

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт Фундаментальной Биологии и Биотехнологии
Кафедра геномики и биоинформатики

УТВЕРЖДАЮ
Заведующий кафедрой
_____ И.Е. Ямских
подпись инициалы, фамилия
« _____ » _____ 20 ____ г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

«Изучение генетической адаптации в популяциях лиственницы сибирской с использованием данных полногеномного генотипирования»

06.04.01 Биология

06.04.01.06 Геномика и биоинформатика

Научный руководитель		<u>доцент, к.б.н.</u>	<u>Орешкова Н.В.</u>
	подпись, дата	должность, ученая степень	инициалы, фамилия
Выпускник			<u>Тимофеева Д.Б.</u>
	подпись, дата		инициалы, фамилия

Красноярск 2021

РЕФЕРАТ

Магистерская диссертация «Изучение генетической адаптации в популяциях лиственницы сибирской (*Larix sibirica* Ledeb.) с использованием данных полногеномного генотипирования» состоит из введения, 4 глав, заключения, списка использованных источников. Работа изложена на 56 страниц текстового документа, содержит 10 рисунков, 6 таблиц, 68 использованных источника.

ГЕНЕТИЧЕСКОЕ РАЗНООБРАЗИЕ, ЛИСТВЕННИЦА, ГЕНЕТИЧЕСКИЕ МАРКЕРЫ, ОДНОНУКЛЕОТИДНЫЕ ПОЛИМОРФИЗМЫ, АДАПТИВНАЯ ИЗМЕНЧИВОСТЬ.

Целью работы является проведение популяционно-генетического анализа выборок из популяций лиственницы сибирской с использованием данных полногеномного генотипирования.

Объектами исследования послужили пять выборок из популяции лиственницы сибирской.

В работе проведен популяционно-генетический анализ пяти выборок из популяции лиственницы сибирской, рассчитаны основные показатели внутри- и межпопуляционной изменчивости, генетической дифференциации популяций, выявлены гены-кандидаты для последующего поиска ассоциаций с дендрофенотипами.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
1 ХАРАКТЕРИСТИКА ВИДА ЛИСТВЕННИЦЫ СИБИРСКОЙ (<i>LARIX SIBIRICA</i> LEDEB.).....	6
1.1 Морфологическое описание.....	6
1.2 Экология и ареал распространения	10
1.3 Таксономическое положение лиственницы сибирской	13
2 ИССЛЕДОВАНИЕ ГЕНЕТИЧЕСКОЙ АДАПТАЦИИ У ХВОЙНЫХ С ИСПОЛЬЗОВАНИЕМ ДАННЫХ ПОЛНОГЕНОМНОГО ГЕНОТИПИРОВАНИЯ	16
2.1 Принцип метода RADseq.....	16
2.2 Особенности метода ddRADseq и его преимущества	21
2.3 Анализ данных ddRADseq, полученных на хвойных видах	25
3 МАТЕРИАЛЫ И МЕТОДЫ	32
3.1 Характеристика объектов исследования.....	32
3.2 Пробоподготовка образцов перед секвенированием	35
3.3 Биоинформатическая обработка полученных данных после секвенирования.....	37
4 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ	39
4.1 Очистка и демультимплексирование полученных прочтений.....	39
4.2 Картирование на референсный геном лиственницы Ошибка! Закладка не определена.	
4.3 Поиск однонуклеотидных полиморфизмов	Ошибка! Закладка не определена.
4.4 Расчёт основных внутри- и межпопуляционных генетических показателей	Ошибка! Закладка не определена.
4.5 Поиск генов-кандидатов	Ошибка! Закладка не определена.
ЗАКЛЮЧЕНИЕ	41
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	42

ВВЕДЕНИЕ

Одними из основных лесообразующих пород в России являются представители рода *Larix* (Mill). Лиственница, как и другие хвойные, обладает высоким уровнем генетической изменчивости, однако адаптивная компонента изменчивости вида слабо изучена, а также отсутствуют данные о генетических механизмах, контролирующие важные адаптивные и селекционные признаки.

Лиственница сибирская (*Larix sibirica* Ledeb.) является экологически пластичным видом, обладает высокой холодостойкостью, что делает ее удобным объектом для изучения адаптивной генетической изменчивости в разнообразных условиях среды, в частности – в условиях периодической засухи.

Использование предложенного подхода, который базируется на современных методах полногеномного генотипирования, для выявления генетических особенностей популяции лиственницы сибирской, позволит приблизиться к пониманию конкретных генетических механизмов адаптации (выявление генов и участков генома, содержащих SNP аутоаутлонов (далее по тексту «генов-кандидатов»)).

Наша работа является частью большого проекта, в котором планируется проанализировать генетические основы индивидуальной реакции деревьев лиственницы сибирской в лесостепной зоне Республики Хакасия на периодические засухи на основе геномно-дендрохронологического анализа.

Целью настоящей работы является проведение популяционно-генетического анализа выборок из популяций лиственницы сибирской (*Larix sibirica* Ledeb) с использованием данных полногеномного генотипирования.

В задачи исследования входило:

1. Подготовить ДНК библиотеки методом ddRADseq для последующего секвенирования пяти выборок лиственницы сибирской.
2. Осуществить первичную биоинформатическую обработку: очистку и фильтрацию по качеству сырых данных секвенирования.

3. Провести поиск однонуклеотидных полиморфизмов (SNP) и на основании этих данных изучить генетическое разнообразие исследованных выборок.
4. Выполнить поиск генетических маркеров адаптации (генов-аутлаеров), которые в дальнейшем будут использованы для ассоциаций с дендрофенотипами.

Автор выражает искреннюю благодарность научному руководителю Орешковой Н.В. за неоценимую помощь на всех этапах работы. Так же автор благодарит сотрудников лаборатории геномных исследований и биотехнологии ФИЦ КНЦ СО РАН за помощь в биоинформатической обработке полученных данных, ценные указания и советы. Огромную признательность автор выражает сотрудникам Научно-образовательной лаборатории "Дендрозкология и экологический мониторинг" Хакасского технического института — филиал СФУ (ХТИ) (Бабушкиной Е.А., Жирновой Д.Ф., Белокопытовой Л.В.) за предоставленные биологические образцы и помощь в описании экологических характеристик объектов исследования.

Магистерская диссертация выполнена в лаборатории лесной геномики (зав. лаб. к.б.н. Крутовский) на кафедре геномики и биоинформатики (зав. каф. д.б.н. И.Е. Ямских) ИФБиТ СФУ.

1 ХАРАКТЕРИСТИКА ВИДА ЛИСТВЕННИЦЫ СИБИРСКОЙ (*LARIX SIBIRICA* LEDEB.)

1.1 Морфологическое описание

Лиственница сибирская - это листопадное дерево, которое, в зависимости от условий произрастания, может достигать высоты 45 м при диаметре ствола 1,5–1,8 м за период до 450 лет, с полнодревесным стволом, в старости утолщённым в нижней части. Древостои возрастом 200-300 лет не являются редкостью [1]. Для неё характерен исключительно сильный рост, в первые три года вырастает чуть выше метра, а начиная с трех-, четырехлетнего возраста дает ежегодно метровый прирост, активно растёт до 100 лет, но и после рост не прекращается вплоть до значительного возраста дерева. В юношескую стадию переходит в возрасте от 8 до 25 лет. Образует как «чистые» древостои, так и смешанные с другими деревьями, такими как: ель, пихта, сосна и кедр [2]. Относится к гелиофитовым деревьям первого яруса, светлохвойная [3]. На севере потребность в свете выше, чем в южных широтах. В молодости может вынести значительные затенения. Малотребовательно к влажности воздуха.

Крона яйцевидно-коническая, раскидистая с толстыми горизонтальными ветвями, у молодых деревьев тонкая, узкая, к старости широкая с тупой вершиной, с горизонтальными длинными сучьями и приподнятыми концами ветвей. Однолетние побеги светлые, соломенно-желтые. Сеянцы лиственницы сохраняют свою хвою на протяжении года [4]. Кора молодых деревьев тонкая, голая, иногда с редкими волосками, буро-серая, мелкотрещиноватая, в то время как у старых деревьев она тёмная, красная на изломе, с глубокими трещинами, а так с очень толстой коркой (10-20 см), отделяющейся крупными кусками. В сомкнутых древостоях стволы ровные, высоко очищенные от сучьев, даже у молодых деревьев при густом стоянии крона занимает не более ¼ высоты.

Верхушечная почка ростовых побегов ширококонической формы крупнее боковых овальных, многочешуйчатая. Почки мужских генеративных побегов 3-5 мм в диаметре, шаровидные, а женские конусовидно-сферические, внешне похожи на вегетативные. Хвоя мягкая, линейная и узкая, ярко-зелёная с сизоватым налётом. На удлинённых побегах – одиночная, а на укороченных (брахибластах) – растёт пучками в количестве 10 – 50 штук, прямая или изогнута в виде серпа (рисунок 1). Длина иглы колеблется в пределах 10–58 мм в зависимости от экологических условий [5].

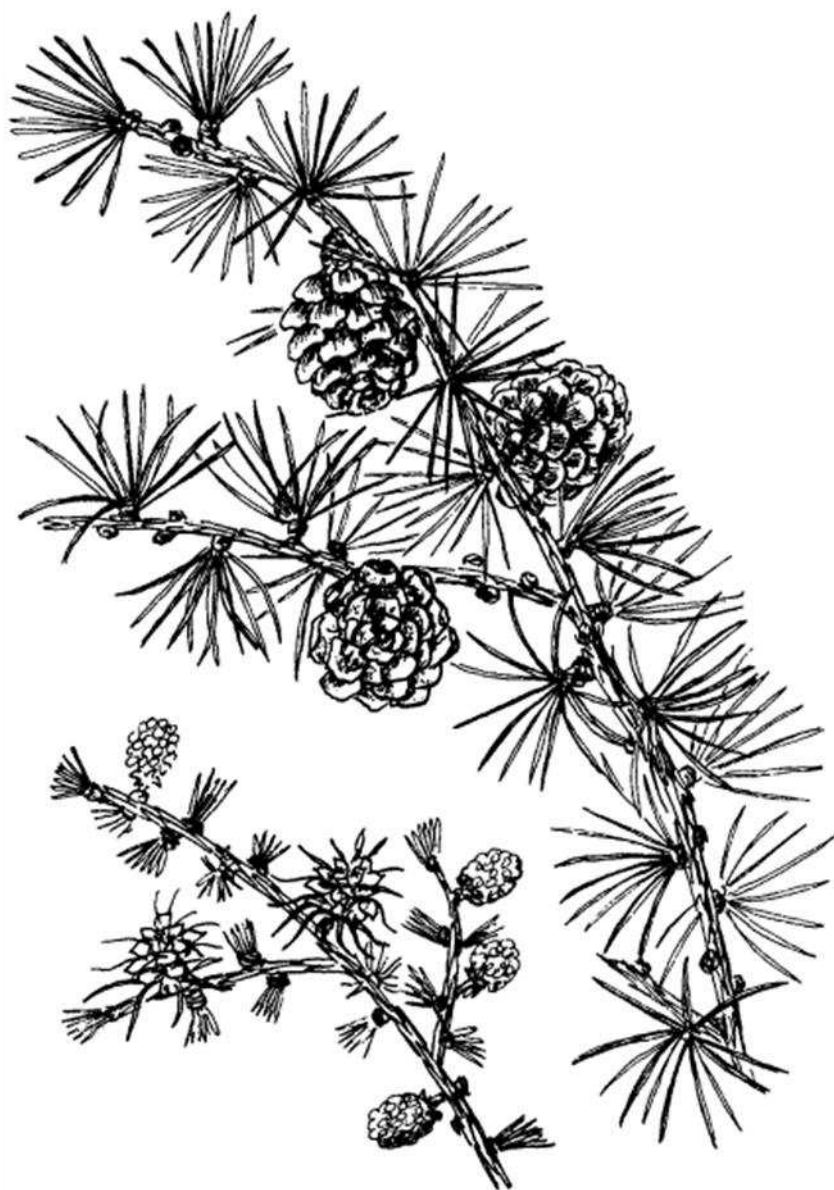


Рисунок 1 – Лиственница сибирская [6]

Дерево имеет мощную, хороша разветвлённую корневую систему с большим количеством развитых боковых корней и стержневым корнем. Произрастая на почвах с высоким увлажнением и с тяжёлым механическим составом, а также почв с залеганием горизонта вечной мерзлоты может образовывать мелкую поверхностную корневую систему, что в свою очередь приводит к ветровалу. Также такое может происходить на песчаных почвах. По сравнению с другими породами, лиственница лучше использует питательные вещества из почвы [7].

