

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.ЛОМОНОСОВА

ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

**Эволюционные сценарии изменения
типов наследования митохондриальных
ДНК.**

**Evolution of mitochondrial DNA
inheritance patterns.**

Курсовая работа студента 2-го курса

Кравченко Павла Андреевича

Научный руководитель:

с.н.с., к.б.н. Д.А. Кнорре

_____ **Д.А. Кнорре**

подпись руководителя

Москва

2018 г.

СОДЕРЖАНИЕ

Введение.....	стр. 2
Цели и задачи исследования.....	стр. 3
Обзор литературы.....	стр. 5
Глава I (Экспериментальная часть).....	стр. 8
§1.1. (Дизайн эксперимента).....	стр. 9
§1.2. (Сбор информации).....	стр. 12
§1.3. (Алгоритм).....	стр. 12
§1.4. (Материалы и методы).....	стр. 14
§1.4.1. (Программное обеспечение).....	стр. 14
§1.4.2. (Базы данных).....	стр. 15
§1.5. (Результаты экспериментальной работы).....	стр. 16
§1.6. (Обсуждение).....	стр. 27
Заключение.....	стр. 28
Выводы.....	стр. 29
Использованная литература.....	стр. 30
Приложения.....	стр. 32

ВВЕДЕНИЕ

Митохондрии - двумембранные органоиды, произошедшие путём эндосимбиоза от безъядерных организмов. В большинстве случаев митохондрии сохранили (сильно редуцированный) собственный геном. У большинства видов митохондриальная ДНК (мтДНК) передается лишь по материнской линии: от матери потомкам, но за время изучения проблемы были открыты механизмы отцовского, двуродительского наследования, а также — двойного многородительского наследования (doubly uniparental).

В процессе эволюции эукариот тип наследования митохондриальной ДНК многократно менялся. Более того, механизм реализации материнского наследования у разных, иногда достаточно близких, видов различается. Это указывает на то, что путем исключения отцовской мтДНК из зародышевой линии живые организмы решают проблемы, природа и значение которых пока не ясна. Задачей нашего исследования являлось изучение изменений типов наследования митохондриальной ДНК в крупных группах эукариот.

В работе мы предлагаем способ предсказания митохондриального типа наследования в локальных популяциях организмов. Для нескольких видов нам удалось подтвердить имеющуюся информацию и проследить некоторые эволюционные сценарии изменения типов наследования митохондриальных ДНК.

ЦЕЛИ И ЗАДАЧИ ИССЛЕДОВАНИЯ

1. Собрать и систематизировать имеющуюся в литературе информацию о экспериментально подтвержденных типах наследования митохондриальной ДНК у различных видов.
2. Отработать полуколичественный метод определения типа митохондриального наследования на основании полиморфизмов последовательностей мтДНК различных особей одного вида.
3. Дополнить существующую систему органического мира информацией о предсказанных типах наследования мтДНК.

ОБЗОР ЛИТЕРАТУРЫ

Исследование цитоплазматической наследственности началось с открытия в 1909 году наследственного материала в хлоропластах. На пестролистных растениях герани (*Pelargonium zonale*), было обнаружено необычное неменделевское расщепление окраски, которое впоследствии было объяснено именно цитоплазматической наследственностью [1]. Открытие наследственного материала в митохондриях произошло немногим позже благодаря Борису Эфрусси, работавшему с неспособными к дыханию клетками дрожжей.

Несмотря на эти ранние открытия, долгое время считалось, что митохондрии наследуются лишь от матери потомкам, но со временем накопилось достаточное количество наблюдений, свидетельствующих о существовании иных путей наследования митохондрий (см. Приложения Таблица 1).

Наша работа основана на представлениях о современной системе многоцарственного мира как об искусственной системе, отражающей лишь основные филогенетические связи. Нами была использована трёхдоменная многоцарственная система, записанная в NCBI Taxonomy, и система, принятая в проекте [2].

