

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
И.Е. Ямских
«__» июня 2018 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

06.04.01 Биология
06.04.01.06 Геномика и биоинформатика

ПОИСК И АНАЛИЗ ПАЛИНДРОМНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В
ГЕНОМЕ ЛИСТВЕННИЦЫ СИБИРСКОЙ

Выпускник: ББ16-06Б, 041625623 _____ Тараненко Е.А.
подпись

Руководитель: д. ф.–м. н., в.н.с. _____ Садовский М.Г.
подпись

Рецензент: д.б.н. Орлов Ю.Л.

Красноярск 2018

АННОТАЦИЯ

Повторяющиеся элементы играют важную роль в структуре генома: они являются регуляторными элементами, структурируют геном, участвуют в регуляции экспрессии генов. Их изучение играет ключевую роль в понимании детальных механизмов функционирования генома в целом. В магистерской диссертации проводился поиск повторяющихся элементов в лиственнице сибирской, сравнение этих последовательностей с уже известными повторами и описание статистических свойств повторов.

Магистерская диссертация по теме «Поиск и анализ последовательностей в геноме лиственницы сибирской» содержит 39 страниц текстового документа, 3 иллюстрации, 11 таблиц, 3 диаграммы, 3 приложения, 14 использованных источника.

Ключевые слова: повторяющиеся элементы, повторы, палиндром, геном, лиственница сибирская.

Содержание

АННОТАЦИЯ	2
ВВЕДЕНИЕ	4
1 Обзор литературы	6
1.1 Структура генома хвойных	6
1.2 Повторяющиеся последовательности в геномах хвойных	7
1.2 CRISPR	8
1.3 Строение локусов CRISPR	8
1.4 Онлайн ресурсы для поиска CRISPR	9
1.5 Палиндромные повторы	10
1.6 Редакционное расстояние	10
2 Материалы и методы	12
2.1 Материалы	12
3 Результаты	15
3.1.1 Проверка на присутствие известных CRISPRs	15
3.1.2 Повторы и палиндромы	16
3.1.3 Проверка BLAST	24
3.2 Вторичная сборка	24
3.2.1 Выборка контигов	24
3.2.2 Проверка на присутствие известных CRISPR	25
3.2.3 Повторы и палиндромы	25
3.2.4 Частота повторов в сборке	26
3.2.5 Проверка BLAST'ом на NCBI библиотеку	27
3.2.5.1 Выравнивание контигов на базу NCBI	28
3.2.5.2 Повторы на базу NCBI	32
4 Обсуждение	37
ЗАКЛЮЧЕНИЕ	39
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	40
ПРИЛОЖЕНИЕ	42

ВВЕДЕНИЕ

Данная работа посвящена поиску палиндромных повторов и повторов с параметрами, аналогичными повторам CRISPR, в геноме лиственницы сибирской (*Larix sibirica* Ledeb.). Повторяющиеся элементы играют важную роль в структуре генома: они являются регуляторными элементами, структурируют геном, участвуют в регуляции экспрессии генов. Их изучение играет ключевую роль в понимании детальных механизмов функционирования генома в целом. Такие элементы крайне слабо изучены для геномов хвойных, что и определяет актуальность диссертационного исследования.

Данная работа была выполнена под руководством д. ф.-м. н., в.н.с. ИВМ СО РАН Садовского М.Г. 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)

Целью работы является изучение особенностей повторяющихся элементов в геноме лиственницы сибирской.

Для достижения заданной цели были поставлены следующие задачи:

- поиск повторов определенной структуры в контигах заданной длины;
- сравнение выделенных повторов с аналогами, содержащимися в различных базах данных;
- описание статистических свойств выделенных повторов.

Предметом исследования является геном лиственницы сибирской, ранее полученный в лаборатории лесной геномики.

Автор выражает искреннюю благодарность Орешковой Н.В. за пробоподготовку и секвенирование, Кузьмину Д.А. и Шарову В.В. за сборку и асемблирование генома, Путинцевой Ю.А., Садовскому М.Г. и Крутовскому К.В. за идею исследования и общее руководство на всех этапах работы. Так же

автор выражает благодарность всем членам лаборатории лесной геномики за участие в обсуждении работы и ценные советы по ней.