Хоть лиственница и однодомное дерево, оно имеет как мужские, так и женские цветки, нередко на одной ветке. Цветёт лиственница в период развёртывания листьев. На цветение отрицательно влияют дожди, низкая влажность и низкие температуры. Мужские колоски и женские шишки расположены на укороченных побегах [7]. Шишки длинной могут достигать 4 см, яйцевидные, семенные чешуи опушённые, ложкообразные. Раскрываются в марте, после выпадения семян могут оставаться на дереве около трёх лет. Крылатые семена до 6 мм длинной [6].

Опыление приходится на конец апреля – начало мая с помощью ветра. Пыльца может лететь на расстоянии 6-15 м, иногда только отдельные пылинки на 100 м. В смешанных насаждения ей препятствует крона других деревьев. Для оплодотворения лиственнице необходимо перекрёстное опыление. Мужские колоски полушаровидные, 5-6 мм в диаметре, желтоватые. Женские колоски широко-яйцевидно-конические, 10-15 мм длиной, чаще пурпурные и розовые, реже бледно-зелёные и беловатые (рисунок 2). Впервые появляются у свободно стоящих деревьев на 15-ом году жизни, а в плотных насаждениях на 30-50-м году [8]. Зрелые женские шишки сравнительно мелкие, 2,5-4 см длиной, образованы 25-50 сросшимися чешуями в 5-7 рядах. Типичная форма зрелых шишек - широкояйцевидная. В процессе созревания шишек угол наклона семенной чешуи колеблется от 20 до 50 градусов. Семенные чешуи мягкокожистые, цельнокрайние, как правило, имеют заметное красноватое опушение у основания или на значительной части поверхности. В основании шишки хорошо различимы

узкие кроющие чешуи [5]. Созревание семян происходит в сентябре – октябре. Семена 3-7 мм длиной с крылом 8-16 мм, в твёрдой оболочке с одной блестящей стороной и другой светло-коричневой с тёмными пятнами. Хорошо видны защитные чешуйки. Средний вес 1000 семян может составлять 4-10 г. [9].

Рисунок 2 – *Larix sibirica*. 1 – многолетний удлиненный побег с укороченными



побегами и хвоей, 2 – побег без хвои, 3 – однолетний побег, 4 – поперечный срез хвои, 5 – мужские колоски, 6 – женские колоски, 7 – незрелая шишка, 8 – старые шишки, 9 – семя с крылом [8]

1.2 Экология и ареал распространения

Лиственницы – наиболее распространённая лесобразующая порода в России. Лиственничные леса занимают площадь примерно 278 млн. га (около 40% всех лесов страны). Насаждения *L. sibirica* занимают примерно 14% всех лиственничных лесов. Наиболее частыми отмечаются в континентальных горных районах Южной Сибири [10]. Произрастает на севере европейской части России, Северном Урале, Западной Сибири, Алтае и Саянах (рисунок 3) [11].

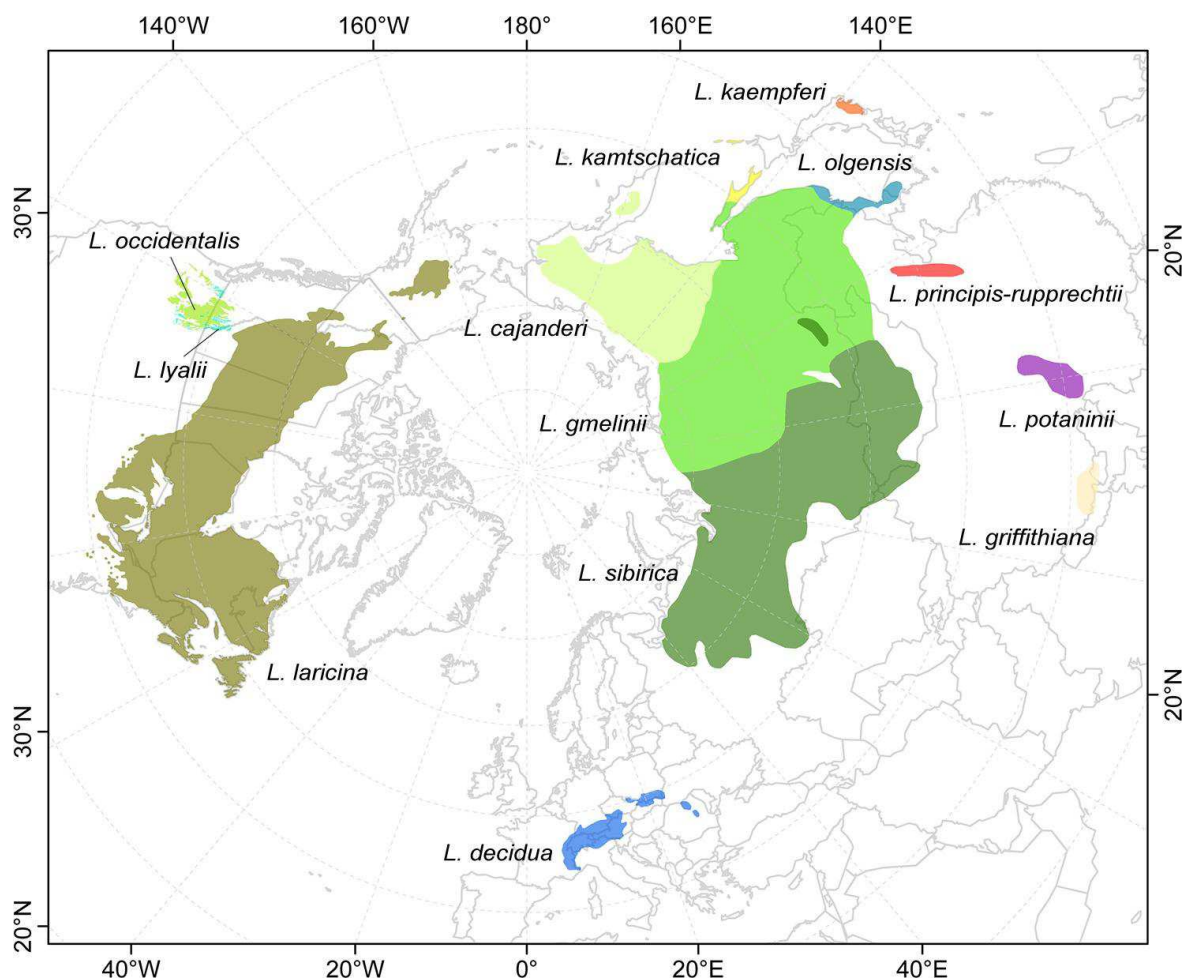


Рисунок 3 – Текущее глобальное распространение таксонов *Larix*, лиственница сибирская (*L. sibirica*) обозначена тёмно-зелёным цветом [12]

Для расселения лиственницы сибирской благоприятны склоновые участки с маломощными эродлируемыми почвами, гари с минерализованной поверхностью почвы, где устранена конкуренция со стороны травяной растительности и кустарников [2].

Лиственница неприхотлива: может произрастать на вечной мерзлоте (корни в этом случае располагаются горизонтально, втискиваясь в тонкий слой почвы). Произрастает на почвах с разным механическим и химическим составом. Почвы могут быть суглинистыми и супесчаными, часто хрящеватыми, с большим количеством дресвы. На севере Европейской России растёт на известняках и гипсах, в предгорьях Саян предпочитает местообитание с залежами карбонатными породами [2]. Достигает высокой производительности только на плодородных, хорошо дренированных и свежих почвах. При заболачивании почвы образует придаточные корни. Древесина очень тяжёлая, сразу после сруба не пригодна для сплава (тонет в воде), очень тяжела в обработке, но при этом она очень долговечная и не поддаётся гниению [4].

Одни из специфических адаптационных механизмов у лиственницы сибирской можно отметить: наиболее эффективную систему терморегуляции фотосинтетического аппарата, энергичную транспирацию на холодных почвах, способность формировать мощный ассимиляционный аппарат хвои при минимальных затратах энергии на питание, интенсивное развитие поверхностной корневой системы и придаточных корней на мерзлоте [13]. *L. sibirica* занимает ареал со значительными колебаниями температур, количества осадков, толщине снежного покрова, влажности воздуха, поэтому она является одной из наименее требовательных к климатическим условиям пород. В пределах ареала она встречается как на многолетней мерзлоте, так и на болотистой местности. Лиственница сибирская – порода континентального климата, малотребовательна к теплу, может переносить низкие температуры, мириться с коротким летним периодом, так и с жарким сухим климатом. Также может переносить значительную сухостью почвы, однако избегает бедных песчаных почв, предпочитая известняки и гипсы. Наибольшей же производительности она достигает на плодородных, хорошо дренированных суглинистых и супесчаных карбонатных почвах или на подзолах, подстилаемых карбонатной мореной [7].

На севере *L. sibirica* вместе с елью сибирской образует полярную границу распространения древостоев на территории между реками Обь и Енисей. Самая верхняя северо-восточная граница ареала этого вида находится в бассейне реки Пясина на полуострове Таймыр. Там он образует редкие насаждения, на которых кустарниковые формы этого вида растут рядом с отдельными деревьями.

Западная граница ареала простирается от городов Курган и Ялуторовск до северо-восточного течения реки Ишим, а по долине переходит в реку Иртыш. Далее граница поворачивает на юго-запад и впадает в бассейн реки Тобол. Далее следует немного западнее города Челябинск и идет на юг в сторону города Оренбург. Имея общее направление на юго-восток, западная граница лиственницы сибирской проходит через южные районы Новосибирской области и Алтайского края.

На юге Алтая доходит до Холзунского хребта и Катунских гор. Далее граница поворачивает на юг в сторону Республики Казахстан по склонам Южного Алтая и хребта Саура и пересекает бассейн озера Маркаголь. Далее идет на территорию Монгольской республики по главному хребту Монгольского Алтая. Отдельные насаждения сибирской лиственницы доходят до гор Байтак-Ула в Монголии.

На крайнем северо-востоке своего распространения, доходя до верхнего водораздела реки Пясина, лиственница сибирская растет по западному макросклону гор Путорана, проникая в долины рек Дудинка, Хантайка и Курейка. Затем восточная граница пересекает реку Нижняя Тунгуска и проходит по ее водоразделу с рекой Подкаменная Тунгуска на территории Эвенкии, Иркутской области и Республики Саха (Якутия). Восточная граница ареала этого вида доходит до верхнего течения бассейна реки Лена. Затем резко поворачивает на юго-запад в сторону западного берега озера Байкал. С восточного берега граница направлена на юго-восток. В Республике Бурятия простирается в сторону города Улан-Удэ. Южнее, на территории Читинской области.

