

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

Институт фундаментальной биологии и биотехнологии  
Кафедра биофизики

УТВЕРЖДАЮ  
Заведующий кафедрой

« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ г.

**БАКАЛАВРСКАЯ РАБОТА**

06.03.01 - Биология

Выявление и анализ митохондриальных последовательностей сосны  
сибирской кедровой (*Pinus sibirica*)

Руководитель	_____	Путинцева Ю.А.
Научный консультант	_____	д-р физ.-мат. наук М.Г. Садовский
Выпускник	_____	Кириченко А.Д.

Красноярск 2019

## РЕФЕРАТ

Выпускная квалификационная работа по теме «Выявление и анализ последовательностей митохондриального генома сосны сибирской кедровой (*Pinus sibirica* Du Tour)» содержит 50 страницы текстового документа, приложение, 67 использованных источников, 12 рисунков и 3 таблицы.

ГОЛОСЕМЕННЫЕ, СОСНА СИБИРСКАЯ КЕДРОВАЯ, МИТОХОНДРИАЛЬНЫЙ ГЕНОМ, ПОВТОРЯЮЩИЕСЯ ПОСЛЕДОВАТЕЛЬНОСТИ, ПРЕДПОЧТИТЕЛЬНОСТЬ ИСПОЛЬЗОВАНИЯ СИНОНИМИЧНЫХ КОДОНОВ, РЕДАКТИРОВАНИЕ РНК.

Целью выпускной квалификационной работы является выявление и анализ митохондриальных последовательностей *Pinus sibirica*.

В результате были отобраны последовательности митохондриального генома сосны кедровой сибирской, проведена аннотация генома, классифицированы повторяющиеся последовательности, определены смещения частот используемых кодонов и определены сайты редактирования РНК.

Полученная сборка митохондриального генома кедра была в пределах длины сборок митохондриальных геномов других хвойных. Были обнаружены типичные гены, отвечающие за окислительное фосфорилирование. Также можно сделать несколько предположений: для митохондриальных геномов голосеменных растений не характерно наличие большого количества повторяющихся последовательностей; смещение частот используемых кодонов коррелирует с длиной гена. Также можно отметить присутствие большого количества сайтов редактирования РНК во всех белок-кодирующих генах.

## СОДЕРЖАНИЕ

РЕФЕРАТ.....	1
1. ВВЕДЕНИЕ .....	4
2. ОБЗОР ЛИТЕРАТУРЫ .....	6
2.1. Технологии секвенирования нового поколения.....	6
2.2. Митохондрии растительных клеток.....	7
2.3. Структура митохондриального генома растений.....	8
2.4. Повторяющиеся последовательности в митохондриальных геномах растений.....	10
2.4.1. Тандемные повторяющиеся последовательности .....	10
2.4.2. Диспергированные последовательности ДНК .....	11
2.5. Митохондриальные геномы хвойных .....	12
2.5.1. <i>Picea glauca</i> .....	12
2.5.2. <i>Cycas taitungensis</i> .....	15
2.5.3. <i>Picea abies</i> .....	16
2.6. Предпочтение кодонов .....	17
2.7. Редактирование РНК.....	19
3. МАТЕРИАЛЫ И МЕТОДЫ .....	23
4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ .....	<b>Error! Bookmark not defined.</b>
5. ЗАКЛЮЧЕНИЕ .....	34
6. СПИСОК СОКРАЩЕНИЙ.....	35
7. СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ .....	36
8. ПРИЛОЖЕНИЕ.....	45

## 1. ВВЕДЕНИЕ

Хвойные леса составляют более трети всех лесов планеты. Из многочисленных видов хвойных, более ста видов представлены соснами. Сосновые леса широко распространены в нашей стране – и в Европейской части, и в Сибири, и на Кавказе. К хвойным деревьям относится около 560 видов из рода ель, лиственница, тис, секвойя, кипарис и другие. Хвойные леса имеют большое экономическое и экологическое значение и играют огромную роль в регуляции климата. Поэтому на данный момент важны мониторинг и сохранение генетических ресурсов пород хвойных растений.

Для выполнения этих задач требуется знание о генах хвойных растений, ведь они контролируют адаптацию и устойчивость к болезням. При наличии полностью расшифрованного генома, можно определить полный набор генов и на его основе разработать генетические маркеры (микросателлитные локусы и однонуклеотидные полиморфизмы – SNP), которые могут быть использованы в генетических исследованиях популяций, в обнаружении генетической связи между генетической и адаптационной изменчивостью и прочими аспектами. Все это важно для понимания генетического контроля адаптивных, селекционных и экономических признаков [1].

Молекулярно-генетические маркеры могут эффективно применяться для определения происхождения древесины, оценки адаптивной способности к изменчивости климата, мониторинга изменчивости хвойных лесов.

Изучение геномов растений – задача сложная. Это связано с огромными размерами геномов (в частности, геномов хвойных), резким колебанием количества хромосом у различных растений (не удастся выявить корреляцию между количеством хромосом и размером генома), аллополиплоидией, чрезвычайно большим количеством некодирующей ДНК и неполным морфологическим, генетическим и физическим картированием хромосом [1].

Отдельный интерес вызывают митохондриальные геномы растений, которые значительно варьируют по длине генома и различаются по структуре. Как правило, митохондриальный геном наследуется по материнской линии и имеет гораздо более сложную структуру, чем хлоропластный геном, особенно у эволюционно древних хвойных растений. Также большая его часть имеет различные некодирующие участки, включая длинные повторяющиеся последовательности, создающие возможность внутримолекулярной рекомбинации.

Целью работы является выявление и анализ митохондриальных последовательностей *Pinus sibirica*, а также поиск и классификация повторяющихся последовательностей.

Для достижения цели были поставлены следующие задачи:

1. Отобрать митохондриальные скаффолды
2. Провести первичную аннотацию
3. Найти и классифицировать повторяющиеся последовательности в митохондриальном геноме *Pinus sibirica*
4. Изучить смещение частот используемых кодонов
5. Предсказать сайты редактирования РНК.

## **2. ОБЗОР ЛИТЕРАТУРЫ**

### **2.1. Технологии секвенирования нового поколения**

Новые технологии секвенирования (NGS) позволили усовершенствовать открытый более 30 лет назад, основанный на электрофорезе, метод Сэнгера, недостатками которого являются низкая производительность и относительно высокая стоимость [2]. Технологии NGS позволяют секвенировать одновременно тысячи молекул ДНК, тем самым повышая скорость исследования и увеличивая объем получаемых данных, при этом снижая себестоимость анализа. Принцип NGS основан на массовом параллельном секвенировании специальным образом подготовленных однонитевых библиотек фрагментированной ДНК исследуемых образцов. В настоящее время для высокопроизводительного секвенирования доступны несколько коммерческих платформ: HiSeq, MiSeq и NextSeq 500 (производитель Illumina), Roche 454 GS, IonTorrent (ThermoFisherScientific) и SOLiD (AppliedBiosystems) [3-7]. Большинство технологий высокопроизводительного секвенирования включают в себя следующие этапы: подготовку библиотек, непосредственно секвенирование и анализ полученных данных.

Подготовка библиотек включает в себя фрагментирование ДНК до 300-500 нуклеотидных пар, лигирование (присоединение) секвенсовых адаптеров (синтезированные олигонуклеотиды с известной последовательностью) с концами изучаемых фрагментов и амплификацию полученных библиотек. Метод амплификации может отличаться для различных платформ. Например, платформа IonTorrent от ThermoFisherScientific использует метод эмульсионной ПЦР для амплификации одноцепочечных фрагментов на микросферах, в то время как технология фирмы Illumina основана на амплификации фрагментов методом мостиковой ПЦР и формировании кластеров на проточной ячейке [8, 9].

Секвенирование производится путем синтеза новых фрагментов ДНК на одноцепочечных ДНК-библиотеках, которые играют роль матрицы. Нуклеотиды встраиваются в новую цепь в определенном порядке, соответствующем матричной цепи, который записывается в цифровом виде. После включения в цепь каждого последующего нуклеотида прибор регистрирует сигнал. Разные платформы для NGS используют различные механизмы детекции для встроенных в новую цепь нуклеотидов: детекцию флуоресцентного сигнала после включения в цепь комплементарного матрице нуклеотида (Illumina, SOLiD Applied Biosystems), изменение pH раствора в микрореакторе, связанное с выделением в среду ионов водорода в ходе синтеза ДНК ферментом (Ion Torrent ThermoFisher Scientific), или регистрацию светового сигнала после выделения пирофосфата, активирующего каскад химических реакций [9].

После секвенирования полученные данные могут быть обработаны методами биоинформатики. Данные проходят несколько этапов обработки: исключение ридов с низким качеством прочтения, выравнивание данных относительно референсной последовательности или сборку последовательности *de novo*, аннотация полученных последовательностей и их последующий анализ, позволяющий определять тип генетических вариантов, включая их наследственный характер, оценить уровень экспрессии генов, идентифицировать новые гены и регуляторные элементы.

## **2.2. Митохондрии растительных клеток**

Митохондрии – «силовые» станции клетки, в них локализована большая часть реакций дыхания (аэробная фаза). В митохондриях происходит аккумуляция энергии дыхания в аденозинтрифосфате (АТФ). Энергия, запасаемая в АТФ, служит основным источником для физиологической деятельности клетки. Количество митохондрий в клетках может варьировать в широких пределах – от 500 до 1000 органелл [10] – это зависит от функции, которую выполняет клетка в той или иной ткани или

органе. Митохондрии имеют удлиненную палочковидную форму длиной 4-7 мкм и диаметром 0,5-2 мкм [10].

