

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
**ИНСТИТУТ
МОЛЕКУЛЯРНОЙ И КЛЕТОЧНОЙ
БИОЛОГИИ**
СИБИРСКОГО ОТДЕЛЕНИЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ИМКБ СО РАН)

пр. Академика Лаврентьева, д. 8/2, Новосибирск, 630090
телефон (383) 3639042, факс (383) 3639078
e-mail: info@mcb.nsc.ru
<http://www.mcb.nsc.ru>

ОКПО 30781167, ОГРН 1115476157070,
ИНН / КПП 5408291757 / 540801001

26.08.2024 № 15318-01-2Н/153
На № _____ от _____



УТВЕРЖДАЮ
Директор ИМКБ СО РАН
доктор биологических наук
Демаков Сергей Анатольевич
«26» августа 2024 г.

ОТЗЫВ

ведущей организации на диссертационную работу Кирова Ильи Владимировича
**«Особенности организации повторяющихся элементов генома растений,
выявленные с помощью новых омиксных подходов»**, представленную на соискание
ученой степени доктора биологических наук по специальности 1.5.7 – Генетика

Актуальность темы исследования

Повторенные последовательности ДНК широко представлены в геномах про- и эукариот и играют ключевую роль в широком спектре биологических и эволюционных процессов, при этом они остаются самой малоизученной частью генома. Из-за технических проблем с секвенированием, сборкой генома и аннотацией насыщение геномной ДНК повторяющимися последовательностями может приводить к ошибкам в базах данных, что делает разработку методов молекулярного и биоинформатического подходов для системного изучения повторяющихся элементов актуальной задачей современной биологии. До недавнего времени не существовало биоинформатических инструментов анализа данных секвенирования длинными ридами, позволяющих выявлять и анализировать повторенные последовательности ДНК, тогда как методы анализа коротких прочтений сильно ограничены в их поиске. В свете этого, особое значение имеет разработка инструментов анализа длинных прочтений.

Геномы растений в значительной степени представлены высококопийными сателлитными ДНК и мобильными генетическими элементами. Сателлитные повторенные

элементы являются главными структурными элементами центромер, теломер и гетерохроматина и могут служить удобными цитогенетическими маркерами отдельных хромосом, тогда как мобильные элементы являются драйверами эволюции и селекции, а создание методов для контролируемой активации нативных мобильных элементов генома растений позволяет расширить генетическое разнообразие, а также ускорить селекцию и развитие функциональной геномики. Исследование повторенных последовательностей ДНК растений важно для изучения варибельности геномов, возникающей в естественных условиях и в условиях биотехнологического размножения растений и селекции. Поиск и аннотация повторяющихся элементов являются важной частью современных проектов по секвенированию и сборке геномов растений, что позволяет ускорять процесс аннотации генов и проводить интегрирование собранных последовательностей с физическими хромосомами. Разработка и внедрение новых методов полногеномного анализа активности мобильного и идентификации новых инсерций мобильных элементов являются крайне необходимы для современной биологии, особенно в свете изучения важных сельскохозяйственных и культурных видов растений.

Значимость, полученных автором результатов, для науки и практики

В рамках диссертационной работы разработан комплекс биоинформатических (pyTanFinder, nanoTRF, NanoCasTE, nanotei и DRAWID) и молекулярно-биотехнологических методов (CANS), направленных на идентификацию новых повторяющихся элементов геномов растений и изучение генетической варибельности, обусловленной этими элементами, включая поиск новых инсерций мобильных элементов, а также направленных на ускорение цитологического и молекулярно-цитогенетического анализа.

Впервые на геномном и постгеномном уровнях изучены повторяющиеся элементы генома важных сельскохозяйственных (подсолнечник, гречиха, лук, тритикале, роза) и модельных (арабидопсис, мох (*Physcomitrium patens*)) растений. Установлены закономерности распределения соматических инсерций ретротранспозонов в геноме. Разработанные методы детекции новых инсерций мобильных элементов могут быть использованы в биотехнологии растений для ускорения процесса создания новых генотипов культурных растений, несущих новые функционально и фенотипически значимые инсерции.