Юго-восточная окраина ареала достигает города Петровск-Забайкальский в бассейне реки Ингода. Затем он покидает территорию Российской Федерации в сторону Республики Монголия [9].

1.3 Таксономическое положение лиственницы сибирской

Вид лиственница сибирская (*Larix sibirica*) принадлежит к классу хвойных (*Coniferopsida*), семейству сосновых (*Pinaceae*), подсемейству лиственничных (*Laricceidae*), роду лиственница (*Larix*) [14]. Род лиственницы был выделен впервые в 1754 г. английским ботаником Ф. Миллером из рода *Pinus*, позднее в 1845 г. Ф. И. Рупрехт описал три вида, произрастающих в России лиственниц – *L. ledebouri*, *L. gmelinii*, *L. kamtschatika*. К.Ф. Ледебур признал выделенную Рупрехтом *L. ledebouri* и отметил ее распространение не только в европейской части России, но и в Сибири, и даже в Забайкалье. Ледебур отнёс описанную им в 1833 году *L. sibirica* к синониму этого вида.

Систематическая классификация рода *Larix* до сих пор является предметом споров. Противоречивые утверждения о количестве видов для этого рода и их таксономический статус были обусловлены различными взглядами на историческое формирование рода, а также высоко выраженным полиморфизмом на фоне слабой репродуктивной изоляции. Наличие внутривидовых гибридов в естественных и индуцированных популяциях привело к значительным расхождениям во взглядах ученых даже на количество видов. До сегодняшнего дня не существует единого мнения о числе видов лиственницы [10; 9]. Так, например, В.Н. Сукачев выделял 14 видов, В.Л. Комаров — 25, Н.В. Дылис — 20, Е.Г. Бобров — 16 [10]. По данным И.Ю. Коропачинского [15] на территории бывшего СССР произрастает не более 6 видов, а в Азиатской части России — 4 вида.

Наиболее подробно вопросы систематики лиственницы в СССР разработаны и освоены Н.В. Дылисом. В 1947 году Дылис [16] из рода *Larix sibirica* выделил ещё один самостоятельный род – лиственницы Сукачева (*L.*

sukaczewii), что долгое время подвергалось сомнению, в связи с мнением об отсутствии морфологических, генетических, географических отличий [7]. В.Н. Сукачев выделил эту лиственницу как подвид – *L. sibirica ssp. rossica* (Regel), занимающий северо-восточную часть европейской России и граничащей с ареалом лиственницы сибирской. Изучение её фенотипической изменчивости на территории обширного ареала от Онежского озера и Белого моря на западе до низовья Енисея и Южного Забайкалья на востоке, показало, что лиственница, распространенная на севере европейской части России, по всему Уралу и отчасти за Уральским Хребтом до долины рек Оби и Иртыша, имеет существенные морфологические и эколого-биологические особенности, отличающие ее от лиственницы сибирской [16]. Позже были получены многочисленные данные о том, что этот вид отличается от лиственницы сибирской по морфологическим, анатомическим и биохимическим признакам [17]. Согласно А.И. Ирошникову, лиственница Сукачева хорошо выраженный вид, характеризующийся небольшой генетической дифференциацией и слабым полиморфизмом. М.А. Михайленко [18] сообщили, что лиственница Сукачева и сибирская различаются по строению покровных тканей и некоторым биохимическим признаком, в частности, составу эфирных масел.

Однако видовой статус лиственницы Сукачева отвергается некоторыми учеными. Е.Г. Бобров не признает целесообразности обособления лиственницы Сукачева в ранг самостоятельного вида по морфологическим и географическим признакам. Ссылаясь на работу Симака [19], он отметил морфологическую идентичность хромосом *L. sibirica* и *L. sukaczewii* при изучении их кариотипов.

Сравнительно-кариологический анализ лиственниц Сукачева и сибирской проводился также Л.И. Милютиным [20], полученный результат свидетельствует, что кариотипы лиственниц Сукачева и сибирской во многом сходны. Исходя из полученных результатов, были сделаны выводы, что нет оснований для выделения лиственницы Сукачева в качестве самостоятельного вида. Отрицание видовой самостоятельности лиственницы Сукачева является, по мнению ряда исследователей, субъективным и не учитывает важных морфологических и

биологических особенностей этого вида. В связи с отсутствием единого мнения по этой проблеме вопрос о видовом статусе лиственницы Сукачева остается открытым.

На данный момент *L. sibirica* имеет следующие разновидности: *rossica* (северные и северо-восточные районы Европейской части России), *obensis* (бассейн реки Обь, за исключением Алтая), *altaica* (Алтай), *jenisseensis* (бассейн Енисея), *sajanensis* (Восточные Саяны), *polaris* (северная Сибирь) *lenensis* (бассейн верхней Лены и районы Иркутского Приангарья), *baicalensis* (северо-западное и северо-восточное побережье Байкала) и *transbaicalensis* (горы западного Забайкалья) [10].

2 ИССЛЕДОВАНИЕ ГЕНЕТИЧЕСКОЙ АДАПТАЦИИ У ХВОЙНЫХ С ИСПОЛЬЗОВАНИЕМ ДАННЫХ ПОЛНОГЕНОМНОГО ГЕНОТИПИРОВАНИЯ

2.1 Принцип метода RADseq

Последние технические достижения в области геномики стимулировали исследования в области экологии и эволюции и способствовали интеграции этих двух областей. Высокопроизводительное генотипирование диких популяций немодельных организмов открывает беспрецедентные возможности для раскрытия генетического материала, ведущего к фенотипическим изменениям и адаптации [21]. Только накопив такие исследования, мы сможем понять генетику адаптации и получить комплексное представление об окружающей среде, фенотипе и генотипе. В то время как секвенирование следующего поколения генерирует огромное количество геномных данных, интерпретация таких данных представляет собой серьезную проблему для ученых. Лимит обычно больше не технический, а скорее сочетание времени, усилий и денег. Анализ сложных полных геномов является дорогостоящим и требует много времени. В последнее время было разработано несколько методов сокращения генома до репрезентативной, но более управляемой фракции. Эти методы сокращенного представления генома используют рестрикционные ферменты для разрезания и фрагментирования генома с последующим целевым секвенированием этих фрагментов. Из мутаций, идентифицированных в последовательностях этих фрагментов, могут быть обнаружены от сотен до десятков тысяч однонуклеотидных полиморфизмов (SNP), которые служат в качестве генетических маркеров, используемых для идентификации генетической структуры и адаптивной изменчивости в популяциях [22].

Производить гигабазы последовательности ДНК в короткие сроки и с минимальными затратами стало возможным с помощью платформ Illumina. Благодаря этому полные геномы могут быть секвенированы с нуля в рамках

обычного исследовательского гранта. Однако стандартные методы для подобных платформ не подходят для популяционных генетических исследований, где обнаружение и использование генетических маркеров у многих индивидуумов имеет первостепенное значение.

Метод секвенирования ДНК ассоциированный с сайтом рестрикции (RADSeq) может идентифицировать и оценивать тысячи генетических маркеров, случайным образом распределенных по целевому геному, от нескольких индивидуумов популяции, используя технологию Illumina. RADSeq может быть использован для проведения популяционных генетических исследований видов, которые не имеют или ограничены в существующих данных о последовательности, а также имеет ряд преимуществ по сравнению с другими методами обнаружения маркеров. Он сродни анализу с использованием полиморфизмов длины рестриционных фрагментов (RFLPs) и амплифицированных полиморфизмов длины фрагментов (AFLPs) в том, что он уменьшает сложность генома путем субсэмплирования только в определенных участках, определенных ферментами рестрикции. RADSeq превосходит эти методы в своей способности идентифицировать, проверять и оценивать маркеры одновременно (вместо того, чтобы требовать обширного процесса разработки) и надежно определять, какие маркеры происходят от каждого сайта. RADSeq может быть использован на скрещиваниях любой конструкции и в диких популяциях, позволяя не только генотипировать и обнаруживать SNP, но и проводить более сложные анализы [23].

RADSeq сочетает в себе два простых метода молекулярной биологии с секвенированием Illumina: использование рестриционных ферментов для разрезания ДНК на фрагменты и использование молекулярных идентификаторов (MID) для ассоциирования считывания последовательностей с конкретными индивидуумами (рисунок 4). ДНК разрезается выбранным рестриционным ферментом, образуя набор липких концевых фрагментов. Чтобы быть секвенированными на машине Illumina, эти фрагменты должны быть привязаны к адаптерам, которые свяжутся с проточной ячейкой Illumina.

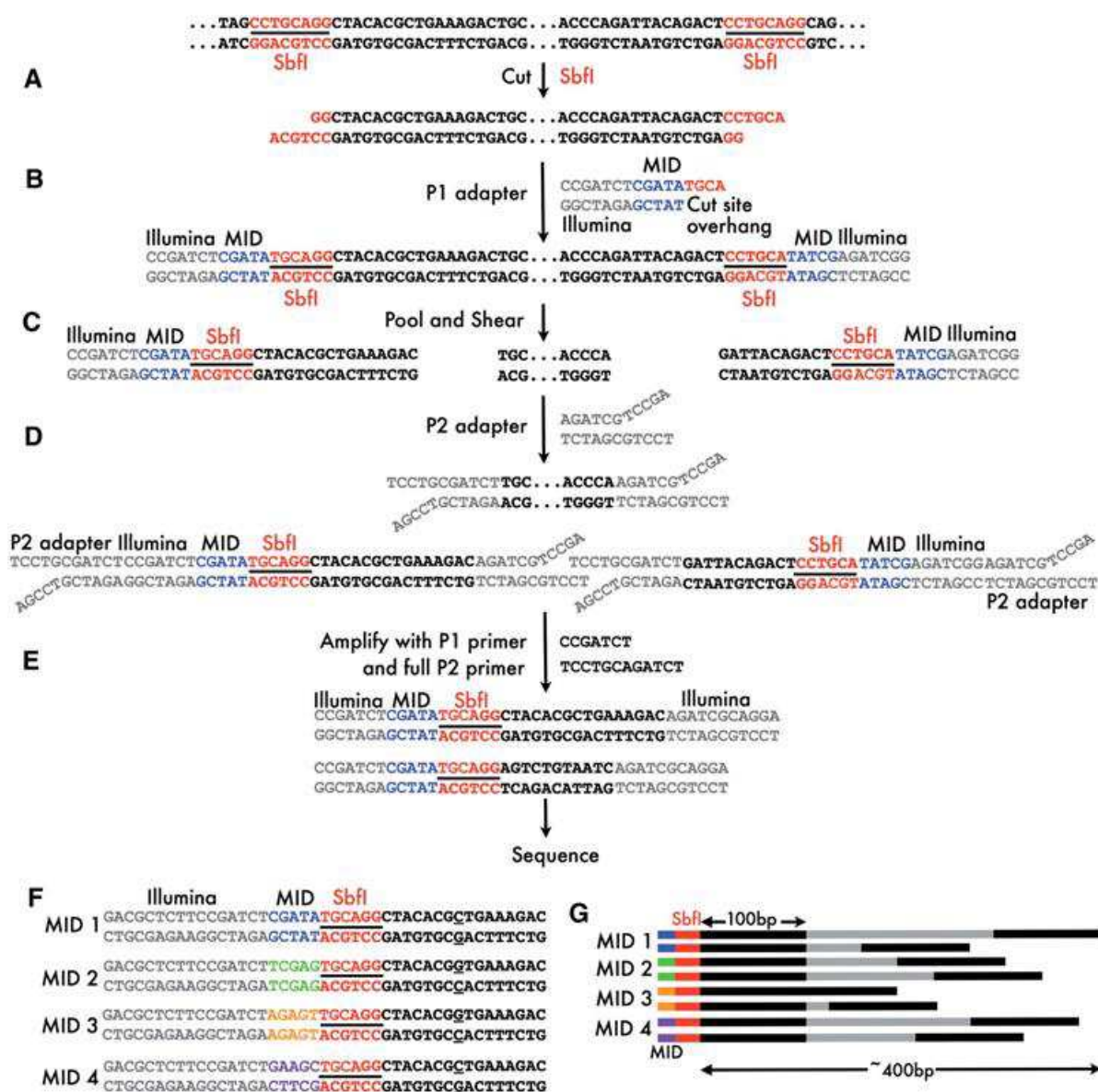


Рисунок 4 – Процесс RADSeq. А – ДНК разрезается рестрикционным ферментом. В – Лигирование адаптера P1 с молекулярным идентификатором. С – Помеченные фрагменты объединяются в пул, а затем разрезаются. D – Лигирование адаптера P2. E – ПЦР-амплификация с праймерами P1 и P2. F – Объединенные образцы с разными MID разделяются биоинформатически. G – Поскольку фрагменты срезаются случайным образом, парные концевые последовательности из каждого секвенированного фрагмента будут охватывать область 300–400 п.н. ниже сайта рестрикции [23]