Митохондрии – это белково-липидные органеллы. Содержание белка в них составляет 60-65% от общей массы, липиды – около 30%, РНК – приблизительно 1% и менее 0,5% – ДНК. В митохондриях имеется не только ДНК, но и вся система синтеза белка, в том числе рибосомы [10].

Свойства митохондрий (белки, структура) закодированы частично в ДНК митохондрий, а частично в ядре. Митохондриальный геном кодирует белки рибосом и частично систему дыхательной цепи переноса электронов, а в геноме ядра закодирована информация о белках-ферментах цикла Кребса. Митохондрии, как и пластиды, считаются полуавтономными органеллами, т.к. они имеют собственную ДНК и собственную белок-синтезирующую систему. Митохондриям свойственна цитоплазматическая наследственность по материнской линии, т.к. инициальные частицы митохондрий локализованы в яйцеклетке.

### **2.3. Структура митохондриального генома растений**

На апрель 2019 года в базе данных нуклеотидных последовательностей RefSeq NCBI (<http://www.ncbi.nlm.nih.gov/refseq>) были представлены полностью секвенированные митохондриальные геномы 169 видов и подвидов высших растений от мхов до цветковых. В базе данных GenBank NCBI (<http://www.ncbi.nlm.nih.gov/genbank/>) были доступны около четырех тысяч митохондриальных нуклеотидных последовательностей растений. Число геномов в двух базах различается, так как GenBank представляет собой архив, в который выкладывается тотальное большинство секвенируемых последовательностей, тогда как RefSeq является курируемой базой данных референсных генетических последовательностей, исключая избыточность данных [22]. В животном царстве митогеномы изучены значительно лучше: в базе RefSeq можно найти несколько тысяч записей полных митохондриальных геномов. С накоплением данных становится ясно, что компактная кольцевая митохондриальная ДНК



животных является нетипичной для остальных эукариот и представляет собой пример высокоспециализированной структурной организации [11,12].

Ранее считалось, что растения также имеют кольцевую структуру мтДНК и в некоторых случаях, находятся в сверхспирализованном состоянии [23, 24, 25]. Однако, *in vivo* структура мтДНК растений намного сложнее, чем предполагала простая кольцевая модель хромосомы. Во-первых, мтДНК в сосудистых растениях обычно содержит повторяющиеся последовательности, которые являются рекомбинантно-активными и служат постоянным источником перегруппировок [26, 27]. В результате, митохондриальный геном растений может существовать как популяция альтернативных структур, даже в пределах одной особи. Поэтому мтДНК растений обычно описывается как «главное кольцо» разделенное на много частей, содержащее полную геномную последовательность и набор субгеномных колец, генерируемый повторами.

Во-вторых, карты кольцевых геномов не обязательно подразумевают существование кольцевых молекул, поскольку альтернативные структуры, такие как конкатемеры и перекрывающиеся линейные фрагменты (то есть кольцевые линейные молекулы), могут также отображаться как кольца. Но на данный момент, есть мнение, что модель «мастер-кольцо» не подходит для описания структуры мтДНК растений [65].

Секвенирование ДНК и сравнение сестринских таксонов показали, что частота митохондриальных синонимичных мутаций в растениях существенно ниже, чем в митохондриях животных или ядре [40-43]. Напротив, после секвенирования цельных геномов, стало ясно, что порядок генов сохраняется очень редко даже между близкими таксонами [44, 45], вероятно, из-за частого восстановления ДНК, которое происходит посредством гомологичной рекомбинации и негомологичного соединения концов [46, 47, 48]. Большинство растительных митохондриальных геномов содержат значительную долю низкоконсервативной ДНК с неизвестными функциями, а также некоторые растения резко увеличили размеры своего генома до

миллионов пар нуклеотидов, хотя все еще кодируют только десятки генов [49, 50, 51]. МтДНК обладает сложной и динамичной структурой, включая линейные и разветвленные молекулы и они могут представлять множественные изоформы генома, поэтому модель мастер-кольца не может описать достоверно структуру [52].

Митохондриальные геномы растений, как правило, имеют небольшое количество нетандемных прямых или инвертированных повторяющихся последовательностей длиной несколько килобаз. Они могут часто и симметрично рекомбинировать, изомеризуя геном [53]. Некоторые геномы собираются в более чем одну независимую молекулу [54], хотя это также может быть следствием применяемых методов и параметров сборки [49]. Есть также рассеянные повторы до нескольких сотен пар оснований, которые рекомбинируют с относительно низкой частотой в растениях дикого типа, но рекомбинируют чаще (и асимметрично) в поддержании и репарации ДНК мутантов [55, 56]. Такие повторы не всегда аннотированы, но их вклад в геномные перестройки важен и начинает проявляться с ходом эволюции [47].

Для достоверного определения структуры применяются новые технологии для решения проблемы сборки геномов с ответвлениями и перестройками путем определения полных последовательностей митохондриального генома, объединяют данные с короткими прочтениями, чтобы определить последовательности, соединения, перегруппировки и стехиометрию с большей точностью. Физическая структура геномов визуализируется с помощью флуоресцентной микроскопии [65].

## **2.4. Повторяющиеся последовательности в митохондриальных геномах растений**

Повторяющиеся последовательности – это участки ДНК, включённые в геном, последовательность которых состоит из повторяющихся фрагментов. Они подразделяются на тандемные и диспергированные повторы.

### **2.4.1. Тандемные повторяющиеся последовательности**

Под этим термином подразумевается характерный компонент эукариотического генома, состоящий из tandemно-организованных повторов. В зависимости от размера tandemные повторы подразделяют на три класса: сателлитная ДНК, минисателлиты и микросателлиты.

Сателлитная ДНК не кодирует белки и локализована в конститутивном гетерохроматине хромосом. Сателлитная ДНК состоит из рядов идентичных повторяющихся единиц (мономеров), число которых может варьировать от нескольких сот до более миллиона.

Микросателлиты имеют длину мономера не более 10 п.н., тогда как у минисателлитов она колеблется в пределах 10–100 п.н.. Остальные повторы с длиной мономера выше 100 п.н. относятся к сателлитной ДНК. Для минисателлитной и сателлитной ДНК характерно более локализованное распределение на концах хромосом или в центромерных районах.

#### ***2.4.2. Диспергированные последовательности ДНК***

Известно, что основная масса диспергированных повторяющихся последовательностей ДНК возникла в результате активности мобильных элементов. Мобильные элементы генома подразделяют на два класса: элементы класса I (ретротранспозоны), которые перемещаются с помощью механизма «копирования–встраивания» с использованием РНК-посредников; и элементы класса II (ДНК-транспозоны), которые используют механизм «вырезания–встраивания», с образованием либо одно-, либо двухцепочечных разрывов ДНК.

- **Ретротранспозоны**

Ретротранспозоны делятся по наличию/отсутствию длинных концевых повторов (LTR) на LTR-содержащие (LTR ретротранспозоны) и не содержащие ретротранспозоны. Последние в свою очередь делятся на длинные диспергированные (LINE) и короткие диспергированные (SINE) элементы. Длина этих элементов варьирует от нескольких тыс.п.н. до 25 тыс.п.н., а размер LTR – от нескольких сотен п.н. до нескольких тыс.п.н.

- **ДНК-транспозоны**

ДНК транспозоны, как и ретротранспозоны, обнаружены почти у всех эукариот. Они обычно представлены в небольшом количестве. Класс ДНК-транспозонов содержит два подкласса: первый объединяет элементы, которые перемещаются с помощью классического механизма «вырезания–встраивания» с образованием двухцепочечного разрыва ДНК, а второй подкласс использует для перемещения механизм «катящегося кольца» [14].

## **2.5. Митохондриальные геномы хвойных**

К настоящему времени среди хвойных растений митохондриальные геномы были опубликованы для *Picea glauca* – 5,9 млн.п.н. [15], *Picea abies* – 4,3 млн.п.н. [17], *Pinus taeda* – 1,2 млн.п.н. [28], *Pinus lambertiana* – 3,9 млн.п.н. и *Pinus sylvestris* – 986 тыс.п.н. ([https://www.ncbi.nlm.nih.gov/assembly/GCA\\_900143225.1/](https://www.ncbi.nlm.nih.gov/assembly/GCA_900143225.1/)). Так же имеются данные для митохондриального генома голосеменного растения *Sycas taitungensis* – 415 тыс.п.н. [16].

### **2.5.1. *Picea glauca***

Митохондриальный геном *P. glauca* был опубликован Jackman et al. в статье посвященной сборке и аннотации органелльного генома [15]. Секвенирование митохондриального генома осуществлялась на платформе Illumina HiSeq 2000. Ассемблирование нуклеотидных последовательностей осуществлялось при помощи ABySS. Потенциальные митохондриальные последовательности отделялись от ядерных последовательностей по их длине, глубине покрытия и GC-контенту, используя кластеризацию k-means в R. Предполагаемые митохондриальные контиги были собраны в скаффолды с использованием ABySS-scaffold, был получен 71 скаффолд. Финиширование сборки осуществлялось с использованием программы Sealer для скаффолдинга и LINKS scaffold 1.1, что уменьшило количество скаффолдов до 58. После этого авторы использовали Connector, с его помощью сборку сократили до 38 скаффолдов.

Собранные скаффолды выравнивались на полнонуклеотидную базу GenBank NCBI (nt) посредством BLAST для проверки на совпадения с уже ранее опубликованными митохондриальными последовательностями и для исключения загрязнения. Аннотация проведена с использованием MAKER и Prokka, два набора полученных аннотаций были объединены с использованием BEDTools и GenomeTools. Повторяющиеся последовательности были идентифицированы с использованием RepeatMasker и RepeatModeler.