Впервые выделены и изучены наборы новых высококопийных тандемных повторов (*Allium fistulosum*, *A. cepa*, *Rosa wichurana*, *R. chinensis* и *P. patens*), т.е. получен уникальный инструмент для молекулярных и эволюционных исследований генома этих видов растений. Детально изучена организация центромер *A. cepa* и *A. fistulosum*. Разработана система хромосомных маркеров для видов луковых и роз, что может быть использовано для

интегрирования физических и генетических карт, а также улучшения полногеномной сборки этих видов.

Впервые показано, что десятки LTR ретротранспозонов и сателлитные повторы экспрессируются как в нормальных, так и в стрессовых условиях. Более того, показано, что экспрессия сателлитных повторов характерна для филогенетически удалённых растений, включая лук и мох. В работе выявлены отличительные особенности геномной и транскриптомной организации экспрессирующихся ретротранспозонов, что вносит важный вклад в изучение фундаментальной проблемы роли мобильных элементов в формировании транскриптома растений.

Автором адаптировано нанопоровое секвенирование для полногеномного анализа внехромосомных кольцевых ДНК (вкДНК) растений, что позволило впервые изучить состав и структуру вкДНК. Это позволило расширить список известных мобильных элементов с доказанной мобильной активностью для тритикале, подсолнечника, рапса и арабидопсиса. Создана научная основа для дальнейшего использования методов контролируемой активации мобилома для создания инсерционных коллекций, а также изучения влияния новых инсерций на структуру генома и эпигенома растений.

Впервые показано, что мобильные элементы могут кодировать белки разной природы, включая белки с известной функцией, а также неизвестные белки, функции которых только предстоит понять в будущем.

Обоснованность и достоверность выводов

Достоверность полученных результатов основывается на применении адекватных поставленным задачам современных молекулярно-биологических и биоинформатических методов. Исследование проведено на большом количестве материала и включало все необходимые контрольные эксперименты. Результаты прекрасно иллюстрированы, осмыслены и подробно обсуждены. По теме работы опубликовано 18 работ в высокорейтинговых международных журналах, зарегистрирован один патент. Результаты апробированы на 12 российских и международных конференциях, а все разработанные программы выложены на общедоступные биоинформатические ресурсы.

Структура и содержание работы

Диссертационная работа состоит из глав «Введение», «Обзор литературы», «Материалы и методы», «Результаты и обсуждение», «Заключение», «Выводы», «Список сокращений», «Список публикаций», «Список литературы». Работа изложена на 266 страницах, содержит 79 рисунков, 14 таблиц. Список литературы содержит 475 источников.

«Введение» содержит основные необходимые подразделы. Автор со ссылкой на современную научную литературу обосновывает актуальность избранной темы, научную новизну и значимость исследования. Цель, задачи и положения, выносимые на защиту, сформулированы четко.

В главе «Обзор литературы» автор приводит классификацию повторяющихся элементов генома, включая сателлитные повторы и мобильные элементы, описывает особенности геномной организации, структуры и жизненного цикла LTR ретротранспозонов растений. Автором подробно освещена тема ретротранскриптома растений, отмечены закономерности организации, состав и методы их изучения. Проанализированы имеющиеся данные по мобилому растений. В обзоре также дан детальный анализ современных методов идентификации сателлитных повторов в геномах растений.

В части «Материалы и методы» описаны впечатляющие своим разнообразием основные цитогенетические, молекулярно-биологические, протеомные, транскриптомные, статистические и биоинформатические подходы, использованные автором для получения результатов. Детально описан использованный растительный материал.

Раздел «Результаты и обсуждения» разбит на пять подразделов. Каждый подраздел включает краткое введение, собственно результаты и обсуждение, разделенные дополнительного на части, и небольшое заключение.

Подраздел 4.1 «Цитогеномный анализ новых сателлитных повторов растений» включает подробное описание биоинформатических программ *pyTanFinder*, *nanoTRF* и *DRAWID*, разработанных автором диссертационной работы. Автор не только обосновывает необходимость разработки этих инструментов, но и проводит оценку результатов их работы с использованием нескольких наборов данных. Далее приводятся данные по выявлению и молекулярно-цитогенетической характеристике тандемных повторов *A. fistulosum*, *R. wichurana* и *R. chinensis* и *P. patens*. Важно подчеркнуть, что в каждом случае автор использует как разработанные другими авторами инструменты для поиска повторенных последовательностей генома, так и инструменты собственной разработки.

