

Генетическая структура и географическая дифференциация популяций лиственницы сибирской (*Larix sibirica* Ledeb.) на основе генотипирования генома путем секвенирования

С. В. НОВИКОВА^{1, 2}, Н. В. ОРЕШКОВА^{1, 2, 3}, В. В. ШАРОВ^{1, 2}, В. Л. СЕМЕРИКОВ⁴, К. В. КРУТОВСКИЙ^{1, 5, 6, 7}

¹Сибирский федеральный университет
660041, Красноярск, просп. Свободный, 79

²Федеральный исследовательский центр “Красноярский научный центр СО РАН”
660036, Красноярск, Академгородок, 50

³Институт леса им. В. Н. Сукачева ФИЦ КНЦ СО РАН
660036, Красноярск, Академгородок, 50/28

⁴Институт экологии растений и животных УрО РАН
620144, Екатеринбург, ул. 8 Марта, 202

⁵Гёттингенский университет им. Георга-Августа
Германия, 37077, Гёттинген, ул. Бюсенвег, 2

⁶Институт общей генетики им. Н. И. Вавилова РАН
119991, Москва, ул. Губкина, 3

⁷Воронежский государственный лесотехнический университет им. Г. Ф. Морозова
394087, Воронеж, ул. Тимирязева, 8

Статья поступила 23.03.2023

После доработки 30.03.2023

Принята к печати 13.04.2023

АННОТАЦИЯ

Приведены результаты исследования генетической дифференциации популяций лиственницы сибирской (*Larix sibirica* Ledeb.) в широтном градиенте климатических условий, полученные на основе генотипирования генома с помощью высокопроизводительного секвенирования геномных районов ДНК, ассоциированных с сайтами рестрикции (ddRADseq). Изучена также корреляция пяти основных климатических переменных с изменчивостью 47 929 генетических маркеров – однонуклеотидных полиморфизмов или “снийпов” (от английского SNPs – single nucleotide polymorphisms). Всего изучено 125 деревьев: 61 дерево в четырех популяциях вдоль западной географической трансекты и 64 дерева в четырех популяциях вдоль восточной географической трансекты. Выявлен 21 SNPs с признаками отбора, включая 9 SNPs аутлайеров, чья изменчивость не может быть объяснена селективно-нейтральными процессами, и 12 SNPs, чья изменчивость коррелировала с изменчивостью некоторых климатических факторов. Семь

SNPs расположены в интронах митохондриальных генов, три расположены вблизи митохондриальных генов, кодирующих NAD2 и рибосомальные белки S7 и S11, один на отдалении от ядерного гена, кодирующего белок, гомологичный связанному с микротрубочками futsch-подобному белку *Arabidopsis thaliana*, два в белковых генах неизвестной природы и три в контигах, не содержащих гены, и для которых не найдены гомологичные последовательности в NCBI GenBank.

Ключевые слова: популяция, генетическая структура, *Larix sibirica*, адаптация, климатические переменные, хвойные, ddRADseq, SNP.

ВВЕДЕНИЕ

Лиственница сибирская (*Larix sibirica* Ledeb.) является одним из ключевых хвойных видов бореальных лесов Сибири, играет очень важную экологическую и экономическую роль [Abaimov et al., 2002; Семериков и др., 2007; Abaimov, 2010]. Это экологически пластичный вид, обладающий высоким уровнем фенотипической изменчивости, генетический контроль которой еще слабо изучен. Кроме того, изучение данного вида затруднено огромным размером генома – ~12 млрд пар нуклеотидных оснований [Ohri, Khoshoo, 1986].

Хвойные породы бореальных лесов, такие как ель, сосна и лиственница, обладают обширными ареалами с широким диапазоном климатических условий, которые определяются в основном широтой местности. Опыты с географическими культурами и практика лесоразведения показали, что наиболее адаптированные к условиям данной местности оказываются деревья, выращенные из семян, происходящих из районов, близких к ней по географической широте, т. е. очевидно, что происходит генетическая адаптация к условиям произрастания [Lu et al., 2016; Bucharova et al., 2017; Barton et al., 2020; Wadgyr et al., 2022].

Ранее работы по изучению генетических механизмов адаптации к условиям произрастания были проведены для ряда организмов (см. для обзора [Balkenhol et al., 2019; Крутовский, 2022; Lasky et al., 2023], включая лиственницу сибирскую [Novikova et al., 2023]), и сейчас, с развитием методов высокопроизводительного секвенирования, такие исследования стали доступны и для организмов с большими и сложными геномами. Более того, широкий спектр методов полногеномного генотипирования позволяет провести анализ большой совокупности генов и их взаимодействий на полногеномном уровне.

На данный момент для изучения адаптации организмов к условиям произрастания, биоклиматическим факторам и воздействию

стрессовых факторов широко используется метод высокопроизводительного секвенирования геномных районов ДНК, ассоциированных с сайтами рестрикции (т. н., restriction site-associated DNA sequencing – RADseq), который позволяет получить тысячи маркеров – однонуклеотидных полиморфизмов (SNPs), равномерно представленных на большей части генома [Peterson et al., 2012].

Целью представленного здесь исследования являлось проведение популяционно-генетического анализа выборок популяций лиственницы сибирской вдоль широтного климатического градиента в двух разных независимых трансектах с использованием данных геномного генотипирования.

Основная задача заключалась в поиске генов-аутлайеров, чья изменчивость не может быть объяснена селективно-нейтральными процессами, а также генов, изменчивость которых коррелирует с биоклиматическими переменными в широтном градиенте климатических условий, таких как длина вегетационного сезона, длина светового дня и его динамика, убывающие к северу ресурсы тепла и усиление дефицита влаги на юге. Обнаруженные гены-аутлайеры и гены, чья изменчивость коррелирует с биоклиматическими переменными, могут находиться под отбором, и их изменчивость, вероятно, связана с адаптацией к локальным климатическим условиям произрастания.

МАТЕРИАЛ И МЕТОДЫ

Растительный материал

Хвоя для выделения ДНК и последующего генотипирования была собрана с 125 деревьев лиственницы сибирской из восьми популяций, произрастающих вдоль двух географических трансект в градиенте биоклиматических факторов, таких как длина вегетационного сезона и длина светового дня в течение вегетационного сезона. С целью учета биогео-

Географическое расположение исследованных популяций лиственницы

Популяция (число деревьев)	Географический район	Координаты	Высота над уровнем моря, м
Западная трансекта (WT)			
SIB (15)	хр. Ирендык, вблизи г. Сибай (Республика Башкортостан)	52° 47' с. ш. 58° 16' в. д.	574
SEV (16)	пос. Северка, окрестности г. Екатеринбурга (Свердловская область)	56° 52' с. ш. 60° 19' в. д.	289
IVD (16)	Окрестности г. Ивдель (Свердловская область)	60° 42' с. ш. 60° 20' в. д.	180
LAB (14)	Окрестности г. Лабытнанги (Ямало-Ненецкий автономный округ)	66° 40' с. ш. 66° 24' в. д.	83
Восточная трансекта (ET)			
TUR (16)	г. Туран, в 65 километрах к северо-западу от г. Кызыла, южные отроги Западного Саяна (Республика Тыва)	51° 55' с. ш. 94° 19' в. д.	1128
KRAS (16)	Окрестности г. Красноярска (Красноярский край)	55° 56' с. ш. 92° 15' в. д.	417
NBSK (16)	Окрестности г. Ноябрьска (Ямало-Ненецкий автономный округ)	63° 06' с. ш. 75° 21' в. д.	146
NU (16)	Окрестности г. Н. Уренгой (Ямало-Ненецкий автономный округ)	66° 05' с. ш. 76° 43' в. д.	38

графического фактора сборы проводились в пределах произрастания западной расы лиственницы сибирской – западная трансекта

(WT), и в пределах восточной расы – типовой лиственницы сибирской – восточная трансекта (ET) (табл. 1, рис. 1).

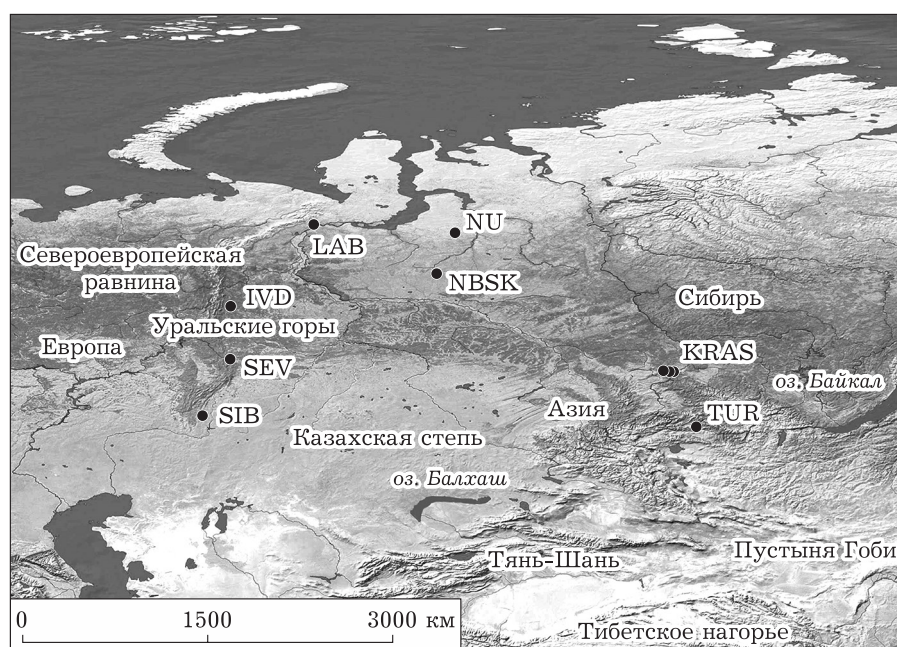


Рис. 1. Географическое расположение исследованных популяций лиственницы сибирской: SIB, SEV, IVD, LAB, TUR, KRAS, NBSK и NU

Биоклиматические переменные

На основании измеряемых на местах сбора географических координат для каждого дерева получены значения 11 климатических переменных из базы данных WorldClim [Fick, Hijmans, 2017] в разрешении 2,5 минуты (примерно 2,6 км): среднегодовая температура (Temp), изотермальность (Isothermality), температурная сезонность (TempSeas), средняя температура самого теплого квартала (MTofWQ – mean temperature of warmest quarter), средняя температура самого холодного квартала (MTofCQ – mean temperature of coldest quarter), среднегодовые осадки (Precp), осадки самого влажного месяца (PofWM – precipitation of wettest month), осадки самого сухого месяца (PofDM – precipitation of driest month), высота над уровнем моря (Elevation), уровень солнечной радиации ($\text{кДж} \cdot \text{м}^{-2} \cdot \text{сут}^{-1}$, Srad – Solar Radiation) и средняя скорость ветра ($\text{м} \cdot \text{с}^{-1}$, Wind).

