

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт фундаментальной биологии и биотехнологии

Базовая кафедра защиты и современных технологий мониторинга лесов

УТВЕРЖДАЮ  
Заведующий кафедрой

подпись \_\_\_\_\_  
инициалы, фамилия  
« \_\_\_\_ » 20 \_\_ г.

## МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**Экспрессия генов у сосны сибирской кедровой (*Pinus sibirica* Du Tour.) и  
сосны обыкновенной (*Pinus sylvestris* Ledeb.) с аномальным  
морфогенезом кроны**

Направление подготовки 04.06.01 – Биология, по профилю 06.04.01.06 –  
Геномика и биоинформатика

Научный руководитель \_\_\_\_\_ д-р физ.-мат. наук М. Г. Садовский  
подпись, дата  
Выпускник \_\_\_\_\_ С. В. Новикова  
подпись, дата  
Рецензент \_\_\_\_\_ в.н.с., канд. биол. наук, Д.А. Афонников  
подпись, дата должность, ученая степень инициалы, фамилия

Красноярск 2018

## АННОТАЦИЯ

В настоящее время механизмы возникновения «ведьминых метел» непатогенного типа остаются невыясненными. Предположительно, такие участки кроны со слабо выраженным верхушечным доминированием и обильным ветвлением являются фенотипическим проявлением соматических мутаций в вегетативных почках растения. Ключевая задача работы – найти и охарактеризовать гены с различием в уровне экспрессии у «ведьминых метел» и участков нормальной кроны у двух хвойных видов: сосны сибирской кедровой и сосны обыкновенной. Был проведен поиск точечных мутаций в транскриптах «ведьминых метел» с использованием подхода для поиска однонуклеотидных полиморфизмов.

Для пяти деревьев было секвенировано 20 образцов матричной РНК, собраны транскриптомы сосен. Для сосны сибирской кедровой найдено 10 генов с различиями в уровне экспрессии, для сосны обыкновенной – 7 генов. Также найдено 5 однонуклеотидных полиморфизмов, характерных для образцов «ведьминых метел».

**Ключевые слова:** *апикальный рост, ведьмины метлы, дифференциальная экспрессия генов, однонуклеотидные полиморфизмы, точечная мутация, транскриптом.*

## **СОДЕРЖАНИЕ**

<b>ВВЕДЕНИЕ .....</b>	<b>4</b>
<b>ОСНОВНАЯ ЧАСТЬ.....</b>	<b>7</b>
1.Аномальный морфогенез кроны у сосновых.....	7
1.1.Паразитарные ВМ .....	7
1.2.Непаразитарные ВМ .....	8
<b>МАТЕРИАЛЫ И МЕТОДЫ .....</b>	<b>12</b>
1.Получение данных.....	12
2.Фильтрация прочтений.....	13
3.Получение и модификации транскриптомных сборок .....	15
4.Экспрессия генов.....	17
5.Поиск однонуклеотидных полиморфизмов (ОНП) .....	19
<b>РЕЗУЛЬТАТЫ.....</b>	<b>21</b>
1.Транскриптомные сборки.....	21
2.Анализ дифференциальной экспрессии.....	22
3.Поиск однонуклеотидных полиморфизмов в транскриптомных данных сосны обыкновенной .....	26
<b>ЗАКЛЮЧЕНИЕ .....</b>	<b>28</b>
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....</b>	<b>30</b>

## ВВЕДЕНИЕ

Диссертационная работа посвящена изучению дифференциальной экспрессии генов у сосны сибирской кедровой (*Pinus sibirica* Du Tour.) и сосны обыкновенной (*Pinus sylvestris* Ledeb.) с аномальным морфогенезом кроны. Была предпринята попытка найти генетические механизмы, обуславливающие образование «ведьминых метел» (ВМ) – фрагментов кроны растения с аномальным морфогенезом. Предположительно, такие локальные системы ветвления с замедленным ростом и интенсивным ветвлением могут возникать ввиду соматических мутаций в меристеме побегов растения.

Цель работы — поиск дифференциально экспрессируемых генов и нуклеотидных замен, потенциально связанных с аномальным морфогенезом кроны у сосны сибирской и сосны обыкновенной. Для достижения цели были поставлены следующие задачи:

- получить транскриптомные сборки сосны сибирской кедровой и сосны обыкновенной;
- выявить гены с различиями в уровне экспрессии между исследуемыми образцами и описать закономерности взаимодействия таких генов;
- провести поиск точечных мутаций, коротких вставок и делеций, которые могут привести к отклонениям в процессе морфогенеза.

В настоящее время селекция хвойных пород в России осуществляется по нескольким направлениям: на быстроту роста и продуктивности лиственницы, ели, сосны, пихты и можжевельника; на смолопродуктивность сосны; на урожайность кедровых сосен; на резонансную способность (звуковые качества) древесины ели; на декоративность и устойчивость к промышленным выбросам сосны, ели, лиственницы, пихты, можжевельника, туи и др.