1 Обзор литературы

1.1 Структура генома хвойных

Хвойные растения имеют очень большие геномы. Их размеры варьируют от 6,5 Гб у Новозеландской *Lepidothamnus intermedius* до 37 Гб у Сосны белой мексиканской (*Pinus ayacahuite*) [1]. Анализ изменения размера их генома указывает на динамическое эволюционное развитие этого таксона, хотя хвойные и считаются древними и даже реликтовыми организмами. Также было замечено недавнее и весьма сильное увеличение генома рода сосновых (*Pinus*) в размере, что не наблюдалось у остальных представителей голосеменных [2]. Сравнительное изучение ортологов ели и сосны выявило заметно более низкую скорость нуклеотидных замен, чем у ортологов цветковых растений. Синонимичные замены и не-синонимичные (смысловые) используются для определения типа отбора (нейтрального и селективного, соответственно). При попарном сравнении ортологов отношение не-синонимичного расстояния (т.е. количество замен на не-синонимичный сайт) к синонимичному расстоянию дает общую оценку модели и силы отбора: при сравнении ортологов цветковых растений с ортологами ели и сосны. Это может говорить о высокой степени адаптации этого таксона, а также об уникальном пути эволюции хвойных [3]. Однако, причина и функция такого большого генома, как у представителей хвойных, до сих пор неизвестны.

Большинство генов в хвойных представлены в небольшом числе копий. Число функциональных генов хвойных сопоставимо с ними же у цветковых растений [4]. Большое число низкокопийных генов, вероятнее всего, есть результат дивергенции в древних повторах и дупликации генов и псевдогенов. Так, например, результаты исследования десяти сосны ВАС-клонов *Pinus taeda* показали, что псевдогены в пять раз более распространены в геноме, чем потенциально белок-кодирующие гены [5]. Из этого следует, что большой объем генома, вероятно, состоит из паралогов и псевдогенов.

1.2 Повторяющиеся последовательности в геномах хвойных

При анализе больших геномов хвойных и других растений встречается очень много повторяющихся последовательностей. Исследователи стараются фокусироваться на участках генома, не относящихся к повторам. Однако, повторы играют важную роль во внутреннем функционировании организма, выполняют регуляторную и структурирующую функции. Они очень распространены в растениях, высоко вариабельны и имеют тенденцию быстро изменяться. Эпигенетическая регуляция так же осуществляется за счет активности повторяющихся элементов. Предполагается, что эпигенетический контроль появился как способ замалчивания повторяющихся участков и вирусов, что в дальнейшем развилось в полномасштабную геномную регуляцию как некодирующих, так и кодирующих участков [6]. Также оспаривается предположение о случайном расположении повторов в геноме: повторы, в зависимости от их вида, располагаются либо вблизи генов, либо в отдалении. Однако, непонятно, какое влияние первично – локация повтора на его функцию или функциональность повтора на её локацию [7].

Анализ повторяющихся фрагментов генома хвойных показал, что при строгих параметрах гибридизации только 24% генома сосны *Pinus taeda* почти идентичен. Однако, если ослабить параметры, то до 80% генома является почти идентичным. Это подтверждает теорию о низкой копийности идентичных повторов и высокой копийности групп повторов, чьи копии успели измениться с течением времени. Такие группы повторов включают в себя до 100 тыс. копий одной группы [5]. Ретроэлементы и неохарактеризованные повторы составляют до 75% генома *Picea abies* [8]; еще большее значение было выявлено у *Taxodium distichum* (90%) [9]. Интересно, что такое изменение объема генома в отношении к ретротранспозонам специфично именно для хвойных.

Хлоропластная ДНК так же отличается от встречающейся в цветковых растениях. Так, наследование хлоропластов у хвойных проходит по отцовской линии, тогда как у цветковых наследование проходит преимущественно по материнской линии [10]. Объем хлоропластов ДНК хвойных меньше в объеме,

чем у цветковых: в них недостает некоторых элементов, свойственных наземным растениям.

