

ПОИСК ИНСЕРЦИЙ МОБИЛЬНЫХ ГЕНЕТИЧЕСКИХ ЭЛЕМЕНТОВ И ХРОМОСОМНЫХ ПЕРЕСТРОЕК, ВЛИЯЮЩИХ НА ИЗМЕНЕНИЕ ЭКСПРЕССИИ ГЕНОВ, В ЛИНИЯХ *D. melanogaster* С НАРУШЕНИЕМ КОНТРОЛЯ ТРАНСПОЗИЦИИ РЕТРОТРАНСПОЗОНА *gypsy*

© 2024 г. И. В. Кукушкина^{a, *}, А. Р. Лавренов^{a, b}, П. А. Миляева^{a, c},
А. И. Лавренова^a, И. В. Кузьмин^a, Л. Н. Нефедова^{a, **}, А. И. Ким^{a, c}

^aБиологический факультет, Московский государственный университет им. М.В. Ломоносова,
Москва, 119234 Россия

^bИнститут проблем экологии и эволюции им. А.Н. Северцова Российской академии наук,
Москва, 119071 Россия

^cFaculty of Biology, Longgang District, Shenzhen MSU-BIT University, Shenzhen, 518172 China

*e-mail: vladimirova-bph@yandex.ru

**e-mail: nefedova@mail.bio.msu.ru

Поступила в редакцию 11.11.2023 г.

После доработки 27.03.2024 г.

Принята к публикации 02.04.2024 г.

Мобильные генетические элементы (МГЭ) способствуют повышению частоты спонтанных мутаций в геноме, а также способны изменять структуру генов или влиять на их экспрессию, поэтому важно иметь представление об активности МГЭ и их положении в геноме. В представленной работе показано преимущество сочетания двух методов секвенирования для поиска инсерций МГЭ и хромосомных перестроек: полногеномного нанопорового секвенирования и секвенирования транскриптома методом Illumina. Полногеномное нанопоровое секвенирование позволяет обнаружить инсерции МГЭ, а секвенирование транскриптома на платформе Illumina – оценить влияние инсерций на экспрессию генов. В работе использованы данные секвенирования геномов линий *Drosophila melanogaster* с фенотипом flamenco SS (w^l , мутант *flamenco*) и MS (w^l , мутант *flamenco*, активная копия *gypsy*). В качестве контроля использована лабораторная линия дикого типа Д32. В линиях SS, MS и линии дикого типа обнаружены инсерции и делеции МГЭ в эухроматиновых районах генома и в интронах генов относительно референсного генома (NCBI GCF_000001215.4). В анализируемых геномах проведен поиск инсерций и делеций в генах системы РНК-интерференции и в генах, дифференциально экспрессируемых у мух линий SS и MS. В результате обнаружены инсерции МГЭ в различных регионах генов *AGO3*, *CG17147*, *Su(var)3-3*, *Gasz*, *CG43348*, *moody*, *CG17752*. Не обнаружено корреляции между изменением положения МГЭ и уменьшением или увеличением экспрессии большинства анализируемых генов. В гене *vig* найдена хромосомная перестройка, затрагивающая 3'-нетранслируемую область. На основании результатов секвенирования длинными прочтениями получена сборка генома линии MS *de novo*. Установлено, что повышение экспрессии генов *CR45822* и *pst* в линиях SS и MS связано с трипликацией, а не с изменением регуляторных последовательностей генов или инсерцией МГЭ.

Ключевые слова: *Drosophila melanogaster*, мобильные генетические элементы, эволюция, Oxford Nanopore, Illumina, сборка генома

DOI: 10.31857/S0026898424050055, EDN: HUPDPY

ВВЕДЕНИЕ

Мобильные генетические элементы (МГЭ) являются структурной частью генома эукариот, они составляют значительную его долю и влия-

ют на активность множества генов. События рекомбинации между последовательностями МГЭ приводят к геномным перестройкам, а транспозиция МГЭ в ген – к нарушению его работы [1]. Мутации в генах системы контроля подавления

активности МГЭ приводят к увеличению вероятности транспозиции этих элементов [2].

Известно несколько путей РНК-интерференции, используемых для подавления экспрессии МГЭ на посттранскрипционном уровне, в каждом из которых функционируют разные гены и разные источники антисмысловой РНК [3]. Один из основных механизмов сайленсинга транспозонов в клетках зародышевого пути у дрозофилы — путь piРНК-интерференции [4], в котором piРНК при участии белков PIWI подавляют ретротранспозоны [5, 6]. piРНК это РНК длиной 23–30 нуклеотидов, которые обычно закодированы в специализированных геномных регионах — в кластерах piРНК, обогащенных дефектными последовательностями транспозонов [7]. Нарушение процессов синтеза и сплайсинга транскриптов кластеров снижает разнообразие и общее количество piРНК, что может привести к повышению уровня транскрипции МГЭ [8]. В то же время, в соматических клетках активность МГЭ в основном контролируется эндогенными малыми интерферирующими РНК (endo-siРНК), за формирование которых отвечают белки Dicer-2 (Dcr-2) и Argonaute 2 (AGO2) [9].

Удобным экспериментальным объектом для исследования МГЭ является *Drosophila melanogaster*. Геном дрозофилы содержит несколько десятков описанных и классифицированных семейств МГЭ [10]. Базы данных GenBank (<http://www.ncbi.nlm.nih.gov>) [11], FlyBase (flybase.org) [12], MSTE [13] и другие содержат информацию о геноме *D. melanogaster*, что значительно облегчает поиск повторяющихся последовательностей. Например, в базе FlyBase [12] в референсном геноме (NCBI GCF_000001215.4) указаны положения полноразмерных копий или фрагментов МГЭ.

Активность МГЭ на уровне транскрипции и биогенез piРНК часто изучают с использованием высокопроизводительного секвенирования. Описаны изменения транскрипционной активности МГЭ при стрессе (температурные факторы, радиация и т.д.) или нарушении работы системы РНК-интерференции [14, 15].

С целью поиска новых положений МГЭ в геноме используют секвенирование длинными прочтениями [16, 17]. К недостаткам секвенирования длинными прочтениями методом Oxford Nanopore относится высокая частота ошибок (5–15%) [18], но это не влияет на определение положения МГЭ и крупных хромосомных перестроек. Для выравнивания прочтений, полученных методом нанопорового секвенирования, с исследуемым геномом, используют такие программы, как BWA-MEM [19],

Minimap2 [20], NGMLR [21] и другие [22–24]. Несмотря на то, что программа BWA-MEM рекомендована для работы с длинными прочтениями, она неточно определяет границы структурных вариаций (Structural variations (SVs)) [20]. Одна из наиболее простых, быстрых и эффективных программ при работе с длинными прочтениями — программа Minimap2, эффективная в случае последовательностей длиной до ~100 млн п.н. [20].

В свою очередь, сопоставление *de novo* сборки исследуемого генома и транскриптома позволяет провести корреляционный анализ изменения экспрессии генов, вблизи которых произошла инсерция МГЭ [16].

Ранее в работах нашей лаборатории было показано нарушение процессов сплайсинга транскриптов локуса *flamenco* у линий SS и MS [25], что может быть причиной нарушения контроля активности МГЭ (фенотип flamenco). Линии SS и MS характеризуются генетической нестабильностью, которую связывают с присутствием мутантного аллеля *flamenco* [26]. Линия MS получена путем введения в линию SS активной копии ретротранспозона с длинными концевыми повторами (ДКП-ретротранспозона) *gypsy* [27]. Линия MS характеризуется повышенной экспрессией *gypsy*, снижением репродуктивных способностей самок, связанным с комплексными нарушениями морфологии яичников, а также снижением экспрессии генов, ответственных за формирование яйца [26, 27]. Анализ транскриптома линий SS и MS не выявил значимых мутаций в экзонах генов системы РНК-интерференции. Обнаруженные изменения экспрессии генов не объясняли природу фенотипа *flamenco* [28]. Можно предположить, что в исследуемых линиях эти изменения связаны с мутациями в регуляторных районах генов или с инсерциями МГЭ, которые невозможно обнаружить секвенированием транскриптома короткими прочтениями.