Результаты, полученные авторами для *Picea glauca*:

- Полная митохондриальная последовательность ДНК *P. glauca* составила 5,94 млн.п.н.
- Было идентифицировано 106 генов (51 белок-кодирующий ген), более 6 тысяч ORF (открытых рамок считывания), 29 тРНК-генов и рРНК-генов (рисунок ).
- повторяющиеся элементы включали только 390 тыс.п.н. (6,6%) митохондриального генома (рисунок 2).

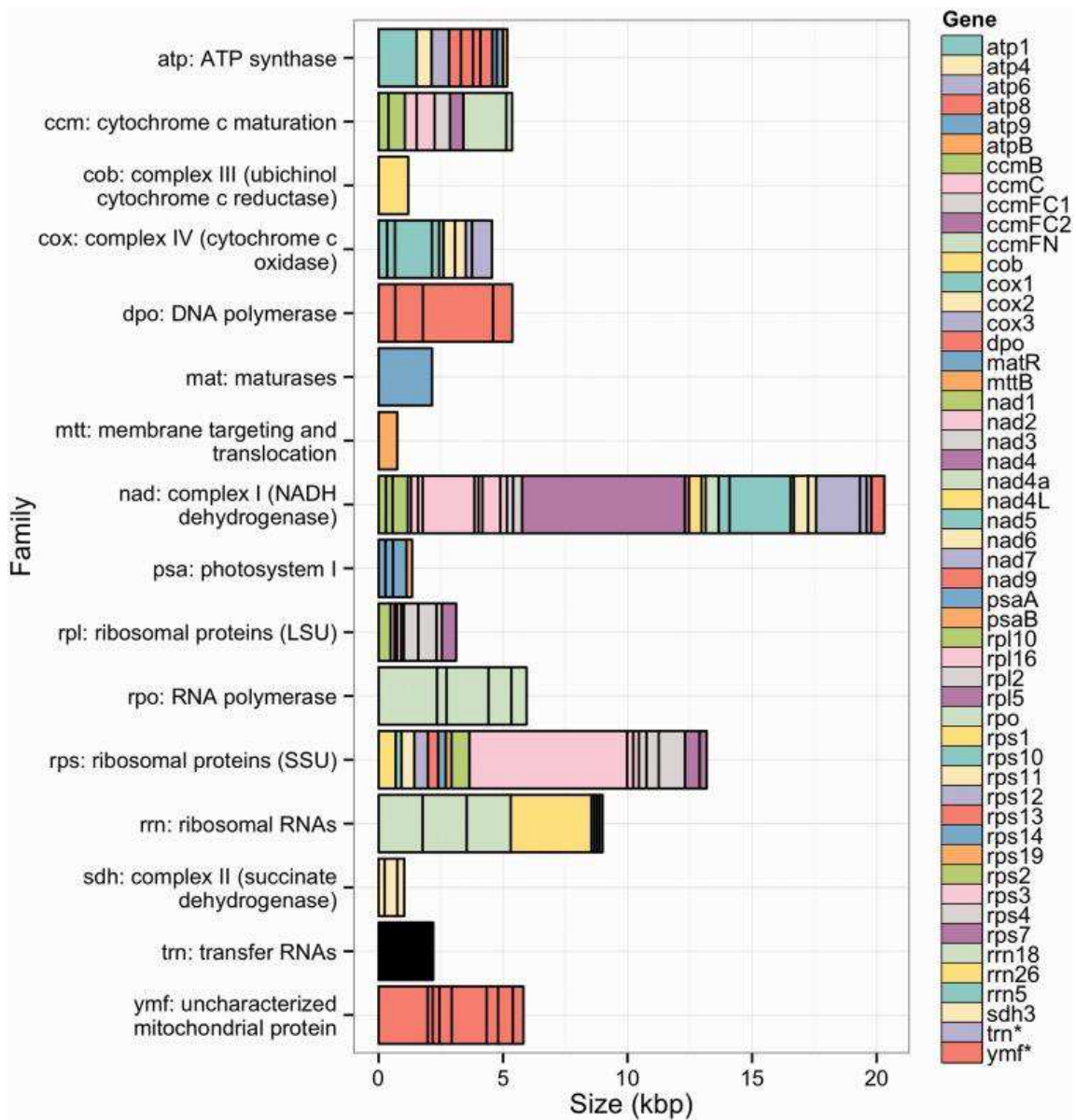


Рисунок 2 – Содержание генов митохондриального генома белой ели, сгруппированных по семейству генов. Каждый прямоугольник пропорционален размеру гена, включая интроны [15]

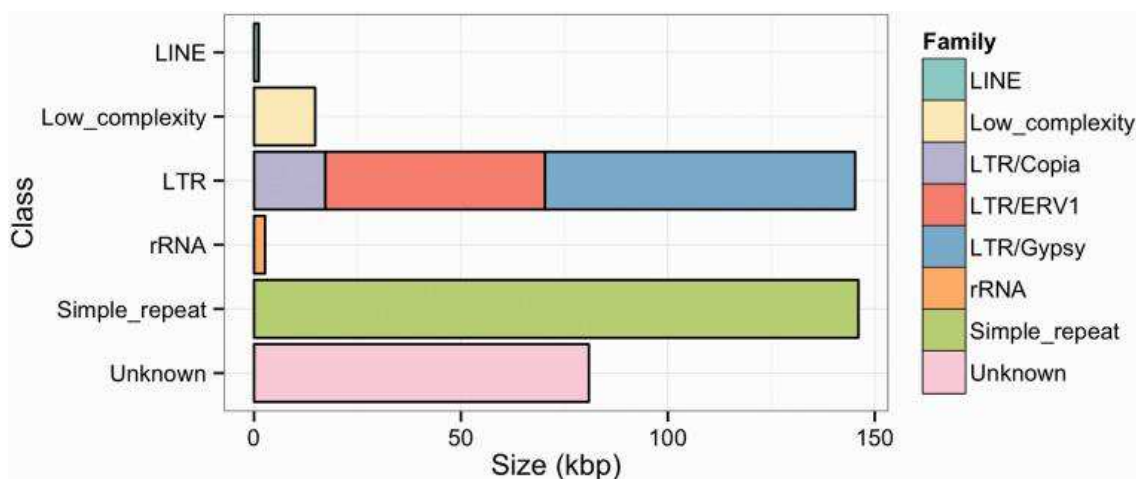


Рисунок 2 – Содержание повторяющихся последовательностей митогенома белой ели [15]

### 2.5.2. *Cycas taitungensis*

Полная последовательность мтДНК *Cycas* была описана в статье [16] и депонирована в ДНК-банке данных Японии под номером NBVI01000001. Точные границы генов и экзонов определялись путем выравнивания на гомологичные гены из доступных и уже аннотированных мтДНК растений. Множественные выравнивания были выполнены с помощью MAP2. Гены тРНК аннотировали с использованием программы tRNAscan-SE. Поиск открытых рамок считывания выполнялся с помощью веб-поиска NCBI ORF Finder. Типы интронов были определены на основе их последовательностей и вторичных структур. Идентификация повторов проводилась с использованием веб-интерфейса REPuter. Информация о тандемных повторах была получена с использованием tandemrepeatsfinder.

Авторами были получены следующие результаты:

- Полная мтДНК *C. taitungensis* представляет собой кольцевую молекулу 414 903 п.н. Обнаружено 39 белок-кодирующих генов (рисунок 3).
- Повторяющиеся последовательности, которые присутствуют в митохондриальном геноме, занимают 15,1%, равномерно распределены по всему геному и в основном присутствуют в некодирующих областях.

- Тандемные повторы занимают 4,97% всего митохондриального генома [16].

