

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
НАУЧНОЕ УЧРЕЖДЕНИЕ

**«ФЕДЕРАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР
ИНСТИТУТ
ЦИТОЛОГИИ И ГЕНЕТИКИ
СИБИРСКОГО ОТДЕЛЕНИЯ
РОССИЙСКОЙ АКАДЕМИИ НАУК»
(ИЦиГ СО РАН)**

Пр-т. Академика Лаврентьева, д. 10, Новосибирск, 630090
Телефон: (383) 363-49-80
Факс (383) 333-12-78
E-mail: icg-adm@bionet.nsc.ru
<https://www.icgbio.ru>
ИНН 5408100138/КПП 540801001
ОКПО 03533895 ОГРН 1025403657410

УТВЕРЖДАЮ

Директор ФГБНУ «Федеральный
исследовательский центр Институт
цитологии и генетики Сибирского
отделения Российской академии наук»
(ИЦиГ СО РАН) академик РАН

Кочетов А.В.

2024 г.



от 26.02.2024 № 15345-29-38/222
на № _____ от _____

ОТЗЫВ

ведущей организации на диссертационную работу **Бондар Евгении Ивановны на тему «Аннотация генома и предсказание сайтов начала транскрипции для лиственницы сибирской (*Larix sibirica* Ledeb.)»**, представленной соискание ученой степени кандидата биологических наук по специальности 1.5.7 – Генетика

Актуальность темы диссертационной работы

Диссертационная работа Бондар Е.И. посвящена аннотации полного генома лиственницы сибирской *Larix sibirica* Ledeb., а также ее расширение на основе полногеномного предсказания сайтов начала транскрипции. Актуальность работы обоснована сразу несколькими факторами.

Информация о геноме является важнейшей основой для изучения функционирования организма на основе расшифровки генов, выявления механизмов их регуляции, изучения особенностей эволюции генов при сравнении с генами других видов. Перспективность такого подхода, с опорой на геномные данные, продемонстрирована при изучении модельных организмов животных (мышь, крыса, мушка дрозофила и др.) и растений (*Arabidopsis thaliana*, рис). Для растений информация о геномных последовательностях позволяет реконструировать их эволюцию и на основе сравнения с геномами других растительных видов, обитающих в разных условиях, выявлять механизмы адаптации к внешней среде и стрессовым факторам.

Знание лишь последовательности геномной ДНК, однако, недостаточно для решения указанных выше задач. Необходима расшифровка структуры генома, поиск участков, кодирующих различные гены, регуляторные области, а для растений еще важно идентифицировать повторенные участки, которые преобладают в их геномах. Основным подходом к решению этих задач является использование методов биоинформатики для

компьютерной аннотации геномов, распознавания регуляторных районов и повторов. Их производительность и точность достаточны для аннотации многих белок-кодирующих генов и мобильных элементов. Эти методы широко используются при изучении геномов немодельных организмов.

В качестве объекта исследования выбран геном лиственницы сибирской (*Larix sibirica* Ledeb.), листопадного хвойного дерева. Это растение занимает около 40% лесистой территории России, используется в качестве источника ценной древесины. Исследование генома этого вида представляет интерес как с фундаментальной, так и с прикладной точки зрения. Результаты анализа пополняют наши знания о структуре геномов и их изменчивости у хвойных растений. Информация о последовательностях маркеров в геноме лиственницы сибирской и родственных видов позволит идентифицировать биологический материал и проводить исследования разнообразия представленности лиственниц различных видов в лесах России.

Научная новизна и практическая значимость работы

В работе впервые проведена подробная аннотация геномной последовательности лиственницы сибирской. С помощью методов биоинформатики установлены структуры генов, определены области мобильных элементов. У хвойных растений рода *Larix* такая полная аннотация получена впервые. На основе геномных последовательностей лиственницы сибирской разработаны полиморфные SSR маркеры, которые могут быть использованы в популяционных исследованиях как лиственницы сибирской, так и лиственниц других видов.

Впервые для геномов трех видов семейства Pinaceae проведен широкомасштабный биоинформатический анализ по идентификации положения сайтов старта транскрипции (TSS). Для решения этой задачи предложен комплексный метод на основе оценки целого ряда параметров, включая распределения длин 5' нетранслируемой области, свободной энергии ДНК дуплексов, позиционного распределения сайтов связывания транскрипционных факторов.