Перспективный и многообещающий путь — поиск соматических мутаций у хвойных. Характерные признаки ВМ, замедленный рост и обильное ветвление не являются полезными для селекции хвойных как источника

древесины, но фенотип ВМ привлекателен для селекции на урожайность. Так как кедровые сосны являются одним из важнейших орехоплодных видов, актуальна их селекция на низкорослость и скороплодность. ВМ незаменимы как исходный генетический материал для таких селекционных программ, ведь некоторые такие деревья обладают ценными признаками: высокой жизнеспособностью, замедленным апикальным ростом, скороплодностью, обильным плодоношением и нормальным качеством шишек. Многообразие форм «ведьминых метел» дает селекционерам большие возможности для выведения различных декоративных сортов хвойных, которые успешно выращиваются в ботанических садах и используются в ландшафтном дизайне.

Исследователи отмечают важность понимания механизмов образования мутационных ВМ для дальнейшей селекции и формирования банка клонов, однако уровень изученности мутационных «ведьминых метел» до сих пор остается крайне низким.

Кроме того, апикальное доминирование является характерной чертой деревьев и изучение механизмов его определяющих является фундаментальной биологической проблемой. В связи с этим ВМ являются прекрасной моделью для изучения этих механизмов.

В данной работе помимо установления структуры различий между нормальной кроной (НК) и ВМ, а также структуры разнообразия последних, с помощью оценки дифференциальной экспрессии генов в парах ВМ/НК была предпринята попытка установить, в какой функциональной части генома произошло нарушение, приводящее к формированию ВМ. Секвенирование транскриптома дает наиболее полную информацию о содержании транскриптов в клетках. Белки, синтезирующиеся благодаря РНК, отвечают за формирование признаков анатомии и физиологии организма, поэтому расшифровка транскриптома является первым ключевым шагом к пониманию природы мутационных ВМ.

Для исследования в качестве объектов были выбраны клоны двух видов, ВМ которых существенно различаются между собой — сосна сибирская

кедровая и сосна обыкновенная. Возможно, в формировании фенотипа ВМ у этих видов могут быть задействованы разные, хотя и подобные по проявлению механизмы. Сравнительный анализ дифференциальной экспрессии генов может показать, произошли ли изменения на одном участке генома, или же это соматическая мутация имеет однотипную природу.

## **ОСНОВНАЯ ЧАСТЬ**

### **1. Аномальный морфогенез кроны у сосновых**

«Ведьмины метлы» (ВМ) – фрагменты кроны (локальные системы ветвления) с аномальным морфогенезом – подавлением апикального роста и доминированием латерального роста ветвей – встречаются у многих видов растений, преимущественно у древесных форм. Обильное ветвление приводит к образованию множества укороченных ветвей, которые часто формируют плотные скопления – бесформенные или в виде шара. В настоящее время в фитопатологии выделяют два типа ВМ: паразитарные (патогенные) и мутационные (непатогенные), различающиеся по причинам возникновения, характеру ветвления и типу распространения в лесных экосистемах. Хотя механизмы в основе формирования этих типов не являются взаимоисключающими. Теоретически, патогенные факторы могут приводить к соматическим как эпигенетическим так и генетическим мутациям генов, связанных с апикальным и латеральным ростом ветвей.

#### **1.1. Паразитарные ВМ**

ВМ могут возникать на растении в результате жизнедеятельности фитоплазм, микоплазм, ржавчинных грибов, аскомицетов, грибов рода *Taphrina*, вирусов и даже других растений, паразитирующих на деревьях (карликовая омела) [1]. Следующие особенности характеризуют паразитарные ВМ: болезненный вид, угнетение репродуктивной функции, очаговый характер распространения: на пораженном дереве имеется несколько ВМ, а такие деревья располагаются группами.

На протяжении многих лет приоритетным считается изучение патогенных ВМ, так как из-за заражения агрокультур страдает урожайность.

Исследования паразитирующих видов и механизмов их влияния на растения ведутся давно, однако развитие геномики вывело эти исследования на новый уровень. Геномы различных патогенных грибов и бактерий, вызывающих образование ВМ сейчас активно исследуются, был секвенирован митохондриальный геном патогенного гриба *Moniliophthora roreri* [2], изучаются метаболические пути патогенов, эволюция плазмид бактерий [3]; ведется поиск генов, ответственных за патогенность [4]. Исследуются организмы, позволяющие осуществлять биоконтроль над распространением ВМ, вызываемых базидиомицетами – *Clonostachys rosea* и *Trichoderma spp.* [4, 5].

Отмечается [6], что проявление многих симптомов у зараженных бактериями и грибами растений возникает в результате взаимодействия патогенов с некоторыми фитогормонами, в то время как другие фитогормоны являются частью иммунного ответа растения на эти патогены. В частности, некоторые фитогормоны стимуляторы роста вызывают развитие симптомов патогенной ВМ. Например, ауксины не только стимулируют деление и рост здоровых клеток, но и становятся причиной образования опухолей при бактериальных инвазиях [6].

## 1.2. Непаразитарные ВМ

Такие новообразования возникают предположительно из-за мутационных или эпимутационных изменений в меристеме побегов. От паразитарных ВМ их отличает нормальная жизнеспособность, высокая долговечность, спорадическое распространение, отсутствие патогенов или следов их жизнедеятельности. ВМ мутационного типа у хвойных видов встречается приблизительно у одного дерева на 5000 [7].

Первые упоминания ВМ мутационного типа у европейских видов сосны, ели и лиственницы появились в конце XIX века в Германии [8]. В этих работах

ВМ описываются как вполне жизнеспособные и относительно долговечные фрагменты кроны с обильным ветвлением и укороченными побегами.

