

УТВЕРЖДАЮ

Проректор по учебной работе

ФГАОУ ВО «Сибирский  
федеральный университет»

Денис Сергеевич Гуц



*[Handwritten signature]*

« 14 » июня 2023 г.

## ЗАКЛЮЧЕНИЕ

Федерального государственного автономного образовательного учреждения  
высшего образования «Сибирский федеральный университет»

Диссертация «Аннотация генома и предсказание сайтов начала транскрипции для лиственницы сибирской (*Larix sibirica* Ledeb)» выполнена на кафедре геномики и биоинформатики.

В период подготовки диссертации соискатель Бондар Евгения Ивановна обучалась в аспирантуре Федерального государственного автономного образовательного учреждения высшего образования «Сибирский федеральный университет» и работала в должности инженера-исследователя лаборатории лесной геномики Федерального государственного автономного образовательного учреждения высшего образования «Сибирский федеральный университет», а также в должности младшего научного сотрудника лаборатории геномных исследований и биотехнологии Федерального исследовательского центра «Красноярский научный центр Сибирского отделения Российской академии наук» (ФИЦ КНЦ СО РАН).

В 2018 году Бондар Е.И. окончила «с отличием» магистратуру ФГАОУ ВО «Сибирский федеральный университет» по направлению подготовки 06.04.01 Биология.

В 2022 году окончила аспирантуру ФГАОУ ВО «Сибирский федеральный университет» по направлению подготовки 06.06.01 Биологические науки.

Справка о сдаче кандидатских экзаменов по английскому языку и по философии науки выдана ФГАОУ ВО «Сибирский федеральный университет» в 2023 году. Справка о сдаче кандидатского экзамена по генетике выдана Федеральным государственным бюджетным учреждением науки Институт общей генетики им. Н.И. Вавилова Российской академии наук в 2023 году.

Научный руководитель — Татарина Татьяна Валерьевна, PhD, доцент кафедры биологии Университета Ла Верна, заведующий кафедрой вычислительной биологии Флетчера Джонса.

На заседании присутствовали:

Шишов Владимир Валерьевич, д-р техн. наук, доцент; Иванова Елена Анатольевна, д-р биол. наук, доцент; Ямских Ирина Евгеньевна, д-р биол. наук, доцент; Садовский Михаил Георгиевич, д-р физ.-мат. наук; Крутовский Константин Валерьевич, канд. биол. наук; Орешкова Наталья Викторовна, канд. биол. наук; Кириченко Наталья Ивановна, д-р биол. наук.; Кузьмин Дмитрий Александрович, канд. техн. наук, доцент; Субботина Татьяна Николаевна, канд. биол. наук; Трусова Мария Юрьевна, канд. биол. наук; Дорошков Алексей Владимирович, канд. биол. наук; Гусейнов Олег Аладдинович, канд. биол. наук; Ибе Алексей Александрович, канд. с.-х. наук; Шаров Вадим Витальевич, науч. сотр.; Путинцева Юлия Андреевна, старший преподаватель; Новикова Серафима Валерьевна, мл. науч. сотр.; Рябова Ксения Константиновна, аспирант.

Были заданы следующие вопросы:

Что такое аннотация, какое точное определение у этого термина? Можно ли на основании последовательности генов отличить листовницу Каяндера и листовницу Гмелина? Можно ли детектировать период вегетативного роста у растений по уровню экспрессии генов? Почему в сравнении среди видов не

было *Pinus sylvestris*? Как проводилось оценивание времени вставки LTR ретротранспозонов? Откуда были взяты модель замен и значение скорости накопления замен для расчета оценки времени вставки LTR? Почему разброс в оценке времени вставки такой большой? Что могло повлиять на оценку времени встраивания ретротранспозонов? Как можно сделать оценку времени встраивания LTR ретротранспозонов более точной? Проверяли ли в работе взаимное расположение генов и повторов? Каков был личный вклад автора на каждом из этапов работы? Есть ли работы, связывающие глобальные климатические изменения с генетическими особенностями или изменением в экспрессии генов у хвойных деревьев? Как физически организована митохондриальная ДНК в клетках лиственницы? Как отличить митохондриальную ДНК от ядерной в общем пуле ДНК?

На все вопросы соискатель дал исчерпывающие ответы.