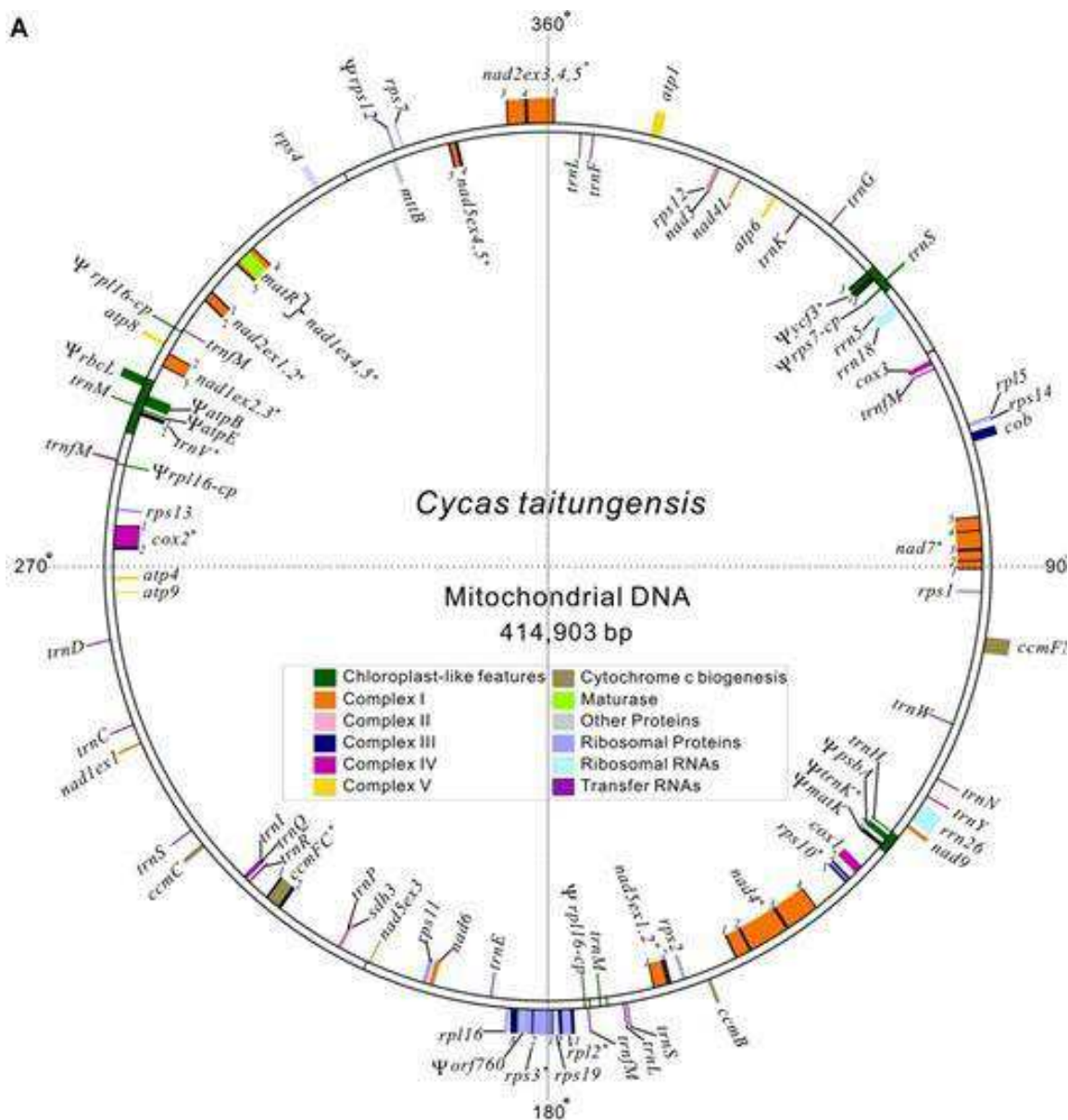


Рисунок 3 – Схема митохондриального генома *Cycastaitungensis*. Гены разбиты на 12 групп в соответствии с их биологическим значением

### 2.5.3. *Picea abies*

Предполагаемые митохондриальные контиги отбирались из полногеномной сборки. Исходя из графиков длина – GC – покрытие,



визуализирующих сборку, они отбирали контиги которые имели длину не менее 1 тыс.п.н.; GC-контент не меньше 40% (среднее содержание GC в ядерном геноме 37,9%) и глубину покрытия, которая превышала в 20 раз глубину среднего покрытия.

Контиги выравнивались при помощи BLAST на базу покрытосеменных растений: *Physcomitrel lapatens*, *Megaceros aenigmaticus*, *Pleurozia purpurea*, *Phaeo ceroslaevis*, *Marchantia polymorpha*, *Arabidopsis thaliana*, *Cucumis sativus*, *Cycastaitun gensis* и единственного, на тот момент, представителя голосеменных *Cycas taitungensis*.

Тотальная сборка мтДНК *Picea abies* составила 4,3 млн.п.н., что авторы сочли недостаточно полной сборкой. В результате был обнаружен 41 ген, причем все 39 генов *Cycas taitungensis* были обнаружены и у *P. abies* [17].

## 2.6. Предпочтение кодонов

Во время трансляции гена в белок, информация, содержащаяся в форме кодонов, транслируется в аминокислотные последовательности. Последовательности белка состоят только из 20 стандартных аминокислот, однако имеется 64 различных кодона (61 кодон, кодирующий аминокислоты и три стоп-кодона), которые кодируют эти 20 аминокислот, следовательно, одну аминокислоту может кодировать несколько кодонов. Этот аспект упоминается как «вырождение» генетического кода и различные кодоны, кодирующие одну и ту же аминокислоту, называются «синонимичными кодонами».

Многочисленные исследования показали, что использование синонимичных кодонов является неслучайным процессом. Было обнаружено, что некоторые кодоны используются чаще по сравнению с другими синонимичными кодонами во время декодирования гена в белок. Эта особенность преимущественного использования определенных кодонов в

процессе трансляции известна как «codon usage bias» или «предпочтение кодонов».

Исследование смещения в частоте использования кодонов приняло новый оборот с появлением полногеномного секвенирования геномов многих видов, как прокариот, так и эукариот. В настоящее время исследования направлены на целостный анализ предпочтительности синонимичных кодонов по всему геному, а не на изучение конкретных генов или наборов генов, чтобы лучше понять паттерны предпочтения кодонов внутри и между геномами. Внутри генома отдельные гены демонстрируют различия в предпочтении кодонов, степень которого зависит от различных факторов, таких как: уровень экспрессии, длина гена, GC-состав, скорость рекомбинации и стабильность РНК. Говорят, что ген сильно смещен в схеме использования кодонов, если он использует определенный синонимичный кодон большую часть времени, когда определенная аминокислота включена в белок. Такие кодоны часто называют оптимальными кодонами или частыми кодонами, поскольку они часто используются при трансляции гена. Вариация оптимизации кодонов среди генов обеспечивает дифференциальную эффективность, а также точность трансляции генов [29].

Помимо ядерного генома, предпочтительность кодонов была идентифицирована и в органелльном геноме. Частота синонимичных кодонов варьируется не только между организмами, но даже между генами внутри организма [30]. Считается, что предпочтительность кодонов отражает чистый результат мутаций, генетического дрейфа и естественного отбора [31, 32].

Эволюция растений всегда следует за эволюцией их геномов, а дупликация и рекомбинация ДНК являются основными факторами эволюции генома у растений [33]. Последние исследования показывают, что предпочтительность кодонов в ядерном геноме тесно связана с эволюцией растений [34], более того, недавно обнаружили, что существует также связь между предпочтением кодонов пластидного генома и эволюцией растений [35].

Появление или потеря интрона, ключевой компонент эволюции геномов [33], в значительной степени обусловлены рекомбинацией и формированием инделов (вставок и делеций). В ядерном геноме предпочтение кодонов, по-видимому, коррелирует с числом интронов [34]. Однако, недавнее исследование показало, что корреляция частоты предпочтения кодонов с числом интронов отсутствует в пластидном геноме [35].

В 2015 году было опубликовано исследование [34] с названием «предпочтение синонимичных кодонов в генах митохондриальных растений связано с количеством интронов и зеркальная эволюция». Выборка состояла из 24 видов растений, включающих четыре хлорофита, четыре харофита, шесть бриофитов (включая два вида *anthocerotophyta*, два вида *marchantiophyta*, два вида *bryophyta*), один птеридофит, один голосеменной, четыре однодольных вида и четыре двудольных вида.

Исследователи обнаружили, что среди 61 кодона А- и Т-концевые кодоны (NNA и NNT) встречаются чаще, чем С- и G-концевые кодоны (NNC и NNG) в растительных митохондриальных генах.

Также авторы проверили, влияет ли количество интронов на предпочтительность кодонов. Для этого они посчитали отношения NNC/G к NNA/T в генах, несущих интроны. Для генов голосеменных *C. taitungensis* отношение увеличивалось с ростом числа интронов с нуля до двух, но уменьшилось при увеличении числа интронов до четырех [34].

Помимо стартового кодона ATG и трех стоп-кодонов, в некоторых митохондриальных генах присутствуют нетипичные старт (не ATG) и стоп (не TAA, TGA и TAG) кодоны [36-38]. У *C. taitungensis* и *A. thaliana* был обнаружен атипичный старт-кодон GGG [34].

## 2.7. Редактирование РНК

Редактирование РНК – это запрограммированное изменение нуклеотидных оснований путем химических модификаций в РНК относительно кДНК.

У покрытосеменных и голосеменных растений большинство митохондриальных транскриптов подвергаются множественным сайт-специфическим изменениям во время созревания мРНК – редактированию РНК. Для некоторых растений описано до 500 сайтов редактирования митохондриальной РНК по типу C-to-U (замена цитозина на урацил). Большинство этих изменений связаны с первым или вторым положением кодонов, и поэтому большинство из них изменяют природу кодируемой аминокислоты [57]. У большинства видов растений U-to-C превращения являются редкими уникальными событиями. Сайты редактирования РНК в основном обнаруживаются в кодирующих областях мРНК и реже в интронах и других нетранслируемых областях. В некоторых случаях редактирование РНК в молекулах тРНК восстанавливает необходимые пары оснований. В этих случаях только полное редактирование РНК позволяет сформировать правильную структуру и далее обрабатывать созревающих предшественников тРНК. Рибосомальная РНК в органеллах растений, по-видимому, подвергается очень незначительному редактированию РНК [58].

Основная сложность в понимании редактирования РНК состоит в том, чтобы понять, как механизм достигает требуемой специфичности, другими словами, как цель редактирования – цитозин, распознается среди всех других цитозинов.

В ряде исследований были выявлены существенные мотивы последовательности в непосредственной близости от сайтов редактирования как в хлоропластах, так и в митохондриях. Цель редактирования C определяется наличием цис-элемента в пределах 20 нуклеотидов выше по последовательности. Эта последовательность распознается белками из семейства белков с пентатрикопептидными повторами (PPR). Каждый мотив PPR, состоящий из двух альфа-спиралей, взаимодействует с нуклеотидом в цис-элементе [59].

Белки PPR являются факторами специфичности. Прорыв, касающийся природы факторов редактирования, был сделан группой профессора

Т. Шиканаи, работающей над хлорореспираторными мутантами в *Arabidopsis thaliana* [60].