Предположение о том, что изменения, вызывающие образование таких ВМ происходят, в первую очередь, на геномном уровне, в современной научной литературе по исследованию генетических закономерностей этого процесса не имеет строгого подтверждения. Большинство работ направлено на изучение морфологических и анатомических особенностей ВМ мутационного типа – ввиду высокой вариативности проявления признаков. Сравнительный морфологический анализ ВМ и образца нормальной кроны (НК) того же дерева у сосны сибирской кедровой показал, насколько отличны образцы ВМ с разных деревьев – от полностью нежизнеспособных ввиду высокой плотности до отличающихся от НК лишь скоростью образования новых осей ветвлений. Изучаемые образцы были взяты из как можно более сходных условий произрастания, чтобы исключить неравномерность влияния внешних факторов. Образцы были попарно (ВМ и НК того же дерева) привиты пятилетним саженцам кедра в 2017 году, и после восьми лет произрастания, с учетом на высоту прививки, диаметр ствола в месте прививки, ширину кроны и порядок ветвлений измерялись ежегодно следующие параметры:

1. Длина самого высокого терминального побега.
2. Максимальная длина бокового побега.
3. Коэффициент верхушечного доминирования (отношение самого длинного терминального побега к самому длинному боковому).
4. Длина самого короткого побега с хотя бы одной почкой.
5. Длина самого короткого побега с дихотомическим ветвлением.
6. Средняя длина хвои.

В результате удалось выделить значения параметров, при которых ВМ можно достоверно отличить от НК. «Ведьмины метлы» были классифицированы по выраженности этих признаков [9].

Пятилетние клоны ВМ не образовывали мужских шишек, что можно отнести к отклонениям от нормальной работы репродуктивной системы кедра,

однако на шести из десяти образцов формировались женские шишки, в то время как на клонах НК образовалась только одна шишка.

Сравнение анатомических характеристик ВМ и НК у пихты сибирской (*Abies sibirica* Ledeb.) показало, что в клетках хвои ВМ более интенсивно протекают процессы метаболизма. В связи с этим, такая хвоя короче, шире и толще, чем у НК, а масса увеличена в среднем на 20 %. Анатомические изменения устьичных аппаратов обусловлены стремлением уменьшить поверхность испарения воды, увеличены оба слоя мезофильных клеток, более развита транспортная система [10]. Исследования хвои кедра показали сходные результаты, авторы также подчеркивают статистически значимую зависимость исследуемых особенностей хвои от общей площади смоляных ходов (трахеид) и ширины клеток эндодермы у ВМ и отсутствие такой зависимости у НК [11]. Таким образом, вероятно, мутация влияет на полярный транспорт или рецепцию ауксинов, что снижает апикальное доминирование и вызывает изменения в ксилогенезе – в годичном кольце закладывалось большее число трахеид, длина трахеид уменьшалась, но увеличивались диаметр трахеид, площадь просвета и толщина клеточной стенки. Примечательно, что только ВМ кроны сосны алеппской (*Pinus halepensis*) имеют хвою, более длинную, чем хвоя нормальной кроны того же дерева [12].

Цитогенетическое изучение различных популяций сосны обыкновенной, включающее в себя кариологический анализ и анализ хромосомных нарушений, показало интересные результаты [13]. В кариотипе сосны обыкновенной имеется 24 хромосомы ( $2n = 24$ ); данный вид является диплоидом; длина одной хромосомы составляет 10-18 мкм. У сосны обыкновенной с «ведьминой метлой» число геномных и хромосомных мутаций выше, чем у нормальных деревьев на болоте [14]. Деревья с «ведьминой метлой» в отдельных случаях содержали триплоидные и тетраплоидные клетки. В митозе отмечены отстающие и хаотически расходящиеся хромосомы. Сосна с «ведьминой метлой» отличается еще большей активностью ядрышкообразующих зон, выражющейся в увеличении

числа вторичных перетяжек до 6 на хромосому, а также нарушением структуры и функций ядрышек в интерфазных ядрах. Были обнаружены аномальные формы ядрышек и так называемое «остаточное ядрышко» в метафазе митоза.

Был проведен всесторонний анализ роста и генеративного развития нескольких десятков ВМ кедра сибирского, а также их вегетативного и семенного потомства в сравнении с потомством нормальной части кроны тех же деревьев. При размножении прививкой характерные для «ВМ» свойства (специфические рост, морфогенез и плодоношение) полностью сохраняются. В семенном потомстве ВМ наблюдается расщепление сеянцев на 2 класса: нормальные растения и растения с замедленным ростом и интенсивным ветвлением в соотношении 1:1. При этом нормальная половина потомства «ВМ» ничем не отличается от потомства нормальной кроны того же дерева [7]. На основании данных о распределении фенотипических проявлений в потомстве, было сделано предположение, что ВМ представляют собой доминантную соматическую мутацию.

Возможно, морфогенез, приводящий к образованию ВМ, обусловлен мутациями в генах фитогормонов, отвечающих за нормальное развитие растений. Так, например, у карликовой формы арабидопсиса (*Arabidopsis thaliana*), найдена мутация в гене, кодирующем трансмембранный белок AtPGP1, вовлеченный в транспорт ауксинов. Мутантный аллель подавляет транспорт этого фитогормона [15].