Подраздел 4.2 «Транскриптомные особенности ретротранспозонов растений» представляет результаты полногеномного анализа транскрипции RTE у подсолнечника и тритикале, выполненного с использованием данных Illumina и нанопорового секвенирования. Приведены результаты по изучению экспрессии LTR ретротранспозонов подсолнечника под действием эпигенетического стресса.

Подраздел 4.3 «Активные мобильные элементы растений и особенности их инсерций» начинается с описания разработанной автором программы *nanotei* для

идентификации нереперенсных инсерций транспозонов по данным полногеномного нанопорового секвенирования, проверки качества ее работы и применения для детекции новых инсерций активных мобильных элементов *A. thaliana*. Далее приводятся результаты адаптации и оптимизации метода CANS для растений, а также результаты разработки биоинформатического подхода для анализа данных CANS и анализа инсерций мобильных элементов в геноме растений. Для анализа данных CANS и выявления сайтов инсерций автором была разработана программа NanoCasTE, описание которой также дано в подразделе 4.3. Разработанные инструменты автор применяет для изучения организации инсерций LTR ретротранспозонов EVD и ONSEN в геноме *A. thaliana*. Автор представляет результаты изучения мобильной активности экспрессирующихся ретротранспозонов подсолнечника и анализа ретротранскриптома развивающейся зерновки тритикале. В последней части автор описывает ретротранспозоны *Fagopyrum tataricum* и *F. esculentum* и доказывает их влияние на эволюцию размера генома этих видов.

Подраздел 4.4. «Структура и состав внехромосомных кольцевых ДНК (вкДНК) LTR ретротранспозонов растений» помимо описания вкДНК, содержит результаты изучения структуры и состав вкДНК у *A. thaliana* и *Brassica napus* с помощью нанопорового секвенирования вкДНК и использовать полученные сведения для идентификации новых активных транспозонов в геноме этих видов.

В подразделе 4.5 «Мобильные элементы и протеом растений» приводятся и обсуждаются результаты работы по прямой детекции белков мобильных элементов методами масс-спектрометрии на примере мутанта *A. thaliana ddm1*.

Все полученные результаты всесторонне рассматриваются и обсуждаются. Проводится обширный анализ литературы с целью сопоставления полученных в диссертационной работе данных с результатами исследований других научных коллективов.

В «**Заключении**» автор логично объединяет все полученные результаты и выстраивает ясную научную взаимосвязь между отдельными главами работы.

В разделе «**Выводы**» автор четко формулирует выводы исследования, которые соответствуют поставленным целям и задачам, и отражают все результаты проведенной работы.

Автореферат соответствует установленным требованиям и полностью отражает основное содержание диссертационной работы.

Использование результатов работы

Результаты, полученные автором, имеют значимое фундаментальное и практическое значение. Они могут использоваться специалистами разных специальностей, занимающихся поиском и анализом повторенных последовательностей генома, а также включаться в теоретические и практические образовательные программы при обучении студентов университетов и ВУЗов, интересующихся исследованием структуры и функции геномов.

Замечания и комментарии

Работа написана хорошим языком и прекрасно иллюстрирована. Как и в любой работе, имеются небольшие недостатки в оформлении, например, встречаются точки в заголовках, единичные грамматические, пунктуационные и стилистические ошибки, англицизмы («сиквенсы»), отдельные недописанные или неудачно построенные предложения (например, на с. 27 «Первым обнаруженным элементом Tc1/mariner был Tc1 при изучении полиморфизма штаммов рестрикционных фрагментов у *Caenorhabditis elegans* в 1983 г. (Emmons et al., 1983)»; на с. 83 «Чтобы соединить два и более фрагментов одной хромосомы и создать идиограмму, мы также приступили к выполнению специального задания.»; на с. 119 «... что объясняет, почему нет четких на этой хромосоме можно обнаружить центромерный район»).