Очевидно, повторов в геноме объекта данного исследования так же много, как и в других хвойных. Было решено работать в направлении таких повторов, по структуре напоминающих последовательности CRISPR в виду повышенного интереса к самим CRISPR.

1.2 CRISPR

Короткие палиндромные повторы, регулярно расположенные группами (CRISPRs) – это такие последовательности ДНК бактерий, архей и плазмид, состоящие из прямых повторяющихся последовательностей, разделенных уникальными последовательностями (спейсерами). На сегодняшний день CRISPR-*Cas* система является наиболее распространенным инструментом генетической модификации, где *Cas* – белок, специфичный CRISPR (CRISPR-associated protein). Она позволяет направленно редактировать геном с высокой точностью.

1.3 Структура локусов CRISPR

Локус CRISPR обычно представлен несколькими элементами (Рис. 1). Первой располагается лидирующая последовательность, преимущественно состоящая из слабых азотистых оснований тимина и аденина, однако, более ничем не консервативна. Далее идет сам массив CRISPR (CRISPR array), состоящий из повторяющихся последовательностей и спейсеров между ними.

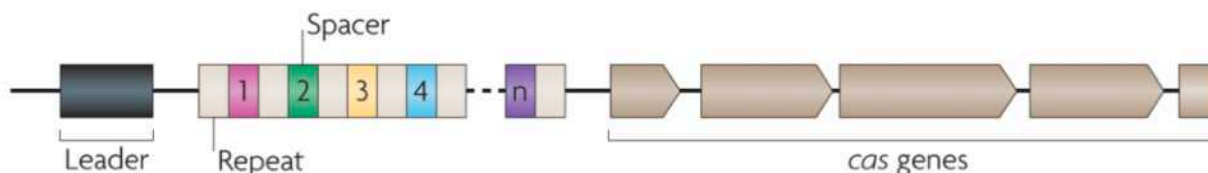


Рис. 1 Структура локуса CRISPR.

Локус отличается от массива CRISPR наличием *cas* генов. Элементы CRISPR: лидирующая последовательность (черный); повторяющиеся последовательности (серые участки); спейсеры (пронумерованные участки); и *cas* гены.

Характеристики этих двух элементов будут описаны более подробно. Еще один элемент, свойственный исключительно бактериям, археям и плазмидам, – гены *cas*, которые располагаются в непосредственной близости к массиву CRISPR как перед ним, так и после него. Эти гены консервативны, выделены несколько типов и семейств генов, ответственных за кодирование белкового аппарата, составляющего рабочую CRISPR-*cas* систему.

Повторы, или повторяющиеся последовательности (direct repeats, DRs), – последовательности ДНК длиной от 23 до 47 нуклеотидов [11]. Количество повторов может достигать нескольких сотен, но обычно наблюдается до 16 повторов.

Спейсеры – вставки между повторяющимися последовательностями. В первую очередь это уникальные последовательности, близкие по длине, но отличные по нуклеотидной последовательности. Последовательности спейсеров заимствованы из чужеродных генетических элементов, с которыми сталкивался организм (бактериофаг, плазида и т.д.) [12]. Таким образом организм создает генетическую базу для быстрого реагирования в случае вторичной встречи с патогеном.

1.4 Онлайн ресурсы для поиска CRISPR

Для проверки нуклеотидной последовательности на предмет наличия в них известных CRISPRs имеются несколько онлайн-ресурсов, совмещенных с собственными базами данных. Одной из таких является ресурс CRISPRFinder (<http://crispr.i2bc.paris-saclay.fr/Server/>), принимающий на вход FASTA файл с последовательностью нуклеотидов анализируемой последовательности. Еще один аналогичный ресурс CRF: CRISPR Finder by Random Forest (<http://bioinfolab.miamioh.edu/crf/home.php>) так же имеет свою базу данных и производит поиск повторов относительно неё. Оба ресурса имеют свои специфичные базы повторов, и способны находить только уже известные повторы, классифицируя их как “проверенные” и “предполагаемые” (не прошедшие проверку BLAST или найденные только в одном организме).