По 11 переменным с помощью R пакета *vegan* [Oksanen et al., 2022] проведена проверка факторов инфляции дисперсии (VIF) переменных-предикторов, используемых в работе, с пороговым значением $VIF \leq 10$. Значения корреляции и результаты анализа главных компонент (PCA) для климатических переменных были также получены с помощью R пакета *vegan*.

Выделение ДНК и подготовка ddRAD библиотек

Из хвои образцов лиственницы сибирской восьми популяционных выборок была выделена ДНК с использованием СТАВ метода [Devey et al., 1996]. Для приготовления ddRAD-Seq библиотек использовали модифицированную версию протокола Peterson et al. [2012] с одновременной обработкой ДНК двумя рестрикционными ферментами: *EcoRI* и *MseI* [Parchman et al., 2012]. К каждой библиотеке перед секвенированием было добавлено 25 % библиотеки Phi X. Секвенирование ddRADseq библиотек осуществлялось на двух дорожках проточной ячейки S1 секвенатора NovaSeq 6000 (Illumina Inc., San Diego, CA, USA) с длиной одноконцевого прочтения 100 н. о.

Биоинформатическая обработка и поиск SNPs

Полученные прочтения были отфильтрованы и обрезаны согласно показателям каче-

ства с помощью программы Trimmomatic-0.39 [Bolger et al., 2014]. Основываясь на баркодированных последовательностях *EcoRI*-адаптеров, данные в библиотеке разделили на отдельные образцы с использованием программы *process_radtags*, входящей в пакет *Stacks* [Catchen et al., 2013].

Отфильтрованные прочтения каждого образца выравнивали на референсный геном лиственницы сибирской [Kuzmin et al., 2019] с помощью программы Bowtie2 v.2.3 [Langmead, Salzberg, 2012] в режиме картирования “local”.

Поиск однонуклеотидных полиморфных позиций осуществлялся с использованием программы Gstacks из пакета Stacks с фильтрацией по качеству выравнивания прочтений “min-mapq 20”. Полученный после выравнивания набор локусов подвергали нескольким этапам дополнительной фильтрации с помощью программы Populations из пакета Stacks таким образом, чтобы сохранить только локусы, которые представлены по крайней мере у 80 % всех образцов и у 80 % образцов в каждой выборке. Максимальный уровень наблюдаемой гетерозиготности по нуклеотидной позиции не должен превышать 0,6, а частота минорного аллеля не менее 0,01 и его покрытие не менее 10.

Так как большинство методов, используемых в работе, чувствительно к наличию пропущенных значений, методом LD-kNNi (k-nearest neighbor genotype imputation method) в программе TASSEL v.5.0 [Bradbury et al., 2007] были сгенерированы пропущенные частоты аллелей.

Также для того, чтобы понять пороговые значения изученных популяций, в основе которых лежат селективно-нейтральные процессы (генетический дрейф и изоляция), был создан набор из 2145 снипов, находящихся вне кодирующих областей генома и картированных на разные контиги для расчета соответствующих параметров генетической изменчивости и дифференциации. Этот набор условно “нейтральных” SNPs был использован также для определения генетической структуры популяций лиственницы сибирской, в основе которой лежат селективно-нейтральные процессы для сравнения со структурой, оцененной по маркерам, чья изменчивость предположительно находится под влиянием отбора.

Поиск маркеров-аутлайеров

Поиск генетических маркеров с резко выделяющимися значениями генетической дифференциации и изменчивости, которые не могут быть объяснены только селективно-нейтральными процессами, так называемых генов-аутлайеров, проводился с использованием следующих трех популяционно-генетических подходов, имплементированных соответственно в трех следующих программах: BayeScan [Foll, Gaggiotti, 2008], Arlequin [Excoffier, Lischer, 2010] и PCAdapt [Privé et al., 2020].

В BayeScan, Arlequin и PCAdapt расчеты проведены отдельно для трансект ET и WT. Контроль частоты ложноположительных результатов (FDR) осуществлялся с помощью q -значений с параметром $\alpha = 0,05$ [Storey, Tibshirani, 2003]. По результатам работы трех программ для каждой трансекты отобраны SNPs, отмеченные как достоверные всеми тремя программами.

Поиск генетических маркеров адаптации, связанных с переменными окружающей среды

Поиск локусов, чья изменчивость предположительно связана с адаптацией к средовым факторам, осуществляли с использованием следующих трех популяционно-генетических подходов, имплементированных соответственно в следующих программах: Bayescenv [de Villemereuil, Gaggiotti, 2015], LFMM [Caye et al., 2019] и RDA [Capblancq, Forester, 2021].

Все расчеты проведены отдельно для трансект ET и WT. В Bayescenv и LFMM2 каждая климатическая переменная рассматривалась отдельно и независимо, при анализе RDA расчеты для пяти климатических переменных (Temp, Precp, Srad, Wind и PofDM) проводились одновременно. Контроль частоты ложноположительных результатов для Bayescenv и LFMM2 осуществлялся с помощью q -значений с параметром $\alpha = 0,05$; для RDA применен контроль FDR по двустороннему значению $p = 0,0027$ (три стандартных отклонения). По результатам работы трех программ для каждой трансекты отобраны SNPs, отмеченные как достоверные всеми тремя программами.

Также для расчета генетических статистик создан набор из 11097 снипов-аутлайеров

и связанных с изменчивостью средовых факторов (условно “адаптивных” SNPs), найденных по крайней мере одной из программ, использованных в нашей работе для поиска адаптивных маркеров, т. е. маркеров, чья изменчивость предположительно находится под влиянием отбора.

Генетический анализ изменчивости и структуры популяции

Для анализа генетической изменчивости использовались пакеты R adegenet [Jombart, 2008] и poppr [Kamvar et al., 2014]. Для каждой выборки рассчитаны: аллельное разнообразие (A_R), наблюдаемая (H_o) и ожидаемая (H_e) гетерозиготность, индекс фиксации (FIS).

Для выявления популяционной структуры нами также выполнен анализ главных компонент (PCA) и дискриминантный анализ главных компонент (DAPC) с использованием пакета R ade4 [Dray, Dufour, 2007].

Структура популяции изучена с помощью алгоритма Admixture, реализованного в программе AdmixPipe [Mussmann et al., 2020], используемого для оценки максимального правдоподобия выделения генетических кластеров на основе данных генотипирования. Для этого был осуществлен поиск наиболее вероятного числа кластеров K путем проверки значения параметра K от 1 до 10, в 20 повторностях для каждого K . Наиболее вероятное значение K выбирали на основании значений ошибки кросс-валидации (CV error) и методом ΔK [Evanno et al., 2005], рассчитанных в онлайн-сервисе Clumpak [Kopelman et al., 2015].

Матрица попарных генетических различий для проведения теста Мантел строилась в программе TASSEL и рассчитывалась как 1-IBS, где IBS – это мера идентичности на основе сходства генотипов (identity by state). Матрицу попарных географических расстояний и матрицу попарных расстояний между параметрами окружающей среды определяли, используя евклидово расстояние с помощью R пакета vegan. Тест Мантел также проводили в vegan с количеством пермутаций 999 и значением $p = 0,001$. Иерархический анализ молекулярной вариации (AMOVA) проведен с помощью программы Arlequin v. 3.5 [Excoffier, Lischer, 2010], с расчетом попарных коэффициентов F_{ST} , с количеством пермутаций 1000 и уровнем значимости 0,05.

Аннотация SNPs

Для анализа геномных участков, в которых расположены значимые маркеры, использована аннотация геномной сборки лиственницы сибирской [Bondar et al., 2022] и поиск гомологов в базе “nr” NCBI GenBank [Clark et al., 2016]. Генные модели были выровнены на базу “nr”, отфильтрованную для видов *Embryophyta* с использованием таксономического идентификатора. Поиск белковых доменов производился с помощью InterProScan [Blum et al., 2021]. Аннотация отобранных маркеров проводилась с помощью программы SNPdat [Doran, Creevey, 2013]. Дополнительно проведено выравнивание участков (интервалов размером 20000 н. о.), содержащих маркеры-кандидаты, с помощью сервиса NCBI BLAST на нуклеотидную (megablast) и белковую (blastx) базы NCBI GenBank.