RADSeq использует модифицированные адаптеры Illumina, которые позволяют связывать и усиливать только фрагменты сайта рестрикции. Фрагменты липкого конца лигируются с первым адаптером, который содержит соответствующий липкий конец и молекулярный идентификатор (короткая

последовательность, которая идентифицирует индивидуума). Помеченные рестрикционные фрагменты из нескольких особей объединяются в пул, а затем случайным образом нарезаются для получения фрагментов со средней длиной в несколько сотен пар оснований, далее лигируются со вторым адаптером и амплифицируются с помощью ПЦР с использованием праймеров. Второй адаптер имеет дивергентную структуру "Y", который не будет связываться со вторым праймером, если не будет завершена амплификация с первым адаптером. Эти нарезанные, готовые для секвенатора фрагменты затем выбираются по размеру, и эта библиотека RADSeq секвенируется на платформе Illumina. Последовательность генерируется из MID в первом адаптере и через сайт рестрикционного фермента [23].

Если доступен эталонный геном, считанные необработанные последовательности можно выравнивать с эталонной последовательностью, а SNP и инделы идентифицировать с помощью существующих инструментов биоинформатики. Сопоставление с эталонным геномом автоматически исправляет низкий уровень ошибки секвенирования при считывании. Если эталонная последовательность недоступна, то можно анализировать *de novo*.

Кроме основного протокола RADseq существуют три его модификации: ddRADseq, ezRAD и 2bRAD. Протокол 2bRAD [24] использует эндонуклеазу рестрикции PstI-типа для вырезания фрагментов длиной 36 п.н., содержащих сайт узнавания из 6 оснований и прилегающих 5' и 3' пар оснований. К этим фрагментам лигируются адаптеры с двойными штрих-кодами, целевую полосу вырезают из агарозного геля после ПЦР. В этом методе нет промежуточных стадий очистки и выбора размера фрагмента. Процедура может быть настроена для представления меньшего количества локусов в геноме с помощью основно-селективных адаптеров. Протокол ezRAD [25] использует два высокочастотных изохиомизирующих фермента-реза (для одного и того же участка разреза) для геномной ДНК. Затем расщепленную ДНК вставляют непосредственно в коммерческий набор для подготовки библиотеки Illumina TruSeq. Используя набор Illumina, ДНК репарируется на концах, а адаптеры с использованием

одиной или двойной индексации лигируются на фрагменты генома. Затем образцы объединяются и выбираются по размеру [26].

Double Digest Restriction Associated DNA Sequencing [27], также называемый ddRAD, является вариацией протокола секвенирования RAD [28], который используется для обнаружения SNP и генотипирования (рисунок 5).

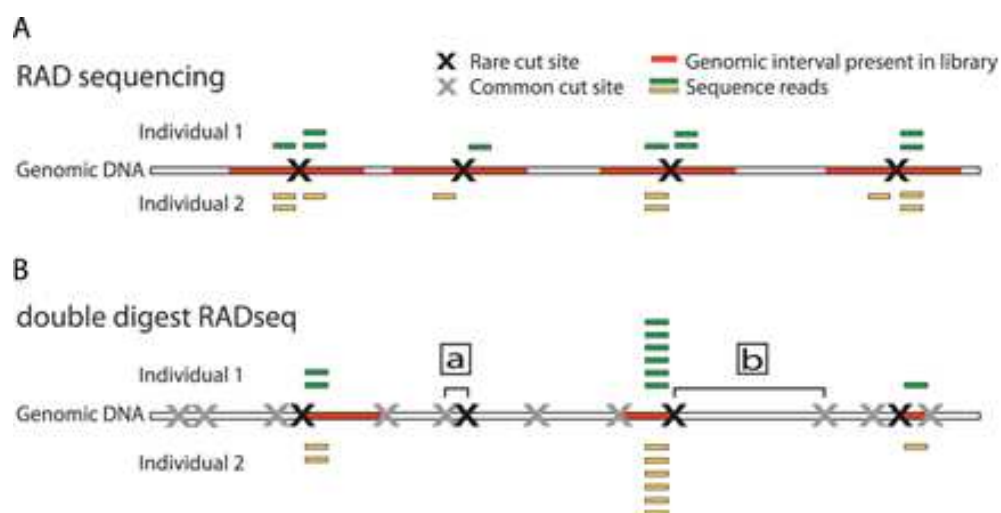


Рисунок 5 – Double digest RADseq. А – стандартный метод секвенирования ДНК – RADseq. В – метод ddRADseq [27]

При традиционном секвенировании ДНК по методу RADseq, используют один фермент рестрикции в сочетании со вторичной случайной фрагментацией и широким выбором размера для генерации библиотек, а при секвенировании по протоколу ddRADseq, напротив, используют два фермента рестрикции с последующим точным выбором размера, который исключает области, окруженные либо очень близкими, либо очень отдаленными сайтами распознавания ферментов рестрикции, восстанавливая библиотеку, состоящую только из фрагментов, близких к целевому размеру (красные сегменты на Рисунке 2 (В)). ddRADseq состоит из обработки геномной ДНК двумя различными ферментами рестрикции, обычно один с высокочастотным участком и один с низкочастотным участком. Следующий шаг – лигирование адаптеров, эффективность лигирования зависит от общего количества адаптеров и количества концов фрагментов в расщепленном образце. Было обнаружено, что 2-10-кратный избыток адаптеров к дополнительным липким концам приводит к

более эффективному лигированию получившихся фрагментов с адаптером. Образцы с индивидуальным штрих-кодом с помощью уникального адаптера могут быть объединены после этапа лигирования и очищены. Задача данного этапа состоит в том, чтобы объединить каждого индивидуума с уникальным штрих-кодом в конечный образец, чтобы далее на всём пуле продуктов лигирования можно было выбрать размер фрагментов. Далее проводят ПЦР-амплификацию лигированных фрагментов для создания библиотек секвенирования Illumina и прикрепления индексов, выбора размера, объединения образцов и, наконец, секвенирования объединенных фрагментов с использованием высокопроизводительной платформы секвенирования. Мультиплексирование для ddRADseq обычно достигается путем лигирования адаптеров с уникальными линейными штрих-кодами к разрезанной геномной ДНК, а затем добавления уникальных индексов во время ПЦР. Адаптеры для лигирования штрихкодов должны синтезироваться парами и могут быть дорогостоящими, специфичными для рестрикционного фермента и поэтому могут применяться только с помощью ddRADseq [27; 29].

2.2 Особенности метода ddRADseq и его преимущества

Секвенирование ДНК, ассоциированной с рестрикционным сайтом, стало мощным и полезным подходом в молекулярной экологии, и в настоящее время молекулярным экологам доступно несколько различных опубликованных методов, ни один из которых нельзя считать лучшим вариантом во всех ситуациях. Существует мнение, что оригинальный протокол RAD [30] превосходит все другие варианты RAD, поскольку предполагаемые дубликаты ПЦР могут быть идентифицированы, тем самым уменьшая влияние дубликатов на оценки частоты аллелей [31]. Но имея представление о сильных и слабых сторонах различных протоколов RAD, можно принять более обоснованное решение при выборе протокола RAD.

ddRADseq улучшает глубину охвата каждого локуса за счет оптимизации секвенирования и сокращения отсутствующих генотипов. Этот гибкий протокол позволяет легко оптимизировать различные организмы, размеры генома, генетическое разнообразие и научные вопросы.

Метод ddRAD предлагает наибольшую степень настройки. В зависимости от выбранных ферментов (один набор уникально штрихкодированных "гибких адаптеров" работает по крайней мере с пятью парами ферментов) и диапазона выбранных размеров фрагментов можно получить сотни SNP на индивидуума при очень низких затратах (например, достаточной для базового анализа структуры популяции), тысячи SNP для экспериментов по отображению QTL при умеренных затратах или десятки тысяч SNP для более точного отображения ассоциаций. Таким образом, исследования, требующие меньшего количества фрагментов для получения надежных выводов, или желающие оптимизировать количество фрагментов или индивидуумов, охваченных на разумную глубину с ограниченным числом считываний секвенирования, могут быть экономичными.

Как и в случае любого протокола, который избегает сдвига и настраивает номера фрагментов с выбором размера, изучение гистограмм, полученных образцов на ранней стадии проекта позволяет идентифицировать, а затем исключать из библиотек слишком частые фрагменты (например, транспозоны). Эта процедура может быть очень ценной при изучении организмов с большими несеквенированными геномами [26].

Использование выбора размера фрагмента для настройки количества выбранных локусов может привести к вариативному представлению некоторых локусов среди библиотек. Это можно минимизировать с помощью инструментов точного выбора размера. А также метод может быть особенно чувствителен к выпадению аллелей [32], и это следует учитывать при выполнении чувствительных популяционных генетических анализов. Возможно, для ddRAD требуется геномная ДНК самого высокого качества из всех методов RAD. Правильное лигирование фрагментов зависит не только от полного переваривания двух ферментов, но и от полностью интактных 5'- и 3'-концов.

Метод исключает случайный сдвиг и конечную репарацию геномной ДНК. Вместо этого используется два рестрикционных фермента, которые приводят по меньшей мере к пятикратному снижению стоимости производства библиотеки. Кроме того, исключение нескольких этапов с высокой потерей ДНК позволяет создавать библиотеки ddRAD из 100 нг или менее исходной ДНК. Также стоит отметить точный отбор геномных фрагментов по размеру, что позволяет более точно контролировать долю регионов, представленных в конечной библиотеке. Точный и повторяемый выбор размера дает еще два преимущества. Во-первых, поскольку только малая часть рестрикционных фрагментов попадет в режим выбора размера мишени, вероятность отбора проб в обоих направлениях от одного и того же сайта рестрикции мала. Это уменьшает «повторяющуюся» (то есть непосредственно соседнюю) выборку области, что фактически вдвое снижает количество считываний, необходимых для достижения высоконадежной выборки SNP, связанного с заданным участком рестрикции. Во-вторых, общее смещение в представлении регионов в пользу фрагментов, наиболее близких к среднему значению выборки по размеру, в свою очередь, смещает независимые образцы в сторону восстановления тех же геномных регионов. Из-за этого коррелированного восстановления области «заполняются» считываниями примерно в одном и том же порядке по всем отдельным образцам. Оба эти свойства делают метод ddRADseq устойчивым к недостаточной выборке в отношении количества считываний, что является часто наблюдаемой проблемой, возникающей из-за неравномерного представления считываний в отдельных образцах в экспериментах по объединенному секвенированию [27].