1.5 Палиндромные повторы

Комплементарные палиндромные повторы, или комплементарные палиндромы, – это такие последовательности ДНК, состоящие из инвертированных повторов, одинаково считывающихся в прямом направлении одной цепи и в обратном направлении другой цепи.

Совершенным палиндромом называется последовательность из четного числа нуклеотидов, одинаково читающаяся в разных направлениях в комплементарном виде. Так, последовательность 3'-GATC-5' является совершенным палиндромом, так как её комплементарная цепь 5'-CTGA-3' читается одинаково в направлении 3' - 5'.

Несовершенный палиндром – это палиндром, включающий в себя участки, комплементарные друг другу, а также участки, сформированные случайными или некомплементарными нуклеотидами. Такие последовательности позволяют образовываться петлям ДНК, характерным, например, структуре тРНК. Например, 3'-GANNTC-5' последовательность имеет центральную петлю, состоящую из любых трех нуклеотидов. Такие последовательности могут быть как четной, так и нечетной длины.

1.6 Редакционное расстояние

Редакционное расстояние – это мера близости двух строк: это минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую. Выделяют несколько редакционных расстояний: расстояние Хэмминга, расстояние Левенштейна, расстояние Дамерау-Левентшейна и другие. Различие между расстоянием Хэмминга и Левенштейна состоит в том, что расстояние Хэмминга описывает количественно позиции несоответствия, а расстояние Левенштейна позволяет вводить дополнительные модификации – ввод и удаление символа, что, порой, значительно сближает последовательности.

Цена каждой операции имеет свое определенное значение. Так, изначально заданные значения любых операций равны единице. Такая оценка

операций может быть применима относительно более-менее равнозначных операций, таких как вставки и удаления. Однако, на практике известно, что частота замен разных нуклеотидов на отличные сильно варьирует.

Расстояние Хэмминга d_H является наиболее простым и дает информацию о количестве несовпадений между двумя последовательностями. Положение несоответствий никак не отражается в этом значении. Это единственное редакционное расстояние, принимаемое используемым в работе ПО. Оно позволяет заранее установить максимальное количество несовпадений, допустимое между искомыми повторами.

2 Материалы и методы

2.1 Материалы

В работе использовалась сборка лиственницы сибирской *Larix sibirica*, предварительно полученная в Лаборатории лесной геномики СФУ в рамках проекта «Геномные исследования основных бореальных лесообразующих хвойных видов и их наиболее опасных патогенов в Российской Федерации» под руководством К.В. Крутовского.

2.2 Методы

2.2.1 Отбор контигов

Таблица 1 Отобранные контиги. Список контигов представлен в Приложении 1.

Кол-во контигов	25
Максимальная длина, н.о.	265,649
Минимальная длина, н.о.	100,215

Результаты сборки генома выявили 13 млн. 798 тыс. 666 контигов длиной в 8 308 543 397 нуклеотидных оснований. Максимальная длина контига составила 265 тыс. 649 нуклеотидов, значения N50 и N90 составили 1 069 и 232 пар оснований, соответственно. Исходя из приведенной статистики можно сделать вывод, что количество относительно максимального длинных контигов мало, а большая часть контигов оказалось длиной около 1 000 нуклеотидов.

Результаты повторной (улучшенной) сборки генома выявили 11 млн. 326 тыс. 214 последовательностей длиной в 12 342 196 774 нуклеотидных оснований. Максимальная длина последовательности 354 тыс. 326 нуклеотидов, минимальная 120 нуклеотидов.

Исходя из литературных данных о длине и строении CRISPR, длины таких повторов могут достигать 1 000 и более нуклеотидов, следовательно, большая часть имеющихся контигов нам не подходит. Для дальнейшего анализа использовались контиги не менее 100 000 нуклеотидов; в дальнейшем значение понижено до 50 000 нуклеотидов.