## **МАТЕРИАЛЫ И МЕТОДЫ**

### **1. Получение данных**

Образцы тканей почек сосны и кедра были предоставлены Институтом мониторинга климатических и экологических систем СО РАН (г. Томск). Исследуемые клонды были созданы д.б.н. С. Н. Горошкевичем вместе с сотрудниками с помощью прививок в 2007 г. и выращены в рядах на научном стационаре «Кедр» ИМКЭС СО РАН, источниками материала для которых были деревья с мутационными ВМ, происходящие из двух районов Западной Сибири.

Сбор материала проводился для каждого дерева в мае и сентябре 2017 года, на участках дерева с аномальным морфогенезом (ВМ) и на морфологически нормальных участках (НК). Всего было собрано 20 образцов: 8 для двух деревьев сосны обыкновенной и 12 для трех деревьев сосны сибирской кедровой. Ниже в таблице 1 представлены условные названия образцов в соответствии с типом ткани и датой сбора.

Таблица 1 – Информация об исследуемых образцах

<b>Образец</b>	<b>Тип строения</b>	<b>Дерево</b>	<b>Дата сбора</b>
<b>Сосна обыкновенная <i>P. sylvestris</i></b>			
010	Аномальное ветвление (ВМ)	Дерево 1	Май 2017
011	Участок нормальной кроны (НК)		Сентябрь 2017
010.2	Аномальное ветвление (ВМ)		
011.2	Участок нормальной кроны (НК)		
014	Аномальное ветвление (ВМ)	Дерево 2	Май 2017
015	Участок нормальной кроны (НК)		Сентябрь 2017
014.2	Аномальное ветвление (ВМ)		
015.2	Участок нормальной кроны (НК)		
<b>Сосна сибирская кедровая <i>P. sibirica</i></b>			
032	Аномальное ветвление (ВМ)	Дерево 1. Рыхлая метла	Май 2017
033	Участок нормальной кроны (НК)		Сентябрь 2017
032.2	Аномальное ветвление (ВМ)		
033.2	Участок нормальной кроны (НК)		

036	Аномальное ветвление (ВМ)	Дерево 2. Плотная метла	Май 2017
037	Участок нормальной кроны (НК)		
036.2	Аномальное ветвление (ВМ)		Сентябрь 2017
037.2	Участок нормальной кроны (НК)		
054	Аномальное ветвление (ВМ)	Дерево 3. Метла средней плотности	Май 2017
055	Участок нормальной кроны (НК)		
054.2	Аномальное ветвление (ВМ)		Сентябрь 2017
055.2	Участок нормальной кроны (НК)		

Секвенирование парно концевых библиотек (paired-end) осуществлялось на приборах Illumina HiSeq (17 образцов), Illumina MiSeq (3 образца), длина вставки составила 120 н.о.

## 2. Фильтрация прочтений

Полученные сырье парные риды проходили несколько этапов очистки и фильтрации. Удаление адаптеров, фильтрация по качеству, обрезка участков ридов с низким качеством и последующая фильтрация по минимальной длине проводились ПО Trimmomatic [16]. Сопутствующая процессу визуализация статистик была сделана при помощи ПО FastQC [17].

Далее проводилась проверка транскриптомных ридов на наличие митохондриальных, хлоропластных и рибосомальных примесей. Согласно методике извлечения матричной РНК за поли-А хвосты с использованием магнитных бусин, в качестве примесей после экстракции в образце может содержаться до 40% рибосомальной РНК [18]. Для очистки ридов была использована база рибосомальной РНК Silva [19], содержащая информацию о нуклеотидных последовательностях малой и большой субъединиц рРНК для архей, бактерий и эукариот. На момент последнего релиза Silva содержит около 7 млн последовательностей рРНК. В рамках исследования эта база была дополнена последовательностями рРНК хвойных, найденными в базе NCBI. Проверка прочтений на наличие органелльного генетического материала проводилась с использованием RefSeq Organelle Genome Resources NCBI [20] – базы данных, содержащей транскриптомы митохондрий и пластид.

Парные прочтения, отдельно для каждого образца, были картированы на полученные базы с использованием ПО Bowtie2 [21]. Фильтрация проходила в два этапа: сначала нами были отобраны парные риды, не имеющие аналогов в органелльной базе, затем эти риды использовались для картирования на базу рРНК. На обоих этапах фильтрации сохранялись только те пары ридов, которые не выровнялись на базу конкордантно, т.е. ни прямое, ни обратное прочтение не имеет идентичной последовательности в базе. В результате, около 55% прочтений были убраны из дальнейшего анализа как возможные примеси: в среднем, 23% ридов не прошли фильтрацию как органелльные транскрипты и 32% как фрагменты рибосомальной РНК (Таблица 2).