В ходе исследования сравнили данные секвенирования длинными прочтениями генома *D. melanogaster* линий SS, MS, линии дикого типа Д32 с референсным геномом. Проведена сборка генома MS и осуществлен поиск генетических перестроек, которые могли повлиять на изменение экспрессии генов. Описаны подходы, позволяющие избежать ошибок, вызванных несовершенством стандартных способов обработки данных.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Линии *D. melanogaster* и условия культивирования. В работе использовали изогенные линии с фенотипом flamenco SS (w^I) и MS (w^I , *forked*, содержит искусственно внесен-

ный МГЭ *gypsy*) [27], и линию дикого типа Д32 из коллекции кафедры генетики МГУ. Мух культивировали в стандартных условиях: на пищевом субстрате (агар, манная крупа, сахар, изюм, дрожжи) в стеклянных пробирках при температуре 25°C и влажности около 65%. Мух пересаживали на свежий корм в новую пробирку каждые 18 дней.

Выделение РНК и высокопроизводительное секвенирование. РНК выделяли из пяти самок линий SS, MS и Д32 с помощью комплекта реактивов ExtractRNA (“Евроген”, Россия). Концентрацию РНК измеряли с помощью флуориметра Qubit (“Thermo Scientific”, США), целостность ($RIN > 7$) оценивали с помощью капиллярного электрофореза на приборе Bioanalyzer 2100 (“Agilent”, США). Подготовка библиотек выполнена с использованием набора TruSeq RNA Library Prep Kit v.2 (“Illumina”, США). Перед секвенированием концентрацию библиотек оценивали с помощью флуориметра Qubit и ПЦР в реальном времени (праймеры: I-qPCR-1.1 AATGATACGGCGACCACCGAGAT и I-qPCR-2.1 CAAGCAGAAGACGGCATACGA). Библиотеки были разведены до концентрации 11 пМ и секвенированы на приборе Illumina HiSeq 2000 (длина прочтения – 50 нуклеотидов). Глубина секвенирования в случае линии MS составила 5 млн прочтений, линий Д32 и SS – 12 млн прочтений на образец. Качество секвенирования контролировали с помощью программы FastQC [29]. Для тримминга секвенированных последовательностей использовали программу Trimmomatic v.0.32 [30]. Доля прочтений с качеством не менее 20 (по системе оценки качества PHRED) составила 98.9–99.5%. Короткие прочтения были сопоставлены с референсным геномом *D. melanogaster* сборки BDGP6 (версия BDGP6.94) с использованием программы Tophat v.2.1.0 [31]. Доля картированных прочтений составила 94–98%. Подсчет прочтений производили с помощью программы HTSeq v.0.6.1 [32].

Дифференциальную экспрессию генов оценивали методом DESeq [33]. Транскриптомные линии SS, MS и Д32 (контрольная линия) сравнивали попарно. Выявленные дифференциально экспрессируемые гены ($p_{adj} < 0.05$) в линии SS дополнительно валидировали путем сравнения транскриптома линии SS с линиями Oregon-R и w^{1118} (глубина секвенирования 11–17 млн прочтений на образец, серии экспериментов GSE99574), которые доступны в NCBI GEO (<https://www.ncbi.nlm.nih.gov/geo>) и процессированы согласно описанному выше протоколу [26, 28].

Выделение ДНК и нанопоровое секвенирование. Для выделения геномной ДНК использовали по восемь самок линий MS,

SS, Д32. Мух гомогенизировали тefлоновым пестиком в микропробирке. Затем добавляли 500 мкл лизирующего раствора (0.1 М Трис-НСl pH 9.0, 0.1 М EDTA, 1% SDS, 1% DEPC) и инкубировали в течение 30 мин при 70°C. ДНК из раствора осаждали ацетатом калия (1 М) и изопропанолом (0.5 объема). После промывки 70%-ным этанолом осадок ДНК растворяли в деионизированной воде. Концентрацию ДНК измеряли спектрофотометрически на приборе NanoDrop (Pqlab). Образцы ДНК были обработаны РНКазой А (“Thermo Scientific”). Соотношения оптической плотности при 260/280 и 260/230 нм составляли более 2.1. Подготовку библиотеки для нанопорового секвенирования проводили с использованием набора Ligation Sequencing kit (“Oxford Nanopore Technologies”, Великобритания). Библиотеки секвенировали на приборе MinION (ячейка FLO-MIN106D). Итоговое количество прочтений составило 2.6×10^6 при среднем размере прочтения 3800 п.н. Максимальная длина прочтения – около 48000 п.н. Общее количество прочитанных нуклеотидов 6.7×10^9 п.н. Процедуру “бейс-коллинга” проводили с помощью программного пакета Guppy. Прочтения выравнивали на геном *D. melanogaster* сборки BDGP6 (версия BDGP6.94) с помощью программы Minimap2 [20]. Получено 20-кратное значение покрытия генома линии MS. В случае линий SS и Д32 покрытие было 5–7-кратным. Последующую сборку генома *de novo* выполняли с помощью программы Canu с параметрами по умолчанию [34]. Полученные контиги картировали на геном *D. melanogaster* сборки BDGP6 с помощью программы Minimap2 [20]. Покрытие контигов относительно референсного генома не менее 65%.

Дополнительные программы. Для визуализации прочтений использовали программу IGV [35]. Программа визуализирует выровненные прочтения на референсный геном с указанием границ генов. Цветовая палитра, предусмотренная программным продуктом IGV, позволяет увидеть замены нуклеотидов, делеции и вставки с указанием размера в парах нуклеотидов (п.н.). Секвенирование длинными прочтениями не позволяет определить положение МГЭ с точностью до нуклеотида в связи с высокой ошибкой секвенирования [18]. В данной работе рассмотрены хромосомные перестройки, инсерции и делеции МГЭ, обнаруженные в гомозиготном состоянии, так как только они могут быть характеристикой линии.

Количественная ОТ-ПЦР. Различия в экспрессии отдельных генов подтверждали с помощью количественной ПЦР на матрице кДНК. Образцы РНК обрабатывали ДНКазой I (“Thermo Scientific”). Обратную транскрип-

цию проводили с использованием случайного праймера по протоколу набора MMLV-RT Kit (“Евроген”, Россия). Амплификацию проводили в амплификаторе Mini-Opticon Real-Time-PCR System (“Bio-Rad Lab”, США) с использованием реакционной смеси qPCRmix-HS SYBR (“Евроген”), содержащей флуоресцентный краситель SYBR Green I. Параметры амплификации: предварительная денатурация — 95°C, 5 мин, затем 40 циклов ПЦР (денатурация — 95°C, 15 с, отжиг праймеров — 55°C, 45 с, элонгация — 72°C, 60 с). Эффективность праймеров оценивали с помощью серии кратных разведений ДНК или кДНК. Результаты ПЦР анализировали с помощью пакета программ Bio-Rad CFX Manager (версия 1.6.541.1028). В качестве референсных использовали гены *αTub84D*, *RpL40*, *EloB*, *eIF-1A*. Уровень экспрессии генов рассчитывали по методу ΔCt (уровень экспрессии гена интереса) = $2^{Ct[\text{среднее арифметическое референсных генов}] - Ct[\text{ген интереса}]}$ на матрице кДНК. Количество копий гена в геноме рассчитывали по методу ΔCt (количество копий гена интереса) = $2^{Ct[\text{референсный ген}] - Ct[\text{ген интереса}]}$ на матрице ДНК. Последовательности праймеров приведены в табл. 1.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Поиск МГЭ и хромосомных перестроек на основании данных секвенирования генома длинными прочтениями и транскриптома

В анализе использованы данные нанопорового секвенирования генома и секвенирования транскриптома методом Illumina изогенных линий SS и MS *D. melanogaster* с нарушенным контролем транспозиции МГЭ *gypsy*. В качестве контроля использованы линия дикого типа Д32 и референсный геном. Мух содержали в одинаковых стандартных условиях на протяжении долгого времени. Таким образом при анализе транскриптома исключается влияние стресса и условий содержания на экспрессию генов. Секвенирование генома длинными прочтениями позволяет увидеть хромосомные перестройки и новые инсерции МГЭ в изогенных линиях, а секвенирование транс-

криптома позволяет оценить дифференциальную экспрессию генов.