2.2.2 Проверка присутствия известных CRISPR'ов

Выше были описаны существующие базы данных с имеющейся информацией по известным и проверенным повторам. В данной работе использовались два онлайн ресурса: CRISPRFinder и CRF. Для поиска повторов в CRISPRFinder были выставлены следующие параметры: 1 Максимальные повторы: длина повтора – 23 -55 н.о.; размер вставки между повторами – 25 - 60 н.о.; количество допустимых несовпадений – 1; 2 свойства CRISPR: размер спейсера – повтор*(0.6 - 2.5); подобие спейсера – 60 %; количество допустимых несовпадений в повторах – 20% от длины повтора.

Для поиска в ресурсе CRF были выставлены следующие параметры: минимальная длина повтора – 20 н.о.; максимальная длина повтора – 55 н.о.; минимальная длина спейсера – 20 н.о.; максимальная длина спейсера – 100 н.о.

2.2.3 Поиск повторов и палиндромов

Для поиска повторяющихся элементов в контигах была выбрана программа REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) [13]. Программа состоит из трех программ: REPfind; REPselect используется для отбора нужных контигов по заданным критериям; и REPvis для визуализации полученных данных.

Программа узнает четыре типа повторов (Рис. 3): прямые (А), обратные (В), комплементарные (С), и палиндромные (D).

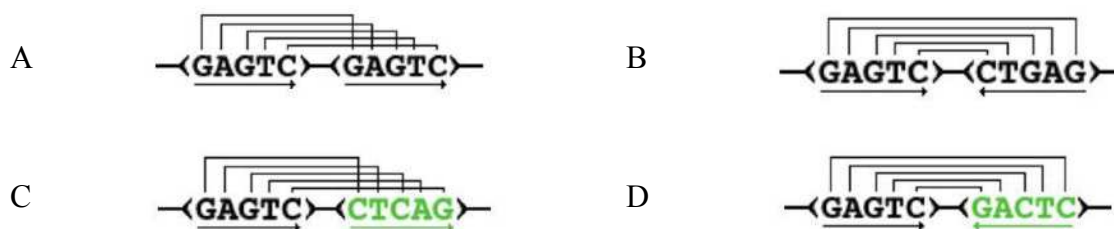


Рис. 2 Схема разных типов повторов, узнаваемых программой REPuter: А – прямые, В – обратные, С – комплементарные, D – палиндромные. Зеленым выделены элементы комплементарной цепи.

Для первоначального запуска были выбраны все типы повторов. Другие характеристики: мин. длина повтора – 20 н.о.; расстояние Хэмминга – 0 (т.е.,

идентичные повторы); макс. анализируемое кол-во повторов – 50. В дальнейшем рассматривались только прямые и палиндромные повторы.

В результате работы программы получались пары повторов с заданным количеством несоответствий, значением e-value и координатами повторов в контигах. Значение e-value (expected value) характеризует уникальность найденного повтора относительно общей последовательности. Чем меньше значение e-value, тем менее вероятно нахождение подобного повтора в последовательности, и, наоборот, чем оно больше, тем больше подобных повторов в последовательности. Данное значение сильно зависит от длины искомого повтора. Так, чем меньше длина повтора тем выше будет e-value. Это связано с тем, что короткие слова встречаются в последовательности много чаще, чем длинные слова, и, чем длиннее искомый повтор, тем более вероятна его уникальность.

Отбор повторов в дальнейшем проходил по нескольким критериям: отношение длины повтора к длине спейсера и палиндромность повтора (только для прямых повторов).

2.2.4 Проверка BLAST

Предполагается, что спейсеры выступают в роли генетической памяти. Информация о нуклеотидной последовательности, заключающаяся в спейсере, может либо подтвердить, что этот спейсер принадлежит CRISPR, либо отвергнуть это предположение. Выравнивание проводилось на локальную копию базы нуклеотидных последовательностей NCBI, доступную на сервере лаборатории лесной геномики.

3 Результаты

3.1 Первая сборка

Для работы были отобраны 25 уникальных контигов (Таблица 2, Прил. 1) из первоначальной сборки. Проведенный анализ повторов выявил одинаковые повторы в 23 контигах, после чего зрительный просмотр контигов выявил высокое сходство участков этих контигов, указывая на возможность усовершенствования сборки.