РЕЗУЛЬТАТЫ SNP-маркеры

Для 125 деревьев путем ddRADseq секвенирования получено 1,7 млрд нуклеотидных прочтений длиной 100 н. о. После первичной обработки и фильтрации по качеству

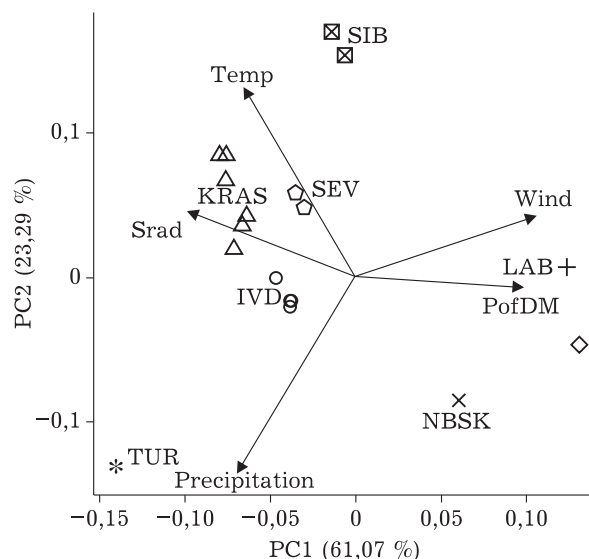


Рис. 2. Распределение пяти средовых факторов в пространстве главных компонент (PCA). Temp – среднегодовая температура; Srad – солнечная радиация; Precipitation – среднегодовые осадки; Wind – средняя скорость ветра, PofDM – осадки самого сухого месяца. Разными символами обозначены группы деревьев, принадлежащих восьми популяциям лиственницы сибирской (см. табл. 1 для их обозначения)

для дальнейшего анализа отобрано около 1,2 млрд прочтений длиной от 32 до 92 н. о., в среднем по $9,637,201 \pm 401,017$ прочтений на образец. В среднем 97,86 % прочтений для каждого образца были успешно картированы на референсный геном лиственницы сибирской [Kuzmin et al., 2019].

При поиске однонуклеотидных полиморфных позиций отобрано 87 тыс. локусов, соответствующих критериям фильтрации и содержащих 47 929 биаллельных снипов (SNPs) для 125 образцов. Процент пропущенных значений на образец составил в среднем $5,98 \pm 0,41$ %. Пропущенные аллели были сгенерированы с помощью процедуры импутации.

Биоклиматические переменные

Из 11 климатических переменных на основании значений VIF и показателей корреляции отобрано пять наиболее значимых переменных: среднегодовая температура (Temp), среднегодовые осадки (Prср), осадки самого сухого месяца (PofDM), солнечная радиация (Srad) и средняя скорость ветра (Wind). Значения VIF для этих пяти отобранных переменных варьировали от 1,684 до 8,229 и не превышали порогового значения, что указывает на то, что мультиколлинеарность среди этих предикторов не повлияет на результаты дальнейшего анализа [Zuur et al., 2010].

Чтобы дополнительно проверить взаимосвязь между климатическими переменными, выполнен анализ главных компонент (PCA). Первая компонента объяснила 61,1 % изменчивости, связанной, в первую очередь, с показателями средней скорости ветра (Wind) и солнечной радиации (Srad) – более высокие показатели скорости ветра и низкое количество солнечного излучения соответствуют положительным значениям PC1. Вторая компонента объяснила 23,3 % вариации, высокие значения среднегодовой температуры (Temp) и низкие показатели влажности (PofDM) здесь соответствуют положительным значениям PC2 (рис. 2).

Маркеры-аутлайеры

С помощью программы PCAdapt для поиска маркеров-аутлайеров сначала был проведен PCA-анализ на большом количестве глав-

ных компонент ($K = 15$) и построен график убывания значений дисперсии, объясненной каждой компонентой от 1 до 15, чтобы выбрать наиболее значимые компоненты. На основании этих данных для обеих трансе́кт вы́брано три главных компоненты. Затем для трансе́кты ЕТ найдено 168 SNP-аутлайеров, а для трансе́кты WT – 551 SNP-аутлайеров. Из них 38 SNP-аутлайеров были общими для обеих трансе́кт.

С помощью программы Arlequin для трансе́кты ЕТ найден 1971 SNP-аутлайер, а для трансе́кты WT – 3885 SNP-аутлайеров. Из них 263 SNP-аутлайера были общими для обеих трансе́кт.

Аналогично с помощью программы Bayescan для трансе́кты ЕТ найдено 104 SNP-аутлайера, для трансе́кты WT – 286 аутлайеров. Из них 13 были общими для обеих трансе́кт.

После сопоставления результатов трех программ для каждой трансе́кты были сформированы конценсусные наборы аутлайеров, обнаруженных всеми тремя программами: 36 высокодостоверных SNP-аутлайеров для трансе́кты ЕТ и 225 для трансе́кты WT. Сравнение данных наборов показало наличие 9 SNP-аутлайеров, общих для обеих трансе́кт.

Маркеры, связанные с переменными окружающей среды

С помощью программы LFMM2 для трансе́кты ЕТ найдено 2167 маркеров, чья изменчивость коррелировала с изменчивостью одной или нескольких из пяти климатических переменных, из них 357 (17 %) коррелировали с изменчивостью всех пяти средовых факторов. Для трансе́кты WT найдено 3725 таких SNPs, из них 2113 (около 57 %) оказались общими для всех пяти переменных. Всего выявлено 196 SNPs, общих для обеих трансе́кт.

Также с помощью программы Bayescenv для трансе́кты ЕТ обнаружено 154 значимых маркера, чья изменчивость коррелировала с изменчивостью одной или нескольких из пяти климатических переменных, из них 32 коррелировали с изменчивостью всех пяти средовых факторов. Для данных трансе́кты WT обнаружено 424 таких SNPs, из которых 84 коррелировали с изменчивостью всех пяти средовых факторов. Всего программой Bayescenv выявлено 32 SNPs, общих для обеих трансе́кт.

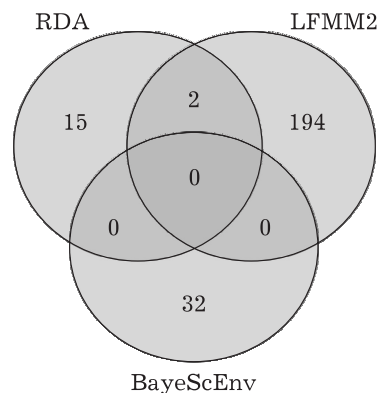


Рис. 3 Диаграмма Венна, отображающая результаты поиска значимых маркеров, связанных с переменными окружающей среды, тремя методами: LFMM2, Bayescenv и RDA

По результатам RDA-анализа выявлено 887 значимых адаптивных SNPs для трансе́кты ЕТ и 769 SNPs – для трансе́кты WT, 17 из них оказались общими для двух трансе́кт.

Всего было найдено только пять высокодостоверных SNPs для трансе́кты ЕТ, обнаруженных всеми тремя программами и коррелирующих хотя бы с одним средовым фактором, и аналогично также пять SNPs для трансе́кты WT. Маркеров, обнаруженных всеми тремя методами и общих для обеих трансе́кт, найдено не было. Из маркеров, связанных с переменными окружающей среды в обеих трансе́ктах, два SNPs были найдены как RDA, так и LFMM2 (рис. 3).

Генетическая изменчивость, структура и дифференциация популяций

Для определения генетической структуры восьми популяционных выборок лиственницы сибирской с помощью алгоритма Admixture проведен поиск наиболее вероятного числа генетических кластеров K (субпопуляций) на основе генотипов 47 929 сипов в 125 деревьях. Различные методы выбора K показали, что наиболее вероятное число кластеров K равно 2, 4 или 8 (рис. 4).

Вклад (“admixture”) каждого из кластеров в индивидуальные деревья (Q-значения) приведен на рис. 5.

При количестве кластеров $K = 2$ в целом разделение на кластеры происходит согласно принадлежности изучаемых образцов к географической трансе́кте, однако образцы

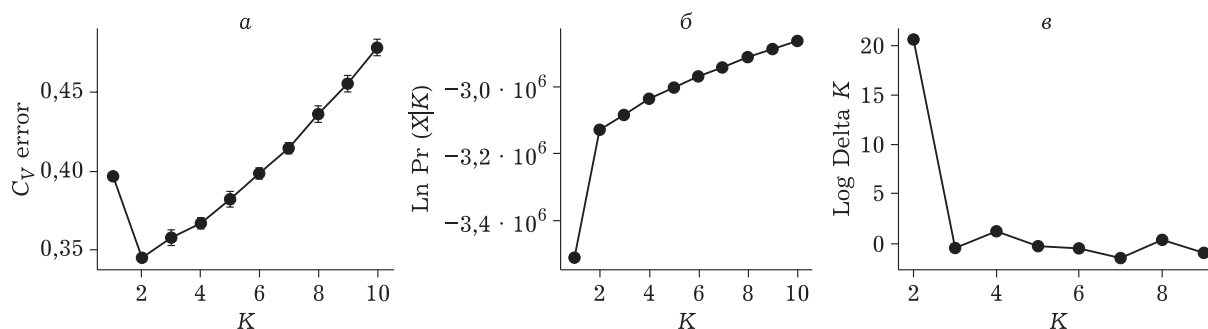


Рис. 4. Анализ наиболее вероятного числа кластеров K

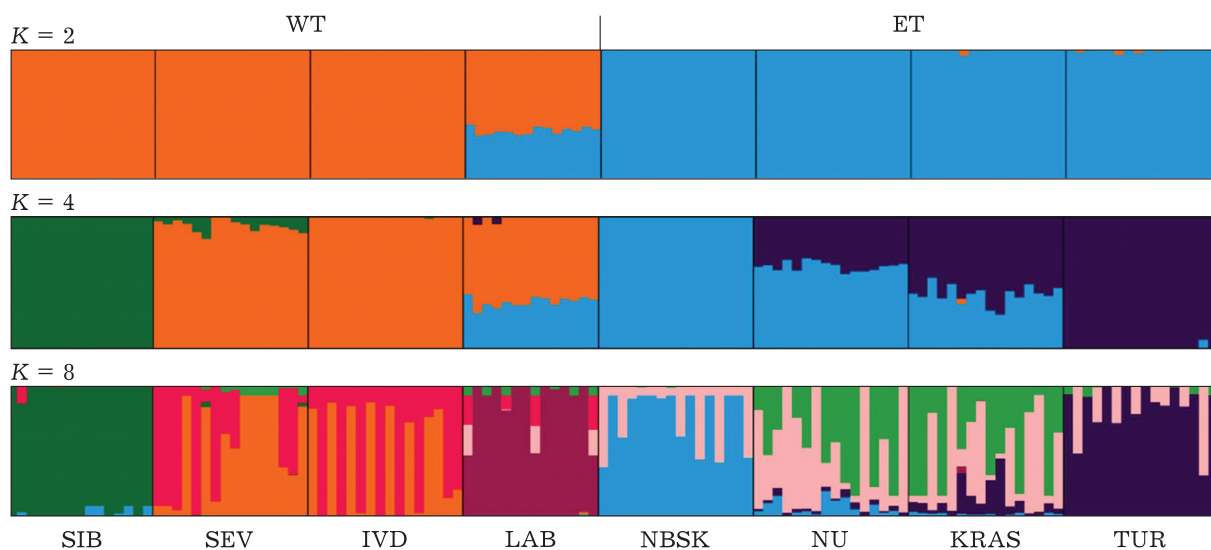


Рис. 5. Вклад ("admixture") каждого из кластеров при K = 2, 4 и 8 в индивидуальные деревья (Q-значения), обозначенный разными цветами