Разработанные адаптеры для секвенирования ddRAD, одновременно включают комбинаторный встроенный штрих-код и стандартный индекс мультиплексированного чтения Illumina. Вкратце, небольшое количество адаптеров со штрих-кодом лигируется отдельно с отдельными образцами в формате микропланшета. Затем эти образцы объединяются после лигирования, но перед выбором размера. Выбор размера выполняется для каждого пула индивидуумов, и полученные библиотеки амплифицируются с помощью

праймера, который вводит индекс, который будет считываться при отдельном мультиплексном считывании в соответствии со стандартным протоколом мультиплексированного секвенирования парных концов Illumina. После ПЦР с однозначно проиндексированными праймерами можно объединить несколько пулов, и индивидуумы, которые используют одни и те же встроенные штрих-коды, различающиеся на основе комбинации штрих-кода адаптера и мультиплексного считывания индексы. Таким образом, эта двухуровневая схема индексации позволяет экспоненциально увеличивать количество однозначно идентифицируемых образцов на пул, избегая при этом дополнительных затрат на синтез олигонуклеотидов и секвенирование, связанных с большим количеством более длинных уникальных штрих-кодов.

Малозатратный ddRADseq, успешно применяется для обнаружения вариантов SNP, а также генерирует достаточное количество гомологических локусов для построения филогении для нескольких предполагаемых видов. Число гомологичных локусов в отдаленно родственных клонах, может быть улучшено за счет большей глубины секвенирования или более узкого выбора длин последовательностей. Следовательно, этот метод особенно эффективен для выявления внутривидового разнообразия. Однако полиморфизм сайтов рестрикции, вероятно, ограничит эффективность ddRADseq для филогенетической реконструкции за пределами уровня рода и семейства. Кроме того, гипотеза риска мутации предполагает, что меньшее количество гомологичных или общих локусов будет восстановлено для локусов с более высоким риском мутаций. Следовательно, более длинные чтения, которые имеют большую вероятность мутаций, создающих новые сайты рестрикции в пределах прочтения, могут с меньшей вероятностью распространяться среди видов, чем внутри видов. Простые мутации в сайтах рестрикции могут также увеличивать вариацию длины чтения для более длинных чтений по сравнению с более короткими чтениями, но могут быть некоторые последовательности с сайтами рестрикции, включенными в середину последовательности, что позволяет предположить, что изменение длины также может быть вызвано неэффективным

использованием ферментов. Вариации длины, вызванные неэффективным использованием ферментов или мутациями в сайтах рестрикции, могут вызывать более низкие скорости кластеризации, но не должны влиять на филогенетический анализ, поскольку пробелы рассматриваются как недостающие данные. Обрезка считываний до более коротких длин или нацеливание на более короткие считывания, вероятно, минимизирует потерю кластеров или локусов из-за вариации длины между образцами и внутри образцов разных видов. Однако использование более длинных чтений дает более информативные сайты, которые могут компенсировать уменьшение количества общих локусов и потенциально могут разрешать более глубокие расхождения лучше, чем более короткие чтения. Более длинные чтения также могут быть более полезными для последующего генетического анализа популяции, такого как картирование генома, реконструкция гаплотипа и связывание с адаптивными локусами [29].

2.3 Анализ данных ddRADseq, полученных на хвойных видах

Геномные ресурсы позволяют изучать организацию генома, реконструировать филогению и демографическую историю видов, а также исследовать генетическую структуру и их функциональные признаки. Геномные ресурсы интенсивно разрабатывались для модельных видов, и с недавним развитием технологий секвенирования следующего поколения (NGS) они также быстро расширяются на немодельные виды. Применение доступных в настоящее время платформ NGS с высокой пропускной способностью сильно ограничены малой длиной считывания. Такие короткие считывания проблематичны для сборки генома, что приводит к сильно фрагментированным, несмежным геномным реконструкциям. Эти ограничения дополнительно усугубляются при сборке геномов с высоким содержанием повторов или с высокими уровнями гетерозиготности. Хвойные деревья обладают большими и сложными геномами в диапазоне от 20 до 30 Гб, особенно относительно большие межгенные области.

Оценки частоты повторения последовательностей в геномах хвойных деревьев колеблются от 23% до 90%, при этом наиболее значительный вклад вносят мобильные элементы. Относительно высокий уровень гетерозиготности, который обычно характерен для древесных пород, наряду с обширными дупликациями генов, особенно в механизмах защиты от вредителей, еще больше усугубляет проблему высокого содержания повторов [33]. Хвойные деревья являются доминирующим компонентом лесных экосистем Северного полушария, и в связи с их экономическим и экологическим значением значительные ресурсы для секвенирования генома были направлены на несколько репрезентативных видов хвойных (например, сосна ладанная (*Pinus taeda*) [34;35], сосна ламберта (*Pinus lambertiana*) [36]). Поэтому при секвенировании нового немодельного вида важным моментом является отсутствие потребности в референсном геноме, такую возможность даёт секвенирование по методу ddRADseq, так как он обладает большей гибкостью и надежностью при восстановлении SNP. А также избежание случайного сдвига ДНК из-за использования двух ферментов рестрикции (редкорезкий и часторезкий). ddRAD-seq был успешно применен к различным организмам для построения карт связей, анализа QTL, сравнительной геномики и сборки генома.

Чтобы выбрать подходящую комбинацию ферментов рестрикции и диапазон размеров фрагментов продуктов рестрикции, можно провести анализ *in silico*. Учитывая тяжелое метилирование повторяющихся элементов в растениях, ожидаемо, что использование чувствительного к метилированию фермента рестрикции снизит содержание повторяющихся элементов в библиотеках. В работах Юцин Цзинь [37] по хвойному виду плосковеточник восточный (*Platycladus orientalis*) было задействовано 139 образцов из мегагаметофита и один образец из материнского дерева. Целью включения образца диплоидного материнского дерева является использование его генотипа в качестве эталона для проверки правильности определения генотипа мегагаметофитов и повышения точности фильтрации данных. Из-за гаплоидной природы мегагаметофитов каждый уникальный локус должен содержать один из двух возможных аллелей

диплоидного генотипа материнского дерева. Локусы с количеством более двух аллелей были идентифицированы как повторяющиеся и были отфильтрованы. Локусы с двумя расщепляющимися аллелями в популяции мегagamетофитов были идентифицированы как полиморфные. Среди полиморфных локусов были оставлены только те локусы, которые соответствовали материнскому генотипу. Авторами была показана эффективность метода ddRAD-seq для идентификации общегеномного полиморфизма в геноме хвойного дерева на примере вида *P. orientalis*, а также его полезность и потенциал для других видов, не имеющих эталонного генома для генерации большого количества маркеров для генетических исследований. В исследовании были собраны 397 226 локусов, из которых 27,4% (108 683) были полиморфными и 72,6%-мономорфными. После строгой фильтрации получили 23 926 высококачественных локусов для построения карты сцепления [37].

В исследовании китайских учёных по древнему виду кипарисовых деревьев – криптомерия японская, или кипарис японский (*Cryptomeria japonica* var. *sinensis*) [38] так же использовали технологию ddRAD-seq, основанную на 122 образцах из семи древних популяций деревьев в Китае, чтобы оценить уровень генетического разнообразия, изучить генетическую структуру среди текущие популяции древних деревьев и оценить демографическую историю. После отбора и фильтрации было получено 922 SNP, которые использовались для оценки генетического разнообразия семи популяций. Локусы не отступили от нейтралитета. Количество аллелей в каждой популяции колебалось от 1,550 до 1,939, в среднем 1,789. В итоге было подтверждено низкое генетическое разнообразие среди криптомерий японских, не только на территории Японии, но также на территории Китая.

Сосна кипарисовая белая (*Callitris glaucophylla*) вечнозеленое хвойное дерево, устойчивое к засухе, относится к австралийской кипарисовой сосне (*Callitris columellaris*) как к видовому комплексу. Комплекс состоит из пяти близкородственных морфовидов, которые встречаются в широком диапазоне биоклиматических регионов Австралии. Экологическая геномика комплекса дает

возможность идентифицировать маркеры, связанные с адаптацией к окружающей среде, и расширит понимание процесса видообразования. Для этого приняли подход картирование сцепления с одним деревом в сочетании с высокопроизводительным секвенированием ДНК, ассоциированным с сайтами рестрикции (RAD). В частности, использовали изменённый протокол ddRADseq с использованием двух ферментов рестрикции: EcoRI и BglII, что позволило обнаружить 7560 SNP в 11 группах сцепления (соответствующих числу гаплоидному набору хромосом $2n = 22$). После исключения маркеров, которые значительно отклонялись от ожидаемого соотношения сегрегации окончательная карта сцепления состояла из 4284 генетических маркеров, из которых 4279 были однонуклеотидные замены [39].

Изучая роль гибридизации при экологической дивергенции сосны белой Мексиканская (*Pinus strobiformis*) и сосны мягкой (*P. flexilis*) Мэнон и Бэгли [40] так же использовали метод ddRADseq. Они выделяли геномную ДНК из 445 особей, взятых из популяций обоих видов. Пять библиотек ddRADseq были обработаны двумя рестрикционными ферментами EcoRI и MseI с последующим лигированием адаптеров, штрих-кодов и праймеров. После ПЦР были отобраны фрагменты ДНК размером 300–400 п.н. В результате всего процесса было получено 51 633 однонуклеотидных полиморфизма (SNP), которые использовались в качестве исходного набора данных для всех последующих анализов. Исследование показало отсутствие повышенных островков дивергенции, что не обязательно указывает на отсутствие адаптивной дивергенции во время видообразования с потоком генов. Отсутствие островков дивергенции ожидается у хвойных пород, учитывая преобладание полигенных архитектур, определяющих непрерывную вариацию признаков на границах видов, и ожидаемую распространенность мягких вытеснений. Проводили пониженную выборку общего набора данных SNP для простоты вычислений на основе сводной генетической статистики популяции, а затем произвольно отобрали один SNP для каждого собранного контига, чтобы получить окончательный набор данных из 4800 SNP. Из апостериорного распределения

были обнаружены 3193 локуса-выброса, из которых 570 (17,9%) имели повышенную вероятность происхождения сосны мягкой (*P. Flexilis*), а 2623 (82,1%) имели повышенную вероятность происхождения сосны белой Мексиканской (*P. strobiformis*). Также было идентифицировано меньшее количество локусов с избыточным происхождением, но локусы с избыточным происхождением отдавали предпочтение сосне мягкой (204 локуса – 71,1%), а не сосне белой Мексиканской (83 локуса, 28,9%). Эти результаты подтвердили обилие гибридов передовых поколения и отсутствие локусов, демонстрирующих резкий переход частоты аллелей по всей гибридной зоне.

Данные, полученные Фридлином для вида сосна Бальфура (*Pinus balfouriana* Grev. & Balf.) [41], так же были получены на основе протокола ddRADseq. Геномную ДНК выделяли из каждого спасенного мегагаметофита, методом двойной рестрикции с использованием EcoRI и MseI с последующим односторонним секвенированием на платформе Illumina HiSeq. Было обнаружено 349 542 предполагаемых SNP. Эти 349 542 SNP были расположены в 83 051 уникальном контиге (35,94% от общего числа), в среднем 4 SNP на контиг.

Закономерности местной адаптации в пространственных масштабах играют центральную роль в понимании того, как протекает эволюция, и имеют важное значение для эффективного управления экономически и экологически важными лесными древесными породами. Линд [42] использовал анализ генетических данных для описания мелкомасштабной адаптации в восьми популяциях сосны белой (*Pinus albicaulis* Engelm.) по всей локальной протяженности экологически гетерогенного бассейна озера Тахо, США. Используя к геномной ДНК два рестрикционных фермента MseI и EcoRI, было обнаружено 116 231 SNP, из них 107 354 (92,4%) сопоставлены с эталонным геномом сосны Ламберта (*P. lambertiana*), что придает достоверность данным о последовательности. Было показано, что, несмотря на высокую общую генетическую вариабельность ($F_{ST} = 0,0069$), имеются убедительные доказательства адаптации к дождевой тени.