По итогам обсуждения принято следующее заключение:

#### **Актуальность исследования**

У более чем 70 % видов всех растений не было секвенировано ни одного участка ДНК, не говоря уже о полном геноме. Из геномов 973 уникальных видов растений, опубликованных в NCBI Genbank по состоянию на август 2022 г., 854 относятся к покрытосеменным, 80 — к водорослям, 12 — к мхам, 14 — к папоротникам и 19 — к голосеменным. Таксономическое распределение публично доступных геномов растений довольно смещено в сторону сельскохозяйственных культур. Многие опубликованные геномы, хотя и достаточно полные на уровне последовательности, имеют очень фрагментарные сборки. Обилие псевдогенов, увеличенное число генных семейств и пролиферирующая активность мобильных элементов затрудняют корректную сборку и аннотацию многих растительных геномов.

Хвойные — древняя группа голосеменных растений, представленная более чем 600 видами, играющими значительную роль в экосистемах бореальных лесов. Очень большие размеры их геномов и высокое содержание повторов в

них требуют для расшифровки геномов хвойных больше времени и усилий, чем для многих других видов растений. Недавно были секвенированы и собраны до чернового состояния несколько мегагеномов хвойных видов. Несмотря на то, что такие аннотации являются предварительными, они дают возможность структурного и функционального анализа. Даже неполный геном является ценным источником данных, и может помочь в понимании регуляторных отношений между элементами генома.

Лиственница сибирская *Larix sibirica* — морозостойкое листопадное хвойное дерево, произрастающее на востоке и северо-востоке европейской части России, Урале, Западной и Восточной Сибири. Она образует обширные хвойные леса, часто произрастающие вместе с сосной обыкновенной, елью сибирской и кедром сибирским, занимающие почти 263 млн га, что составляет около 40 % покрытой лесом территории России. Лиственница сибирская известна своей морозостойкостью, относительно быстрым ростом и устойчивой к гниению древесиной, что делает ее особенно ценной в строительстве. Хотя геном лиственницы сибирской был впервые опубликован в 2019 году (Kuzmin et al. 2019), для нее до сих пор не существует публично-доступной генной аннотации.

### **Цель работы**

Основной целью данного исследования было получение аннотации полного генома лиственницы сибирской *Larix sibirica* Ledeb., а также ее улучшение с помощью полногеномного предсказания сайтов начала транскрипции.

### **Научная новизна исследования**

Впервые представлена подробная аннотация генов и мобильных элементов генома лиственницы сибирской. Данная аннотация является первым публично доступным ресурсом для рода *Larix*. Была получена комплексная видоспецифическая библиотека повторов, которая может использоваться для поиска мобильных элементов в геномах других голосеменных растений. Были

разработаны и протестированы полиморфные SSR-маркеры для лиственницы сибирской, подходящие также для популяционных исследований лиственниц Гмелина и Каяндера. Для трёх видов семейства Pinaceae были предсказаны сайты начала транскрипции с помощью вычислительных подходов, основанных на методе максимизации ожидания и классификации нейронной сетью; был опробован метод валидации предсказаний *de novo* на основе распределения длин 5'-нетранслируемой области, профиля распределения свободной энергии ДНК дуплексов и позиционного распределения сайтов связывания транскрипционных факторов. Эта работа является первым полногеномным предсказанием TSS в геномах размером более 10 млрд.п.н.

### **Теоретическая и практическая значимость работы**

Теоретическая значимость работы обусловлена тем, что черновые сборки и аннотации геномов хвойных представляют собой ценный ресурс для дальнейших геномных и генетических исследований. Текущее состояние аннотаций геномов хвойных позволяет наблюдать различия между видами голосеменных и покрытосеменных на геномном уровне, которые проявляются в обилии генов в различных функциональных категориях. Разработанные в данной работе полиморфные SSR-маркеры позволяют оценивать уровень генетического разнообразия и дифференциации популяций лиственницы сибирской. Данные маркеры могут успешно применяться для изучения также лиственниц Гмелина и Каяндера и специально разработаны с учётом их возможного использования также в лабораториях, где отсутствует техническая возможность проведения капиллярного электрофореза. Предсказанные TSS и соответствующие им промоторные области являются основой будущей экспериментальной проверки и представляют собой ценный ресурс для лучшего понимания регуляции генов и исследования эволюционных отношений между голосеменными и покрытосеменными растениями. Идентификация TSS может найти своё применение в генетической селекции и редактировании генома, предоставляя возможности более точного картирования мутаций и полиморфизмов в функциональных областях генома и

локусах количественных признаков (QTL), связанных с экспрессией генов и адаптивными признаками, такими как скорость роста, устойчивость к холоду и засухе, резистентность и устойчивость к инвазии патогенов. Все данные, полученные в данной работе, включая файлы аннотации и комплексную библиотеку повторов, доступны публично в figshare (DOI: 10.6084/m9.figshare.19785913) и в репозитории суперкомпьютерного центра СФУ. Геномные последовательности, треки с генными моделями, предсказания TSS и данные о покрытии РНК-секвенирования доступны в геномном браузере Persephone (<https://web.persephonesoft.com>).