популяции LAB, принадлежащие трансекте WT, занимают промежуточное положение. При K = 4 и 8 хорошо выделяются популяции SIB, LAB и TUR.

По результатам анализа популяционной структуры с помощью PCA и DAPC выявлено, что исследуемые выборки образуют четыре основных кластера: четыре выборки, принадлежащие трансекте ET (NBSK, NU, KRAS и TUR), формируют один кластер, выборки SEV и IVD – второй, популяции SIB и LAB занимают отдельное положение (рис. 6).

Средние значения основных популяционно-генетических параметров для всех восьми популяций на основе 47929 снипов представлены в табл. 2.

Уровень генетической дифференциации по параметру F_{ST} между трансектами по всем маркерам был довольно высоким ($F_{ST} = 0,136$) и варьировал попарно между всеми восемью

популяциями от минимального значения 0,004 между популяциями NU и KRAS до максимального 0,189 между SIB и NBSK (рис. 7), в среднем $F_{ST} = 0,099 \pm 0,013$. Все полученные значения были статистически значимы ($p \ll 0,01$). Среднее значение F_{ST} между популяциями трансекты WT было равно $0,040 \pm 0,008$, между популяциями трансекты ET – $0,016 \pm 0,004$.

Для набора из 2145 "нейтральных" SNPs значения F_{ST} мало отличались от таковых, полученных по всем маркерам: между трансектами $F_{ST} = 0,133$, а при попарном сравнении популяций F_{ST} варьировало в диапазоне от 0,004 между популяциями NU и KRAS до 0,177 между SIB и NBSK, в среднем $F_{ST} = 0,097 \pm 0,013$ ($p \ll 0,01$). Среднее значение F_{ST} между популяциями внутри трансекты WT было равно $0,040 \pm 0,008$, между популяциями внутри трансекты ET – $0,016 \pm 0,004$.

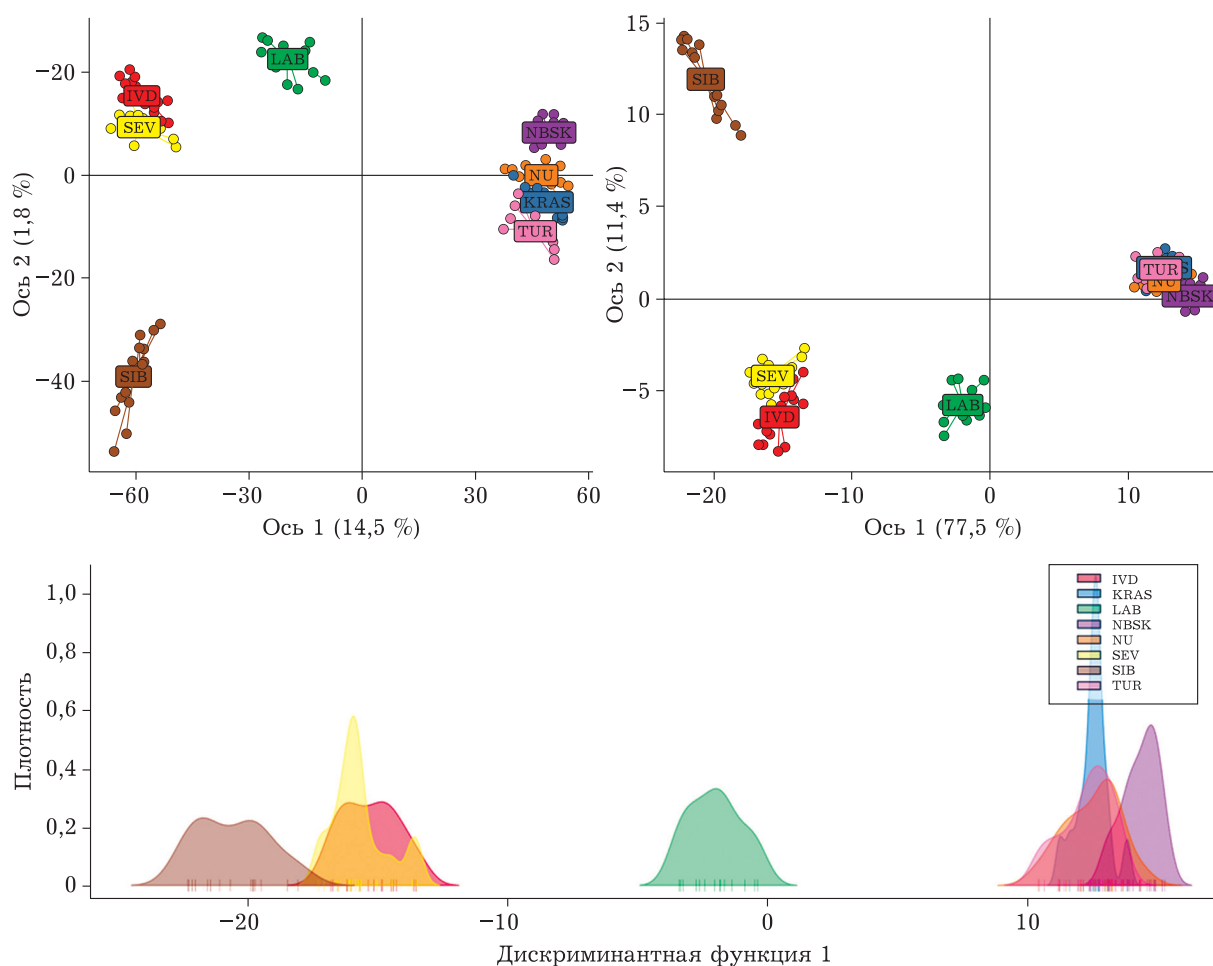


Рис. 6. Результаты анализа главных компонент (слева) и дискриминантного анализа главных компонент (справа), график плотности образцов вдоль первой дискриминантной функции (внизу)

Т а б л и ц а 2

Средние значения параметров генетической изменчивости $\pm$ SE по 47929 SNPs для восьми популяций лиственницы сибирской

Трансекта	Популяция	Количество деревьев	A_R	H_o	H_e	F_{IS}
WT	SIB	15	$1,579 \pm 0,002$	$0,137 \pm 0,001$	$0,141 \pm 0,001$	$0,022 \pm 0,002$
	SEV	16	$1,621 \pm 0,002$	$0,138 \pm 0,001$	$0,142 \pm 0,001$	$0,020 \pm 0,001$
	IVD	16	$1,614 \pm 0,002$	$0,135 \pm 0,001$	$0,139 \pm 0,001$	$0,025 \pm 0,001$
	LAB	14	$1,791 \pm 0,002$	$0,150 \pm 0,001$	$0,157 \pm 0,001$	$0,032 \pm 0,001$
ET	NBSK	16	$1,716 \pm 0,002$	$0,150 \pm 0,001$	$0,157 \pm 0,001$	$0,034 \pm 0,001$
	NU	16	$1,725 \pm 0,002$	$0,145 \pm 0,001$	$0,151 \pm 0,001$	$0,034 \pm 0,001$
	KRAS	16	$1,739 \pm 0,002$	$0,151 \pm 0,001$	$0,159 \pm 0,001$	$0,040 \pm 0,001$
	TUR	16	$1,721 \pm 0,002$	$0,147 \pm 0,001$	$0,157 \pm 0,001$	$0,047 \pm 0,001$
Среднее		15,6	$1,688 \pm 0,026$	$0,144 \pm 0,002$	$0,150 \pm 0,003$	$0,032 \pm 0,003$

П р и м е ч а н и е. A_R – аллельное разнообразие; H_o – наблюдаемая гетерозиготность; H_e – ожидаемая гетерозиготность; F_{IS} – индекс фиксации (все значения F_{IS} значимы с $p < 0,001$).