Результаты секвенирования длинными прочтениями анализировали с использованием программы Minimap2. Мы обратили внимание, что Minimap2 по-разному интерпретирует данные прочтений с гибридной структурой. Прочтениями с гибридной структурой мы называем прочтения, состоящие из последовательности МГЭ и геномного окружения инсерции.

При выравнивании на референсный геном прочтения с гибридной структурой, включающего последовательность МГЭ с геномным окружением (последовательность генома + МГЭ + последовательность генома) (рис. 1а), программа Minimap2 будет интерпретировать данные как новую инсерцию. Если гибридное прочтение начинается или заканчивается последовательностью МГЭ, которой нет в референсном геноме в данном положении (последовательность генома + начало/конец МГЭ) (рис. 1б), то Minimap2 разделяет прочтение на два положения в геноме: один фрагмент прочтения будет сопоставлен с соответствующим участком референсного генома, а другой, содержащий фрагмент МГЭ, будет ошибочно картирован в случайную последовательность этого МГЭ в референсном геноме. Таким образом, у гомозиготы по инсерции МГЭ будет показана инсерция или набор прочтений, обрывающихся в области инсерции. У гетерозиготы мы можем наблюдать прочтения без инсерции, прочтения с инсерцией (при хорошем покрытии) и набор прочтений, обрывающихся в области инсерции.

Единственное, что позволяет предположить факт некорректного выравнивания, это увеличение покрытия для МГЭ в референсном геноме в 2–5 раз. По анализу прочтений в районе МГЭ в референсном геноме можно установить координаты расположения других фрагментов прочтения, содержащих участок генома, и тем самым установить локализацию инсерции МГЭ.

В анализе транскриптома методом Illumina использовали программу Tophat. Короткие прочтения не могут иметь гибридную струк-

Таблица 1. Олигонуклеотидные праймеры для ОТ-ПЦР

Ген	Прямой праймер (5'-3')	Обратный праймер (5'-3')	Ампликон, п.н.
<i>CR45822</i>	GTTCCGTGGTCAGTACAGC	GTCGAGCAAGTGATACAAATG	81
<i>αTub84D</i>	GTGCATGTTGTCCAACACCAC	AGAACTCTCCCTCCTCCATA	131
<i>RpL40</i>	CTGCGTGGTGGTATCATTG	CAGGTTGTTGGTGTGTCC	147
<i>EloB</i>	GCACAAACATACACACTCACG	TTTCCTACTTCGCTTGACCC	142
<i>eIF-1A</i>	TCGCTCTGGAGGCAATGTG	GAGTCCTGGTAGTCACGC	126
<i>forked</i>	GGATACGGAGGACTACATGC	GTGTGGCTTCTGGGGCTTGT	265

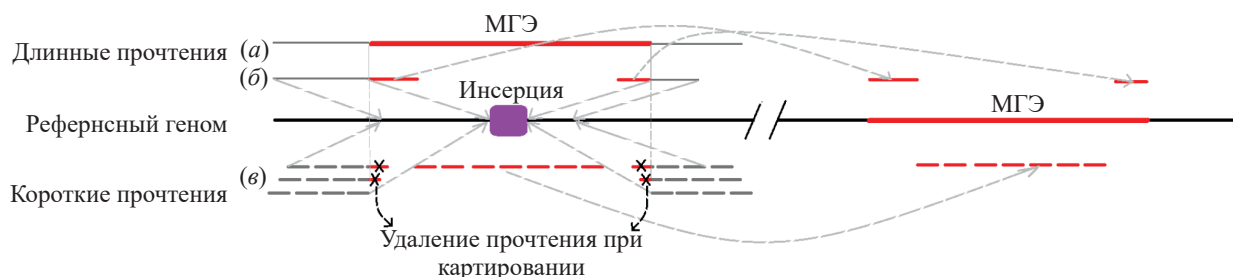


Рис. 1. Схема сопоставления прочтений с референсным геномом, включающих последовательность МГЭ, для секвенирования длинными прочтениями и секвенирования короткими прочтениями. Красным цветом изображен МГЭ в исследуемом геноме, черным — референсный геном с указанием положения инсерции (фиолетовый). При выравнивании длинных прочтений, включающих последовательность МГЭ, окруженную последовательностями генома, данные будут интерпретированы как новая инсерция (а). Короткие фрагменты МГЭ будут ошибочно перенесены в случайную последовательность этого МГЭ в референсном геноме (б, в).

туру “последовательность генома + МГЭ + последовательность генома”, так как инсерции МГЭ или его фрагментов чаще всего имеют размер более 300 п.н., что больше размера короткого прочтения (рис. 1в). При структуре короткого прочтения “последовательность генома + начало/конец МГЭ” программа для сопоставления прочтений с референсным геномом не будет использовать при выравнивании такие прочтения из-за превышения параметра несовпадающих нуклеотидов.

Верификация инсерции ретротранспозона gypsy в ген forked линии MS D. melanogaster

В анализе использовали линию MS с инсерцией ретротранспозона gypsy в ген *forked*. В начале исследования точное положение данной инсерции было неизвестно. Анализ секвенирования генома линии MS *D. melanogaster* длинными прочтениями подтвердил инсерцию ретротранспозона gypsy размером около 7450 п.н. (рис. S1, анализ последовательности S1, см. Дополнительные материалы в электронном виде по DOI статьи и на сайте http://www.molecbio.ru/downloads/2024/5/supp_Kukushkina_rus.pdf) в ген *forked*.

Зная точное положение инсерции МГЭ, можно подтвердить мутацию методом ПЦР с использованием ДНК как матрицы. Праймеры при этом подбирают в окружении и внутри МГЭ [36]. Для подтверждения инсерции использовали отводку линии MS, прошедшую через насыщающее скрещивание, чтобы исключить возможную гетерогенность линии по инсерции gypsy в *forked*. Ретротранспозон gypsy относится к классу ретротранспозонов с длинными концевыми повторами (ДКП-ретротранспозон), он содержит два однонаправленных ДКП размером 482 п.н. Для подтверждения инсерции использовали праймеры в окружении МГЭ согласно [36].

Так как размер МГЭ составляет более 7 т.п.н., а время элонгации равно 60 с, то ПЦР-продукт не должен синтезироваться. При этом праймеры на ген *forked* без инсерции должны амплифицировать один продукт размером 265 п.н.

Однако в результате ПЦР на геномной ДНК линии MS с соответствующими праймерами вокруг инсерции МГЭ получен продукт размером около 750 п.н., который далее был секвенирован по методу Сэнгера (анализ последовательности S2, см. электронное приложение). Обнаруженный фрагмент 750 п.н. соответствовал участку гена *forked* со вставкой фрагмента МГЭ gypsy размером 481 п.н. При этом последовательность инсерции совпадала по размерам с ДКП МГЭ (482 п.н.). Такой ПЦР-фрагмент мог получиться аналогично процессам, происходящим при ПЦР для сайт-направленного мутагенеза [37] (рис. 2).

Таким образом, при подтверждении инсерций методом ПЦР следует учитывать возможность ложноотрицательного результата, если МГЭ относится к классу ДКП-ретротранспозонов. Такие МГЭ с обоих концов фланкированы длинными прямыми повторами, которые могут отжигаться друг на друга за счет взаимной комплементарности во время ПЦР.

Поиск хромосомных перестроек в генах системы рiРНК-интерференции и в дифференциально экспрессируемых генах

Инсерции и делеции МГЭ являются характеристикой линий с нарушением контроля транспозиции [38–40]. В связи со значительным количеством генетических перестроек, обнаруженных при поиске ассоциаций с фенотипом flamenco, мы решили остановиться на анализе инсерций и делеций в генах системы рiРНК-интерференции и генах, дифференциально экспрессируемых в линиях SS и MS [26, 28].