### **Положения, выносимые на защиту**

1. Получена подробная структурная и функциональная аннотация ядерного, митохондриального и хлоропластного геномов для вида *Larix sibirica*. Размер митохондриального генома составил 11,7 млн.п.н., что на текущий момент является самым большим митогеномом из известных.
2. Оценка доли повторов в геноме лиственницы составляет порядка 66%. Вероятный период массированного встраивания ретротранспозонов в геноме лиственницы может быть оценен порядка 4-5 млн лет назад.
3. 14 полиморфных микросателлитных маркеров, разработанных в данном исследовании для лиственницы сибирской, могут также использоваться для популяционно-генетических исследований лиственницы Гмелина и Каяндера.

### **Личный вклад автора в проведенные исследования**

Автором работы выполнены лично: аннотация хлоропластного генома, проверка сборки и анализ повторов в митохондриальном геноме, черновая аннотация митохондриального генома, идентификация и анализ повторов в ядерном геноме, оценка эволюционного времени расхождения (дивергенции) или вставки дублированных длинных концевых повторов-ретротранспозонов (LTR-RT) внутри вида, функциональная аннотация ядерного генома, сравнительный анализ представленности категорий генной онтологии,

предсказание сайтов начала транскрипции и анализ статистических свойств геномов хвойных, подготовка данных к публикации и написание рукописей статей.

Секвенирование транскриптома лиственницы сибирской проводилось сотрудниками лаборатории лесной геномики СФУ под руководством Орешковой Н. В. в рамках гранта Правительства РФ (договор № 14.Y26.31.0004, руководитель проекта проф. К. В. Крутовский). Сборка транскриптомных данных выполнена сотрудником лаборатории лесной геномики СФУ Бирюковым В.В. Секвенирование и сборка митохондриального генома лиственницы сибирской проводились в рамках гранта РФФИ № 16-04-01400 (руководитель проекта проф. К. В. Крутовский), выделение интактных митохондрий и обогащенной митохондриальной ДНК проводилась в СИФИБР СО РАН в лаборатории генетической инженерии растений под руководством Константинова Ю.М., работы по аннотации митохондриального генома выполнены совместно с Путинцевой Ю.А. Идентификация генов с лейцин богатыми повторами проводилась Мирошниковой К.А. Организация запуска программного конвейера для аннотации на 448 ядерном вычислительном кластере СФУ проводилась совместно с сотрудниками кафедры высокопроизводительных вычислений СФУ под руководством Кузьмина Д.А. и Феранчука И.С. Тестирование микросателлитных маркеров проводилось под руководством Орешковой Н. В. Образцы лиственницы сибирской для тестирования микросателлитных локусов предоставлены сотрудниками отдела мониторинга состояния лесных генетических ресурсов Центра защиты леса г. Красноярск.

**Достоверность результатов обеспечивается** использованием адекватных современных молекулярно-генетических и биоинформатических методов работы; обсуждением основных положений работы на научных всероссийских и международных конференциях. Результаты исследования в полном объеме были изложены в статьях в специализированных международных журналах, содержащихся в перечне Высшей Аттестационной Комиссии Российской

Федерации и индексирующихся базами Web of Science и Scopus, что обеспечивает критическое рассмотрение и рецензирование публикуемых работ научной общественностью. Сформулированные в диссертации положения, выводы и заключения отражают большой объем проделанной диссертантом работы.

### **Апробация работы**

Промежуточные и итоговые результаты работы были представлены на российских и международных конференциях: III Всероссийская конференция «Высокопроизводительное секвенирование в геномике» (19–24 июня 2022 г, Новосибирск), 6-ая Международная научная конференция «Plant Genetics, Genomics, Bioinformatics, and Biotechnology (PlantGen2021)» (14–18 июня 2021 г, Новосибирск), Международная конференция американской ассоциации RASA Global 2020 (2020, онлайн), 12-ая международная конференция «Биоинформатика регуляции и структуры генома\системная биология BGRS» (06-10 июля 2020 г, Новосибирск), 6-ая международная конференция-совещание, 11-я международная конференция «Биоинформатика регуляции и структуры генома\системная биология BGRS» (2018 г, Новосибирск).

### **Полнота изложения материалов диссертации в работах, опубликованных соискателем.**

Результаты исследований Бондар Е.И. изложены в 5 статьях, опубликованных в международных и отечественных рецензируемых изданиях, входящих в перечень ВАК и индексируемых в базах Web of Science и Scopus, а также 14 научных работ опубликованных в сборниках материалов российских и международных конференций.