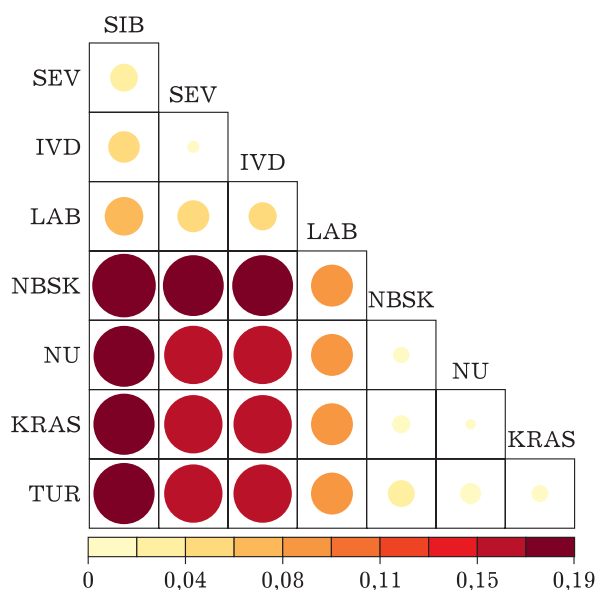


Рис. 7. Попарные значения индекса фиксации F_{ST} , рассчитанного между восемью популяциями

В то же время для набора из 11097 “адаптивных” SNPs значения F_{ST} значительно отличались от значений, полученных по всем маркерам или по селективно-нейтральным SNPs: между трансектами $F_{ST} = 0,196$, а при

попарном сравнении популяций F_{ST} варьировало в диапазоне от 0,012 между популяциями SEV и IVD до 0,296 между SIB и NBSK, среднее $F_{ST} = 0,154 \pm 0,019$ ($p < 0,01$). Среднее значение F_{ST} между популяциями внутри трансекты WT было равно $0,075 \pm 0,016$, между популяциями внутри трансекты ET – $0,041 \pm 0,009$.

Аналогично иерархический анализ молекулярной дисперсии (AMOVA) показал высокое значение индекса фиксации Райта между трансектами ($F_{CT} = 0,130$) в целом для всех выборок ($F_{ST} = 0,155$) и относительно низкое между популяциями внутри трансект ($F_{SC} = 0,029$) по всем маркерам (табл. 3). Близкие значения получены и по “нейтральным” SNPs. А вот по “адаптивным” SNPs значения были гораздо выше ($F_{CT} = 0,185$, $F_{SC} = 0,059$, $F_{ST} = 0,234$). Все значения достоверно отличались от нуля ($p < 0,001$, см. табл. 3).

Географически ограниченные разности пыльцы и расселение семян могут создать эффект изоляции расстоянием (Isolation by Distance, IBD) и способствовать формированию генетической структуры популяций, при которой возникает корреляция между генети-

Т а б л и ц а 3
Результаты анализа молекулярной дисперсии (AMOVA)

Источник изменчивости	Суммы квадратов	Дисперсионная составляющая	Доля в общей изменчивости, %	F-index
Все 47929 SNPs				
Между трансектами	76049,662	553,451	13,0	$F_{CT} = 0,130$
Между популяциями внутри трансект	41421,479	105,955	2,5	$F_{SC} = 0,029$
Между всеми популяциями	869762,116	3594,062	84,5	$F_{ST} = 0,155$
Суммарно	987233,257	4253,468		
2145 “нейтральных” SNPs				
Между трансектами	3224,994	23,431	12,7	$F_{CT} = 0,127$
Между популяциями внутри трансект	1785,596	4,538	2,5	$F_{SC} = 0,028$
Между всеми популяциями	37715,174	155,848	84,8	$F_{ST} = 0,152$
Суммарно	42725,764	183,817		
11097 “адаптивных” SNPs				
Между трансектами	28699,698	209,079	18,5	$F_{CT} = 0,185$
Между популяциями внутри трансект	15465,185	54,835	4,9	$F_{SC} = 0,059$
Между всеми популяциями	209269,251	864,750	76,6	$F_{ST} = 0,234$
Суммарно	253434,134	1128,665		

Коэффициенты корреляции (r) по результатам Мантел-теста

Популяция	Между генетическим и географическим расстояниями	Между генетическим и климатическим расстояниями	Между географическим и климатическим расстояниями
Все образцы	0,49*	0,23*	0,70*
Трансекта ЕТ	0,08*	0,07**	0,95*
Трансекта WT	0,40*	0,39*	0,91*

П р и м е ч а н и е. * $p < 0,001$, ** $p < 0,01$.

ческой и географической дистанциями. Градиентное влияние экологических условий на генетическую дифференциацию и локальная адаптация также могут приводить к изоляции средой (Isolation by Environment, IBE). Для выявления подобных корреляций использован Мантел-тест, результаты которого показали наличие достоверной линейной связи генетического расстояния с географическим и экологическим расстояниями (табл. 4).

Аннотация SNPs

Для девяти общих SNP-аутлайеров, а также для 12 SNPs, изменчивость которых достоверно коррелировала с изменчивостью средовых факторов, проведен анализ геномных участков, содержащих эти SNPs. Для всех 21 SNPs извлечены последовательности интервалом ± 10000 н. о. Выравнивание этих последовательностей на нуклеотидную и белковую базы NCBI GenBank показало наличие в базе гомологичных сиквенсов для 18 SNPs. Пять SNPs расположены в межгенных областях, а три SNPs оказались в контигах, не содержащих гены, и в NCBI GenBank для них не найдены гомологичные последовательности. Результаты аннотации суммированы в табл. 5.

Выравнивание нуклеотидных последовательностей, содержащих SNP-аутлайеры, на митохондриальный геном лиственницы сибирской [Putintseva et al., 2020] показало, что 10 найденных маркеров расположены в межгенном пространстве митохондриального генома: в интронах генов *NAD1* и *NAD2*, в интронах генов рибосомальных протеинов *L2*, *L5*, *S7*, *S11*, *S14* рядом с генами *COX2*, *NAD4* и *ATP4*. Из них пять снупов (*LS_5154937_1260*, *LS_5119432_22927*, *LS_5156723_4944*, *LS_4695581_16666* и *LS_4400751_1818*) находят-

ся в интроне 1 митохондриального гена рибосомального белка *L2*, т. е. сцеплены, четыре из них выявлены как аутлайеры в WT и ЕТ, но *LS_5156723_4944*, находящийся на расстоянии 69126 н. о. от ближайшего снупа, показал отличное от четырех других меркеров поведение, что может быть объяснено большим размером интрона и удаленностью от других снупов, и идентифицирован как аутлайер только в ЕТ, но его изменчивость коррелировала в ЕТ с *Prp*, *Srad* и *PofDM*.

На основании данных геномной аннотации установлено, что маркер *LS_28474_5984*, изменчивость которого коррелирует с изменчивостью показателя *Temp* для трансекты ЕТ, расположен на расстоянии 11 тысяч н. о. от гена, кодирующего связанный с микротрубочками futsch-подобный белок, вовлеченный в регуляцию процесса метаболизма азотистых соединений, а также процессы роста и развития тканей растения [Wang et al., 2008].

Для последовательностей, содержащих SNP-маркеры *LS_4137682_158*, *LS_132659_5239*, *LS_132659_5292*, *LS_1251904_144* и *LS_3903199_5631*, ассоциированные с климатической переменной *Temp* для трансекты WT, найдены гомологичные участки в неохарактеризованных белках *Picea sitchensis* и *Shorea leprosula*.

Для маркеров *LS_7403_79464*, *LS_3963342_9449* и *LS_4465_9359*, коррелирующих со средовыми факторами, поиск гомологов в базах NCBI GenBank не дал достоверных результатов.

ОБСУЖДЕНИЕ

Результаты представленного анализа геномной дифференциации популяций лиственницы сибирской в целом не согласуются с предыдущими выводами об относительно слабой структуре популяций хвойных,

Т а б л и ц а 5
Результаты аннотации 19 значимых SNPs

SNP	Локализация и аннотация (позиция в NCBI GenBank нуклеотидной последовательности и ее accession number)	Аутлайер и коррелирующий средовой фактор в трансектах WT и/или ET
LS_4447298_1210	9375 н.о. от начала митохондриального гена, кодирующего рибосомальный белок S11 (2195303 н.о., MT797188.1)	Аутлайер в WT и ET
LS_4806638_6052	2818 н.о. от конца митохондриального гена, кодирующего рибосомальный белок S7, и на расстоянии генов 30124 и 30704 н.о. перед генами рибосомальных белков L5 и S14 соответственно (889123 н.о., MT797189.1)	
LS_4766550_9636	Интрон 1 митохондриального гена <i>NAD1</i> (1651865 н.о., MT797188.1)	
LS_4800638_59265	55579 н.о. после митохондриального гена <i>NAD2</i> (548253 н.о., MT797190.1)	
LS_5154937_1260 LS_5119432_22927 LS_5156723_4944 LS_4695581_16666 LS_4400751_1818	Интрон 1 митохондриального гена рибосомального белка L2 (позиция в MT797187.1: 806147, 1077727, 1146853, 2293241, 2322428 соответственно)	Все аутлайеры в WT и ET, кроме LS_5156723_4944, который аутлайер только в ET, а также его изменчивость коррелировала с Prcp (ET), Srad (ET) и PofDM (ET)
LS_4586690_9126	Интрон 4 митохондриального гена <i>NAD2</i> (1957472 н.о., MT797188.1)	Аутлайер в WT и ET
LS_28474_5984	11 тысяч н.о. перед ядерным геном, кодирующим белок, гомологичный связанному с микротрубочками futsch-подобному белку <i>Arabidopsis thaliana</i> (NP_201061.1)	Temp (ET)
LS_4137682_158	В ядерном гене, кодирующем неизвестный белок <i>Picea sitchensis</i> (ABR16307.1), гомологичный транспозону <i>Prunus dulcis</i> (BBG96345.1)	Temp (ET)
LS_132659_5239 LS_132659_5292	После ядерного гена, кодирующего неизвестный белок <i>Shorea leprosula</i> (GKV31644.1), гомологичный транспозону <i>Glycine soja</i> (RZC29890.1)	Аутлайер в WT; PofDM (WT)
LS_1251904_144	В ядерном гене, кодирующем неизвестный белок <i>Picea sitchensis</i> (QHR90813.1), гомологичный zinc finger domain-containing protein бактерии <i>Stenotrophomonas</i> sp. (MBD3828844.1), хотя встречается и в других хвойных видах, что может быть связано с бактериальным загрязнением или какой-то эндогенной бактерией	Аутлайер в WT; Temp (WT)
LS_3903199_5631	После ядерного гена, кодирующего неизвестный белок <i>Shorea leprosula</i> (GKV44695.1), гомологичный ретровирусному полипротеину Pol транспозона TNT 1–94 <i>Cajanus cajan</i> (KYP38095.1)	Temp (WT)
LS_3983248_4557	В участке, гомологичном гену, кодирующему неизвестную mRNA <i>Picea glauca</i> (BT118153.1)	PofDM, Prcp, Srad, Temp (WT и ET), Wind (WT)
LS_2939623_746	В участке, гомологичном гену, кодирующему бактериальный белок retropepsin-like aspartic protease (WP_160144215.1) и неизвестный гипотетичный белок <i>Picea sitchensis</i> (QHR91704.1)	Prcp, Srad, Temp (WT и ET), PofDM (ET)
LS_7403_79464 LS_3963342_9449 LS_4465_9359	Контиги этих трех снипов не содержат гены, и в NCBI GenBank для них не найдены гомологичные последовательности	PofDM (WT) Temp (ET) Temp (ET)