Цитоплазматическая наследственность как явление связана, в общем случае, с двумя типами органоидов цитоплазмы – митохондриями и хлоропластами [3]. Митохондрии - двумембранные органоиды, произошедшие путём эндосимбиоза от безъядерных организмов, ассоциированных с альфа протеобактериями [4].

Молекулярно-филогенетические данные указывают на то, что данное событие произошло однажды: в нём участвовал родственник архей, и безъядерный организм, близкий к альфа протеобактериям.

Приобретение митохондрий происходило, судя по всему, постепенно. Существует несколько косвенных свидетельств, являющихся основой теорий симбиогенеза. [5,6,7,8]

Митохондрии присутствуют практически во всех эукариотических клетках, за несколькими исключениями. В норме митохондрии выполняют энергетическую функцию в клетке за счёт работы электронтранспортной цепи. У некоторых

представителей группы Protista митохондрии видоизменяются в гидрогеносомы или митосомы, в которых обычно не поддерживаются процессы реализации наследственной информации и дыхания, но осуществляется окислительное фосфорилирование в бескислородной среде [9]. В митохондриях могут кодироваться короткие пептиды, выполняющие важную функцию в ретроградной сигнализации. Также с функционированием митохондрий связаны процессы апоптоза [10], клеточной сигнализации [11], гомеостаза [12].

Митохондриальные геномы широко используются при выделении таксонов в современной системе органического мира наравне со строением крист, устройством жгутиков и количеством мембран пластид. Митохондриальная ДНК является хорошим маркером, в большинстве случаев, позволяющим качественно кластеризовать организмы по крупным и средним систематическим группам [13]. Часто данные критерии могут использоваться с хорошей точностью, но существуют факторы, делающие реконструкцию менее качественной. Например, несоблюдение общепринятого представления о материнском наследовании и гомоплазии митохондрий или увеличенная скорость мутирования в некоторых группах живого [14].

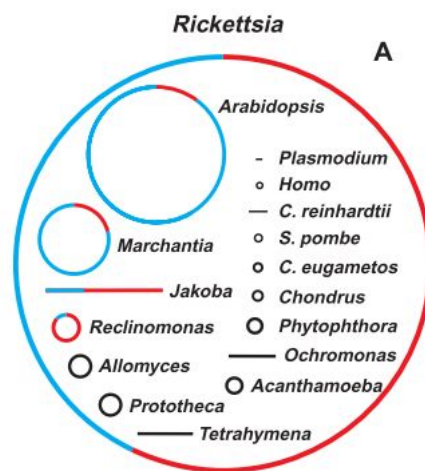


Рис. 1. Пример митохондриальных геномов представителей разных надцарств [8].

Одновременно с приобретением органоидов возникла потребность в их передаче потомству. Долгое время считалось, что мтДНК наследуется только лишь по

материнской линии, от матери потомкам, но из правила материнского наследования митохондрий можно найти ряд исключений (см. Приложения Таблица 1).

Согласно литературным данным, все типы наследования можно разделить на **монородительское** и **биродительское** наследование. Отдельно выделяют **двойное биродительское наследование**. Все ссылки на источники для каждого организма приведены в таблице 1 (см. Приложения Таблица 1).

1. **Монородительское наследование** - митохондрии передаются от одного из родителей всем потомкам.
 - a. Митохондрии могут передаваться только лишь от отца потомству - **отцовское наследование**, как, например, у Секвойи вечнозелёной (*Sequoia sempervirens*).
 - b. Митохондрии могут передаваться только лишь от матери потомству - **материнское наследование**. Такой тип наследования митохондрий характерен для Розы душистой (*Rosa odorata*).

При монородительском наследовании возможна “утечка” нетипичной к передаче мтДНК, то есть, иногда происходят события передачи и рекомбинации мтДНК родителей, как это было показано, например, для мыши и человека.