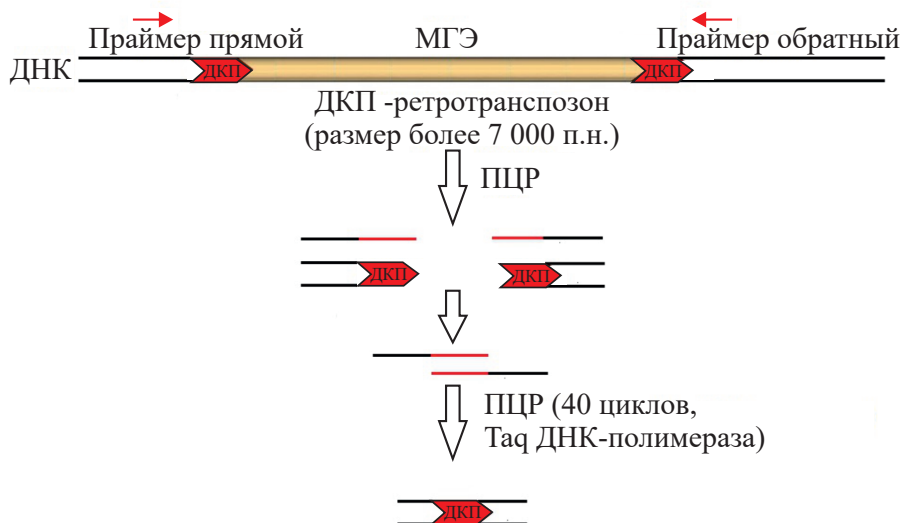


Рис. 2. Схема амплификации ДКП ДКП-ретротранспозонов, имеющих в своей структуре два сонаправленных ДКП. В ходе ПЦР синтезируются продукты, содержащие на 3'-концах два гомологичных и сонаправленных ДКП. Достаивание фрагмента происходит за счет перекрытия гомологичных участков.

Анализ последовательностей и промоторных областей 74 генов системы *pi*РНК-интерференции [41] и 65 генов, дифференциально экспрессируемых в линиях SS и MS [26, 28], позволил обнаружить инсерции МГЭ или другие перестройки генома, которые могут нарушать работу гена (табл. 2).

В табл. 3 представлены результаты определения дифференциальной экспрессии генов в линиях SS и MS с фенотипом *flamenco* относительно линии Д32 ($p_{adj} \leq 0.05$), содержащие инсерции/делеции МГЭ или другие перестройки генома. Гены из таблицы 2 с $p_{adj} > 0.05$ не приведены.

Один из основных белков *pi*РНК-интерференции — Argonaute 3 (AGO3). В экзонах гена *AGO3* мутаций не обнаружено, но в отличие от референсного генома интроны данного гена в анализируемых линиях содержат несколько инсерций МГЭ. Линии SS, MS и Д32 имеют свой уникальный набор МГЭ. Инсерции МГЭ X и G4 представлены в линиях SS и MS, а также в линии дикого типа Д32 (рис. 3а, табл. 2). Ген *AGO3* находится на границе с кластером *80F*, что может характеризовать данный район как горячую точку транспозиции МГЭ. Аналогичные инсерции обнаружены и в линии дикого типа, поэтому можно сделать вывод, что эти мутации не ассоциированы с фенотипом *flamenco*.

В линиях SS и MS (2L: 15064123) найдена инсерция около 2250 п.н. в интроне гена *vasa* (*vas*) (табл. 2). Анализ этой инсерции показал, что это не МГЭ, а дупликация фрагмента интрона данного гена (2L: 15061758–15064119). Эта дупликация в гене *vasa* затрагивает 3'-не-

транслируемую область транскриптов *vig-Rd* и *vigRB* гена *vig*, расположенного на противоположной цепи ДНК относительно гена *vasa*. Мутация в 3'-нетранслируемой области может нарушать экспрессию гена на уровне белка. *VIG* участвует в организации гетерохроматина и его положение частично перекрывается с *HP1*. Мутанты по гену *vig* демонстрируют истощение *HP1* и *HP2* в изолированных ядрах [42]. Ранее в линиях SS и MS обнаружили снижение уровня экспрессии гена *rhino* (*hpld*), который кодирует белок из семейства *HP1* [25].

Vasa (*Vas*) играет решающую роль в широком спектре клеточных процессов, включая биогенез рибосом [43, 44], инициацию трансляции [45], развитие ооцита [46, 47], сплайсинг пре-мРНК и распад мРНК [44]. Мутации в гене *vasa* обычно приводят к частичной или полной стерильности, а также к повышению экспрессии МГЭ [41, 46, 47]. Так как дупликация в интроне гена *vasa* обнаружена в линиях SS и MS, можно предположить, что мутация не приводит к нарушению репродуктивных способностей самок [26]. При этом анализ транскриптомных данных не подтвердил появления нового экзона или изменения транскрипции гена [28].

В линии MS обнаружена инсерция МГЭ в гене *Gas2* (табл. 2). *Gas2* кодирует фактор первичного биогенеза *pi*РНК. Инсерция ретро-транспозона 17,6 в 3'-нетранслируемой области может нарушать экспрессию гена на уровне белка. *Gas2* локализуется на внешней мембране митохондрий вместе с *Daed* и *Zuc* и участвует в созревании *pi*РНК на поверхности митохондрий. Истощение *Gas2* приводит к накопле-

Таблица 2. Мутации в генах рiРНК-интерференции и генах, дифференциально экспрессируемых в линиях SS, MS и линии дикого типа Д32

Ген	Положение инделов, п.н.**	Положение, инсерция/делеция и размер, п.н.**			МГЭ	MS	SS	Д32
<i>AGO3*</i>	3L: 23572954	Интрон	Инсерция	5050	<i>BS</i>		+	
	3L: 23610908	То же	То же	4700	<i>Doc</i>			+
	3L: 23618761	—//—	—//—	910	<i>roo</i>		+	
	3L: 23610908— 23620600	—//—	Делеция (гетерозигота)	9700	—			+
	3L: 23663464	—//—	Инсерция (гетерозигота)	6150	<i>X</i>	+	+	+
	3L: 23684000	—//—	Инсерция (гетерозигота)	2050	<i>G4</i>	+		+
<i>vasa*/vig*</i>	2L: 15064123	—//—	Инсерция	2350	Дубликация района гена <i>vasa</i> 2L: 15061758— 15064019	+	+	
<i>Su(var)3-3*/ CG17147</i>	3L: 20225800		То же	7400	<i>blood</i>	+	+	
<i>Gasz*</i>	2R: 8170000	3'-Нетранслируемая область	—//—	6700	<i>17.6</i>	+		
<i>CG43348</i>	2L: 1199692	3'-Нетранслируемая область	—//—	9700	<i>Max</i>	+		
<i>moody</i>	X: 2014730	3'-Нетранслируемая область	—//—	5200	<i>copia</i>	+		
<i>CG17752</i>	3R: 19622486	Интрон	—//—	7600	<i>roo</i>	+		
<i>CR45822/ pst</i>	3L: 7355650— 7363640		—//—	16000	Трипликация локуса	+	+	
<i>Mur2B</i>	X: 1567717— 1569443	Интрон	Делеция	1740	<i>I</i>	+	+	+
	X: 1628000— 1635500	Интрон	Делеция	7400	<i>gtwin</i>	+	+	+
<i>su(w[a])</i>	X: 1033407— 1040982	5'-Нетранслируемая область	Делеция интрона	7580	<i>412</i>	+	+	+
<i>Scgδ</i>	X: 1737166— 1746272	5'-Нетранслируемая область	Делеция интрона	9112	<i>roo</i>	+	+	+
<i>forked</i>	X: 17279030	Экзон	Инсерция	7400	<i>gypsy</i>	+		

+ — Инсерции или делеции в анализируемых линиях.