3.1.1 Проверка на присутствие известных CRISPRs

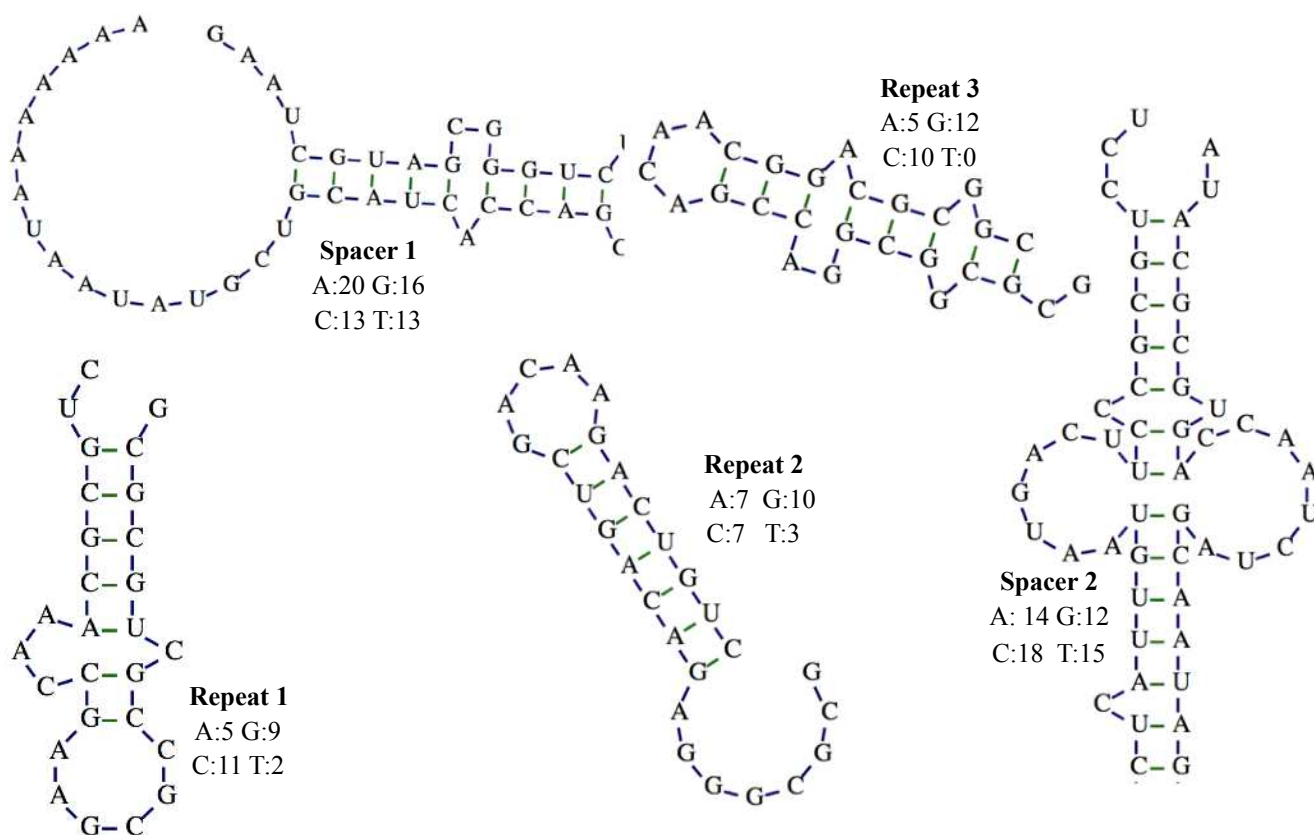


Рис. 3 Вторичная структура CRISPR, представленная отдельно для повторяющихся участков и спейсеров.