Таблица 2 – Результаты фильтрации прочтений

Образец	Количество прочтений			Осталось прочтений, %
	До фильтрации	После фильтрации органелльной РНК	После фильтрации рРНК	
010	4 915 882	4 142 660	2 445 082	49.7
011	4 813 001	3 906 232	2 356 305	49.0
010.2	93 702 967	76 202 257	43 880 751	46.8
011.2	34 410 083	24 770 317	14 695 172	42.7
014	2 745 771	2 295 272	1 468 096	53.5
015	3 068 852	2 208 815	952 583	31.0
014.2	3 871 235	2 366 039	817 850	21.1
015.2	75 713 232	58 667 377	33 705 176	44.5
032	25 373 965	16 755 968	8 601 099	33.9
033	61 551 579	40 387 039	12 567 981	20.4
032.2	4 610 092	2 965 970	2 097 731	45.5
033.2	4 168 894	3 176 295	2 030 765	48.7
036	2 726 968	2 409 483	1 503 842	55.1
037	2 676 144	2 132 523	1 425 885	53.3
036.2	87 427 824	68 499 795	36 276 122	41.5
037.2	84 694 462	68 371 216	39 836 560	47.0
055	5 638 051	4 708 394	1949863	34.6
054.2	82 659 434	63 078 469	35 176 574	42.6
055.2	91 296 159	74 541 258	49 580 639	54.3

В образце ВМ сосны сибирской кедровой 054 по результатам проверки был обнаружен высокий уровень контаминации человеческой и бактериальной РНК. Образец был исключен из дальнейшего анализа.

### **3. Получение и модификации транскриптомных сборок**

Сборка транскриптома осуществлялась программой Trinity [22, 23], версия 2.5.1. Программа Trinity широко используется для *de novo* сборки транскриптомов на данных секвенирования нового поколения. Разработчики выделяют три этапа работы программы: первый, InchWorm, собирает прочтения RNA-seq в уникальные последовательности транскриптов, часто генерируя полноразмерные транскрипты (контиги) для доминирующей изоформы. На втором этапе, Chrysalis распределяет эти контиги в кластеры и конструирует полные графы де Брейна для каждого кластера. Каждый кластер представляет собой совокупность вариантов транскрипции гена (или наборов генов с идентичными последовательностями). Затем Chrysalis разбивает набор входных прочтений на эти графы. Далее, третий этап – Butterfly обрабатывает отдельные графы параллельно, отслеживая пути, которым соответствуют риды и пары ридов, в конечном счете находя полные транскрипты для альтернативно сплайсированных изоформ и разделяя транскрипты, соответствующие парalogическим генам. Программа Trinity была использована для создания сборок транскриптомов, что, в нашем случае, является необходимым первым шагом установления величин экспрессии генов. Две сборки, одна для образцов сосны и одна для кедра, были сгенерированы программой с использованием стандартных значений параметров без дополнительной нормализации. Для проверки качества полученных транскриптомов использовалась программа BUSCO v3 [24]. BUSCO проводит количественные измерения для оценки сборки генома, набора генов и полноты транскриптома на основе эволюционно обоснованных ожиданий по содержанию универсальных однокопийных генов-ортологов,

выбранных из базы OrthoDB v9 [25]. В данном исследовании был использован набор генов-ортологов высших растений (*Embryophyta*), содержащий последовательности генов 1440 ортологических групп по 31 виду растений. BUSCO создает отчет о представленности этих групп в транскриптомной сборке: информация о количестве генов, найденных в одной копии, дуплицированных (причем программа учитывает наличие изоформ), а также представленных фрагментарно или не найденных в сборке.

Улучшение транскриптомных сборок проводилось с помощью ПО EvidentialGene, скрипт tr2aacds [26]. Эта программа работает как классификатор, распределяя последовательности транскриптов по шести группам:

- транскрипты с альтернативными изоформами,
- транскрипты без альтернативных изоформ,
- альтернативные изоформы с высоким или средним уровнем выравнивания на первичный транскрипт,
- альтернативные изоформы, имеющие фрагментированные участки, по сравнению с первичным транскриптом,
- альтернативные последовательности, отличающиеся от первичного транскрипта по показателям аминокислотного выравнивания кодирующего региона (CDS),
- короткие, неполные и фрагментированные альтернативные последовательности.

В итоге, это разделение на группы, а также результаты выравнивания на белковые базы данных и качество полученных белковых последовательностей (размер, полнота) были использованы для присвоения транскриптам значений «okay» (сохраняемые последовательности) и «drop» (последовательности, исключаемые из дальнейшего анализа). Класс «drop» содержит избыточные и неинформативные транскрипты, включая избыточные точные копии транскриптов из «okay» группы или точные фрагменты этих транскриптов и

последовательности изоформ, имеющие идентичные с первичным транскриптом кодирующие регионы.

Для оценки качества транскриптомных сборок использовались следующие классические в биоинформатике показатели: средняя и максимальная длина контига, общая длина – суммарное количество нуклеотидных оснований в сборке, GC-состав - доля гуанина и цитозина среди всех нуклеотидных оснований транскриптома, величина N50. Статистика N50 определяется как самый короткий контиг при суммировании длин минимального числа контигов для получения суммы, большей или равной половине длины сборки.

#### **4. Экспрессия генов**

Анализ дифференциальной экспрессии генов был проведен с помощью программ Trinity и Blast2GO [27]. Это многоэтапный процесс, состоящий из сборки тотального транскриптома для всех образцов, выравнивания на этот транскриптом прочтений разных образцов по отдельности, оценки избыточности транскриптов, создания матрицы со значениями встречаемости транскриптов в образцах и попарное сравнение значений для всех образцов.

В настоящее время существует несколько методов оценки избыточности транскриптов без использования геномного референса: методы на основе выравнивания ридов на транскриптомную сборку и методы без выравнивания – сравнение уровня представленности  $k$ -меров в ридах и в тотальной сборке, отличающиеся, помимо основного принципа, уровнем чувствительности и временем работы. В данном исследовании использовался метод на основе выравнивания с использованием алгоритмов программ RSEM [28] и Bowtie2. RSEM позволяет получить оценку уровней экспрессии генов и изоформ для данных RNA-seq секвенирования. Для этого, с помощью Bowtie2 риды выравниваются на транскриптом, и на основе этих выравниваний считается представленность транскрипта и доверительные интервалы.