* Гены системы рiРНК-интерференции.

** Округленное значение, п.н.

нию предшественников транскриптов *flamenco* и уменьшению количества зрелой рiРНК [48]. В линии SS инсерции в этом гене не обнаружено. Следовательно, данные мутации не являются причиной возникновения фенотипа *flamenco*.

Анализ дифференциально экспрессирующихся генов выявил инсерцию МГЭ *blood* в области

(chr3L:20225801) между генами *CG17147* и *Su(var)3-3* на расстоянии 478 п.н. после гена *CG17147* (рис. 3б, табл. 2). В гене *CG17147* обнаружена инсерция 210 п.н. в интрон, аналогичная инсерция представлена и в линии дикого типа. По данным транскриптомного анализа экспрессия этого гена повышена в 6 раз (табл. 3) [28]. Возможно, эта инсерция могла

Таблица 3. Дифференциальная экспрессия генов, содержащих инсерцию/делецию МГЭ или хромосомную перестройку, у линий SS, MS относительно Д32 (метод DESeq)

Ген	flybase_id	SS относительно Д32		MS относительно Д32	
		Log ₂ FC SS/Д32	p _{adj} SS/Д32	Log ₂ FC MS/Д32	p _{adj} MS/Д32
<i>CG17147</i>	FBgn0260393	4.431	0.050	3.104	0.042
<i>CG43348</i>	FBgn0263080	−6.871	0.000	−4.463	0.000
<i>moody</i>	FBgn0025631	1.626	0.043	1.256	0.030
<i>CG17752</i>	FBgn0038718	1.813	0.011	1.061	0.047
<i>CR45822</i>	FBgn0267472	2.286	0.035	1.857	0.046
<i>pst</i>	FBgn0035770	1.690	0.002	1.169	0.040

Примечание. p_{adj} — скорректированное значение p ; Log_2FC — $\log_2$ отношения уровней экспрессии генов в сравниваемых линиях.

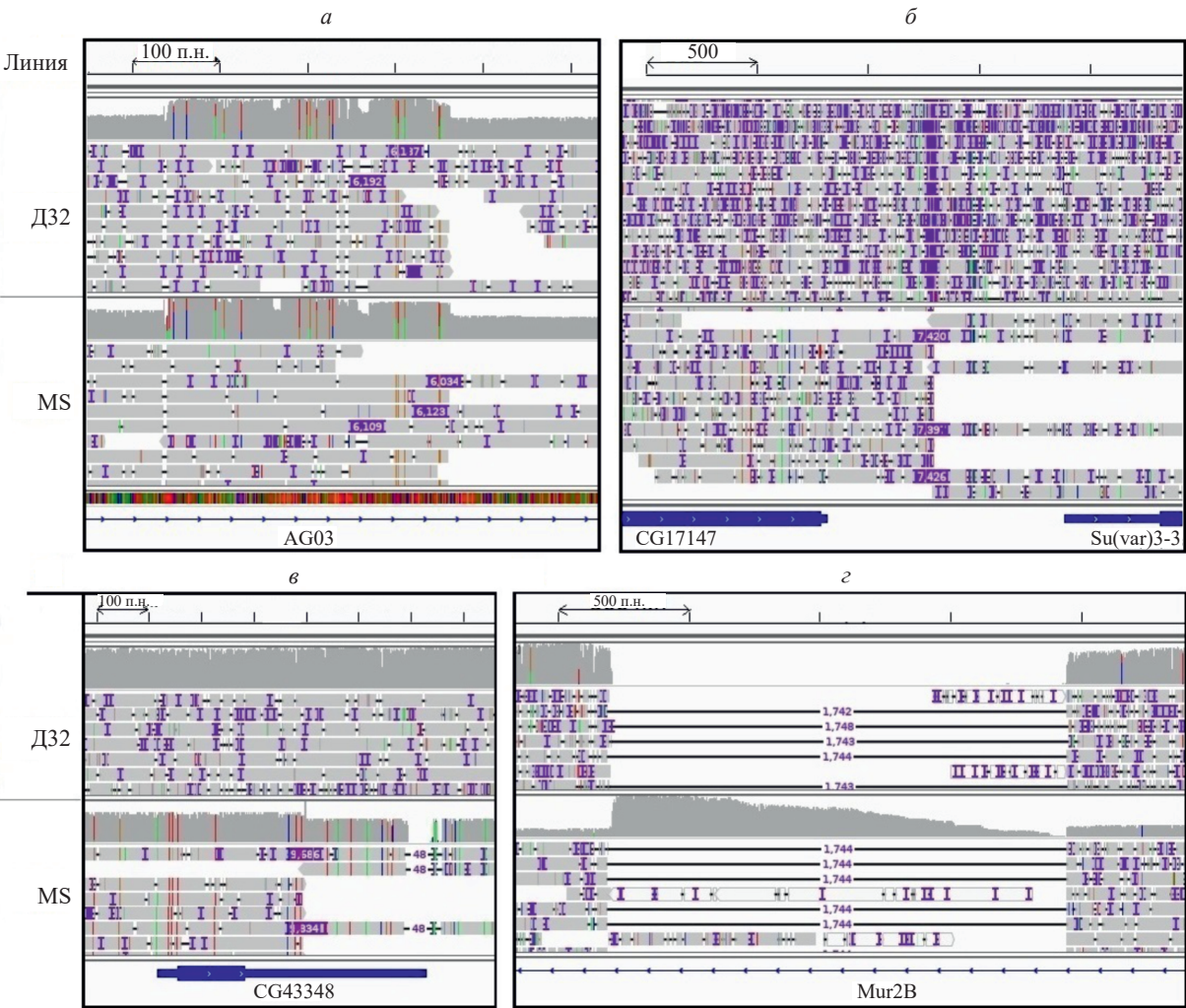


Рис. 3. Визуальное отображение результатов анализа нанопорового секвенирования в программе IGV. Серым показаны прочтения, соответствующие референсному геному. Красным, желтым, зеленым отмечены однонуклеотидные полиморфизмы, соответствующие замене определенного нуклеотида. Черные штрихи в прочтении — неопределенные нуклеотиды. Прочтения имеют границы и направление чтения. Фиолетовый прямоугольник с цифрами в прочтениях соответствует инсерциям в генах *AGO3*, *CG43348* и в области между генами *CG17147* и *Su(var)3-3* с указанием длины вставки (а, б, в). Длинные черные линии с цифрами соответствуют делеции в интроне гена *Mur2B* в линиях с указанием длины (г).

повлиять на изменение экспрессии гена, если затронула или принесла энхансерный участок. Интересно, что в 580 п.н. от обнаруженной инсерции расположен ген *Su(var)3-3*. На изменение экспрессии гена *Su(var)3-3* или сплайсинг его мРНК инсерция МГЭ не повлияла. *Su(var)3-3* участвует в образовании гетерохроматина, облегчая метилирование H3K9 [42]. Нокаут гена *Su(var)3-3* приводит к повышению экспрессии определенных групп МГЭ [41]. Не выявлено прямого участия *Su(var)3-3* в процессах сплайсинга транскриптов локусов рiРНК. При этом процесс формирования гетерохроматина с помощью гистон-метилтрансфераз, необходимых для правильной транскрипции локусов, не до конца изучен.

При анализе дифференциально экспрессируемых генов *CG43348*, *moody*, *CG17752* в линии MS обнаружены инсерции МГЭ (рис. 3в, табл. 2). В линии SS не найдено хромосомных перестроек или инсерций МГЭ в этих генах, при этом их экспрессия в линиях MS и SS не отличается. Следовательно, изменения экспрессии генов *CG43348*, *moody*, *CG17752* не связаны с данными мутациями.

Анализ данных секвенирования транскриптома методом Illumina показал увеличение экспрессии генов *CR45822* и *pst* в несколько раз (табл. 3) в линиях SS и MS относительно линии дикого типа Д32 [30, 31]. Методом ОТ-

ПЦР подтверждено увеличение экспрессии гена *CR45822* на уровне транскрипции в несколько раз в линиях SS и MS относительно линии Д32 (тест Манна–Уитни, $p < 0.01$, $N = 6$).