Structure	Sequence	Length, bp
Repeat 1	CTGCGCAAACCGAAGCGCCGCTGCGCG	27
Spacer 1	AAAAAAATAATATGCTGCATCACCAGCGCTGCCATGATGGCGAGATTCTGGGCGATGCTAAG	62
Repeat 2	CTGTCAGAACAGCTGACAGAGGGCGCG	27
Spacer 2	TCCTGCGCCCTTCAGTAATGTTACTCGCTGACGCGATAACGATCTAACCAGTGCGCATA	59
Repeat 3	CGGCGCAGGCAACAGCCAGGCGGCGCG	27

Изъята 21 страница.

4 Обсуждение

Анализ выявленных повторов с помощью BLAST'a позволил выявить интересную закономерность. Ассоциация повтора с последовательностями хвойных и митохондрий коррелирует с высоким числом копий повтора в геноме. Согласно теоретическим данным [7], для хвойных характерно низкое число копий сильно похожих последовательностей и высокое число копий дивергированных со временем последовательностей, однако, в исследованных повторах с расстоянием Хэмминга равным единице и двум, число копий соответствовало последовательностям с большим расстоянием Хэмминга. Две последовательности длиной в 20 – 24 нуклеотида, различающиеся на 1 – 2 нуклеотида сложно назвать сильно дивергированными, из чего следует, что, возможно, это исключение из общего правила для хвойных.

Интересно, что частота встречаемости повторов, выровнявшихся на бактериальные организмы, очень низка. Это может свидетельствовать о низком бактериальном загрязнении генома, хотя результаты выравнивания контигов на базу BLAST'a выявили девять контигов, полностью состоящих из последовательностей, выровнявшихся на бактериальные организмы. С другой стороны, искомые пары повторов были очень похожи (расстояние Хэмминга равное 1 и 2). Учитывая высокую скорость изменчивости бактериальных геномов можно предположить, что повторы в бактериях так же быстро мутируют и отличаются друг от друга на значения большие, чем расстояние Хэмминга 1 и 2.

Особенно интересным случаем является повтор с очень большим числом копий и отсутствием результатов выравнивания на базу NCBI; парный повтор, однако, с отличием на 2 нуклеотида, выровнялся на животные и бактериальные хиты. Природа такой асимметрии не понятна и требует дальнейшего исследования.

Ситуация с митохондриальными результатами выравнивания на базу BLAST не является неожиданной, так как эти повторы могут входить в состав генов домашнего хозяйства митохондрии и высокое число копий этих участком

митохондрий важно для функционирования органеллы. Результаты выравнивания контигов на базу BLAST'a выявили один контиг с большим числом хитов пластиной ДНК. Теоретически число копий повторов этого контига должно было иметь одинаковый порядок с числом копий повторов с митохондриальными хитами, однако, ни один из повторов из этого контига не подошел по искомым параметрам и их частота не рассчитывалась. В дальнейшем можно продолжить эксперимент с большими значениями расстояния Хэмминга (чем больше редакционное расстояние двух последовательностей тем больше пар повторов анализируется в дальнейшем, тем выше вероятность нахождения в них повторов, подходящих по всем искомым параметрам) и проверить данное предположение.

ЗАКЛЮЧЕНИЕ

22 контига из первоначальной сборки имеют протяженную идентичную часть последовательности, что позволило обнаружить избыточность первоначальной сборки и улучшить повторную сборку генома.

В повторной сборке не было найдено ни одного из подтвержденных CRISPR, что может свидетельствовать об отсутствии бактериального загрязнения.

Были найдены прямые повторы и комплементарные палиндромы в геноме лиственницы сибирской. Найдена асимметрия в частоте встречаемости повторов одного комплекса, требующая дополнительного исследования.