изученных по относительно небольшому числу в основном селективно-нейтральных маркеров [Bagnoli et al., 2011; Neale, Wheeler, 2019]. Уровень генетической дифференциации по параметру F_{ST} был довольно высоким как на основе всех маркеров, так и на основе “нейтральных” маркеров и еще выше на основе “адаптивных”: между трансектами $F_{ST} = 0,136$, $0,133$ и $0,196$, в среднем между популяциями $F_{ST} = 0,099$, $0,097$ и $0,154$ соответственно. Для сравнения показатель F_{ST} , полученный при изучении генетической дифференциации популяций сибирской лиственницы на основе 22 аллозимных локусов, был в 2–3 раза ниже и равнялся $0,071$ [Oreshkova, Larionova, 2007], так же как и на основе семи ядерных микросателлитных локусов лиственницы сибирской из Республик Алтай и Тыва, Красноярского края ($F_{ST} = 0,071$) [Орешкова и др., 2013] или восьми ядерных микросателлитных локусов лиственницы сибирской из популяций Урала ($F_{ST} = 0,089$) [Chertov et al., 2021].

Полученные значения генетической дифференциации между трансектами (F_{CT}), между популяциями внутри трансект (F_{SC}) и между всеми популяциями (F_{ST}) согласуются с результатами, полученными для лиственницы ранее по большому числу снипов, полученных при ddRADseq [Novikova et al., 2023]. Большая доля генетической дисперсии, $84,5\%$, сосредоточена внутри изученных выборок, однако 13% варианты приходится на различия между трансектами.

По результатам анализа главных компонент показано, что выборки западной трансекты (WT) более генетически дифференцированы, нежели выборки восточной трансекты (ET). Минимальные значения генетической дифференциации обнаружены между популяциями NU и KRAS восточной трансекты. Для западной трансекты более выражено наличие линейной связи между генетическим, географическим и климатическим расстояниями между изучаемыми образцами. Также стоит отметить, что в популяции LAB выявлена существенная генетическая примесь восточной трансекты. Это согласуется с данными по изоферментам о наличии в крайне северных участках ареала зоны гибридизации западной и восточной рас лиственницы сибирской [Semerikov, Semerikov, 1999].

Поиск адаптивных маркеров-кандидатов с использованием шести различных методов позволил получить набор данных из 21 высокодостоверного маркера, изменчивость 12 из которых коррелировала с изменчивостью средовых климатических факторов. Аннотация геномных участков, содержащих эти 21 SNPs, показала, что большинство из маркеров сосредоточено в межгенных областях митохондриального генома, преимущественно вблизи генов *NAD1*, *NAD2*, *NAD4*, *L2*, *L5*, *S7*, *S11*, *S14*, *COX2* и *ATP4*. Также один из маркеров, LS_28474_5984, ассоциированный со средовой изменчивостью в трансекте ET, расположен недалеко от гена *futsch*-подобного белка, участвующего в метаболизме азотистых соединений. Некоторые маркеры находятся, возможно, в промоторном районе и могут влиять на экспрессию генов.

Локализация большинства из обнаруженных адаптивных маркеров в митохондриальном геноме представляет интерес. Известно, что митохондриальные гены играют решающую роль в локальной адаптации растений [Bock et al., 2014; Sloan, 2015], а найденные маркеры-кандидаты, вероятно, могут находиться в регуляторных областях и влиять на экспрессию данных генов.

Наличием среди обнаруженных маркеров значительного числа митохондриальных маркеров можно также объяснить большую дифференциацию популяций по этим наборам снипов по сравнению с маркерами, представляющими такие ядерные маркеры, как аллозимные и микросателлитные локусы в ранних исследованиях. В значительной мере такая значительная дифференциация митохондриальных маркеров может быть вызвана материнским характером наследования митохондриального генома, следствием чего является пониженная скорость пространственного перемещения митохондриальных генов, переносимых с семенами, по сравнению с ядерными генами, переносимыми и с семенами, и с пыльцой. Анализ митохондриальных снипов одновременно вместе с ядерными может привести к ошибочному выделению их как аутлайеров только на основании их повышенной дифференциации по сравнению с ядерными снипами. В то же время связь изменчивости митохондриальных снипов с климатическими переменными может быть

опосредована высокой корреляцией последних с географическими факторами.

Из маркеров, связанных с переменными окружающей среды в обеих трансектах, только два SNPs найдены как RDA, так и LFMM2. Небольшое число общих для двух трансект адаптивных маркеров, ассоциированных с изменчивостью факторов окружающей среды и подтвержденных более чем одним методом, может быть вызвано либо относительно небольшой выборкой, недостаточной статистически для обнаружения искомых ассоциаций, или же небольшим вкладом наблюдаемой генетической изменчивости в локальную адаптацию лиственницы сибирской в изученном градиенте климатических условий. Необходимо исследовать также эпигенетические механизмы адаптации.

Следует отдельно отметить некоторую генетическую обособленность популяции SIB в Республике Башкортостан от остальных популяций (см. рис. 5, 6). Вероятно, это отражает более выраженную дифференциацию популяций этой части Урала по сравнению с более северными за счет более длительного стабильного существования на Южном Урале, который рассматривается в качестве плейстоценового рефугиума для многих видов деревьев, включая лиственницу [Araki et al., 2008; Semerikov et al., 2013, 2019].

Некоторые авторы пишут об интрогрессивной гибридизации между лиственницей сибирской и лиственницей Сукачева (*Larix sukaczewii* Dyl.) на севере – в Приполярном Урале [Milyutin, Vishnevetskaia, 1995; Putenikhin, 2002; Путенихин, 2003]. Наши данные по снимкам это подтверждают – популяция LAB (г. Лабытнанги, Ямало-Ненецкий автономный округ) занимает промежуточное положение при кластеризации (см. рис. 5, 6), что соответствует также ее географическому расположению на границе между ареалами лиственницы сибирской и лиственницы Сукачева (или границы между западной и восточной расами лиственницы), где лиственница сибирская заходит с востока на ареал лиственницы Сукачева близко к Полярному Уралу.

ЗАКЛЮЧЕНИЕ

Впервые степень генетической дифференциации популяции лиственницы сибирской

в пределах двух географических трансект, соответствующих ареалам западной и восточной расы вида, была оценена по большому числу SNPs, полученных в результате геномного генотипирования методом ddRADseq.

Определено 9 SNPs, локализованных в митохондриальном геноме, потенциально находящихся под действием отбора. Обнаружено 12 адаптивных SNPs, ассоциированных с изменчивостью факторов окружающей среды, по пять для каждой трансекты и два общих для обеих трансект по результатам применения двух методов.

Исследование выполнено в рамках проекта “Изучение геномной и эпигеномной изменчивости, связанной с адаптацией древесных растений к гетерогенной среде”, поддержанного Российским фондом фундаментальных исследований (№ 18-29-13044). Сбор материала и выделение ДНК частично выполнены в рамках госзадания Института экологии растений и животных УрО РАН, № 122021000090-5. Биоинформатическая обработка и апробация полученных результатов частично выполнены в рамках госзадания ФИЦ КНЦ СО РАН № FWES-2022-0003.

Коллектив авторов выражает благодарность кафедре высокопроизводительных вычислений Сибирского федерального университета и ее заведующему, канд. техн. наук, доценту Дмитрию Александровичу Кузьмину за возможность проводить вычисления на высокопроизводительном компьютерном кластере.