Сосновые шишки, которые остаются закрытыми и сохраняют семена до тех пор, пока огонь не заставит шишки раскрыться (серотиния шишек), представляют собой ключевой адаптивный признак у множества видов сосны. Серотинность сосны широкохвойной (*Pinus contorta*) значительно различается по региону Скалистых гор. Это изменение серотины возникло в результате географически различающегося отбора, с последствиями, которые распространяются на лесные сообщества и экосистемы. Понимание генетической архитектуры этого признака представляет интерес из-за далеко идущих экологических последствий серотины, а также из-за повторяющейся эволюции признака в пределах рода. Парчман [43] применил метод ddRADseq для эффективного по времени получения геномных данных, используя рестрикционные ферменты EcoRI и MseI для создания библиотеки фрагментов, которые можно секвенировать с высоким уровнем мультиплексирования. В итоге получили данные для более чем 95 000 однонуклеотидных полиморфизмов у 98 серотиновых и несеротиновых сосен из трех популяций. Вероятность серотинизации варьировала в зависимости от генотипа в 11 локусах, и связь между генотипом и серотиной в этих локусах была последовательной в каждой из трех популяций сосен. Генетическая изменчивость по этим 11 локусам объясняет 50% фенотипической изменчивости серотина.

Пан и другие использовали комбинацию из четырех ферментов рестрикции: EcoRI-MseI, HpaII и PstI, для оптимизации процесса секвенирования для хвойных пород деревьев на трёх видах: сосна красная китайская (*Pinus tabuliformis*), сосна густая (*Pinus densata*) и сосна юньнаньская (*Pinus yunnanensis*). Используя программное обеспечение Stacks и bowtie2-GATK, исследовали количество SNP внутри каждого вида путем отбора девяти особей каждого вида сосны из библиотек PstI, было обнаружено 6917 SNP в 20 049 референсных локусах *P. densata*, 7299 SNP в 20 973 локусах *P. yunnanensis* и 11 268 SNP в 18 295 локусах *P. tabuliformis* [44].

Данные для анализ данных ddRADseq, полученных на хвойных видах, представлены в таблице 1.

Таблица 1 – Данные ddRADseq, полученные на хвойных видах

Рестриктазы		Виды	SNPs	Ссылки на статьи
1	2			
ApeKI	-	<i>Pinus contorta</i> <i>Picea glauca</i>	17 765 (63 388) 17 845 (56 807)	[45]
EcoT22I (NsiI)	-	<i>Pinus contorta</i> <i>Picea glauca</i>	-	
PstI	-		-	
EcoT22I (NsiI)	Pst I		-	
PstI	-	<i>Cedrus atlantica</i>	129 471	[33]
PstI	-	<i>Pinus densata</i> <i>Pinus yunnanensis</i>	6 917 7 299	[44]
EcoRI	MseI	<i>Pinus tabuliformis</i>	11 268	
EcoRI	MseI	<i>Pinus contorta</i>	95 000	[44]
SphI	MluCI	<i>Tsuga mertensiana</i>	171 019	[46]
EcoRV	ScaI	<i>Platycladus orientalis</i>	397 226	[37]
PstI	SphI	<i>Cryptomeria japonica var. sinensis</i>	922	[38]
EcoRI	MseI	<i>Pinus strobiformis</i> <i>Pinus flexilis</i>	51 633	[40]
EcoRI	MseI	<i>Pinus balfouriana</i>	349 542	[41]
MseI	EcoRI	<i>Pinus albicaulis</i>	116 231	[42]
EcoRI	BglII	<i>Callitris glaucophylla</i>	4 279	[39]

3 МАТЕРИАЛЫ И МЕТОДЫ

3.1 Характеристика объектов исследования

Изучены популяции лиственницы, пережившие сильные засухи (менее 250 мм осадков) в 1930, 1932, 1946, 1959, 1965, 1974 гг. Исследуемая территория предгорья отрога Батеневского кряжа располагается в республике Хакасии (Боградском районе) и представляет широтную часть Белдыкского нагорья, протягивающегося между Енисеем и Кузнецким Алатау. Горная система характеризуется небольшим перепадом высот (500-1200 м н.у.м.).

Батеневский кряж является продолжением одной из более крупной части структур горного обрамления впадины – Кузнецкого Алатау, и отделяет Чулымо-Енисейскую котловину от Сыдо-Ербинской и Южно-Минусинской (Рисунок 6). В этом регионе постоянно действующий умеренный дефицит увлажнения постепенно усиливается вследствие быстрого повышения температур вегетационного сезона (на 0,14-0,19°C за декаду). В предгорьях Батеневского кряжа на границе со степями и сельскохозяйственными угодьями наблюдаются как чистые лиственничники, так и смешанные березово-сосново-лиственничные древостои, состоящим из лиственницы сибирской (*Larix sibirica* Ledeb.), сосны обыкновенной (*Pinus sylvestris* L.) и березы повислой (*Betula pendula* Roth.). На более сухих южных и юго-восточных склонах редколесье перемежается степными участками. В связи с ограниченным увлажнением, расположение нижней границы леса определяется ландшафтными условиями, в первую очередь ориентацией склонов: на северных склонах деревья имеют наилучшее жизненное состояние и спускаются ниже всего в связи с меньшими потерями влаги на эвапотранспирацию.

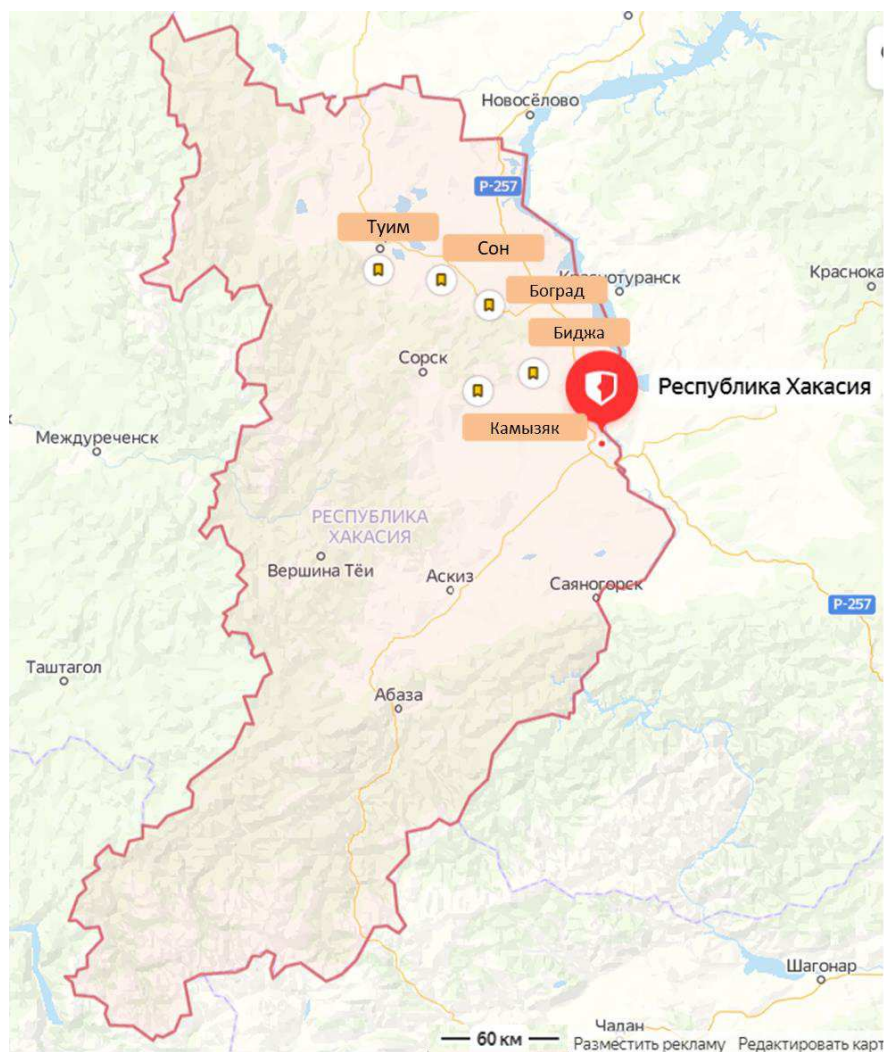


Рисунок 6 – Расположение выборок

Климат региона резко континентальный, с большими сезонными и суточными перепадами температур в зависимости от месяца. Среднегодовая температура около 0°C , годовое количество осадков 470-560 мм, из них около 80% выпадает за сезон положительных температур с максимумом в июле. Безморозный период длится в среднем с мая по сентябрь, в течение этого же интервала среднесуточные температуры превышают $+5^{\circ}\text{C}$. В результате климат вегетационного сезона колеблется между состояниями жарко-сухо и прохладно-влажно, т.е. экстремальными и благоприятными для роста деревьев условиями.

Образцы древесины лиственницы собраны на пяти участках (таблица 2), расположенных через 30-50 км в лесостепной зоне вдоль предгорий

Батеневского хребта. Участки выбраны на хорошо освещенных склонах (ориентированных от юго-запада до юго-востока).

Таблица 2 – Географическое расположение исследованных выборок из популяций лиственницы сибирской

Участок	Растительность	Рельеф	Координаты	Высота над уровнем моря, м
Биджа (BID)	Смешанный лес (сосна обыкновенная – доминант, также лиственница сибирская, береза повислая) с кустарниковым подлеском (кизильник, спирея) и подростом хвойных, злаково-разнотравный травяной покров (ирис, герань лесная, осока, мышиный горошек). Почва слегка каменистая.	Склоны юго-западной ориентации (до 15°) и их подножие	54°00' с.ш. 90°58' з.д.	640-670
Боград (BOGR)	Смешанный лес (сосна обыкновенная, лиственница сибирская, береза повислая), изреженный злаковый травяной покров (после пожара) на каменистой почве: полынь, чабрец, злаки, герани. У подножия склона молодая поросль лиственницы, сбоку есть погибшие от пожара лиственницы и берёзы	Склон южной ориентации (~25°) и гребень холма	54°15' с.ш. 90°41' з.д.	550-620
Камызяк (КАМ)	Лиственничный лес с березой (подножие) и отдельные деревья (склоны), травяной покров злаковый, на склонах скудный из-за очень каменистой почвы, в глубине леса злаково-разнотравный	Подножие северного склона (деревья 5-18), склоны юго-западной и южной ориентации (~20°)	53°55' с.ш. 90°37' з. д.	700-770

Окончание таблицы 2

Сон (SON)	Смешанный лес (лиственница сибирская, береза повислая, отдельные деревья сосны обыкновенной), некоторые деревья несут следы низовых пожаров (сажа на коре). На восточном склоне злаково-разнотравный травяной покров (василисник, косяника, мышиный горошек, клубника), местами подлесок. На западном склоне условия суше, травянистый покров представлен в основном низкорослыми степными злаками, древостой изрежен.	Склон юго-западной (15-29 деревья) и юго-восточной (1-14, 30-35) ориентации, крутизна 20-25° и гребень холма	54°21' с.ш. 90°22' з. д.	530-600
Туим (TUIM)	Отдельно стоящие и группы деревьев лиственницы сибирской на фоне сухостепной растительности (в основном злаки, чабрец), почва местами каменистая со скальными выходами	Равнина и пологие (до 15°) склоны, преимущественно юго-западной ориентации	54°21' с.ш. 89°57' з. д.	550-600

В настоящее время термин «популяция» используется весьма широко, однако общепринятого определения до сих пор нет. Не заостряя внимания на дискуссионных вопросах, касающихся различных взглядов на термин «популяция», отметим, что мы придерживаемся эколого-систематического подхода, тем самым понимая его как ценопопуляцию по А.А. Корчагину [47], или местная популяция (выборка) по К.М. Завадскому [48].