Проверка BLAST'ом выявила большое количество совпадений с последовательностями митохондрий в некоторых контигах, что может быть причиной асимметрии частот повторов. Дальнейшая проверка BLAST'ом самих повторов на базу NCBI выявила закономерность в частоте встречаемости повтора в геноме и результатами BLAST'а, а именно, между высокой копийностью повтора и хитами хвойных и митохондрий. Интересно, что бактериальные хиты не давали такого результата, как и другие хиты. Однако, было и исключение, где без каких-либо найденных хитов число копий повтора было очень высоким. Данное исключение требует дальнейшего исследования.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Plant DNA C-values Database: <http://data.kew.org/cvalues/CvalServlet?querytype=1>* Murray et al., 2010
- 2 JG Burleigh, WB Barbazuk, JM Davis, AM Morse, PS Soltis. Exploring diversification and genome size evolution in extant Gymnosperms through phylogenetic synthesis. *Journal of Botany* 2012, - doi: 10.1155/2012/292857
- 3 E Buschiazzo, C Ritland, J Bohlmann, K Ritland. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC Evolutionary Biology*. - 2012. - 12. - 8
- 4 P Rigault, B Boyle, P Lepage, JEK Cooke, J Bousquet, et al. A white spruce gene catalog for conifer genome analyses. *Plant Physiology*. 2011. - 57:14-28
- 5 A Kovach, II Wegrzyn, G Parra, C Holt, GE Burning, et al. The *Pinus tadel* genome characterised by diverse and highly divergent repetitive sequences. *BMC Genomics*. 2010. - 11:420. doi: 10.1186/1471-2164-11-420.
- 6 D Lisch, JL Benetzen. Transposable element origins of epigenetic gene regulation / [Current Opinion in Plant Biology / V14(2), 2011. - Pp.156–161
- 7 RS Baucom, JC Estill, C Chaparro, N Upshaw, A Jogi, JM Deragon, RP Westerman, PJ SanMiguel, JL Bennetzen. Exceptional Diversity, Non-Random Distribution, and Rapid Evolution of Retroelements in the B73 Maize Genome / PLoS Genetics, 2009, - V5(11),e1000732
- 8 M Morgante, E De Paoli. Towards the conifer genome sequence. In: Plomion C, Bousquet J, Kole C (eds). *Genetics, genomics and Breeding of Conifers*. Science Publishers, Edenbridge. pp.389-403.
- 9 W Liu, S Thummasuwan, SK Sehgal, P Chouvarine, DG Peterson. Characterization of the genome of bald cypress. *BMC Genomics*. 2011. - 12:553
- 10 DB Neale, RR Sederoff. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics*. 77:212-16.
- 11 Grissa, I. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) for the Genotyping of Bacterial Pathogens / I.Grissa, G.Vergnaud,

C.Pourcel // *Molecular Epidemiology of Microorganisms*. – Vol.551. – N 2. – 2009. – DOI: 10.1007/978-1-60327-999-4

12 Mojica, FJ. Intervening sequences of regularly spaced 123 prokaryotic repeats derive from foreign genetic elements / FJ. Mojica, C. Díez-Villasenor, J. Garcia-Martinez, E. Soria // *J Mol Evol.* – 2005.– 60:174–182

13 Kurtz, S. REPuter: The Manifold Applications of Repeat Analysis on a Genomic Scale / S. Kurtz, J.V. Choudhuri, E. Ohlebusch, C. Schleiermacher, J. Stoye, R. Giegerich // *Nucleic Acids Res.*, – 2001. – 29(22):4633-4642


14 RS Baucom, JC Estill, C Chaparro, N Upshaw, A Jogi, JM Deragon, RP Westerman, PJ SanMiguel, JL Bennetzen. Exceptional Diversity, Non-Random Distribution, and Rapid Evolution of Retroelements in the B73 Maize Genome / *PLoS Genetics*, 2009, - V5(11),e1000732

Изъято 3 страницы.

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
 И.Е. Ямских
«22» июня 2018 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

06.04.01 Биология
06.04.01.06 Геномика и биоинформатика

ПОИСК И АНАЛИЗ ПАЛИНДРОМНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ В
ГЕНОМЕ ЛИСТВЕННИЦЫ СИБИРСКОЙ

Выпускник: ББ16-06Б, 041625623



Тараненко Е.А.

подпись

Руководитель: д. ф.-м. н., в.н.с.



Садовский М.Г.

подпись

Рецензент: д.б.н



Орлов Ю.Л.

Красноярск 2018