Важно отметить, что анализ данных секвенирования длинными прочтениями выявил в исследуемом районе крупные инсерции общим размером около 16 т.п.н. Одно прочтение в области, прилегающей к началу гена *CR45822*, другое – на расстоянии около 8 т.п.н. При этом покрытие в данном регионе было заметно выше, чем в остальных областях хромосомы (рис. 4а).

С использованием данных нанопорового секвенирования и программы Canu проведена сборка *de novo* генома мух линии MS. Полученные контиги сопоставлены с последовательностями генов из базы данных FlyBase [12]. В результате найдена трипликация района, охватывающего ген *CR45822* и второй экзон гена *Sec63* размером около 16 т.п.н. (рис. 4б). Этот район не является последовательностью МГЭ и не содержит тандемных повторов.

С целью проверки результатов методом ПЦР в реальном времени определено количество копий гена *CR45822* в геномной ДНК исследуемых линий SS и MS относительно линии дикого типа Д32. В расчетах учитывали эффективность праймеров к исследуемому гену и референсным генам. В результате получено расчетное значение – около трех копий *CR45822*

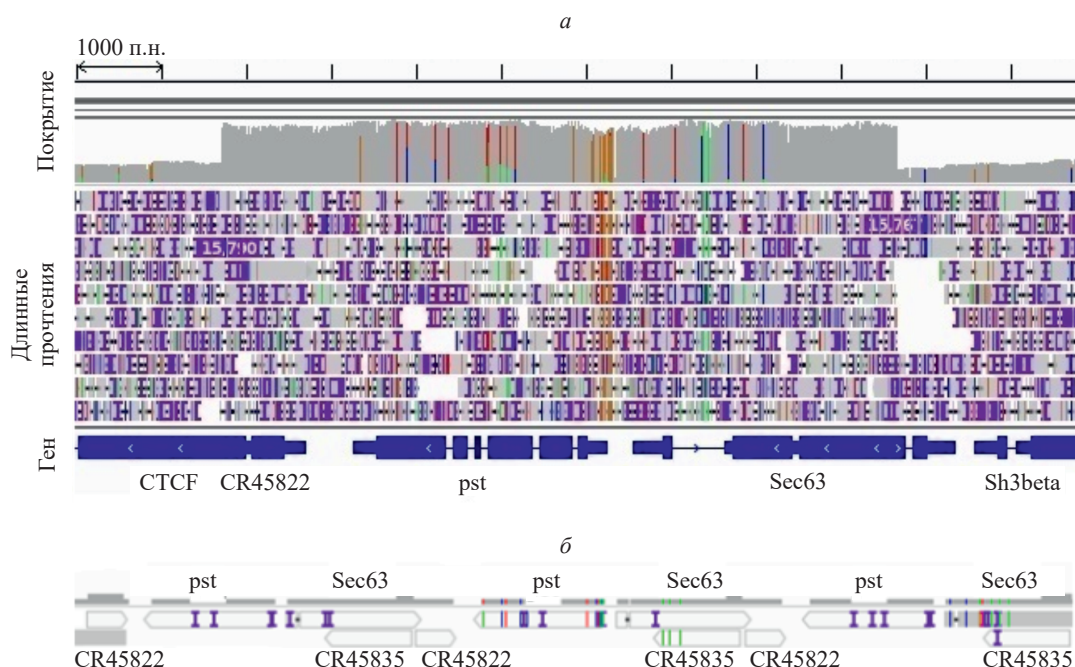


Рис. 4. Трипликация локуса между генами *CR45822* и *CR45835*. Наблюдается увеличение покрытия прочтений между генами *CR45822* и *CR45835* в 2–3 раза (а). На собранном *de novo* участке хромосомы 3L на основе данных нанопорового секвенирования сопоставлены последовательности генов *CR45822*, *CR45835*, *pst*, *Sec63*, *CTCF*. Видна трипликация генов *CR45822*, *CR45835*, *pst*, фрагментов генов *CTCF* и *Sec63*, при этом можно определить собственные полиморфные варианты некоторых генов (б).

в линии MS и родительской линии SS против одной копии в линии дикого типа Д32.

*Инсерции МГЭ в референсном геноме
относительно исследуемых линий
и линии дикого типа*

Проведение анализа в определенных положениях референсного генома привело к обнаружению МГЭ, которых не было в этих локусах в линиях SS и MS и линии дикого типа Д32. Мутации этого типа представлены в интроне гена *Mur2B* (рис. 3г, табл. 2) в линиях SS и MS и линии дикого типа Д32. Можно предположить, что произошли эксцизия и транспозиция МГЭ из одного локуса в другой. Однако после эксцизии ретротранспозонов путем гомологичной рекомбинации чаще всего остается одиночный ДКП [49]. При этом в анализируемых линиях эти локусы не содержат фрагментов МГЭ. Соответственно, референсный геном содержит не делецию МГЭ, а инсерцию относительно исследуемой линии, то есть эта инсерция МГЭ в референсном геноме, характерна исключительно для линии дрозофил, использованной при сборке генома.

Существуют и индивидуальные особенности анализируемых линий. Так, лабораторная линия дикого типа содержит делецию МГЭ, не обнаруженную в линиях MS и SS. Подобные геномные перестройки могут использоваться для проведения филогенетического анализа.

Анализ таких кластеров рiРНК, как *flamenco*, *42AB*, *38C*, *80F* [7], выявил значительные различия между референсным геномом и исследуемыми и контрольными лабораторными линиями. Анализ состава МГЭ в кластерах и определение точного количества обнаруженных инсерций и делеций — задача дальнейших исследований.

Можно предположить, что не каждая обнаруженная инсерция МГЭ в ген приводит к нарушению последовательности мРНК. Предполагается, что интроны могут происходить от транспозонов [50–52], возможно также существование механизмов распознавания МГЭ как интрона [51]. Например, в референсном геноме в 5'-нетранслируемой области мРНК генов *su(w[a])* и *Scgd* (табл. 2) присутствуют МГЭ *412* и *roo*, соответственно, представляющие собой интроны. В исследуемых линиях SS и MS и линии дикого типа Д32 в генах *su(w[a])* и *Scgd* этих интронов нет, как нет и последовательности МГЭ.

ЗАКЛЮЧЕНИЕ

В нашей работе при помощи секвенирования длинными прочтениями проведен поиск мутаций

в генах системы рiРНК-интерференции и генах, дифференциально экспрессирующихся в геноме линий SS и MS *D. melanogaster*. Обнаружены инсерции МГЭ в генах *AGO3*, *CG17147*, *Su(var)3-3*, *Gasz*, *CG43348*, *moody*, *CG17752*, затрагивающие различные структуры генов. Среди проанализированных генов рiРНК-интерференции линий SS и MS можно отметить перспективный маркер фенотипа *flamenco* — гены *vig* и *Su(var)3-3*. Обнаружены хромосомные перестройки генов *vasa*, *CR45822* и *pst*. Инсерция МГЭ возле гена *CG17147* связана, по-видимому, с увеличением его транскрипции в линиях SS и MS. В связи с тем, что большинство обнаруженных инсерций МГЭ не вызывали изменений экспрессии генов, можно предположить, что эта инсерция повлияла на регуляторную область гена *CG17147*. Не обнаружено также корреляции между изменением положения МГЭ и уменьшением или увеличением экспрессии других дифференциально экспрессируемых генов.