Практически все исследованные белки PPR находятся либо в хлоропластах, либо в митохондриях, где было показано, что они играют роль на разных этапах экспрессии генов (транскрипция, сплайсинг, редактирование и трансляция РНК) [61]. Белки PPR состоят из tandemного набора вырожденных 35 аминокислотных повторов. Они делятся на несколько подгрупп в зависимости от длины повторов и наличия дополнительных С-концевых доменов. PPR-белки Р-класса содержат только канонические 35 аминокислотных повторов и присутствуют во всех эукариотах, которые имеют митохондриальный геном. PPR-белки PLS-класса специфичны для растений и в дополнение к стандартным мотивам PPR Р-класса, распознаются по наличию длинных (L) мотивов и коротких (S) мотивов. Этот специфичный для растений класс белков PPR также был назван «растительным комбинаторным и модульным белком» (PCMP)[62]. Почти все белки класса PLS имеют дополнительные С-концевые домены, E (расширенный) домен и домен DYW, названный в честь консервативного конечного трипептида аспаргат (D) -тирозин (Y) -триптофана (W). Все белки PPR, участвующие в редактировании, являются членами подсемейства PLS.

Все белки PPR, участвующие в редактировании, содержат домен E. Этот домен вряд ли выполняет саму редактирующую функцию, поскольку в нем отсутствует какой-либо очевидный каталитический сайт, и поэтому предполагается, что он задействует активность из отдельного домена того же или другого белка. Хотя сейчас получены убедительные доказательства того, что белки PPR являются факторами специфичности, которые связывают РНК и нацелены на редактируемый С, специфичность распознавания/связывания не совсем понятна [63].

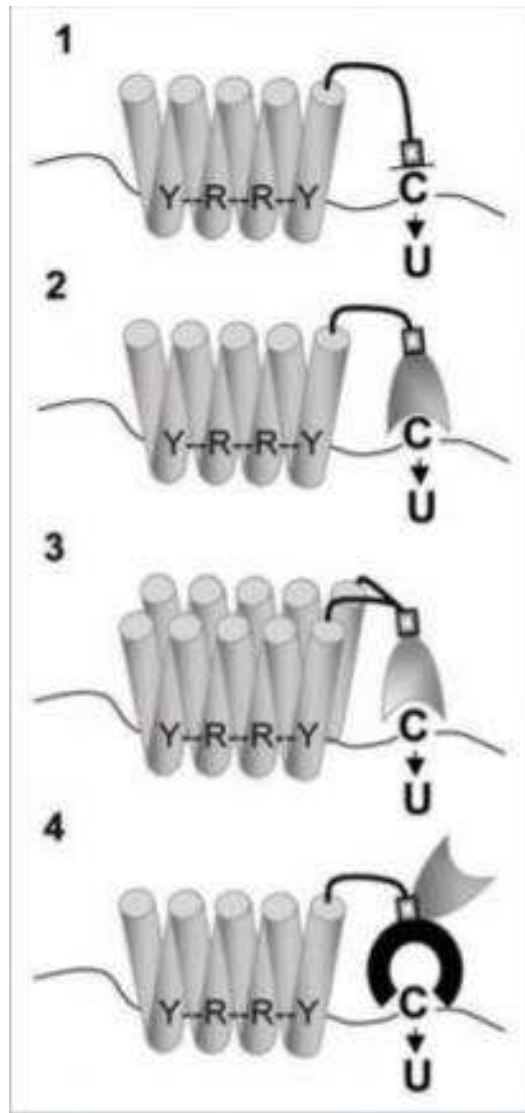


Рисунок 4 – Модели распознавания и дезаминирования цитидина в органеллах растений: 1 - дезаминирование может быть катализировано E-доменом (маловероятно, учитывая отсутствие очевидного активного центра), 2 и 3 - дезаминирование может быть катализировано доменом DYW либо в цис- (2), либо в транс-положении (3) для белков PPR подгруппы E, 4 - дезаминирование может быть осуществлено еще не обнаруженным ферментом, вырабатываемым доменом E, и домен DYW не будет играть никакой роли в редактировании [63].

Исследования по предсказанию редактирования РНК проводились для белой ели, гинкго и саговника. В 39 белок-кодирующих генах *Sucas taitungensis* было обнаружено 1214 сайтов, у *Ginkg obiloba* 1306 сайтов в 41 гене [64] и у *Picea glauca* 1751 позиция в 68 генах. Также авторы отмечают, что редактирование C-to-U РНК может генерировать новые старт- и стоп-кодоны, но не может уничтожить существующие [15].

### 3. МАТЕРИАЛЫ И МЕТОДЫ

Для облегчения секвенирования и сборки митохондриального генома сосны сибирской кедровой (*Pinus sibirica* Du Tour) было проведено обогащение образцов ДНК митохондриальной фракцией в лаборатории генетической инженерии растений СИФиБР СО РАН под руководством проф. д. б. н. Ю. М. Константинова. В лаборатории лесной геномики СФУ было проведено секвенирование полученных образцов на платформе Illumina HiSeq2000. Ассемблирование нуклеотидных последовательностей осуществлялось сотрудником лаборатории лесной геномики при помощи коммерческого пакета компьютерных программ CLC Assembly Cell с последующим использованием программ BESST [66] и Sealer [67] для скаффолдинга и закрытия гэпов. Глобальное выравнивание исходных парных ридов на сборку осуществлялось с помощью программы Bowtie2 [18]. Для визуализации сборки были построены графики в RStudio по трем переменным: GC-контент скаффолдов, длина и среднее покрытие.

Для выявления митохондриальных скаффолдов была проведена кластеризация методом динамических ядер (k-means) по трем переменным длина – GC – покрытие, но ожидаемых результатов она не дала, обособленных кластеров с митохондриальными скаффолдами выявлено не было. Поэтому, для отбора именно митохондриальных скаффолдов проводилось выравнивание геномной сборки при помощи BLAST на базу из 3489 (143,5 млн. п. н.) последовательностей митохондриальных генов и геномов семенных растений, извлеченных из GenBank NCBI, фильтрация результатов осуществлялась вручную, чтобы исключить загрязнения и последовательности неясного происхождения. После первого отсева проводился повторный скаффолдинг и еще одно локальное выравнивание на базу семенных растений. Для создания баз данных и локальных выравниваний использовалась консольная версия nBLAST [19].

Первичная аннотация была проведена при помощи пакета компьютерных программ Mitofy. Дальнейший поиск белок-кодирующих генов осуществлялся при помощи BLAST в ПО CLC Genomic Workbench путем выравнивания сборки на хорошо аннотированные геномы саговника и гинкго. Гены, кодирующие тРНК, были выявлены при помощи tRNA-scan.

Для поиска повторяющихся последовательностей использовался RepeatModeler, было обнаружено 23 повторяющиеся последовательности, из которых классифицировано было только две [20].

Для идентификации классов повторов применяли TEclass – инструмент, позволяющий классифицировать неизвестные элементы в четыре основные категории, которые отражают их режим встраивания в геном: ДНК-транспозоны, LTR, LINE, SINE [21].

Для определения предпочтения кодонов были подсчитаны индексы RSCU (relative synonymous codon usage) и CAI (codon adaptation index).

RSCU – это количество раз, когда наблюдается один конкретный кодон, относительно количества встречаемости всех кодонов, кодирующих одну и ту же аминокислоту.

RSCU определяется по формуле

$$RSCU_i = \frac{X_i}{\frac{1}{n} \sum_{i=1}^n X_i} \quad (1),$$

где  $n$  – количество синонимичных кодонов ( $1 \leq n \leq 6$ ), кодирующих одну и ту же аминокислоту,  $X_i$  – количество кодонов в гене  $i$ .

CAI – измеряет степень, с которой гены используют предпочтительность кодонов.

Для начала, нужно построить таблицу значений RSCU, используя формулу (1) для высоко экспрессируемых генов.



Относительная адаптивность кодона ( $w_i$ ) вычисляется как

$$W_i = \frac{RSCU_i}{RSCU_{max}} \quad (2),$$

где  $RSCU_{max}$  - значение  $RSCU$  для наиболее часто используемого кодона для аминокислоты.

Используя  $w_i$ , рассчитанное по формуле (2), находим CAI для каждого гена

$$CAI = \left( \prod_{i=1}^L w_i \right)^{\frac{1}{L}} \quad (3),$$

где  $w_i$  - относительная адаптивность кодона,  $L$  – количество кодонов в гене [39].

Предсказание места редактирования C-to-U осуществлялось с помощью онлайн-инструмента PREP-Mt с  $C$  (cutoff) равным 0,2 – это означает, что вероятность правильно определенного сайта связывания не менее 20%, было показано, что это значение оптимально для многих голосеменных [15, 64].

[В связи с авторскими правами были изъяты результаты (стр 26-33)  
следующие страницы с 26 по 33.

#### 4. ЗАКЛЮЧЕНИЕ

Были выявлены последовательности митохондриального генома сосны кедровой сибирской. Полученная сборка для сосны сибирской кедровой была в пределах длины сборок митохондриальных геномов хвойных, которые варьируют в широком диапазоне: от 1,2 млн. п.н. у *Pinus taeda* и 3,9 млн. п.н. у *Pinus lambertiana* – до 5,9 млн. п.н. у *Picea glauca*. Считается, что разница в размерах митохондриального генома обусловлена в основном различными размерами некодирующих участков.