Матрица генной экспрессии была построена при помощи скрипта `abundance_estimates_to_matrix`, входящего в пакет Trinity. При создании матрицы учитываются отношения ген-транскрипт-изоформа. Построенная матрица содержит информацию о покрытии ридами генов для каждого образца без нормализации; отдельно создаются матрицы с TPM (Transcript per million) и TMM нормализованными значениями экспрессии.

Матрица с ненормализованными данными является входным файлом на следующем этапе анализа дифференциальной экспрессии, для пакета edgeR, позволяющего выполнять дифференциальный количественный анализ заранее определенных геномных структур (genomic features) [29, 30]. Пакет реализует точные статистические методы для многогрупповых экспериментов, разработанные Робинсоном и Смитом [29], а также статистические методы, основанные на обобщенных линейных моделях [31, 32]. В ходе работы EdgeR происходит попарное сравнение всех образцов, с учетом их статуса исследуемая «группа»/«контроль». Файл с результатами содержит значения logFC – двоичный логарифм разницы в количестве встречаемости транскриптов. Положительные значения logFC указывают на высокую экспрессию транскрипта, а отрицательные на пониженную экспрессию.

Blast2GO – биоинформационическое программное обеспечение для автоматической высокопроизводительной функциональной аннотации секвенированных генетических данных. ПО использует алгоритм BLAST для идентификации последовательностей по существующим базам и последующей привязки ранее полученных функциональных аннотаций изученных последовательностей к новым. Функциональная информация предоставляется базой Gene Ontology (GO) [33]. Помимо возможностей для аннотации последовательностей, в Blast2GO реализована вся последовательность подпрограмм Trinity для анализа дифференциальной экспрессии: создание *de novo* сборки транскриптома, матриц экспрессии, анализ дифференциально экспрессируемых транскриптов.

## **5. Поиск однонуклеотидных полиморфизмов**

Для обнаружения однонуклеотидных замен, вставок и делеций, которые могли бы приводить к аномальному морфогенезу, был использован ряд программ: Bowtie2, SAMtools и BCFtools [34]. Полученные с помощью Bowtie2 файлы выравниваний в формате SAM (Sequence Alignment/Map) конвертировались сортированные в BAM-файлы программой SAMtools. Для анализа производились выравнивания на тотальные транскриптомы, для сосны обыкновенной также использовалась референсная транскриптомная сборка [35]. Программа SAMtools использовалась для обнаружения однонуклеотидных полиморфизмов (ОНП), преобразовывая эти файлы в Pileup – текстовый формат для суммирования результатов поиска ОНП в выровненных ридах на референсной последовательности, объединяя данные по всем образцам. В таком файле каждая строка представляет собой геномную позицию, состоит из:

- названия хромосомы (контига),
- координат нуклеотида,
- самого референсного нуклеотида,
- количества прочтений, выравнивающихся на данную позицию,
- нуклеотидов в данной позиции у выравнивающихся прочтений,
- качества нуклеотидов в прочтениях,
- качества выравниваний прочтений.

Информация о совпадениях, несовпадениях, вставках и делециях, направленности, начале и конце ридов закодирована в отдельном столбце. В финальном варианте файла в VCF (Variant Call Format) формате содержится также информация о генотипах и вероятностях их возникновения. Производилась фильтрация ОНП по трем параметрам:

1. Минимальное среднее квадратичное для качества выравниваний на ОНП – 30.
2. Минимальное покрытие позиции прочтениями – 30.

3. Количество прочтений с альтернативными нуклеотидами на позиции – 10.

Далее нами отбирались только ОНП, вставки и делеции, распределенные по образцам так, что ВМ оказывались гетерозиготами по альтернативному аллелю, а образцы НК не отличалась от референса. Проверка ОНП на принадлежность кодирующему региону транскрипта проводилась онлайн-сервисом ORFfinder [36].

## РЕЗУЛЬТАТЫ

### 1. Транскриптомные сборки

Было получено несколько транскриптомных сборок Trinity: из прочтений до фильтрации рибосомальной и органелльной РНК, после фильтрации. Транскрипты, собранные из прошедших фильтрацию прочтений, были классифицированы ПО EvidentialGene с целью улучшения характеристик сборки: уменьшения количества транскриптов за счет удаления дубликатов и фрагментированных транскриптов, сохраняя при этом полученный уровень полноты сборки. В таблице 3 представлены основные характеристики сборок до фильтрации, а также до и после попытки улучшения.

Как видно из таблицы, после удаления большого количества прочтений у транскриптома кедра увеличилась общая длина и количество самих транскриптов, а для сосны эти показатели в целом остались без изменений. При увеличении количества транскриптов без изменений общей длины сборки мы могли бы предположить, что фильтрация вызвала «раздробление» ранее собранных транскриптов: извлечение части ридов повлияло на способность сборщика соединить участки юнигенов и в новой сборке они представлены как самостоятельные транскрипты.

Таблица 4 содержит результаты работы ПО BUSCO и отображает полноту полученных сборок до улучшения с помощью EvidentialGene и после. При небольшой потере полноты данных заметно уменьшилось количество дубликатов – с 40% до 5,5% в транскриптоме кедра и с 24% до 4% в транскриптоме сосны.

