaud A., Baley C. Flax (*Linum usitatissimum* L.) fibers for composite reinforcement: exploring the link between plant growth, cell walls development, and fiber properties. *Front Plant Sci*. 2019;10:411. doi 10.3389/fpls.2019.00411
- Helbaek H. Domestication of food plants in the Old World. Joint efforts by botanists and archeologists illuminate the obscure history of plant domestication. *Science*. 1959;130(3372):365-372. doi 10.1126/science.130.3372.365
- Kanapin A., Rozhmina T., Bankin M., Surkova S., Duk M., Osyagina E., Samsonova M. Genetic determinants of fiber-associated traits in flax identified by Omics data integration. *Int J Mol Sci*. 2022;23(23):14536. doi 10.3390/ijms232314536
- Karim S., Lundh D., Holmström K.O., Mandal A., Pirhonen M. Structural and functional characterization of AtPTR3, a stress-induced peptide transporter of *Arabidopsis*. *J Mol Model*. 2005;11(3):226-236. doi 10.1007/s00894-005-0257-6
- Lin S.H., Kuo H.F., Canivenc G., Lin C.S., Lepetit M., Hsu P.K., Tildard P., Lin H.L., Wang Y.Y., Tsai C.B., Gojon A., Tsay Y.F. Mutation of the *Arabidopsis* NRT1.5 nitrate transporter causes defective root-to-shoot nitrate transport. *Plant Cell*. 2008;20(9):2514-2528. doi 10.1105/tpc.108.060244
- Liscum E., Reed J.W. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol Biol*. 2002;49(3-4):387-400. doi 10.1023/A:1015255030047
- Meyer A., Eskandari S., Grallath S., Rentsch D. AtGAT1, a high affinity transporter for  $\gamma$ -aminobutyric acid in *Arabidopsis thaliana*. *J Biol Chem*. 2006;281(11):7197-7204. doi 10.1074/jbc.M510766200
- Mhiri C., Borges F., Grandbastien M.-A. Specificities and dynamics of transposable elements in land plants. *Biology*. 2022;11(4):488. doi 10.3390/biology11040488
- Niu X.M., Xu Y.C., Li Z.W., Bian Y.T., Hou X.H., Chen J.F., Zou Y.P., Jiang J., Wu Q., Ge S., Balasubramanian S., Guo Y.L. Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc Natl Acad Sci USA*. 2019;116(14):6908-6913. doi 10.1073/pnas.1811498116
- Pulido M., Casacuberta J.M. Transposable element evolution in plant genome ecosystems. *Curr Opin Plant Biol*. 2023;75:102418. doi 10.1016/j.pbi.2023.102418
- Quesneville H. Twenty years of transposable element analysis in the *Arabidopsis thaliana* genome. *Mob DNA*. 2020;11:28. doi 10.1186/s13100-020-00223-x
- Schrader L., Schmitz J. The impact of transposable elements in adaptive evolution. *Mol Ecol*. 2019;28(6):1537-1549. doi 10.1111/mec.14794
- Takanashi K., Shitan N., Yazaki K. The multidrug and toxic compound extrusion (MATE) family in plants. *Plant Biotechnol*. 2014;31(5): 417-430. doi 10.5511/plantbiotechnology.14.0904a
- Takemoto K., Ebine K., Askani J.C., Krüger F., Gonzalez Z.A., Ito E., Goh T., Schumacher K., Nakano A., Ueda T. Distinct sets of tethering complexes, SNARE complexes, and Rab GTPases mediate membrane fusion at the vacuole in Arabidopsis. *Proc Natl Acad Sci USA*. 2018;115(10):E2457-E2466. doi 10.1073/pnas.1717839115
- Veiseth S.V., Rahman M.A., Yap K.L., Fischer A., Egge-Jacobsen W., Reuter G., Zhou M.M., Aalen R.B., Thorstensen T. The SUV4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in Arabidopsis. *PLoS Genet*. 2011;7(3):e1001325. doi 10.1371/journal.pgen.1001325
- Wang J., Zhang Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genom Proteom Bioinform*. 2021;19(4):629-640. doi 10.1016/j.gpb.2021.08.005
- Xu T., Lee K., Gu L., Kim J.I., Kang H. Functional characterization of a plastid-specific ribosomal protein PSRP2 in *Arabidopsis thaliana* under abiotic stress conditions. *Plant Physiol Biochem*. 2013; 73:405-411. doi 10.1016/j.plaphy.2013.10.027
- Yanofsky M.F., Ma H., Bowman J.L., Drews G.N., Feldmann K.A., Meyerowitz E.M. The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature*. 1990; 346(6279):35-39. doi 10.1038/346035a0
- You F., Cloutier S. Mapping quantitative trait loci onto chromosome-scale pseudomolecules in flax. *Methods Protoc*. 2020;3(2):28. doi 10.3390/mps3020028
- Yu G. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinform*. 2020;69(1):e96. doi 10.1002/cpbi.96
- Yu T., Huang X., Dou S., Tang X., Luo S., Theurkauf W.E., Lu J., Weng Z. A benchmark and an algorithm for detecting germline transposon insertions and measuring *de novo* transposon insertion frequencies. *Nucleic Acids Res*. 2021;49(8):e44. doi 10.1093/nar/gkab010

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 30.07.2025. После доработки 24.09.2025. Принята к публикации 25.09.2025.