Были обнаружены типичные гены, отвечающие за окислительное фосфорилирование. Также можно сделать несколько предположений: для голосеменных растений не характерно большое количество повторяющихся последовательностей в митохондриальных геномах; смещение частот используемых кодонов коррелирует с длиной гена. Также можно отметить присутствие большого количества сайтов редактирования РНК во всех белок-кодирующих генах.

## 5. СПИСОК СОКРАЩЕНИЙ

1. C-to-U – тип редактирования РНК, при котором цитозин заменяется урацилом
2. DNA transposons – ДНК-транспозоны
3. GC-состав – доля гуанина и цитозина среди всех остатков нуклеотидов нуклеотидной последовательности
4. LINE – длинные диспергированные повторы
5. LTR – длинные концевые повторы
6. NGS – методы секвенирования нового поколения
7. ORF ( open reading frame) – открытые рамки считывания
8. PPR белки – семейство белков с пентатрикопептидными повторами
9. SINE – короткие перекрестные ядерные элементы
10. SNP – однонуклеотидные полиморфизмы
11. U-to-C – тип редактирования РНК, при котором урацил заменяется цитозином
12. Контиг – набор перекрывающихся нуклеотидных последовательностей ДНК
13. мтДНК – митохондриальная ДНК
14. П.н. – пар нуклеотидов
15. Референсная последовательность – Последовательность ДНК с уже определенной структурой
16. Риды (reads) – фрагменты ДНК, считываемые секвенатором
17. рРНК – рибосомальная РНК
18. Скаффолд – последовательность контигов, расположенных и ориентированных в порядке их расположения на (теоретической) хромосоме
19. тРНК – транспортная РНК
20. ЭТЦ – дыхательная цепь переноса электронов

## СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Зеленин А.В., Бадаева Е.Д., Муравенко О.В. – Введение в геномику растений/ А.В. Зеленин, Е.Д. Бадаева, О.В. Муравенко// Молекулярная биология. – 2001. – Т. 35. – С. 339-348
2. Sanger F., Nicklen S., Coulson A. – DNA sequencing with chain-terminating inhibitors./F. Sanger, S. Nicklen, A. Coulson// Proceedings of the National Academy of Sciences USA. – 1977. – V. 74. – №. 12. – P. 5463-5467.
3. Cronin M., Ross J. S. – Comprehensive next-generation cancer genome sequencing in the era of targeted therapy and personalized oncology./ M. Cronin, J. S. Ross// Biomarkers in Medicine. – 2011. – V. 5. – №. 3. – P. 293–305.
4. Meldrum C., Doyle M. A., Tothill R W. – Next-generation sequencing for cancer diagnostics: a practical perspective./ C. Meldrum, M. A. Doyle, R. W. Tothill// Clinical Biochemistry Reviews. – 2011. – V. 32. – №. 4. – P. 177–195.
5. Desai A. N., Jere A. – Next-generation sequencing: ready for the clinics?/A N. Desai, A. Jere // Clin Genet. – 2012. – V. 81. – №. 6. – P. 503–510.
6. Ku C. S., Wu M., Cooper D. N. et al. – Technological advances in DNA sequence enrichment and sequencing for germline genetic diagnosis./C. S. Ku, Wu M., D. N. Cooper// Expert Review of Molecular Diagnostics. – 2012. – V. 12. – №. 2. – P. 159–173.
7. Rizzo J. M., Buck M. J. K. – Principles and clinical applications of “next-generation” DNA sequencing/ J. M. Rizzo, M. J. Buck// Cancer Prevention Research. – 2012. – V. 5. – №. 7. – P. 887–900.
8. Berglund E. C., Kiialainen A., Syvänen A. C. – Next-generation sequencing technologies and applications for human genetic history and forensics./ E. C. Berglund, A. Kiialainen, A. C. Syvänen// Investigative Genetics. – 2011. – V. 2. – P. 23.
9. Quail M. A., Smith M., Coupland P. et al. – A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences

and Illumina MiSeq sequencers./ M. A. Quail, M. Smith, P. Coupland et al. // BMC Genomics. – 2012. – V. 13. – P. 341.

10. Якушкина Н.И. – Физиология растений/ Н.И. Якушкина// изд. Владос. – 2004.

11. Kolesnilov A.A., Gerasimov E.C. – Diversity of mitochondrial genome organization/ A.A. Kolesnilov, E.C. Gerasimov// Biochemistry. – 2012. – V. 77. – I. 13. – P. 1424–1435.

12. Михайлова Ю. В., Терентьева Л. Ю. – Гигантские митохондриальные геномы высших растений/ Ю. В. Михайлова, Л. Ю. Терентьева// Успехи современной биологии. – 2017. – Т. 137. – № 3. – с. 237–246.

13. Sloan D. B. – One ring to rule them all? Genome sequencing provides new insights into the ‘master circle’ model of plant mitochondrial DNA structure/ D. B. Sloan// New phytologist. – 2013. – V. 200. – I. 4. – P. 978–985.

14. Щербань А.Б. – Повторяющиеся последовательности ДНК в геномах/ А.Б. Щербань// Вавиловский журнал генетики и селекции. – 2014. – Т. 18. – № 4/1.

15. Jackman S. D., Warren R. L., Gibb E. A. and et. al. – Organellar Genomes of White Spruce (*Picea glauca*): Assembly and Annotation/ S. D. Jackman, R. L. Warren, E. A. Gibb and et. al.// Genome Biology and Evolution. – 2016. – V. 8. – I. 8. – P. 29-41.

16. Chaw S., Shih A. C., Wang D. et. al. – The Mitochondrial Genome of the Gymnosperm *Cycas taitungensis* Contains a Novel Family of Short Interspersed Elements, Bpu Sequences, and Abundant RNA Editing Sites/ S. Chaw, A. C. Shih, D. Wang and et. al.// Molecular Biology and Evolution. – V. 25. – I. 3. – P. 603–615.

17. Nystedt B., Street N. R., Wetterbom A. and et. – The Norway spruce genome sequence and conifer genome evolution/ B. Nystedt, N. R. Street, A. Wetterbom and et. al.// Nature. – 2013. – V. 497. – P. 579–584.

18. Langmead B., Salzberg S.L. – Fast gapped-read alignment with Bowtie 2. / B. Langmead, S.L. Salzberg// Nature Methods. – 2012. V. 4. – I. 9. – P. 357–359.
19. Johnson M., Zaretskaya I., Raytselis Y., Merezhuk Y., McGinnis S., Madden T.L – NCBI BLAST: a better web interface/ M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, T.L. Madden// Nucleic Acids Reseach – 2008. – I. 36(Web Server issue):W5–W9;.
20. Tarailo-Graovac M., Chen N. – Using RepeatMasker to identify repetitive elements in genomic sequences/ M. Tarailo-Graovac, N. Chen.// Current Protocols in Bioinformatics. – 2009. – Ch. 4. – U. 4.10.
21. Abrusán G., Grundmann N., De Mester L., Makalowski W. – TEclass- a tool for automated classification of unknown eukaryotic transposable elements/ G. Abrusán, N. Grundmann, DeMester L, W. Makalowski// Bioinformatics. – 2009. – V. 25. – I. 10. – P. 1329-1330.
22. O'Leary N.A., Wright M.W., Brister J.R., Ciufu S. and et.al. – Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation/ N.A. O'Leary, M.W. Wright, J.R. Brister, S. Ciufu// Nucleic Acids Res. – 2016. – V. 4. – I. 44. – P. 733-745.
23. Sparks R.B., Dale R.M.K. – Characterization of H-labeled supercoiled mitochondrial DNA from tobacco suspension culture cells/ R.B. Sparks, R.M.K. Dale// Molecular Genetics and Genomics. – 1980. – V. 180. – I. 2. – P. 351-355.
24. Palmer J.D., Shields C.R. – Tripartite structure of the *Brassica campestris* mitochondrial genome/J.D. Palmer, C.R. Shields//Nature. – 1984. – V. 307. – P. 437-440.
25. Palmer J.D., Herbon L.A. – Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence/ J.D. Palmer, L.A. Herbon// journal of molecular evolution. – 1988. – V. 180. – I. 2. – P. 351-355.
26. Lonsdale D.M., Hodge T.P., Fauron C.M. – The physical map and organisation of the mitochondrial genome from the fertile cytoplasm of

maize./ D.M. Lonsdale, T.P. Hodge, C.M.Fauron// *Nucleic Acids Research*. – 1984. – V. 12. – P. 9249–9261.

27. Palmer J.D., Shields C.R. – Tripartite structure of the *Brassica campestris* mitochondrial genome/ J.D. Palmer, C.R. Shields// *Nature*. – 1984. – P. 437–440.

28. Zimin A., Stevens K.A., Crepeau M.W., et al. – Sequencing and Assembly of the 22-Gb Loblolly Pine Genome/ A. Zimin, K.A. Stevens, M.W. Crepeau and et al.// *Genetics*. – 2014. – V. 196. – I. 3. – P. 875-890.

Behura S.K., Severson D.W. – Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes/ S.K. Behura, D.W. Severson// *Biological reviews of the Cambridge Philosophical Society*. – 2013. – V. 88. – I. 1. – P. 49-61.

29. Hershberg R., Petrov D.A. – Selection on Codon Bias/ D.A. Petrov, R. Hershberg// *The Annual Review of Genetics*. – 2008. – V. 42. – P. 287-299.

30. Akashi H. – Gene expression and molecular evolution/ H. Akashi// *Curr Opin Genet Dev*. 2001. – V. 11. – P. 660-666.

31. Akashi H, Eyre-Walker A. – Translational selection and molecular evolution/ H. Akashi. A. Eyre-Walker // *Current Opinion in Genetics and Development*. – 1998. – V. 8. – P. 688-893.