ЛИТЕРАТУРА

- Крутовский К. В. Дендрогеномика – новая междисциплинарная область исследований адаптивного генетического потенциала лесных древесных популяций, интегрирующая дендрохронологию, дендрозоологию, дендроклиматологию и геномику // Генетика. Т. 58, № 11. С. 1225–1286. <https://doi.org/10.31857/S0016675822110054> [Krutovsky K. V. Dendrogenomics is a new interdisciplinary field of research of the adaptive genetic potential of forest tree populations integrating dendrochronology, dendroecology, dendroclimatology and genomics // Russ. J. Genet. 2022. Vol. 58, N 11. P. 1273–1286. <https://doi.org/10.1134/S1022795422110059>].
- Орешкова Н. В., Белоконов М. М., Жамъянсурен С. Генетическое разнообразие, популяционная структура и дифференциация лиственниц сибирской, Гмелина и Каяндера по данным SSR-маркеров // Генетика. 2013. Т. 49, № 2. С. 204–213. <https://doi.org/10.7868/s0016675812120090> [Oreshkova N. V., Belokon M. M., Jamiyansuren S. Genetic diversity, population structure, and differentiation of Siberian larch, Gme-

- lin larch, and Cajander larch on SSR-marker data // Russ. J. Genet. 2013. Vol. 49, N 2. P. 178–186. <https://doi.org/10.1134/S1022795412120095>.
- Путенихин В. П. Микроэволюционные аспекты внутривидовой дифференциации лиственницы Сукачева на Урале // Хвойные бореальной зоны. 2003. Вып. 1. С. 21–27 [Putenikhin V. P. Microevolutionary aspects of intra-specific differentiation of the Sukachev larch in the Urals // Conifers of the Boreal Zone. 2003. Vol. 1. P. 21–27. (In Russian with English Abstract)].
- Семериков В. Л., Ирошников А. И., Ласко М. Структура изменчивости митохондриальной ДНК и последняя история лиственницы сибирской (*Larix sibirica* Ledeb.) // Экология. 2007. № 3. С. 163–171 [Semerikov V. L., Iroshnikov A. I., Lascoux M. Mitochondrial DNA variation pattern and postglacial history of the Siberian Larch (*Larix sibirica* Ledeb.) // Rus. J. Ecol. 2007. Vol. 38, N 3. P. 147–154]. <https://doi.org/10.1134/S1067413607030010>
- Abaimov A. P. Geographical distribution and genetics of Siberian Larch species // Permafrost Ecosystems. Ecological Studies: Analysis and Synthesis / Eds.: A. Osa-wa, O. Zyryanova, Y. Matsuura, T. Kajimoto, R. Wein. Netherlands: Springer Dordrecht, 2010. Vol. 209. P. 41–58. https://doi.org/10.1007/978-1-4020-9693-8_3
- Abaimov A. P., Barzut V. M., Berkutenko A. N., Buitink J., Martinsson O., Milyutin L. I., Polezhaev A., Putenikhin V. P., Takata K. Seed collection and seed quality of *Larix* Spp. from Russia, initial phase on the Russian-Scandinavian Larch project // Eur. J. Forest Res. 2002. Vol. 4. P. 39. <http://hdl.handle.net/2115/22142>
- Araki N. H. T., Khatab I. A., Hemamali K. K. G. U., Inomata N., Wang X.-R., Szmidt A. E. Phylogeography of *Larix sukaczewii* Dyl. and *Larix sibirica* L. inferred from nucleotide variation of nuclear genes // Tree Genet. Genom. 2008. Vol. 4. P. 611–623. <https://doi.org/10.1007/s11295-008-0137-1>
- Bagnoli F., Fady B., Fineschi S., Oddou-Muratorio S., Potti A., Sebastiani F., Vendramin G. Neutral patterns of genetic variation and applications to conservation in conifer species // Genetics, Genomics and Breeding of Conifers / Eds.: C. Plomion, J. Bousquet, C. Kole. New York, USA: CRC Press and Edenbridge Science Publishers, 2011. P. 141–195. <https://doi.org/10.1201/b11075-5>
- Balkenhol N., Dudaniec R. Y., Krutovsky K. V., Johnson J. S., Cairns D. M., Segelbacher G., Selkoe K. A., von der Heyden S., Wang I. J., Selmoni O., Joost S. Landscape Genomics: Understanding Relationships Between Environmental Heterogeneity and Genomic Characteristics of Populations // Population Genomics: Concepts, Approaches and Applications / Ed.: O. Rajora. Switzerland: Springer Cham, 2019. P. 261–322. https://doi.org/10.1007/13836_2017_2
- Barton K. E., Jones C., Edwards K. F., Shiels A. B., Knight T. Local adaptation constrains drought tolerance in a tropical foundation tree // J. Ecol. 2020. Vol. 108, N 4. P. 1540–1552. <https://doi.org/10.1111/1365-2745.13354>
- Blum M., Chang H.-Y., Chuguransky S., Grego T., Kandasamy S., Mitchell A., Nuka G., Paysan-Lafosse T., Qureshi M., Raj S., Richardson L., Salazar G. A., Williams L., Bork P., Bridge A., Gough J., Haft D. H., Letunic I., Marchler-Bauer A., Mi H., Natale D. A., Neci M., Orengo C. A., Pandurangan A. P., Rivoire C., Sigrist C. J. A., Sillitoe I., Thanki N., Thomas P. D., Tossatto S. C. E., Wu C. H., Bateman A., Finn R. D. The InterPro protein families and domains database: 20 years on // Nucleic Acids Res. 2021. Vol. 49, N 1. P. 344–354. <https://doi.org/10.1093/nar/gkaa977>
- Bock D. G., Andrew R. L., Rieseberg L. H. On the adaptive value of cytoplasmic genomes in plants // Mol. Ecol. 2014. Vol. 23, N 20. P. 4899–4911. <https://doi.org/10.1111/mec.12920>
- Bolger A. M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data // Bioinformatics. 2014. Vol. 30, N 15. P. 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bondar E. I., Feranchuk S. I., Miroshnikova K. A., Sharov V. V., Kuzmin D. A., Oreshkova N. V., Krutovsky K. V. Annotation of Siberian larch (*Larix sibirica* Ledeb.) nuclear genome – one of the most cold resistant tree species in the only seasonal senescence Genus in Pinaceae // Plants. 2022. Vol. 11, N 15. Article number 2062. <https://doi.org/10.3390/plants11152062>
- Bradbury P. J., Zhang Z., Kroon D. E., Casstevens T. M., Ramdoss Y., Buckler E. S. TASSEL: software for association mapping of complex traits in diverse samples // Bioinformatics. 2007. Vol. 23, N 19. P. 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Bucharova A., Durka W., Hölzel N., Kollmann J., Michalski S., Bossdorf O. Are local plants the best for ecosystem restoration? It depends on how you analyze the data // Ecol. Evol. 2017. Vol. 7, N 24. P. 10683–10689. <https://doi.org/10.1002/ece3.3585>
- Capblancq T., Forester B. R. Redundancy analysis: A Swiss Army Knife for landscape genomics // Methods Ecol. Evol. 2021. Vol. 12, N 12. P. 2298–2309. <https://doi.org/10.1111/2041-210X.13722>
- Catchen J., Hohenlohe P. A., Bassham S., Amores A., Cresko W. A. Stacks: an analysis tool set for population genomics // Mol. Ecol. 2013. Vol. 22, N 11. P. 3124–3140. <https://doi.org/10.1111/mec.12354>
- Caye K., Jumentier B., Lepeule J., François O. LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies // Mol. Biol. Evol. 2019. Vol. 36, N 4. P. 852–860. <https://doi.org/10.1093/molbev/msz008>
- Chertov N., Vasilyeva Y., Zhulanov A., Nechaeva Y., Boronnikova S., Kalendar R. Genetic Structure and Geographical Differentiation of *Larix sibirica* Ledeb. in the Urals // Forests. 2021. Vol. 12, N 10. Article number 1401. <https://doi.org/10.3390/f12101401>
- Clark K., Karsch-Mizrachi I., Lipman D. J., Ostell J., Sayers E. W. GenBank // Nucleic Acids Res. 2016. Vol. 44, N 1. P. 67–72. <https://doi.org/10.1093/nar/gkv1276>
- Devey M. E., Bell J. S., Smith D. N., Neale D. B., Moran G. F. A genetic linkage map for *Pinus radiata* based on RFLP, RAPD and microsatellite markers // Theor. Appl. Genet. 1996. Vol. 92, N 6. P. 673–679. <https://doi.org/10.1007/BF00226088>
- de Villemereuil P., Gaggiotti O. E. A new F_{ST} -based method to uncover local adaptation using environmental variables // Methods Ecol. Evol. 2015. Vol. 6, N 11. P. 1248–1258. <https://doi.org/10.1111/2041-210X.12418>
- Doran A. G., Creevey C. J. Snpdat: Easy and rapid annotation of results from de novo snp discovery projects for model and non-model organisms // BMC Bioinformatics. 2013. Vol. 14, N 1. Article number 45. <https://doi.org/10.1186/1471-2105-14-45>