Изъято 6 страниц.

## **ЗАКЛЮЧЕНИЕ**

Были получены транскриптомные сборки сосны сибирской кедровой (впервые для данного вида) и сосны обыкновенной длиной 218 314 180 н.о. и 116 712 954 н.о. соответственно. Для сосны обыкновенной полнота сборки, согласно представленности генов-ортологов высших растений, составила 75,3%, для кедра – 77,9%, достигнут минимальный уровень дуплицированности транскриптов.

В транскриптоме сосны обыкновенной найдено 7 генов с различиями в уровне экспрессии между образцами ВМ и НК, два с пониженным уровнем экспрессии и 5 с повышенным. Была проведена структурная и функциональная аннотация этих генов, выявлены гены, ответственные за формирование ответа на патогенную инвазию.

В транскриптоме сосны сибирской кедровой было обнаружено 10 дифференциально экспрессирующихся генов, 6 с повышенным уровнем экспрессии и 4 с пониженным. Достоверно определить функции этих генов не удалось.

В образцах ВМ сосны сибирской кедровой обнаружено 5 однонуклеотидных полиморфизмов, потенциально ответственных за формирование аномальных фенотипических признаков. Из них четыре находятся в кодирующих регионах и вызывают несинонимичные замены, приводящие к изменению аминокислотного состава белка.

Проделанная работа является частью проекта ««Ведьмины метлы» мутационного происхождения у российских видов *Pinus*: характер разнообразия и молекулярно-генетическая природа» Института мониторинга климатических и экологических систем СО РАН (руководитель с.н.с., к.б.н. Е. А. Жук, соруководитель проф. К. В. Крутовский),

финансируемого Российским Фондом Фундаментальных Исследований (РФФИ) в рамках Программы инициативных проектов (грант № 16-04-00440).

Магистерская диссертация выполнена в лаборатории лесной геномики СФУ и базовой кафедры защиты и современных технологий мониторинга лесов (зав. каф. д.б.н. И. Е. Ямских) в рамках проекта **«Геномные исследования основных бореальных лесообразующих хвойных видов и их наиболее опасных патогенов в Российской Федерации»**, руководимого проф. К. В. Крутовским и финансируемого Правительством РФ (договор №14.У26.31.0004).

Автор выражает искреннюю благодарность Н. В. Орешковой за пробоподготовку и секвенирование, Д. А. Кузьмину и В. В. Шарову за помощь в сборке транскриптома, Ю. А. Путинцевой, М. Г. Садовскому и К. В. Крутовскому за помощь в компьютерном анализе, за идею исследования и общее руководство на всех этапах работы. Также автор выражает благодарность всем членам лаборатории лесной геномики за участие в обсуждении работы, помощь и ценные советы.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Bendix, C., Lewis, J. D. The enemy within: phloem-limited pathogens // Molecular Plant Pathology. – 2016. doi:10.1111/mpp.12526.
2. Costa, G.G. et al. The mitochondrial genome of *Moniliophthora roreri*, the frosty pod rot pathogen of cacao // Fungal Biol. – 2012. – 116(5). Pp. 551-62. doi: 10.1016/j.funbio.2012.01.008.
3. Ishii, Y. et al. Process of reductive evolution during 10 years in plasmids of a non-insect-transmissible phytoplasma // Gene. – 2009. – 446(2). Pp. 51-7. doi: 10.1016/j.gene.2009.07.010.
4. Gildemberg, A. L. et al. Searching for *Moniliophthora perniciosa* pathogenicity genes // Fungal Biology. – 2010. – 114(10). Pp 842-854. doi: 10.1016/j.funbio.2010.07.009.
5. Loguerio, L. L. et al. Selection of *Trichoderma stromaticum* isolates for efficient biological control of witches' broom disease in cacao // Biological Control. – 2009. – 51(1). Pp. 130–139 doi:10.1016/j.biocontrol.2009.06.005.
6. Ludwig-Müller, J. Bacteria and fungi controlling plant growth by manipulating auxin: Balance between development and defense // Journal of Plant Physiology. - 2015. - 172. Pp 4-12. doi: <https://doi.org/10.1016/j.jplph.2014.01.002>.
7. Ямбуров, М. С. «Ведьмины метлы» кедра сибирского как спонтанные соматические мутации: встречаемость, свойства и возможности использования в селекционных программах/ М. С. Ямбуров, С. Н. Горошкевич // ХБЗ. – 2007. - №2-3.
8. Wittmack, L. Deutsche Garten-Zeitung // Wochenschrift für Gärtner und Gartenfreunde. – 1886.
9. Zhuk, E., Vasilyeva, G. & Goroshkevich, S. Witches' broom and normal crown clones from the same trees of *Pinus sibirica*: a comparative morphological study // Trees. – 2015. doi:10.1007/s00468-015-1187-2.