32. Sharpton T.J., Neafsey D.E., Galagan J.E., Taylor J.W. – Mechanisms of intron gain and loss in *Cryptococcus* / T.J. Sharpton, D.E. Neafsey, J.E. Galagan, J.W. Taylor// *Genome Biology*. – 2008. – V. 9. – I. 1. – R.24.

33. Qin Z., Cai Z., Xia G., Wang M. – Synonymous codon usage bias is correlative to intron number and shows disequilibrium among exons in plants/ Z. Qin, Z. Cai, G. Xia, M. Wang// *BMC Genomics*. – 2013. – V. 14. – I. 56.

34. Qi Y., Xu W., Xing T., Zhao M., Li N., Yan L., et al. – Synonymous codon usage bias in the plastid genome is unrelated to gene structure and shows evolutionary heterogeneity/ Y. Qi, W. Xu, T. Xing, M. Zhao, N. Li, L. Yan, et al. // *Evolutionary Bioinformatics* – 2015. – V. 11. – P. 65–77.



35. Pitsch N.T., Witsch B., Baier M. – Comparison of the chloroplast peroxidase system in the chlorophyte *Chlamydomonas reinhardtii*, the bryophyte *Physcomitrella patens*, the lycophyte *Selaginella moellendorffii* and the seed plant *Arabidopsis thaliana*/ N.T. Pitsch, B. Witsch, M. Baier// BMC Plant Biology. – 2010. – V. 10. – I. 133.
36. Gabrielsen T.M., Minge M.A., Espelund M., Tooming-Klunderud A., Patil V., Nederbragt A.J., et al. – Genome Evolution of a Tertiary Dinoflagellate Plastid./ T.M. Gabrielsen, M.A. Minge, M. Espelund, A. Tooming-Klunderud, V. Patil, A.J. Nederbragt et al.// PLoS One. – 2011. – V. 6. – I. 4.
37. Sanchez-Puerta M.V., Abbona C.C. – The Chloroplast Genome of *Hyoscyamusniger* and a Phylogenetic Study of the Tribe Hyoscyameae (Solanaceae)./ M.V. Sanchez-Puerta, C.C. Abbona// PLoS One. – 2014. – V. 9. – I. 5.
38. Sharp P.M., Li W.H. – The codon adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications./ P.M. Sharp, W.H. Li// Nucleic Acids Reseach. – 1987. – V. 5. – P. 1281–1295.
39. Wolfe K.H., Li W.H., Sharp P.M. – Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs./ K.H. Wolfe, W.H. Li, P.M. Sharp// PloS one. – 1987. – V. 84. – I. 24. – P. 9054–9059.
40. Palmer J.D., Herbon L.A. – Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence./ J.D. Palmer, L.A. Herbon// Journal of Molecular Evolution. – 1988. – V. 28. – I. 1-2. – P. 87-97.
41. Drouin G., Daoud H., Xia J. – Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. / G. Drouin, H. Daoud, J. Xia // Molecular Phylogenetic Evolution.. – 2008. – V. 49. – I. 3. – P. 827–31.
42. Richardson A.O., Rice D.W., Young G.J., Alverson A.J., Palmer J.D. – The “fossilized” mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation

rate./ A.O. Richardson, D.W. Rice, G.J. Young, A.J. Alverson, J.D. Palmer// *BMC Biology*. – 2013. – V. 11. – I. 29.

43. Kubo T., Nishizawa S., Sugawara A., Itchoda N., Estiati A., Mikami T. – The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNACys(GCA)/ T. Kubo, S. Nishizawa, A. Sugawara, N. Itchoda, A. Estiati, T. Mikami// *Nucleic Acids Reseach*. – 2000. – V. 28. – I. 13. – P. 2571–2576.

44. Satoh M., Kubo T., Nishizawa S., Estiati A., Itchoda N., Mikami T. – The cytoplasmic male-sterile type and normal type mitochondrial genomes of sugar beet share the same complement of genes of known function but differ in the content of expressed ORFs./M. Satoh, T. Kubo, S. Nishizawa, A. Estiati, N. Itchoda, T. Mikami // *Molecular Genetics & Genomics*. – 2004. – V. 272. – I. 3. – P. 247-256.

45. Maréchal A, Brisson N. – Recombination and the maintenance of plant organelle genome stability./ A. Maréchal, N. Brisson// *New Phytol*. – 2010. – V. 186. – I. 2. – P. 299–317; doi: 10.1111/j.1469-8137.2010.03195.x.

46. Cole L.W., Guo W., Mower J.P., Palmer J.D. – High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants/ L.W. Cole, W. Guo, J.P. Mower, J.P. Palmer// *Molecular Biology Evolution* – 2018. – V. 35. – I. 11. – P. 2773-2785.

47. Wynn E.L., Christensen A.C. – Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution./ E.L. Wynn, A.C. Christensen// *G3 (Bethesda)*. – 2019. – V. 9. – I. 2. – P. 549-559.

48. Skippington E., Barkman T.J., Rice D.W., Palmer J.D. – Miniaturized mitogenome of the parasitic plant *Viscumcurreloideum* is extremely divergent and dynamic and has lost all nadgenes./ E. Skippington, T.J. Barkman, D.W. Rice, J.D. Palmer// *Proceedings of the National Academy of Sciences*. – 2015. – V. 112. – I. 27. – P. 3515-3524.

49. Rice D.W., Alverson A.J., Richardson A.O., Young G.J., Sanchez-Puerta M.V., Munzinger J., et al. – Horizontal transfer of entire genomes via

mitochondrial fusion in the angiosperm *Amborella*./ D.W. Rice, A.J. Alverson, A.O. Richardson, G.J. Young, M.V. Sanchez-Puerta, J. Munzinger et al// *Science*. – 2013. – V. 342. – I. 6165. – P. 1468-1473.

50. Shearman J.R., Sangsrakru D., Ruang-Areerate P., Sonthirod C., Uthapaisanwong P., Yoocha T., et al. – Assembly and analysis of a male sterile rubber tree mitochondrial genome reveals DNA rearrangement events and a novel transcript./ J.R. Shearman, D. Sangsrakru, P. Ruang-Areerate, C. Sonthirod, P. Uthapaisanwong, T. Yoocha et al.// *BMC Plant Biology*. – 2014. – V. 14. – I. 45.

51. Backert S., Lynn N. B, Börner T. – The mystery of the rings: structure and replication of mitochondrial genomes from higher plants./ S. Backert, B. L. Nielsen, T. Börner // *Trends Plant Science*. – 1997. – V. 2. – I. 12. – P. 477–83.

52. Klein M., Eckert-Ossenkopp U., Schmiedeberg I., Brandt P., Unseld M., Brennicke A., et al. – Physical mapping of the mitochondrial genome of *Arabidopsis thaliana* by cosmid and YAC clones./ M. Klein, U. Eckert-Ossenkopp, I.Schmiedeberg, P. Brandt, M. Unseld, A. Brennicke et al.// *Plant J*. – 1994. – V. 6. – I. 3. – P. 447-455.

53. Shearman J.R., Sonthirod C., Naktang C., Pootakham W., Yoocha T., Sangsrakru D., et al. - The two chromosomes of the mitochondrial genome of a sugarcane cultivar: assembly and recombination analysis using long PacBio reads/ J.R. Shearman, C. Sonthirod, C. Naktang, W. Pootakham, T. Yoocha, D. Sangsrakru et al. // *Scientific Reports*. – 2016. – V. 6.

54. Shedge V., Arrieta-Montiel M., Christensen A.C., Mackenzie S.A. – Plant mitochondrial recombination surveillance requires unusual RecA and MutShomologs./ V. Shedge, M. Arrieta-Montiel, A.C. Christensen, S.A. Mackenzie // *Plant Cell*. – 2007. – V. 19. – I. 4. – P. 1251–1264.

55. Arrieta-Montiel M.P., Shedge V., Davila J., Christensen A.C., Mackenzie S.A. – Diversity of the *Arabidopsis* mitochondrial genome occurs via nuclear-controlled recombination activity./ M.P. Arrieta-Montiel, V. Shedge, J. Davila, A.C. Christensen, S.A. Mackenzie// *Genetics*. – 2009. – V. 183. – I. 4. – P. 1261–1268.

56. Gray M. W. and Covello P. S. – (1993) RNA editing in plant mitochondria and chloroplasts/M. Gray, P.S. Covello // the FACEB. – 1993. – V. 7. – P. 64–71.
57. Brennicke A., Marchfelder A., Binder S. – RNA editing/ A. Brennicke, A. Marchfelder, S. Binder// FEMS Microbiology Reviews. – V. 23. – I. 3. – P. 297–316.
58. Hammani K., Okuda K., Tanz S.K., Chateigner-Boutin A.L., Shikanai T., Small I. – General features of chloroplast RNA editing factors and their target sites gained from a study of new Arabidopsis editing mutants./ K. Hammani, K. Okuda, S.K. Tanz, A.L. Chateigner-Boutin, T. Shikanai, I. Small// Plant Cell. – 2009. – V. 21. – I. 11. – P. 3686–3699.
59. Kotera E., Tasaka M., Shikanai T. A – Pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. / E. Kotera, M. Tasaka, T.A. Shikanai // Nature. – 2005. – T. 433. – C. 326-30.
60. Schmitz-Linneweber C., Small I. – Pentatricopeptide repeat proteins: a socket set for organelle gene expression/ C. Schmitz-Linneweber, I. Small// Trends Plant Science. – 2008. – T. 1. – C. 663-70.
61. Aubourg S., Boudet N., Kreis M., Lecharny A. – In Arabidopsis thaliana, 1% of the genome codes for a novel protein family unique to plants./ S. Aubourg, N. Boudet, M. Kreis, A. Lecharny// Plant Molecular Biology – 2000. – T. 42. – C. 603-613.
62. Chateigner-Boutin A.L., Small I. – Plant RNA editing, RNA / A.L. Chateigner-Boutin, I. Small // RNA Biology– 2010. – T.7. – B. 2. – C. 213-219.
63. Guo W., Grewe F., Fan W., Young G.J., Knoop V., Palmer J.D., Mower J.P. – *Ginkgo* and *Welwitschia* Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial Evolution / W. Guo, F. Grewe, W. Fan, G.J. Young, Knoop V., J.D. Palmer, J.P. Mower// Molecular Biology and Evolution. – 2016. – V. 33. – I. 6. – P. 1448–1460.