Как правило, при анализе геномных и транскриптомных данных принято опираться на референсный геном. Однако при анализе мобильного компонента генома необходимо учитывать, что референсные геномы сделаны на основании данных генома определенной популяции, обладающей характерными для нее позициями инсерций МГЭ, что может отразиться на интерпретации получаемых результатов. В частности, при использовании программного пакета TLDR [53] для поиска новых вставок МГЭ в геном, можно не учесть популяционные особенности линии и референсного генома, с которым исследуемая линия давно разошлась в ходе эволюции. В результате может быть сделан ошибочный вывод, например, об увеличении количества копий МГЭ в исследуемой линии, а также выполнен неправильный популяционный анализ, учитывающий только увеличение числа копий МГЭ и положение новых копий относительно референсного генома. При исследовании анализируемых геномов обнаружены делеции МГЭ, которые являются инсерциями МГЭ, характерными для референсного генома. Очевидно, что одного референсного генома недостаточно для охвата глобального генетического разнообразия мобильного компонента генома, присутствующего в различных популяциях. В свободном доступе есть данные длинных прочтений основных контрольных линий (Canton-S, OregonR, w¹¹¹⁸ и т.д.) [11]. При этом не обязательно секвенировать все известные линии, достаточно использовать линии, являющиеся отводками или “основой” исследуемой линии. Например, линия w¹¹¹⁸ считается базовой для получения большинства трансгенных линий [54].

Исследование транскриптома методом высокопроизводительного секвенирования, сопоставленного только с одним референсным геном, может не отражать все мутационные события. Например, инсерция МГЭ в 5'-нетранслируемую область может нарушать старт инициации трансляции, тем самым полностью меняя последовательность продукта гена [50–52]. Однако, используя короткие прочтения, можно не только пропустить инсерцию МГЭ, но и не заметить изменений в экспрессии гена. Классические методы определения экспрессии МГЭ методом ПЦР в реальном времени тоже не обеспечивают получение точного результата. Две популяционно различные линии могут содержать разное количество копий МГЭ, но экспрессироваться каждый отдельный элемент будет на одном уровне. Анализ транскриптомных данных будет выявлять повышение экспрессии МГЭ, что можно трактовать как повышение активности этого МГЭ или нарушение работы системы контроля транспозиции.

Возможно, что повышение уровня экспрессии МГЭ *gypsy* в линии MS [27] связано только с увеличением количества копий, а не с нарушением системы piРНК-интерференции. Так как линии SS и MS изогенны, то различия в положении новых инсерций наблюдаются даже при небольшом числе проанализированных генов. Это может быть связано с тем, что линия MS получена из одной особи линии SS, которая имела определенный набор инсерций МГЭ. Возможно, новые инсерции связаны с нарушением работы системы piРНК-интерференции и, таким образом, являются следствием фенотипа *flamenco*. Ранее в работах нашей лаборатории были показаны нарушения процессов сплайсинга транскриптов локуса *flamenco* у линий с фенотипом *flamenco* SS и MS [25]. Этот феномен может быть связан с особенностями структуры и состава кластера МГЭ, что представляет интерес для дальнейших исследований.

Работа получила поддержку программы “Молекулярно-генетические механизмы нестабильности генома и мутагенеза у животных и человека” (грант 13-1-21). Номер ЦИТИС: 121032500085-3.

В соответствии с пунктом 3 главы 1 Директивы 2010/63/ЕС от 22 сентября 2010 г. о защите животных, используемых в научных целях, требования биоэтики не распространяются на объект данного исследования.

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Moschetti R., Dimitri P., Caizzi R., Junakovic N. (2010) Genomic instability of I elements of *Drosophila melanogaster* in absence of dysgenic crosses. *PLoS One*. **5**(10), e13142.
2. Coline G., Theron E., Brasset E., Vaury C. (2014) History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in *Drosophila melanogaster*. *Front. Genet.* **5**, 257.
3. Lee Y.S., Nakahara K., Pham J.W., Kim K., He Z., Sontheimer E.J., Carthew R.W. (2004) Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*. **117**(1), 69–81.
4. Brennecke J., Aravin A.A., Stark A., Dus M., Kellis M., Sachidanandam R., Hannon G.J. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*. **128**(6), 1089–1103.
5. Aravin A.A., Hannon G.J., Brennecke J. (2007) The piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*. **318**(5851), 761–764.
6. Siomi H., Siomi M.C. (2009) On the road to reading the RNA-interference code. *Nature*. **457**(7228), 396–404.
7. Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H. (1998) A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev.* **12**(23), 3715–3727.
8. Baumgartner L., Handler D., Platzer S. W., Yu C., Duchek P., Brennecke J. (2022) The *Drosophila* ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. *Elife*. **11**, e80067.
9. Czech B., Malone C.D., Zhou R., Stark A., Schlingehayde C., Dus M., Perrimon N., Kellis M., Wohlschlegel J.A., Sachidanandam R., Hannon G.J., Brennecke J. (2008) An endogenous small interfering RNA pathway in *Drosophila*. *Nature*. **453**(7196), 798–802.
10. Flutre T., Duprat E., Feuillet C., Quesneville H. (2011) Considering transposable element diversification in *de novo* annotation approaches. *PLoS One*. **6**(1), e16526.
11. Benson D.A., Cavanaugh M., Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J., Sayers E.W. (2012) GenBank. *Nucl. Acids Res.* **41**, D36–D42.
12. Larkin A., Marygold S.J., Antonazzo G., Attrill H., dos Santos G., Garapati P.V., Goodman J.L., Gramates L.S., Millburn G., Strelets V.B., Tabone C.J., Thurmond J., FlyBase Consortium (2021) FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucl. Acids Res.* **49**, D899–D907.
13. Rech G.E. (2021) Fasta sequences for the *Drosophila melanogaster* manually curated transposable elements (MCTE) library. https://www.lareferencia.info/vufind/Record/ES_a77af0688d4516b156d77549ef-83f0e4#core
14. Simmons M.J., Thorp M.W., Buschette J.T., Peterson K., Cross E.W., Bjorklund E.L. (2010) Maternal impairment of transposon regulation in *Drosophila melanogaster* by mutations in the genes *aubergine*, *piwi*