3.2 Пробоподготовка образцов перед секвенированием

Из хвои образцов лиственницы сибирской пяти выборок (30 образцов в каждой) была выделена ДНК с использованием СТАВ метода. В свою очередь она была проверена на качественно-количественный состав. Концентрацию ДНК проверяли при помощи настольного флуориметра Qubit 2.0 Invitrogen/Life Technologies USA в сочетании с набором реагентов Qubit TM dsDNA BR Assay

Kit. Первичную качественную оценку ДНК проводили с использованием горизонтального электрофореза в 0,8 – 2 %-ом агарозном геле. Данная процедура проводилась для выявления отсутствия признаков деградации. Оценка частоты и качества выделенной ДНК так же проводили с использованием Implen NanoPhotometr P330. Частоту образца определяли, исходя из соотношения оптических плотностей при длинах волн 230, 260 и 280 нм. Соотношение A260/A280 для чистых нуклеиновых кислот должно лежать в пределах от 1,8 до 2,2, оптимальное значение которого – 1,8. Для дальнейшей работы были отобраны образцы высококачественной ДНК с концентрацией 20-150 нг/мкл, что соответствует протоколу приготовления ddRAD-Seq библиотек.

Разработан и модифицирован протокол для приготовления ddRAD-Seq библиотек на имеющемся материале. Использовали одновременно обработку двумя рестрикционными ферментами: EcoRI – Mse. При отработке методики были внесены модификации, позволяющие создавать библиотеки из исходной ДНК с концентрациями от 20 нг/мкл до 150 нг/мкл.

Был осуществлён дизайн адаптеров с баркодами с сайтами рестрикции вышеуказанных рестриктаз для мультиплексирования образцов при последующем секвенировании. На первом этапе для ddRad-Seq была проведена обработка ДНК рестриктазами. Далее для каждого образца было осуществлено лигирование адаптера, содержащего участок комплементарный сайту рестрикции EcoRI, адаптера – баркода уникального для каждого образца (индивидуального дерева), для его последующей идентификации, адаптера, содержащего участок комплементарный праймеру P1 для амплификации и секвенирования и адаптера, содержащего участок комплементарный сайту рестрикции MseI и участок комплементарный праймеру P2 для амплификации и секвенирования. После лигирования адаптеров была проведена очистка фрагментированной ДНК на шариках AMPure XP Beckman Coulter в соотношении 1:1, для удаления слишком коротких фрагментов ДНК. Далее осуществлялась ПЦР амплификация продуктов лигирования с использованием Q5 Hot Start HF полимеразы. Затем полученные ПЦР продукты выпаривались на

половину и все образцы объединялись в один пул. Выделение из продуктов амплификации фрагментов размером 300-700 bp для последующего секвенирования осуществлялось путём вырезания зоны соответствующей 300-700 bp из 2,5%-ого агарозного геля после электрофореза в течении 2 -2,5 часов при 110 V. Выделение ДНК из геля осуществляли при помощи набора QIAquick Gel Extraction Kit компании Qiagen. Полученные пулы библиотек ddRAD-Seq были проверены на качество при помощи капиллярного электрофореза на приборе Bioanalyzer 2100 Agilent Technologies с использованием набора High Sensitivity DNA. Концентрацию пула измеряли на флуориметре Qubit™ dsDNA BR Assay.

Секвенирование ddRADseq библиотек осуществлялось на платформе NovaSeq 6000 компании Illumina с использованием проточной ячейки S1 (длина прочтений 1x100 п.о.).

3.3 Биоинформатическая обработка полученных данных после секвенирования

С помощью программ Stacks [49] и PLINK [50] был рассчитан ряд популяционных статистик: наблюдаемая и ожидаемая гетерозиготность, коэффициент инбридинга, значения F_{st} , частота минорного аллеля и оценка равновесия Харди-Вайнберга. Также с помощью ПО Arlequin [51] был проведён анализ AMOVA (Analysis of MOlecular VAriance), который представляет собой метод выявления дифференцировки популяции с использованием молекулярных маркеров [52].

Поиск по выявлению предполагаемых участков генома, содержащих SNP-аутлаер (далее называемые гены-кандидаты), проводился с использованием двух различных методов. Первый метод – байесовская структура с методом дифференцирования была использована в BayeScan v2.1 [53]. А также использовался пакет для R OutFLANK [54] для идентификации генов-кандидатов находящихся под действием отбора. Он оценивает распределение F_{ST} значений в

нейтральных или почти нейтральных локусах путем подгонки эмпирических данных к распределению хи-квадрат, после обрезки чрезмерно высоких и низких значений F_{ST} , поскольку эти локусы могут находиться под диверсифицирующим или балансирующим отбором. Затем распределение сравнивается с эмпирическими данными, и выбросы идентифицируются как те, которые находятся вне ожидаемого распределения.

4 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

Проведено массовое секвенирование ddRADseq полученных библиотек, отобранных 5 выборок лиственницы сибирской, ранее пережившие сильные засухи в предгорьях Батеневского кряжа (Туим, Камызяк, Сон, Боград и Биджа). Секвенирование проводили на платформе NovaSeq 6000 Illumina с использованием 3-х ячеек S1 (длина прочтений 1x100 п.о.). Поскольку мы работали с библиотеками, имеющими высокую степень гомологичности, в пул библиотек, для снижения количества ошибок на начальных этапах запуска, была добавлена сбалансированная библиотека PhiX control Illumina. Библиотека PhiX составляла 25% от общего объёма образцов. По результатам секвенирования было получено более 1800 млн. непарных чтений.

4.1 Очистка и демультимплексирование полученных прочтений

Биоинформатическая обработка результатов секвенирования проводилась на высокопроизводительных серверах лаборатории лесной геномики СФУ.

Полученные данные ddRAD секвенирования проходили несколько этапов начальной обработки. Прочтения (длиной 100 н.о.) были отфильтрованы и обрезаны согласно показателям качества. Каждая последовательность проверялась на наличие сайта узнавания рестриктазы EcoRI. Основываясь на баркодированных последовательностях EcoRI-адаптеров, уникальных для каждого образца, было произведено демультимплексирование, т.е. разделение данных на отдельные образцы в библиотеке. Все операции проводились с помощью, входящей в ПО Stacks программы process_radtags, статистика для первоначальных данных была получена с помощью программы FastQC [55]. Длина каждого прочтения после обработки составила 92 н.о.; Phred score после обработки в среднем равнялся 36, что указывает на точность определения нуклеотидов при секвенировании более 99,9%. Статистика прочтений до и после фильтрации представлена в таблице 3.

Изъято 8 страниц

ЗАКЛЮЧЕНИЕ

Проведено полногеномное секвенирование ddRADseq библиотек пяти выборок лиственницы сибирской. По результатам секвенирования было получено более 1800 млн. непарных чтений.

Осуществлена первичная биоинформатическая обработка данных (очистка и фильтрация по качеству сырых данных секвенирования). Также проведено картирование на референсный геном.

В результате фильтрации данных был получен массив из 151 441 однонуклеотидного полиморфизма (SNPs). На его основе были рассчитаны значения основных показателей генетического полиморфизма - наблюдаемой и ожидаемой гетерозиготности ($H_O = 0,1991 \pm 0,0200$; $H_E = 0,2149 \pm 0,0013$). Полученные значения свидетельствуют о невысоком уровне разнообразия исследованных выборок по сравнению с аналогичными данными для других видов хвойных. У четырех из пяти выборок (Биджа, Боград, Камызяк, Туим) был отмечен избыток гетерозиготных генотипов.

Анализ популяционной структуры лиственницы сибирской с помощью F-статистик Райта показал, что более 97,6 % генетической изменчивости реализуется внутри выборок. На долю межпопуляционной составляющей приходится менее 3% ($F_{ST} = 2,4\%$), что свидетельствует о слабой дифференциации популяций.

С помощью программ BayeScan и OutFLANK выполнен поиск генетических маркеров адаптации. В результате чего выявлены 1236 и 829 соответственно генов-аутлаеров, которые в дальнейшем будут использованы для поиска ассоциаций с дендрофенотипами.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Maiti R. Autoecology and ecophysiology of woody shrubs and trees: Concepts and applications / Ratikanta Maiti, Humberto González Rodríguez, Natalya Sergeevna Ivanova // John Wiley & Sons, 2016. – P. 285 – 311. – doi: 10.1002/9781119104452.
2. Рысин Л. П. Лиственничные леса России / Л. П. Рысин. – Москва: Товарищество научных изданий КМК, 2010. – 344 с.
3. Воскобойникова И. В. Дендрология / И. В. Воскобойникова. – Москва, Берлин: Директ-Медиа, 2020. – 140 с.
4. Проскурякова Г. М. Лиственница / Г. М. Проскурякова // Наука и Жизнь. — 1987. — № 11. — С. 158-161.
5. Абаимов, В. Ф. Дендрология: учеб. пособие для студ. высш. учеб. заведений / В. Ф. Абаимов. — 3-е изд., перераб. — Москва: Издательский центр «Академия», 2009. — 368 с.
6. Губанов И. Иллюстрированный определитель растений Средней России / И. А. Губанов. – Москва: Товарищество научных изданий, 2002. — Т. 1. — 527 с.
7. Кашин В. И. Лиственничные леса Европейского севера России / В. И. Кашин, А. С. Козобродов. – Архангельск: АФРГО РАН, 1994. — 215 с.
8. Белосельская З. Г. Деревья и кустарники СССР / З. Г. Белосельская, ред. тома С. Я. Соколов и Б. К. Шишкин. — М.; Л.: Изд-во АН СССР, 1949. — Т. I. Голосеменные. — 464 с.
9. Abaimov A. P. Geographical Distribution and Genetics of Siberian Larch Species / A. P. Abaimov // Permafrost Ecosystems. – 2009. – V. 209. – P. 41-58. – doi: 10.1007/978-1-4020-9693-8_3.
10. Милютин, Л. И. Биоразнообразие лиственниц России / Л. И. Милютин // Хвойные бореальной зоны. — 2003. — № 1. — С. 6–9.
11. Губанов И. А. Дикорастущие полезные растения СССР / И. А. Губанов, И. Л. Крылова, В. Л. Тихонова // Москва: Мысль. – 1976. – 360 с.