*Публикации в рецензируемых научных изданиях, рекомендованных ВАК и в изданиях, индексируемых в базе Web of Science:*

1. **Bondar, E. I.** Annotation of Siberian Larch (*Larix sibirica* Ledeb.) Nuclear Genome – One of the Most Cold-Resistant Tree Species in the Only



Deciduous GENUS in *Pinaceae* / E. I. Bondar, S. I. Feranchuk, K. A. Miroshnikova, V. V. Sharov, D. A. Kuzmin, N. V. Oreshkova, K. V. Krutovsky // *Plants*. – 2022a. – Vol. 11, Iss. 15. – P. 2062.

2. **Bondar, E. I.** Genome-wide prediction of transcription start sites in conifers / E. I. Bondar, M. E. Troukhan, K. V. Krutovsky, T. V. Tatarinova // *International Journal of Molecular Sciences*. – 2022b. – Vol. 23, Iss. 3. – P. 1735.

3. Putintseva, Yu. A. Siberian larch (*Larix sibirica* Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome / Y. A. Putintseva, **E. I. Bondar**, E. P. Simonov, V. V. Sharov, N. V. Oreshkova, D. A. Kuzmin, Y. M. Konstantinov, V.N. Shmakov, V.I. Belkov, M.G. Sadovsky, O. Keech, K. V. Krutovsky // *BMC genomics*. – 2020. – Vol. 21, Iss. 1. – P. 1-12.

4. **Bondar, E. I.** Siberian larch (*Larix sibirica* Ledeb.) chloroplast genome and development of polymorphic chloroplast markers / E. I. Bondar, Y. A. Putintseva, N. V. Oreshkova, K. V. Krutovsky // *BMC bioinformatics*. – 2019. – Vol. 20, Iss. 1. – P. 47-52.

5. Орешкова, Н. В. Разработка ядерных микросателлитных маркеров с длинными (трех-, четырех-, пяти- и шестинуклеотидными) мотивами для трех видов лиственницы на основе полногеномного *de novo* секвенирования лиственницы сибирской (*Larix sibirica* Ledeb.) / Н. В. Орешкова, **Е. И. Бондар**, Ю. А. Путинцева, В. В. Шаров, Д. А. Кузьмин, К. В. Крутовский // *Генетика*. – 2019. – Т. 55, № 4. – С. 418-425.

*Прочие публикации:*

1. **Bondar, E. I.** Annotation of Siberian larch reference genome, the only seasonal senescence genus in *Pinaceae* / E. I. Bondar, S. I. Feranchuk, K. A. Miroshnikova, V.V. Sharov, D.A. Kuzmin, N. V. Oreshkova, K. V. Krutovsky // *Высокопроизводительное секвенирование в геномике : Тезисы III Всероссийской конференции, Новосибирск, 19-24 июня 2022 г.* – 2022. – С. 24.

2. **Bondar, E. I.** Annotation of Siberian Larch Genome Draft Assembly / E.I. Bondar, S.I. Feranchuk, V.V. Birukhov, D.A. Kuzmin, V.V. Sharov, N.V.

Oreshkova, K.V. Krutovsky // Plant Genetics, Genomics, Bioinformatics, And Biotechnology : Abstracts of The 6th International Scientific Conference. – 2021. – P. 40.

3. **Bondar, E. I.** Genome-wide prediction of transcription start site in four conifer species / E. I. Bondar, V. V. Sharov, D. A. Kuzmin, T. V. Tatarinova, K. V. Krutovsky // Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2020). – 2020. – С. 300-301. DOI: 10.18699/BGRS/SB-2020-188/

4. Putintseva, Yu. A. Siberian larch (*Larix sibirica* Ledeb.) Mitochondrial genome, the largest currently known mitogenome / Yu.A. Putintseva, **E.I. Bondar**, V.V. Sharov, E.P. Simonov, N.V. Oreshkova, D.A. Kuzmin, Yu.M. Konstantinov, V.N. Shmakov, V.I. Belkov, M.G. Sadovsky, K.V. Krutovsky // Conservation of Forest Genetic Resources: Proceedings of the 6th International Conference. – Kokshetau publishing house «World of Printing», IE «Ustyugova». – 2019. – P. 19.