Особый интерес представляет результат по реконструкции митохондриального генома лиственницы сибирской, который, несмотря на использование длинных прочтений Oxford Nanopore, не удалось собрать в одну кольцевую молекулу ДНК и размер которого превосходит 11 млн. п.о. Этот геном по оценкам авторов работы является самым большим из известных митохондриальных геномов, при этом функция существенной доли ДНК этого генома оказалось неизвестной. Полученные данные безусловно привлекут внимание научных исследователей и стимулируют дальнейшие исследования в этой области.

Полученные в работе аннотации генома лиственницы сибирской могут служить основой для дальнейшего изучения особенностей структуры и эволюции геномов голосеменных растений и, на более общем уровне, наземных растений. Положения сайтов старта транскрипции позволят более точно идентифицировать молекулярные механизмы регуляции экспрессии генов у этих растений.

Последовательности SSR 14 полиморфных маркеров могут быть использованы для исследования популяций лиственниц сибирской, Гмелина и Каяндера, оценки в них уровня генетического разнообразия, что может быть важно для описания особенностей экосистемы лесов России в целом.

Отметим, что результаты анализа, полученные в работе, были предоставлены в свободный доступ, что будет способствовать развитию геномного ресурса для рода *Larix* и хвойных в целом.

Обоснованность и достоверность научных положений, выводов и заключений

Для получения большинства результатов соискателем были использованы методы биоинформатики. В силу сложности изучения генома экспериментальными методами, о достоверности полученных результатов можно судить на основе (1) адекватного использования наиболее современных биоинформатических методов анализа и (2) использования статистических критериев для проверки гипотез о характеристиках геномных последовательностей.

Эти требования в работе были выполнены. Более того, соискатель использовала целый комплекс методов биоинформатики для решения поставленных задач. Например, для идентификации повторенных последовательностей был использован набор из реконструированной библиотеки повторов, специфических для данного генома, данные из библиотеки RepBase и других баз данных по повторенным геномным последовательностям растений. Для предсказания TSS также использовались несколько методов для оценки различных статистических характеристик геномных последовательностей. Такой комплексный подход, безусловно, обеспечивал необходимую достоверность полученных результатов.

Для реконструкции структур генов использован комплексный подход на основе конвейера MAKER2, успешно зарекомендовавший себя при аннотации других геномов растений. Для полученных оценок статистических характеристик ядерного генома листовницы сибирской и ее пластидного и митохондриального генома проводилось сравнение с известными аналогичными характеристиками для других видов растений, в том числе и хвойными. Для оценки достоверности полученных результатов по реконструкции геномов и аннотации дополнительно использовались различные методы. Так для оценки достоверности выявления генов, кодирующих белки, использовался метод BUSCO, для оценки правильности сборки митохондриального генома был использован метод REAPR.

В случае разработки микросателлитных маркеров, проведена экспериментальная верификация успешности амплификации соответствующих праймеров. Это позволило отобрать локусы, которые демонстрируют наиболее стабильные интерпретируемые спектры. Полученные локусы были протестированы на трех видах листовниц из разных ареалов произрастания.

Таким образом, можно заключить, что полученные автором результаты, выводы, а также положения, выносимые на защиту, являются обоснованными и достоверными.

Оценка содержания диссертационной работы и ее завершенности

Материалы и результаты работы изложены последовательно и аргументировано. Диссертация построена по традиционному плану и включает следующие разделы: введение, обзор литературы, главы «Материалы и методы», «Результаты и обсуждение», разделы «Заключение», «Выводы», список сокращений и условных обозначений, список литературы (360 источников), 7 приложений. Общий объем текста диссертации составляет 151 страницу и включает 26 рисунков и 13 таблиц. Диссертационная работа, как и автореферат, оформлены в соответствии с установленными требованиями. Автореферат в необходимой и достаточной степени отражает результаты диссертационной работы. Необходимо отметить, что текст написан хорошим языком, изложен логично, опечаток в нем мало.

В литературном обзоре рассмотрены существующие на сегодняшний день результаты секвенирования и аннотации геномов зеленых растений с фокусом на геномах хвойных

растений. Охарактеризован вид лиственницы сибирской. Описаны биоинформатические методы, используемые для аннотации геномов: поиска генов, их функциональной аннотации, распознавания TSS. Описана методика разработки микросателлитных маркеров.

В главе «Материалы и методы» описаны источники данных и методы исследования. Они включают: программы и методы выявления повторенных последовательностей, оценки времени вставки ретротранспозонов, методы поиска и функциональной аннотации белок-кодирующих генов, сборки и аннотации геномов органелл (пластид и митохондрий), методы предсказания TSS, методы разработки и апробации микросателлитных маркеров.