- and *Suppressor of variegation* 205. *Genet. Res.* **92**(4), 261–272.
15. Васильева Л.А., Антоненко О.В., Захаров И.К. (2011) Роль мобильных генетических элементов в геноме *Drosophila melanogaster*. *Вавиловский журн. генетики и селекции.* **15**(2), 225–260.
 16. Rech G.E., Radío S., Guirao-Rico S., Aguilera L., Horvath V., Green L., Lindstadt H., Jamilloux V., Quesneville H., González J. (2022) Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nat. Commun.* **13**(1), 1948.
 17. Mohamed M., Dang N.T., Ogyama Y., Burlet N., Mugat B., Boulesteix M., Mérel V., Veber P., Salces-Ortiz J., Severac D., Péllisson A., Vieira C., Sabot F., Fablet M., Chambeyron S. (2020) A transposon story: from TE content to TE dynamic invasion of *Drosophila* genomes using the single-molecule sequencing technology from Oxford nanopore. *Cells.* **9**(8), 1776.
 18. Rang F.J., Kloosterman W.P., de Ridder J. (2018) From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **19**(1), 90.
 19. Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.
 20. Li H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* **34**(18), 3094–3100.
 21. Sedlazeck F.J., Rescheneder P., Smolka M., Fang H., Nattestad M., von Haeseler A., Schatz M.C. (2018) Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods.* **15**(6), 461–468.
 22. Kiełbasa S.M., Wan R., Sato K., Horton P., Frith M.C. (2011) Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**(3), 487–493.
 23. Chaisson M.J., Tesler G. (2012) Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics.* **13**(1), 238.
 24. Xiao C.L., Chen Y., Xie S.Q., Chen K.N., Wang Y., Han Y., Luo F., Xie Z. (2017) MECAT: fast mapping, error correction, and *de novo* assembly for single-molecule sequencing reads. *Nat. Methods.* **14**(11), 1072–1074.
 25. Лавренов А.Р., Нефедова Л.Н., Романова Н.И., Ким А.И. (2014) Экспрессия генов семейства *hp1* и их возможная роль в формировании фенотипа flamenco у *D. melanogaster*. *Биохимия.* **79**(11), 1554–1560.
 26. Кукушкина И.В., Махновский П.А., Нефедова Л.Н., Балакирева Е.А., Романова Н.И., Кузьмин И.В., Лавренов А.Р., Ким А.И. (2020) Исследование фертильности линии *Drosophila melanogaster* MS с нарушением контроля транспозиции мобильного элемента *gypsy*. *Молекуляр. биология.* **54**(3), 412–425.
 27. Kim A.I., Lyubomirskaya N.V., Belyaeva E.S., Shostack N.G., Ilyin Y.V. (1994) The introduction of a transpositionally active copy of retrotransposon GYPSY into the stable strain of *Drosophila melanogaster* causes genetic instability. *Mol. Gen. Genet. MGG.* **242**(4), 472–477.
 28. Кукушкина И.В., Махновский П.А., Нефедова Л.Н., Миляева П.А., Кузьмин И.В., Лавренов А.Р., Ким А.И. (2020) Анализ транскриптома линий *Drosophila melanogaster* с нарушением контроля транспозиции ретротранспозона *gypsy*. *Генетика.* **56**(5), 550–560.
 29. Andrews S., Krueger F., Segonds-Pichon A., Biggins L., Krueger C., Wingett S. (2010) FastQC. A quality control tool for high throughput sequence data. **370**.
 30. Bolger A.M., Lohse M., Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* **30**(15), 2114–2120.
 31. Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., Salzberg S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**(4), 1–13.
 32. Anders S., Pyl P.T., Huber W. (2015) HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics.* **31**(2), 166–169.
 33. Anders S., Huber W. (2012) Differential expression of RNA-Seq data at the gene level—the DESeq package. *Eur. Mol. Biol. Lab. (EMBL), Heidelberg.* **10**, f1000research.
 34. Koren S., Walenz B.P., Berlin K., Miller J.R., Bergman N.H., Phillippy A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**(5), 722–736.
 35. Thorvaldsdottir H., Robinson J.T., Mesirov J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**(2), 178–192.
 36. Cooley M.B., Yoder J.I., Goldsborough A.P., Still D.W. (1996) Site-selected insertional mutagenesis of tomato with maize *Ac* and *Ds* elements. *Mol. Gen. Genet. MGG.* **252**(1–2), 184–194.
 37. Kay M., Hojati Z., Heidari M., Bazi Z., Korbekandi H. (2015) Effects of disruption of the nucleotide pattern in CRID element and Kozak sequence of interferon β on mRNA stability and protein production. *Autoimmunity.* **48**(5), 336–343.
 38. Ким А.И., Беляева Е.С., Ларкина З.Г., Асланян М.М. (1989) Генетическая нестабильность и транспозиции мобильного элемента МДГ4 в мутаторной линии *Drosophila melanogaster*. *Генетика.* **25**(10), 1747–1756.
 39. Prud'homme N., Gans M., Masson M., Terzian C., Bucheton A. (1995) *Flamenco*, a gene controlling the *gypsy* retrovirus of *Drosophila melanogaster*. *Genetics.* **139**(2), 697–711.
 40. Leblanc P. (1997) Invertebrate retroviruses: ZAM a new candidate in *D. melanogaster*. *EMBO J.* **16**(24), 7521–7531.
 41. Czech B., Preall J.B., McGinn J., Hannon G.J. (2013) A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol. Cell.* **50**(5), 749–761.

42. Gracheva E., Dus M., Elgin S.C. (2009) *Drosophila* RISC component VIG and its homolog Vig2 impact heterochromatin formation. *PLoS One*. **4**(7), e6182.
43. Carrera P., Johnstone O., Nakamura A., Casanova J., Jackle H., Lasko P. (2000) Vasa mediates translation through interaction with a *Drosophila* yIF2 homolog. *Mol. Cell*. **5**, 181–187.
44. Jankowsky E. (2011) RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.* **36**(1), 19–29.
45. Styhler S., Nakamura A., Swan A., Suter B., Lasko P. (1998) *vasa* is required for Gurken accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. *Development*. **125**, 1569–1578.
46. Tomancak P., Guichet A., Zavorszky P., Ephrussi A. (1998) Oocyte polarity depends on regulation of *gurken* by Vasa. *Development*. **125**, 1723–1732.
47. Pek J.W., Kai T. (2011) A role for Vasa in regulating mitotic chromosome condensation in *Drosophila*. *Curr. Biol.* **21**, 39–44.
48. Munafò M., Manelli V., Falconio F.A., Sawle A., Kneuss E., Eastwood E.L., Seah J.E., Czech B., Hannon G.J. (2019) Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to the biogenesis machinery. *Genes Dev.* **33**(13–14), 844–856.
49. Stoye J.P. (2001) Endogenous retroviruses: still active after all these years? *Curr. Biol.* **11**(22), R914–R916.
50. Kaer K., Speck M. (2012) Intronic retroelements: Not just “speed bumps” for RNA polymerase II. *Mobile Genetic Elements*. **2**(3), 154–157.
51. Saze H. (2018) Epigenetic regulation of intragenic transposable elements: a two-edged sword. *J. Biochem.* **164**(5), 323–328.
52. Rogers S.O., Bendich A.J. (2023) Introns are derived from transposons. *bioRxiv*, 2023-02.
53. Ewing A.D., Smits N., Sanchez-Luque F.J., Faivre J., Brennan P.M., Richardson S.R., Cheetham S.W., Faulkner G.J. (2020) Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol. Cell*. **80**(5), 915–928.e5.
54. Dietzl G., Chen D., Schnorrrer F., Su K.C., Barinova Y., Fellner M., Gasser B., Kinsey K., Oppel S., Scheiblaue S., Couto A., Marra V., Keleman K., Dickson B.J. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*. **448**(7150), 151–156.

Search for the Insertions and Chromosomal Rearrangements Affecting Changes in Gene Expression in *D. melanogaster* Strains with Impaired Transposition Control of the gypsy Retrotransposon

I. V. Kukushkina^{1, *}, A. R. Lavrenov^{1, 2}, P. A. Milyaeva^{1, 3}, A. I. Lavrenova¹, I. V. Kuzmin¹,
L. N. Nefedova^{1, **}, and A. I. Kim^{1, 3}

¹Lomonosov Moscow State University, Moscow, 119234 Russia

²Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, 119071 Russia

³Faculty of Biology, Longgang District, Shenzhen MSU-BIT University, Shenzhen, 518172 China

*e-mail: vladimirova-bph@yandex.ru

**e-mail: nefedova@mail.bio.msu.ru

Transposable elements (TE) increase the frequency of spontaneous mutations in the genome and are also capable of altering function and affecting gene expression, so it is important to have an idea of their activity and position in the genome. The paper demonstrates the advantage of combining the analysis of two sequencing methods of searching TE insertions and chromosomal rearrangements: full-genome nanopore sequencing allows the detection of TE insertions, and the use of transcriptome sequencing evaluates the effect of insertions on gene expression. The results are presented using SS (*w*¹, *flamenco* mutant) and MS (*w*¹, *flamenco* mutant, active copy of *gypsy*) strains with the flamenco phenotype as an example to investigate the causes of impaired control of TE activity. The laboratory wild type strain D32 was used as a control. Insertions and deletions of TE into the euchromatin regions of the genome and into the introns of genes relative to the reference genome were found in the studied strains and wild-type strains. In the analyzed genomes, a search for insertions and deletions in RNA interference system genes and in differentially expressed genes in SS and MS strains with flamenco phenotype was performed. We have detected TE insertions in various structures of *AGO3*, *CG17147*, *Su(var)3-3*, *Gasz*, *CG43348*, *moody*, *CG17752* genes. For most of the analyzed genes, no correlation between a change in the TE position and a decrease or increase in gene transcription was found. A chromosomal rearrangement affecting the 3'-untranslated region has been detected for the *vig* gene. Based on the results of long-read sequencing, a *de novo* genome assembly for the MS strain was obtained. The increased expression in SS and MS strains for *CR45822* and *pst* genes was found to be associated with triplication, but not with changes in gene regulatory sequences or TE insertion.

Keywords: *Drosophila melanogaster*, mobile genetic elements, evolution, Oxford Nanopore, Illumina, genome assembly