5. **Bondar, E. I.** Sequencing and assembly of mitochondrial genomes in three conifer species *Larix sibirica*, *Pinus sibirica* and *Pinus sylvestris* / E.I. Bondar, A. Kirichenko, V.V. Sharov, Yu.A. Putintseva, N.V. Oreshkova, S.I. Feranchuk, Yu.M. Konstantinov, V.N. Shmakov, V.I. Belkov, D.A. Kuzmin, M.G. Sadovsky, K.V. Krutovsky. // Systems biology BGRS/SB-2018 : The Eleventh International Conference on bioinformatics of genome regulation and structure. – 2018. – P. 153.

6. **Бондар, Е. И.** Разработка микросателлитных маркеров лиственницы сибирской (*Larix sibirica* Ledeb.) на основе полногеномного *de novo* секвенирования / Е. И. Бондар // Биология : Материалы 55-й Международной научной студенческой конференции. – 2017. – С. 156.

7. Sadovsky, M.G. *L. sibirica* Ledeb. chloroplast genome yields unusual seven-cluster structure / M.G. Sadovsky, Yu.A. Putintseva, **E.I. Bondar**, K.V. Krutovsky // IWBBIO 2016 : Proceedings Extended abstracts of International work-conference on Bioinformatics and biomedical engineering, Granada (Spain). – 2016. – P. 360.

8. **Bondar, E. I.** Siberian larch chloroplast genome analysis over triplet frequency distribution / E.I. Bondar, Yu.A. Putintseva, K.V. Krutovsky // Systems

biology : Abstracts of the tenth international conference on bioinformatics of genome regulation and structure. – 2016. – P. 48.

9. **Bondar, E. I.** Comparative studying of multicluster structure of chloroplast genomes / E.I. Bondar, M.G. Sadovsky, Yu.A. Putintseva, M.Yu. Senashova // Systems Biology and Bioinformatics (SBB-2016) : Abstracts of The eighth international young scientists school. – 2016. – P. 12.

10. **Bondar, E. I.** Assembly and annotation of Siberian larch chloroplast genome and the search for single nucleotide polymorphisms / E.I. Bondar // Systems Biology and Bioinformatics : Abstracts of the 7th International Young Scientists School. – 2015. – С. 13.

11. **Бондар, Е. И.** Изучение хлоропластного генома лиственницы сибирской (*Larix sibirica* Ledeb.) и разработка полиморфных хлоропластных маркеров / Е.И. Бондар, Ю.А. Путинцева, Н.В. Орешкова, К.В. Крутовский // Материалы 4-го международного совещания “Сохранение лесных генетических ресурсов Сибири”. – 2015 г. – С. 20-21;

12. **Бондар, Е. И.** Сборка и аннотирование хлоропластного генома лиственницы сибирской (*Larix sibirica* Ledeb.) и поиск маркеров (SNPs) / Е.И. Бондар // Проспект Свободный-2015 : Сборник материалов Международной конференции студентов, аспирантов и молодых ученых, Красноярск, 15-25 апреля 2015 г. – 2015. – С. 6-7.

13. **Бондар, Е. И.** Сборка и аннотирование хлоропластного генома лиственницы сибирской / Е.И. Бондар // МНСК-2015 (Биология) : Материалы 53-й Международной научной студенческой конференции / Новосиб. гос. ун-т. – Новосибирск. – 2015. – С. 178.

14. **Бондар, Е. И.** Сборка и сравнительный анализ хлоропластных геномов хвойных рода *Larix* / Е.И. Бондар // Третья летняя школа по биоинформатике : Сборник тезисов. – 2015. – С. 9.

#### **Соответствие диссертации паспорту специальности 1.5.7 Генетика**

Диссертационная работа соответствует критериям, предъявляемым к диссертационным работам, представленным на соискание ученой степени

кандидата наук. Содержание диссертационной работы соответствует областям исследования «Геномы, их структура и функция», «Структурно-функциональная организация геномов. Структурная и функциональная геномика» согласно паспорту специальности 1.5.7 Генетика Биологические науки (п.2, п.3. и п.7 паспорта научной специальности).

**Заключение:** Диссертация Бондар Евгении Ивановны «Аннотация генома и предсказание сайтов начала транскрипции для лиственницы сибирской (*Larix sibirica* Ledeb)» рекомендуется к защите на соискание учёной степени кандидата биологических наук по специальности 1.5.7 Генетика (Биологические науки).

Заключение принято на расширенном заседании кафедры геномики и биоинформатики.

Присутствовало на заседании 17 чел., из них с правом решающего голоса — 13 чел. Результаты голосования: «за» — 13 чел., «против» — 0 чел., «воздержалось» — 0 чел., протокол № 9 от 30 мая 2023 года.



Ямских Ирина Евгеньевна  
д-р биол. наук, доцент,  
заведующая кафедрой геномики и  
биоинформатики