10. Yamburov, M.S. & Titova, K. Needle anatomy of mutational witches' brooms of siberian fir // World Applied Sciences Journal. – 2013. – 28(7) Pp. 909-913. doi: 10.5829/idosi.wasj.2013.28.07.13834.
11. Vasilyeva, G. & Zhuk, E. Needle structure of mutational witches' brooms in *Pinus sibirica* // Dendrobiology. – 2016. – 75(1). Pp. 79-85. doi: 10.12657/denbio.075.008.
12. Vrgoc, P. Witches' broom of Aleppo pine (*Pinus halepensis* Mill.) and its use for new ornamentals // Proceedings XX EUCARPIA Symposium. – 2002. – Part 2. Pp. 199–205.
13. Муратова, Е. Н., Седельникова, Т.С. Геномные и хромосомные мутации у сосны обыкновенной в экстремальных условиях произрастания / Е. Н. Муратова, Т. С. Седельникова // Хвойные бореальной зоны. – 2004. – Вып. 2. – С. 128-140.
14. Седельникова, Т.С. Кариологические особенности сосны кедровой сибирской (*Pinus sibirica* Du Tour) на болотах Западной Сибири / Т.С. Седельникова, Е.Н. Муратова // Экология. – 2002. – № 5. – С. 323-328
15. Ye, L., Liu, L., Xing, A., Kang, D. Characterization of a dwarf mutant allele of *Arabidopsis* MDR-like ABC transporter AtPGP1 gene // Biochem Biophys Res Commun. – 2013. – 441(4). Pp. 782-6. doi: 10.1016/j.bbrc.2013.10.136.
16. Bolger, A. M., Lohse, M., & Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data // Bioinformatics. – 2014.
17. Andrews, S. FastQC: a quality control tool for high throughput sequence data. [Электронный ресурс]. – Режим доступа: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
18. Nicholas M. et al. Comparison of Three Magnetic Bead Surface Functionalities for RNA Extraction and Detection // Haselton ACS Applied Materials & Interfaces. – 2015. – 7(11). Pp. 6062-6069. doi: 10.1021/am506374t
19. Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., Glöckner F.O. The SILVA ribosomal RNA gene database project: improved data

- processing and web-based tools // Nucl. Acids Res. – 2013. – 41(D1). Pp. D590-D596.
20. NCBI Reference Sequence (RefSeq) project [Электронный ресурс]. – Режим доступа: <https://www.ncbi.nlm.nih.gov/genome/organelle/>
21. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2 // Nature Methods. – 2012. – 9. Pp. 357-359.
22. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-seq data without a reference genome // Nat Biotechnol. – 2011. – 29(7). doi: 10.1038/nbt.1883. PubMed PMID: 21572440.
23. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. // Nat Protoc. – 2013. – 8(8). Pp. 1494-512. doi: 10.1038/nprot.2013.084.
24. Felipe A. Simão, Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs // Bioinformatics. – 2015. doi: 10.1093/bioinformatics/btv351.
25. Zdobnov EM et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs // NAR. – 2016.
26. EvidentialGene: Evidence Directed Gene Construction for Eukaryotes. [Электронный ресурс]. – Режим доступа: <http://arthropods.eugenies.org/EvidentialGene/>.
27. S. Götz et al. High-throughput functional annotation and data mining with the Blast2GO suite // Nucleic Acids Research. – 2008. – 36. Pp. 3420-3435.

28. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome // BMC Bioinformatics. – 2011. – 12. doi:10.1186/1471-2105-12-323.
29. Robinson, MD., and Smyth, GK. Small sample estimation of negative binomial dispersion, with applications to SAGE data // Biostatistics. – 2008. – 9. Pp. 321–332.
30. Chen Y, Lun ATL, and Smyth, GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline // F1000Research. – 2016. – 5. Pp. 1438.
31. Lund, S., Nettleton, D., McCarthy, D., and Smyth, G. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates // Statistical Applications in Genetics and Molecular Biology. – 2012. – 11(8).
32. McCarthy, D.J., Chen, Y., and Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation // Nucleic Acids Research. – 2012. – 40. Pp.4288–4297.
33. Ashburner et al. Gene ontology: tool for the unification of biology // Nat Genet. – 2000. – 25(1). Pp. 25-9.
34. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data // Bioinformatics. – 2011. – 27(21). Pp. 2987-93.
35. Wachowiak W, Trivedi U, Perry A, Cavers S. Comparative transcriptomics of a complex of four European pine species // BMC Genomics. – 2015. – 16(1). Pp. 234. doi:10.1186/s12864-015-1401-z.
36. Open Reading Frame Finder. [Электронный ресурс]. – Режим доступа: <https://www.ncbi.nlm.nih.gov/orffinder/>.

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
**«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

Институт фундаментальной биологии и биотехнологии

Базовая кафедра защиты и современных технологий мониторинга лесов

УТВЕРЖДАЮ  
Заведующий кафедрой  
Лисина Ч.Е.  
подпись инициалы, фамилия  
«22 » июня 2018 г.

**МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ**

**Экспрессия генов у сосны сибирской кедровой (*Pinus sibirica* Du Tour.) и  
сосны обыкновенной (*Pinus sylvestris* Ledeb.) с аномальным  
морфогенезом кроны**

Направление подготовки 04.06.01 – Биология, по профилю 06.04.01.06 –  
Геномика и биоинформатика

Научный руководитель М. Г. Садовский д-р физ.-мат. наук М. Г. Садовский  
подпись, дата

Выпускник С. В. Новикова С. В. Новикова

подпись, дата

Рецензент Д.А. Афонников в.н.с., канд. биол. наук, Д.А. Афонников  
подпись, дата должность, ученая степень инициалы, фамилия

Красноярск 2018