- Dray S., Dufour A.-B. The ade4 Package: Implementing the Duality Diagram for Ecologists // *J. Stat. Soft.* 2007. Vol. 22, N 4. P. 1–20. <https://doi.org/10.18637/jss.v022.i04>
- Evanno G., Regnaut S., Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study // *Mol. Ecol.* 2005. Vol. 14, N 8. P. 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Excoffier L., Lischer H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows // *Mol. Ecol. Resour.* 2010. Vol. 10, N 3. P. 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Fick S. E., Hijmans R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas // *Int. J. Climatol.* 2017. Vol. 37, N 12. P. 4302–4315. <https://doi.org/10.1002/joc.5086>
- Foll M., Gaggiotti O. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective // *Genetics.* 2008. Vol. 180, N 2. P. 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Jombart T. adegenet: a R package for the multivariate analysis of genetic markers // *Bioinformatics.* 2008. Vol. 24, N 11. P. 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Kamvar Z. N., Tabima J. F., Grünwald N. J. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction // *Peer J.* 2014. Vol. 2. Article number 281. <https://doi.org/10.7717/peerj.281>
- Kopelman N. M., Mayzel J., Jakobsson M., Rosenberg N. A., Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across *K* // *Mol. Ecol. Res.* 2015. Vol. 15, N 5. P. 1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- Kuzmin D. A., Feranchuk S. I., Sharov V. V., Cybin A. N., Makolov S. V., Putintseva Y. A., Oreshkova N. V., Krutovsky K. V. Stepwise large genome assembly approach: a case of Siberian larch (*Larix sibirica* Ledeb.) // *BMC Bioinformatics.* 2019. Vol. 20, N 1. Article number 37. <https://doi.org/10.1186/s12859-018-2570-y>
- Langmead B., Salzberg S. L. Fast gapped-read alignment with Bowtie 2 // *Nat. Methods.* 2012. Vol. 9, N 4. P. 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lasky J. R., Josephs E. B., Morris G. P. Genotype–environment associations to reveal the molecular basis of environmental adaptation // *The Plant Cell.* 2023. Vol. 35, N 1. P. 125–138. <https://doi.org/10.1093/plcell/koac267>
- Lu P., Parker W. C., Colombo S. J., Man R. Restructuring tree provenance test data to conform to reciprocal transplant experiments for detecting local adaptation // *J. Appl. Ecol.* 2016. Vol. 53, N 4. P. 1088–1097. <https://doi.org/10.1111/1365-2664.12647>
- Milyutin L. I., Vishnevetskaia K. D. Larch and larch forests of Siberia // *Ecology and Management of Larix Forests: A look ahead. Proceedings of an international symposium. 5–9 October 1992, Whitefish, MT* / Eds.: W. C. Schmidt, K. J. McDonald. U. S. Forest Service, General Technical Report INT-319. 1995. P. 50–53.
- Musmann S. M., Douglas M. R., Chafin T. K., Douglas M. E. AdmixPipe: population analyses in AdmixPipe for non-model organisms // *BMC Bioinformatics.* 2020. Vol. 21, N 1. Article number 337. <https://doi.org/10.1186/s12859-020-03701-4>
- Neale D. B., Wheeler N. C. Neutral Genetic Variation // *The Conifers: Genomes, Variation and Evolution* / Eds.: D. B. Neale, N. C. Wheeler. Switzerland: Springer Cham, 2019. P. 181–224. https://doi.org/10.1007/978-3-319-46807-5_9
- Novikova S. V., Sharov V. V., Oreshkova N. V., Simonov E. P., Krutovsky K. V. Genetic Adaptation of Siberian Larch (*Larix sibirica* Ledeb.) to High Altitudes // *Int. J. Mol. Sci.* 2023. Vol. 24, N 5. Article number 4530. <https://doi.org/10.3390/ijms24054530>
- Ohri D., Khoshoo T. N. Genome size in gymnosperms // *Plant Syst. Evol.* 1986. Vol. 153, N 1. P. 119–132. <https://doi.org/10.1007/BF00989421>
- Oksanen J., Simpson G. L., Blanchet F. G., Kindt R., Legendre P., Minchin P. R., O'Hara R. B., Solymos P., Stevens M. H. H., Szoecs E., Wagner H., Barbour M., Bedward M., Bolker B., Borcard D., Carvalho G., Chirico M., Caceres M. D., Durand S., Evangelista H. B. A., FitzJohn R., Friendly M., Furneaux B., Hannigan G., Hill M. O., Lahti L., McGlinn D., Ouellette M.-H., Cunha E. R., Smith T., Stier A., Braak C. J. F. T., Weedon J. vegan: Community Ecology Package, 2022
- Oreshkova N. V., Larionova A. Ya. Gene diversity and differentiation of Siberian larch (*Larix sibirica* Ledeb.) populations in the middle Siberia // *Proceedings of LARIX 2007: International Symposium of the IUFRO Working Group S2.02.07 (Larch Breeding and Genetic Resources)* “Integrated research activities for supply of improved larch to tree planting: tree improvement, floral biology and nursery production”, Saint-Michel-des-Saints and Québec City, September 16–21, 2007 / Compiled by M. Perron, Ministère des Ressources naturelles et de la Faune du Québec Direction de la recherche forestière: Canada, Québec, 2007. P. 33–38.
- Parchman T. L., Gompert Z., Mudge J., Schilkey F. D., Benkman C. W., Buerkle C. A. Genome-wide association genetics of an adaptive trait in lodgepole pine: ASSOCIATION MAPPING OF SEROTINY // *Mol. Ecol.* 2012. Vol. 21, N 12. P. 2991–3005. <https://doi.org/10.1111/j.1365-294X.2012.05513.x>
- Peterson B. K., Weber J. N., Kay E. H., Fisher H. S., Hoekstra H. E. Double Digest RADseq: An Inexpensive Method for *De Novo* SNP Discovery and Genotyping in Model and Non-Model Species // *PLoS One.* 2012. Vol. 7, N 5. Article number 37135. <https://doi.org/10.1371/journal.pone.0037135>
- Privé F., Luu K., Vilhjálmsson B. J., Blum M. G. B. Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4 // *Mol. Biol. Evol.* 2020. Vol. 37, N 7. P. 2153–2154. <https://doi.org/10.1093/molbev/msaa053>
- Putenikhin V. P. Phenotypic diversity, introgressive hybridization and microevolution of *Larix sukaczewii* Dyl. in the Urals // *Improvement of larch (Larix sp.) for better growth, stem form and wood quality. Proceedings of an International Symposium, Gap (Hautes-Alpes) – Auvergne & Limousin, France, 16–21 September, 2002* / Ed.: L. E. Pâques. Olivet Cedex, France: INRA, Unité d'Amélioration, 2002. P. 25–31.
- Putintseva Y. A., Bondar E. I., Simonov E. P., Sharov V. V., Oreshkova N. V., Kuzmin D. A., Konstantinov Y. M., Shmakov V. N., Belkov V. I., Sadovsky M. G., Keech O., Krutovsky K. V. Siberian larch (*Larix sibirica* Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest

- known mitogenome // BMC Genomics. 2020. Vol. 21, N 1. Article number 654. <https://doi.org/10.1186/s12864-020-07061-4>
- Semerikov V. L., Semerikov L. F. Intra- and interspecific allozyme variability in Eurasian *Larix* Mill. species // Heredity. 1999. Vol. 82, N 2. P. 193–204. <https://doi.org/10.1038/sj.hdy.6884710>
- Semerikov V. L., Semerikova S. A., Polezhaeva M. A., Koshintsev P. A., Lascoux M. Southern montane populations did not contribute to the recolonization of West Siberian Plain by Siberian larch (*Larix sibirica*): a range-wide analysis of cytoplasmic markers // Mol. Ecol. 2013. Vol. 22, N 19. P. 4958–4971. <https://doi.org/10.1111/mec.12433>
- Semerikov V. L., Semerikova S. A., Putintseva Y. A., Oreshkova N. V., Krutovsky K. V. Mitochondrial DNA in Siberian conifers indicates multiple postglacial colonization centers // Can. J. For. Res. 2019. Vol. 49, N 8. P. 875–883. <https://doi.org/10.1139/cjfr-2018-0498>
- Sloan D. B. Using plants to elucidate the mechanisms of cytonuclear co-evolution // New Phytologist. 2015. Vol. 205, N 3. P. 1040–1046. doi: 10.1111/nph.12835
- Storey J. D., Tibshirani R. Statistical significance for genomewide studies // Proc. Natl. Acad. Sci. U.S.A. 2003. Vol. 100, N 16. P. 9440–9445. <https://doi.org/10.1073/pnas.1530509100>
- Wadgymar S. M., DeMarche M. L., Josephs E. B., Sheth S. N., Anderson J. T. Local Adaptation: Causal Agents of Selection and Adaptive Trait Divergence // Annu. Rev. Ecol. Evol. Syst. 2022. Vol. 53, N 1. P. 87–111. <https://doi.org/10.1146/annurev-ecolsys-012722-035231>
- Wang Y., Zhang W.-Z., Song L.-F., Zou J.-J., Su Z., Wu W.-H. Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis* // Plant Physiol. 2008. Vol. 148, N 3. P. 1201–1211. <https://doi.org/10.1104/pp.108.126375>
- Zuur A. F., Ieno E. N., Elphick C. S. A protocol for data exploration to avoid common statistical problems // Methods in Ecology and Evolution. 2010 Vol. 1, N 1. P. 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>

Genetic structure and geographical differentiation of Siberian Larch (*Larix sibirica* Ledeb.) populations based on genome genotyping by sequencing

S. V. NOVIKOVA^{1, 2}, N. V. ORESHKOVA^{1, 2, 3}, V. V. SHAROV^{1, 2}, V. L. SEMERIKOV⁴, K. V. KRUTOVSKY^{1, 5, 6, 7}

¹Siberian Federal University

660041, Krasnoyarsk, Svobodny av., 79

E-mail: serafima_novikova_11@mail.ru

²Federal Research Center “Krasnoyarsk Science Center of the SB RAS”

660036, Krasnoyarsk, Akademgorodok, 50

³V. N. Sukachev Institute of Forest SB RAS

660036, Krasnoyarsk, Akademgorodok, 50/28

⁴Institute of Plant and Animal Ecology UB RAS

620144, Yekaterinburg, 8 March str., 202

⁵Georg-August University of Göttingen

Germany, 37077, Göttingen, Busgenweg, 2

⁶N. I. Vavilov Institute of General Genetics RAS

119991, Moscow, Gubkina str., 3

⁷Voronezh State University of Forestry and Technologies named after G. F. Morozov

394087, Voronezh, Timiryazev str., 8

The genetic differentiation of Siberian larch (*Larix sibirica* Ledeb.) populations in the latitudinal gradient of climatic conditions was studied based on high-throughput double digest restriction-site associated DNA sequencing (ddRADseq) data. We studied the correlation of five main climatic variables with the variability of 47,929 single nucleotide polymorphisms (SNPs). A total of 125 trees were studied: 61 trees in four populations along the western geographic transect and 64 trees in four populations along the eastern geographic transect. 21 SNPs with signatures of selection were identified, including 9 outlier SNPs whose variability cannot be explained by selectively neutral processes, and 12 SNPs whose variability correlated with the environmental factors. Seven SNPs are located in the introns of mitochondrial genes, three are located relatively close to the mitochondrial genes encoding NAD2 and ribosomal proteins S7 and S11, one is located at a distance from the nuclear gene encoding a protein homologous to the microtubule-associated futsch-like protein of *Arabidopsis thaliana*, two in the protein genes of an unknown nature and three in contigs containing no genes, and for which no homologous sequences were found in the NCBI GenBank.

Key words: population, genetic structure, *Larix sibirica*, adaptation, climatic variables, conifers, ddRADseq, SNP.