Большинство рисунков и отдельные таблицы приведены с надписями на английском языке без соответствующих расшифровок в подписях. Их описания не всегда полноценны, например, на Рисунке 4.22 приведена идиограмма с отмеченными FISH сигналами от проб на три tandemных повтора *R. wickurana*, при этом не указано, какой цвет какому повторенному элементу соответствует, а также чему соответствует разное количество «*». Большинство рисунков в диссертационной работе содержат очень мелкие обозначения, нечитаемые даже при 2-кратном увеличении текста. На рисунке 4.56 не подписаны части А и В.

Повсеместно встречается использование разных аббревиатур для одних и тех же терминов или одинаковые аббревиатуры для разных понятий: IN или INT для интегразы, вкДНК или ессDNA для внехромосомной кольцевой ДНК, ORF или OPC для открытой рамки считывания, гРНК для геномной РНК и для гидовых РНК и т.п., при этом в списке сокращений приведен лишь один из вариантов аббревиатуры и/или расшифровки.

Формулировка положений 1 и 8, выносимых на защиту, не очень удачна, так как представляет, скорее, результаты работы. В положении 9 хотелось бы видеть уточнение того, о чьих мобильных генетических элементах идет речь.

Раздел 2.1.1 (с. 14-15). В предложении «Длина мономера лежит в основе классификации tandemных повторов генома и их условного разделения на микросателлиты

(2–7 п.н.), минисателлиты (десятки п.н.) или сателлиты (сотни п.н.)» следовало бы привести ссылку на источник классификации, так как разными авторами выделяются разное количество категорий и устанавливают разные длины мономеров для принадлежности повторенных элементов к той или иной категории. Например, например Ляо и др. (2023), относят к микросателлитам повторённые элементы с размером повторяющейся единицы 1–4 п.н. (Liao X, Zhu W, Zhou J, Li H, Xu X, Zhang B, Gao X. Repetitive DNA sequence detection and its role in the human genome. *Commun Biol.* 2023 Sep 19;6(1):954. doi: 10.1038/s42003-023-05322-y).

Раздел 2.1.1 (с. 21). В предложении «Позднее, Wicker с соавторами предложил иерархическую классификацию всех МЭ эукариот.» нет ссылки на работу. По-видимому, подразумевается работа 2007 года: Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capu P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* 2007 Dec;8(12):973-82. doi: 10.1038/nrg2165. PMID: 17984973.

В разделе 2.5 обзора литературы, в котором описаны методы идентификации сателлитных повторов растений, к сожалению, не упоминается достаточно новый, но очень эффективный инструмент – TideCluster, который является, по сути, модифицированной версией TideHunter, в которую добавлен модуль TAREAN для оценки консенсусной последовательности выявленных повторенных элементов.

Раздел 4.1.2 (с. 75). «Для поиска ТП в «сырых» данных ONT была разработана программа nanoTRF (<https://github.com/Kirovez/nanoTRF>)». Несколько досадно, что столь полезный инструмент выложен на GitHub только в виде пререлиза.

Раздел 4.1.2 (с. 74-75). Из описания авторской программы nanoTRF осталось неясным, какое покрытие генома необходимо для качественного анализа и есть ли этап предварительной случайной фильтрации ридов, как, например, у TAREAN, перед проведением анализа?

В этом же разделе на с. 77 написано, что «... было проведено сравнение результатов nanoTRF с результатами программного обеспечения TAREAN (Novák et al., 2017), которое использует риды Illumina. Для этого были идентифицированы ТП с помощью TAREAN, используя риды Illumina для *D. antarctica* (González et al., 2018)». Непонятно, использовал ли автор данные из статьи Гонзалес с соавт (2018), или проводил анализ с помощью TAREAN самостоятельно. Если анализ был выполнен заново, то следовало бы привести ссылку на использованные данные, например, номер в NCBI, а также указать, какое количество ридов было взято в анализ.

В целом непонятно, почему автор сравнивает на с. 76-79 результаты работы nanoTRF с результатами работы TAREAN, который использует в качестве входных данных короткие прочтения и частичное покрытие генома («Read length should be 100–200 nt and the number of analyzed reads should represent less than 1× genome equivalent (genome coverage of 0.01–0.50× is recommended).» Petr Novák, Laura Ávila Robledillo, Andrea Koblížková, Iva Vrbová, Pavel Neumann, Jiří Macas, TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads, *Nucleic Acids Research*, Volume 45, Issue 12, 7 July 2017, Page e111, <https://doi.org/10.1093/nar/gkx257>), тогда как авторская программа работает с длинными геномными ридами Nanopore и, по-видимому, не имеет ограничения на количество входных данных.