12. Mamet D. Shifting global *Larix* distributions: Northern expansion and southern retraction as species respond to changing climate / Steven D. Mamet, Carissa D. Brown, Andrew J. Trant, Colin P. Laroque // *Journal of Biogeography*. – 2018. – V. 46. – P. 30-44. – doi: 10.1111/jbi.13465.
13. Лучник, З. И. Интродукция деревьев и кустарников в Алтайском крае / З. И. Лучник. — М.: Колос, 1970. — С. 68-70.
14. Крюссман, Г. Хвойные породы / Г. Крюссман; пер. с нем. Н. Н. Непомнящего; под ред. Н. Б. Гроздовой. — Москва: Лесная промышленность, 1986. — 256 с.
15. Коропачинский, И. Ю. Древесные растения Азиатской России / И. Ю. Коропачинский, Т. Н. Встовская // Новосибирск: Изд-во СО РАН «Гео». – 2002. — 707 с.
16. Дылис Н. В. Сибирская лиственница / Н. В. Дылис. – Москва: Московское общество испытателей природы, 1947. – 137 с.
17. Ирошников А.И. Проблема селекции и внутривидовой дифференциации лесных древесных растений в работах В.Н.Сукачева / А. И. Ирошников // Проблемы лесной биогеоценологии. – 1980. – С. 15-33.
18. Михайленко М.А. Развитие и строение покровных тканей у лиственниц / М.А. Михайленко, Р.И. Дерюжкин // Вопросы анатомо-морфологических и эколого-физиологических особенностей строения растений. – 1970. - Т.112. - С.33-45.
19. Simak M. Karyotype analysis of Siberian larch (*Larix sibirica* and *Larix sukaczewii*) / M. Simak // *Studia Forestalia Suesica*. – 1964. – № 17. – P. 15.
20. Милютин Генетико-таксономический анализ популяций лиственниц сибирской и Сукачева / Л. И. Милютин, Е. Н. Муратова, А.Я. Ларионова // Лесоведение. – 1993. – № 5. – С.55-63.
21. Barrett R. Molecular spandrels: tests of adaptation at the genetic level / Rowan Barrett, Hopi Hoekstra // *Nature Reviews Genetics*. – 2011. – V. 12. – P. 767–780. – doi: 10.1038/nrg3015.

22. Emerson K. Resolving postglacial phylogeography using high-throughput sequencing / Kevin Emerson, Clayton Merz, Julian Catchen, Paul Hohenlohe, William Cresko, William Bradshaw, Christina Holzapfel // *Proc Natl Acad Sci U S A.* – 2010. – V. 107. – № 37. – P. 16196-200. – doi: 10.1073/pnas.1006538107.
23. Davey J. RADSeq: next-generation population genetics / John W. Davey, Mark L. Blaxter // *Briefings in Functional Genomics.* – 2010. – V. 9. – I. 5-6. – P. 416–423. – doi:10.1093/bfgp/elq031.
24. Wang S. 2b-RAD: a simple and flexible method for genome-wide genotyping / Shi Wang, Eli Meyer, John McKay, Mikhail Matz // *Nat Methods.* – 2012. – V. 9. – P. 808-810. – doi: 10.1038/nmeth.2023.
25. Toonen R. ezRAD: a simplified method for genomic genotyping in non-model organisms / Robert Toonen, Jonathan Puritz, Zac Forsman, Jonathan Whitney, Iria Fernandez-Silva, Kimberly Andrews, Christopher Bird // *PeerJ.* – 2013. – V. 19. – doi: 10.7717/peerj.203.
26. Puritz J. Demystifying the RAD fad / Jonathan Puritz, Mikhail Matz, Robert Toonen, Jesse Weber, Daniel Bolnick, Christopher Bird // *Molecular Ecology.* – 2014. – V. 23. – P. 5937–5942. – doi: 10.1111/mec.12965.
27. Peterson B. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species / Brant Peterson, Jesse Weber, Emily Kay, Heidi Fisher, Hopi Hoekstra // *PLoS ONE.* – 2012. – V. 7. – P. 11. – doi: 10.1371/journal.pone.0037135.
28. Baird N. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers / Nathan Baird, Paul Etter, Tressa Atwood, Mark Currey, Anthony Shiver, Zachary Lewis, Eric Selker, William Cresko, Eric Johnson // *PLoS ONE.* – 2008. – V. 3. – P. 7. – doi: 10.1371/journal.pone.0003376.
29. Salas-Lizana R. Double-digest RADseq loci using standard Illumina indexes improve deep and shallow phylogenetic resolution of *Lophodermium*, a widespread fungal endophyte of pine needles / Rodolfo Salas-Lizana, Ryoko Oono // *Ecology and Evolution.* – 2018. – V. 8. – P. 6638-6651. – doi: 10.1002/ece3.4147.

30. Miller M. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers / Michael Miller, Joseph Dunham, Angel Amores, William Cresko, Eric Johnson // *Genome Res.* – 2007. – V. 17. – P. 240-248. – doi: 10.1101/gr.5681207.

31. Andrews K. Recent novel approaches for population genomics data analysis / Kimberly Andrews, Gordon Luikart // *Mol Ecol.* – 2014. – V. 23. – P. 1661–1667. – doi: 10.1111/mec.12686.

32. Arnold B. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling / B. Arnold, R. Corbett-Detig, D. Hart, K. Bomblies // *Mol Ecol.* – 2013. – V. 22. – P. 3179-3190. – doi: 10.1111/mec.12276.

33. Karam M-J. Genomic exploration and molecular marker development in a large and complex conifer genome using RADseq and mRNAseq / M-J. Karam, F. Lefèvre, M. Bou Dagher-Kharrat, S. Pinosio, G. G. Vendramin // *Mol Ecol Resour.* – 2015. – V. 15. – P. 601-612. – doi: 10.1111/1755-0998.12329.

34. Neale D. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies / David B. Neale, Jill L. Wegrzyn, Kristian A. Stevens, Aleksey V. Zimin, Daniela Puiu, Marc W. Crepeau // *Genome Biol.* – 2014. – V. 15. – P. 13. – doi:10.1186/gb-2014-15-3-r59.

35. Zimin A. Sequencing and assembly of the 22-gb loblolly pine genome / Aleksey Zimin, Kristian Stevens, Marc Crepeau, Ann Holtz-Morris, Maxim Koriabine // *Genetics.* – 2014. – V. 196. – P. 875-890. – doi: 10.1534/genetics.113.159715.

36. Crepeau M. From Pine Cones to Read Clouds: Rescaffolding the Megagenome of Sugar Pine (*Pinus lambertiana*) / Marc Crepeau, Charles Langley, Kristian Stevens // *G3 (Bethesda).* – 2017. – V. 7. – P. 1563-1568. – doi: 10.1534/g3.117.040055.

37. Jin Y. Genome-Wide Variant Identification and High-Density Genetic Map Construction Using RADseq for *Platycladus orientalis* (*Cupressaceae*) / Yuqing Jin, Wei Zhao, Shuai Nie, Si-Si Liu, Yousry A. El-Kassaby, Xiao-Ru Wang, and Jian-Feng Mao // *G3 Genes|Genomes|Genetics.* – 2019. – V. 9. – P. 3663–3672. – doi: 10.1534/g3.119.400684.

38. Cai M. Population Genetic Diversity and Structure of Ancient Tree Populations of *Cryptomeria japonica* var. *sinensis* Based on RAD-seq Data / Mengying Cai, Yafeng Wen, Kentaro Uchiyama, Yunosuke Onuma, Yoshihiko Tsumura // Forests. – 2020. – V. 11. – P. 16. – doi:10.3390/f11111192.

39. Sakaguchi S. High-throughput linkage mapping of Australian white cypress pine (*Callitris glaucophylla*) and map transferability to related species / Shota Sakaguchi, Takeshi Sugino, Yoshihiko Tsumura, Motom Ito, Michael Crisp, David M. // Tree Genetics & Genomes. – 2015. – V. 11. – P. 12. – doi: 10.1007/s11295-015-0944-0.

40. Menon M. The role of hybridization during ecological divergence of southwestern white pine (*P. strobiformis*) and limber pine (*P. flexilis*) / Mitra Menon, Justin Bagley, Christopher Friedline, Amy Whipple, Anna Schoettle // Molecular Ecology. – 2018. – V. 27. – P. 1245-1260. – doi: 10.1111/mec.14505.

41. Friedline C. The genetic architecture of local adaptation I: the genomic landscape of foxtail pine (*Pinus balfouriana* Grev. & Balf.) as revealed from a high-density linkage map / Christopher Friedline, Brandon Lind, Erin Hobson, Douglas Harwood, Annette Delfino, Patricia Maloney, Andrew Eckert // Tree Genetics & Genomes. – 2015. – V. 11. – P. 46. – doi: 10.1007/s11295-015-0866-x.

42. Lind B. Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA / Brandon Lind, Christopher Friedline, Jill Wegrzyn, Patricia Maloney, Detlev Vogler, David Neale, Andrew Eckert // Molecular Ecology. – 2017. – V. 26. – P. 3168-3185. – doi: 10.1111/mec.14106.

43. Parchman T. Genome-wide association genetics of an adaptive trait in lodgepole pine / Thomas Parchman, Zachariah Gompert, Joann Mudge, Faye Schilkey // Molecular Ecology. – 2012. – V. 21. – P. 2991-3005. – doi: 10.1111/j.1365-294X.2012.05513.x.

44. Pan J. Optimization of the genotyping-by-sequencing strategy for population genomic analysis in conifers / Jin Pan, Baosheng Wang, Zhi-Yong Pei, Wei Zhao, Jie

Gao, Jian-Feng Mao, Xiao-Ru Wang // *Mol Ecol Resour.* – 2015. – V. 15. – P. 711-722. – doi: 10.1111/1755-0998.12342.

45. Chen C. Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform / Charles Chen, Sharon E. Mitchell, Robert J. Elshire, Edward S. Buckler, Yousry A. El-Kassaby // *Tree Genetics & Genomes.* – 2013. – V. 9. – P. 1537–1544. – doi: 10.1007/s11295-013-0657-1.

46. Johnson J. Growth and Its Relationship to Individual Genetic Diversity of Mountain Hemlock (*Tsuga mertensiana*) at Alpine Treeline in Alaska: Combining Dendrochronology and Genomics / Jeremy S. Johnson, Parveen K. Chhetri, Konstantin V. Krutovsky, David M. Cairns // *Forests.* – 2017. – V. 8. – doi: 10.3390/f8110418.

47. Корчагин А.А. Внутривидовой (популяционный) состав растительных сообществ и методы его изучения / А.А. Корчагин // *Полевая геоботаника.* – М.-Л.:Наука, 1966. – Т. 3.

48. Завадский К.М. Вид и Видообразование / К.М. Завадский // *Л.:Наука,* 1968. – 404 с.

49. Catchen J. Stacks: an analysis tool set for population genomics / Julian Catchen, Paul A. Hohenlohe, Susan Bassham, Angel Amores, William A. Cresko // *Molecular Ecology.* – 2013. – V. 22. – P. 3124-3140. – doi: 10.1111/mec.12354.

50. Purcell Sh. PLINK / Shaun Purcell // GNU General Public License. – 2006. – V. 2.

51. Excoffier L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows / L. Excoffier, L. Lischer // *Molecular Ecology Resources.* – 2010. – doi: 10: 564-567.

52. Excoffier L. Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data / Laurent Excoffier, Peter E. Smouse, Joseph Michael // *Genetics.* – 1992. – V. 131. – P. 479-491. – doi: 10.1093/genetics/131.2.479.

53. Foll M. A genome scan method identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective / M. Foll, O.E. Gaggiotti // *Genetics*. – 2008. – V. 180. – P. 977-993. – doi: 10.1534/genetics.108.092221.

54. Whitlock M. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F_{ST} / Michael Whitlock, Katie Lotterhos // *Am Nat*. – 2015. – V. 186. – doi: 10.1086/682949.

55. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (дата обращения 29.03.21)

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт Фундаментальной Биологии и Биотехнологии
Кафедра геномики и биоинформатики

УТВЕРЖДАЮ

Заведующий кафедрой

 И.Е.Ямских

подпись инициалы, фамилия

« 21 » июня 20 21 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

«Изучение генетической адаптации в популяциях лиственницы сибирской с
использованием данных полногеномного генотипирования»

06.04.01 Биология

06.04.01.06 Геномика и биоинформатика

Научный руководитель


подпись, дата

доцент, к.б.н.
должность, ученая
степень

Орешкова Н.В.
инициалы, фамилия

Выпускник


подпись, дата

Тимофеева Д.Б.
инициалы, фамилия

Красноярск 2021