В главе результаты «Результаты и обсуждение» описаны результаты исследования в соответствии с поставленными задачами: анализ повторяющихся элементов генома, оценка времени вставки ретротранспозонов LTR-RT, структурная аннотация ядерного генома и предсказание функций генов, результаты сборки и аннотации геномов органелл, результаты предсказания сайтов начала транскрипции, результаты разработки микросателлитных маркеров и оценки генетического разнообразия лиственниц сибирской, Гмелина и Каяндера. В Заключение подводится итог работы.

Текст работы полностью соответствует поставленным целям исследования. Таким образом, поставленные цели и задачи выполнены в полном объеме, что позволяет говорить о завершенности поставленной работы. Заключение и выводы являются обоснованными и аргументированными.

Замечания по содержанию диссертации

Несмотря на общую высокую оценку диссертационной работы, к работе есть ряд незначительных замечаний, в основном редакционного характера:

1. В обзоре литературы хотелось бы видеть больше иллюстративного материала (он содержит всего 2 рисунка).
2. В тексте диссертации отсутствует ссылка на идентификаторы последовательностей митохондриального генома лиственницы сибирской.
3. Ссылки на данные по результатам работы на сервере СФУ оказались неработающими.
4. На рис 20 панели указаны на русском языке, ссылка на них в тексте – на английском.
5. Что касается сравнения аннотации генов терминами GO у *L. sibirica* и других растений, то в методах не описано, откуда были взяты термины аннотации для генов других растений (стр. 49): из других баз данных или получены автором тем же методом, что и для *L. sibirica*. В первом случае разные методы аннотации разных геномов могут повлиять на результаты сравнения частот встречаемости терминов.
6. Кроме того, насколько понятно из методов, статистика встречаемости GO терминов в разных геномах была подсчитана на основе аннотации отдельных генов. Полагаю, в этом случае, на статистику могут влиять дубликации и потери отдельных генов. Вероятно, более адекватный способ: разбить все гены анализируемых растений на ортогруппы и проводить такой анализ уже по ортогруппам, отражающим функциональное сходство генов из разных видов.

Эти замечания, однако, не снижают высокий уровень представленной работы и основных научных положений и выводов диссертации.

Подтверждение опубликования полученных результатов диссертации в научной печати

По материалам диссертационной работы опубликовано пять статей в рецензируемых научных журналах, входящие в основные библиометрические базы данных (РИНЦ, WoS и Scopus). В трех публикациях соискатель является первым автором. Опубликованные по теме диссертации работы в достаточной степени отражают ее основные положения. Также основные положения и результаты диссертационной работы были представлены на международных и российских научных конференциях и опубликованы в сборниках трудов этих конференций (14 публикаций).

Заключение

Диссертационная работа Бондар Е.И. является законченной и выполнена соискателем на высоком научном уровне. Представленные в работе исследования достоверны, использованные методы и результаты хорошо описаны, обсуждена значимость полученных результатов и их связь с современным представлением о структурах геномов хвойных растений и функциях их генов. Выводы и положения, выносимые на защиту обоснованы.

Диссертационная работа Бондар Евгении Ивановны «Аннотация генома и предсказание сайтов начала транскрипции для лиственницы сибирской (*Larix sibirica* Ledeb.)» представляет собой завершённое самостоятельное исследование, содержание диссертации полностью соответствует п.п. 9-14 «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства Российской Федерации №842 от 24.09.2013 с изменениями, предъявляемым к диссертациям, представленным на соискание степени кандидата биологических наук по, а ее автор Бондар Е.И. заслуживает присуждения искомой степени кандидата биологических наук по специальности 1.5.7 – генетика биологической отрасли науки.

Отзыв был заслушан и утвержден на семинаре Лаборатории эволюционной биоинформатики и теоретической генетики «ФИЦ Институт цитологии и генетики СО РАН» 19 февраля 2024 года.

зав. лабораторией эволюционной
биоинформатики и теоретической генетики
ФГБНУ «Федеральный исследовательский центр
Институт цитологии и генетики Сибирского
отделения Российской академии наук»
(ИЦиГ СО РАН), к.б.н., доцент
(e-mail ada@bionet.nsc.ru)


Афонников Д.А.

Ученый секретарь ИЦиГ СО РАН, к.б.н.





Орлова Г.В.