64. A. Kozik, B. A. Rowan, D. Lavelle et al. – The alternative reality of plant mitochondrial DNA/ Kozik A., Rowan B.A., Lavelle D. et al// bioRxiv. - 2019.
65. K. Sahlin, Vezzi F, Nystedt B., Lundeberg J., ArvestadL. – BESST - Efficient scaffolding of large fragmented assemblies/ Sahlin K., F.Vezzi, B.Nystedt, J. Lundeberg, L.Arvestad// BMC Bioinformatics. – 2014. – T. 15. – C. 281.
66. D. Paulino, R.L. Warren, B.P. Vandervalk, A. Raymond, S.D. Jackman, I. Birol – Sealer: a scalable gap-closing application for finishing draft genomes/ Paulino D., Warren R.L., Vandervalk B.P., Raymond A., Jackman S.D., Birol I.// BMC Bioinformatics. – 2015. – T. 26. – C. 230.

## 6. ПРИЛОЖЕНИЕ

В приложении приведены таблицы значений CAI (codon adaptation index) для различных видов голосеменных растений.

Таблица А. 1 – CAI для *P.taeda*

Ген	Длина в а/к	Кол-во экзонов	CAI
atp9	83	1	0,858532
rps19	94	1	0,837367
NAD4L	101	1	0,795576
rps14	101	1	0,83663
nad3	119	1	0,838461
rps12	126	1	0,809736
sdh3	132	1	0,829847
rps13	145	1	0,838373
rpl16	153	1	0,833408
atp8	161	1	0,840426
rps11	171	1	0,839507
rpl5	192	1	0,865331
rps1	195	1	0,890452
rps7	196	1	0,826305
nad9	197	1	0,884076
nad6	205	1	0,860863
ccmB	220	1	0,871736
rps2	240	1	0,865185
ccmC	240	1	0,907022
mttB	250	1	0,885687
atp6	259	1	0,911716
cox3	266	1	0,87276
cox2	267	2	0,884302
atp4	316	1	0,917451
rps4	351	1	0,921003
cob	398	1	0,924477
rpl2	477	2	0,943617
atp1	508	1	0,942585
cox1	524	1	0,945848
ccmFN	575	1	0,96003
matR	727	1	0,966306
rps10	802	1	0,979887
nad1	958	3	0,979413

nad7	1259	5	0,988681
ccmFC	1335	1	0,988434
nad5	1341	2	0,985371
nad2	1647	3	0,992036
nad4	2272	2	0,993672
rps3	2619	1	0,993841

Таблица А. 2 – САІ для *G. biloba*

Ген	Длина в а/к	Кол-во экзонов	САІ
rrn5	38	1	0,836988
atp9	75	1	0,868549
rps19	94	1	0,84235
rps14	101	1	0,821988
nad4L	101	1	0,845881
rps10	108	1	0,850538
nad3	119	1	0,826604
rps12	126	1	0,821222
rps13	127	1	0,851736
sdh4	132	1	0,843887
sdh3	134	1	0,847972
rpl16	137	1	0,852058
atp8	163	1	0,859396
rps11	173	1	0,895699
rpl5	190	1	0,867291
atp4	198	1	0,870688
nad9	198	1	0,887834
rps1	198	1	0,833711
nad6	202	1	0,89123
rps7	207	1	0,887018
ccmB	213	1	0,886622
rps2	232	1	0,890214
ccmC	237	1	0,877126
cox2	255	3	0,847046
atp6	256	1	0,920705
mttB	260	1	0,894314
cox3	266	1	0,878315
nad1	326	3	0,919705
rps4	347	1	0,905572
nad7	395	5	0,908879
Cob	399	1	0,908103

ccmFc	462	2	0,953505
rpl2	480	2	0,954432
nad2	491	5	0,93129
nad4	496	4	0,94181
atp1	509	1	0,948099
cox1	521	1	0,942033
rps3	579	3	0,966075
ccmFn	605	1	0,959855
rrn18	613	1	0,959769
nad5	668	2	0,95364
matR	710	1	0,970627

Таблица А. 3 – CAI для *C. taitungensis*

Ген	Длина в а/к	Кол-во экзонов	CAI
atp9	75	1	0,879
rps19	94	1	0,8
nad4L	101	1	0,824
rps14	101	1	0,822
rps10	105	1	0,807
nad3	109	1	0,89
rps12	124	1	0,874
rps12	126	1	0,858
sdh3	134	1	0,862
rps13	151	1	0,87
atp8	163	1	0,848
rpl16	178	1	0,848
rpl15	193	1	0,84
rps11	194	1	0,876
nad9	198	1	0,882
atp4	198	1	0,87
nad6	202	1	0,879
ccmB	213	1	0,868
rps2	217	1	0,862
rps7	232	1	0,868
ccmC	240	1	0,875
atp6	256	1	0,895
mttB	260	1	0,887
cox3	266	1	0,862
cox2	269	1	0,882
nad5	326	5	0,916



rps4	349	1	0,916
nad7	385	5	0,9
cob	399	1	0,919
ccmFC	464	1	0,952
rpl2	478	2	0,946
nad2	491	0	0,941
nad4	496	4	0,936
atp1	509	1	0,934
cox1	524	1	0,938
rps3	580	3	0,96
ccmFN	631	1	0,96
matR	705	1	0,97
nad7	1336	5	0,976

Таблица А. 4 – САІ для *L.sibirica*

Ген	Длина в а/к	Кол-во экзонов	САІ
nad7	66	5	0,79
atp9	81	1	0,864
nad5	85	5	0,8
rps19	99	1	0,822
atp4	99	1	0,814
rps14	101	1	0,829
nad3	101	2	0,807
rps10	112	1	0,847
rps12	126	1	0,815
rps13	126	1	0,867
rpl16	145	1	0,828
sdh3	156	1	0,828
rps11	170	1	0,811
nad9	195	1	0,897
rps7	195	1	0,87
ccmB	212	1	0,878
nad6	214	1	0,869
nad3	219	2	0,875
rps2	222	1	0,871
rpl2	235	2	0,884
ccmC	240	1	0,901
atp6	243	1	0,911
mttB	255	1	0,88
cox3	266	1	0,88

rps4	350	1	0,923
cox2	385	1	0,913
Cob	398	1	0,919
atp8	409	1	0,915
nad2	488	5	0,926
atp1	508	1	0,935
nad4	515	4	0,937
cox1	524	2	0,941
rps3	563	1	0,961
ccmFn	576	1	0,96
rps1	690	1	0,885
matR	732	1	0,969
ccmFc	768	2	0,965

Таблица А. 5 – САІ для *P.sibirica*.

Ген	Длина в а/к	Кол-во экзонов	САІ
atp9	83	1	0,838
rps19	94	1	0,842
rps14	101	1	0,837
rps10	105	1	0,823
nad3	119	1	0,809
rps12	126	1	0,808
rps13	130	1	0,851
sdh4	131	1	0,861
sdh3	140	1	0,83
atp8	161	1	0,843
rpl16	162	1	0,834
nad4L	166	1	0,856
rps11	171	1	0,838
nad6	180	1	0,845
rpl5	192	1	0,868
atp4	192	1	0,844
rps7	196	1	0,819
nad9	196	1	0,877
ccmB	220	1	0,874
rps2	223	1	0,855
rps1	229	1	0,876
ccmC	240	1	0,906
mttB	249	1	0,889
cox2	255	3	0,876

cox3	266	1	0,873
atp6	283	1	0,916
rps4	351	1	0,922
nad7	397	5	0,936
cob	398	1	0,923
ccmFc .	466	1	0,955
ccmFc	466	2	0,954
rpl2	475	2	0,94
nad2	490	5	0,934
nad4	496	4	0,936
atp1	508	1	0,942
cox1	532	1	0,948
rps3	557	1	0,963
ccmFn	571	1	0,957
MatR	701	1	0,966

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт фундаментальной биологии и биотехнологии  
Кафедра биофизики

УТВЕРЖДАЮ

Заведующий кафедрой

Иванова В.А. Крагасюк В.А.

подпись                      инициалы, фамилия

«20»                      июня                      2019 г.

**БАКАЛАВРСКАЯ РАБОТА**

06.03.01 Биология

Выявление и анализ митохондриальных последовательностей сосны  
сибирской кедровой (*Pinus sibirica*)

Руководитель:

  
подпись

Путинцева Ю.А.

инициалы, фамилия

Научный  
консультант

  
подпись

д-р физ.-мат. наук М. Г. Садовский

инициалы, фамилия

Выпускник:

  
подпись

Кириченко А.Д.

инициалы, фамилия

Красноярск 2019