Раздел 4.1.4.1. Автор пишет на с. 88, что «Все контиги из кластеров с глобулярной структурой (соответствующей тандемно организованным повторам) сравнивали с известными последовательностями из NCBI с помощью BLASTN». Были ли выявлены кластеры кольцеобразной формы, которые также соответствуют тандемно организованной ДНК? Сравнивали ли их?

Здесь же, на с. 88-89 указано, что «Эти четыре кластера вместе составляют 7,12% генома. Два других кластера, CL58 (Рисунок 4.6А) и CL36 (Рисунок 4.6В), составляют, соответственно, 0,09 и 0,16% от общего числа использованных ридов». Неясно, как число проанализированных ридов соотносится с размером генома.

В раздел 4.1.5.2 при указании ссылки на RepeatExplorer2 на с.101 указывается работа Новак с соавт. (2013), тогда как данный инструмент был представлен в работе того же автора, но в 2020 г. (Novák P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat Protoc.* 2020 Nov;15(11):3745-3776. doi: 10.1038/s41596-020-0400-y. Epub 2020 Oct 23. PMID: 33097925.). Также непонятно, зачем использовали два инструмента, RepeatExplorer2 и TAREAN, если последний интегрирован в RepeatExplorer2 (аналогично на с.106.).

Раздел 4.1.7.1, рисунок 4.24, с. 124. Было бы интересно соотнести, где какой тандемный повтор на графике А и В, тем более что в таблице 4.3., приведенной чуть выше, данных по GC-составу нет. На этом же рисунке приводятся фотографии электрофореза продуктов ПЦР, из которых очевидно, что проводился подбор условий для поиска оптимальной температур отжига праймеров, однако, об этом не упоминается в тексте работы.

Заключение. Суммируя вышесказанное, можно заключить, что диссертационная работа Кирова Ильи Владимировича «Особенности организации повторяющихся

элементов генома растений, выявленные с помощью новых омиксных подходов», представленная к защите на соискание ученой степени доктора биологических наук по специальности 1.5.7 – Генетика, выполнена на высоком методическом уровне и является законченным научно-квалификационным трудом. По содержанию, актуальности, новизне, научному и методическому уровню, практической и теоретической значимости полученных результатов, диссертационная работа полностью соответствует требованиям «Положения о порядке присуждения научных степеней», утвержденного Постановлением Правительства РФ от 24 сентября 2013 г. №842, предъявляемым к диссертациям на соискание ученой степени доктора биологических наук, а ее автор, Киров Илья Владимирович, заслуживает присуждения искомой степени доктора биологических наук по специальности 1.5.7 – Генетика.

Отзыв заслушан и утвержден на заседании Отдела разнообразия и эволюции геномов Федерального государственного учреждения науки Института молекулярной и клеточной биологии Сибирского отделения Российской академии наук 26 августа 2024 г. (Протокол №112).

Ведущий научный сотрудник лаборатории цитогенетики животных Федерального государственного учреждения науки Института молекулярной и клеточной биологии Сибирского отделения Российской академии наук, доктор биологических наук,

Романенко Светлана Анатольевна

26 августа 2024 г.

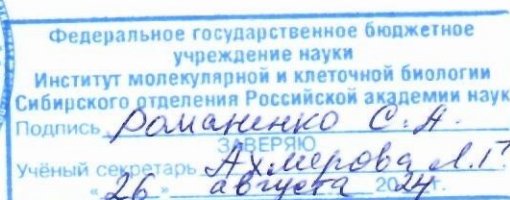
Сведения о лице, утвердившем отзыв:

Демаков Сергей Анатольевич
Директор ИМКБ СО РАН

Адрес: 630090, г. Новосибирск, пр-т Академика Лаврентьева, 8/2

Тел. +7 (383) 373-02-49

Электронная почта: demakov@mcb.nsc.ru



Сведения о составителе отзыва:

Романенко Светлана Анатольевна

Адрес: 630090, г. Новосибирск, пр-т Академика Лаврентьева, 8/2

Тел. +7 (383) 373-02-49

Электронная почта: gosa@mcb.nsc.ru