В целом, в клетках эукариот с монородительским типом наследования мтДНК наблюдается **гомоплазмия** - состояние, когда в каждой отдельной клетке присутствует только один вариант мтДНК

2. **Биродительское наследование** - митохондрии могут передаваться от обоих родителей потомкам. При биродительском типе передачи мтДНК было обнаружено несколько возможных путей:
 - a. Наследуемые мтДНК не рекомбинируют и находятся в дочерних клетках в определённых соотношениях. Если мтДНК наследуется от обоих родителей всем потомкам, говорят, что в клетке наблюдается **гетероплазмия** по мтДНК. При данном типе наследования молекулы мтДНК могут сильно отличаться. У Моркови дикой (*Daucus carota*) была обнаружена гетероплазмия, при которой от одного из родителей наследовалось около 60% мтДНК.

б. мтДНК рекомбинируют как следствие слияния митохондрий.

Получаются новые варианты молекул. Подобные процессы характерны для дрожжей (*Saccharomyces cerevisiae*) и других представителей отдела Аскомикота.

3. **Двойное биродительское наследование.** Описана передача митохондрий от отца только сыновьям, а от матери всем потомкам. Часто встречающаяся, например, у двустворчатых - Синанодонта Вуда (*Anodonta woodiana*).

Данная система определений имеет право на существование, но является неустойчивой и противоречивой. Существуют примеры организмов, для которых известно сразу несколько типов наследования. Они также представлены в таблице 1 (см. Приложения Таблица 1).

Из вышеописанного следует, что тип наследования митохондрий не является качественным, а скорее является количественным признаком. Он может быть определён по тому, сколько по количеству и какие по качеству мтДНК перешли к потомкам.

Изучение типов наследования мтДНК находится лишь в самом начале. Прошло лишь несколько десятилетий с тех пор, как были проведены первые эксперименты и опубликованы первые результаты [15].

Глава I. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Возможно, митохондриальное наследование может быть определено с помощью специальных критериев. Именно поэтому выделение этих критериев и их грамотное использование является одной из задач исследования. Используя консервативные последовательности и большие датасеты, в последнее время становится возможным определение типа наследования митохондрий без скрещиваний организмов.

Достаточно лишь иметь несколько десятков качественных образцов ДНК организмов из одной популяции.

Информация о типе наследования митохондрий может быть очень важна в эволюционной биологии, поэтому было бы очень полезно иметь инструмент, способный хотя бы полуколичественно измерить и предсказать тип наследования.

Возможность предсказывать тип наследования может быть полезна при более точном филогенетическом анализе или баркодировании.

До сих пор остаются вопросы о том, как можно связать имеющиеся данные о наследовании митохондрий с древом современной системы многоцарственного мира.

Мы предлагаем способ детектирования наиболее вероятного типа наследования по молекулярно-генетическим данным. Мы предлагаем критерии, адекватно, на наш взгляд, оценивающие тип наследования. Стоит заметить, что для будущего более точного анализа могут пригодиться данные о различии в скорости мутирования ядерной и митохондриальной ДНК в разных группах живого [14].

Обычно тип наследования определяется путём генетических скрещиваний или привлечением полимеразных цепных реакций (ПЦР). На данный момент уже имеется достаточное для компьютерного анализа количество данных, которые могут быть использованы в работе. С использованием этих данных, тип наследования может быть определён по секвенированным последовательностям мтДНК через подсчёт Linkage Disequilibrium - параметра, характеризующего зависимость или независимость двух вариантов (полиморфизмов) в геноме. Он определяет неслучайное распределение частот единичных полиморфизмов в разных локусах, которое в гипотетической

популяции может быть как выше, так и ниже ожидаемой при случайном распределении частоты.

§1.1. Дизайн эксперимента

1. По литературным данным была создана и индексируется электронная база данных организмов с известными митохондриальными типами наследования (см. Приложения Таблица 1). Из 194 статей была собрана информация о типах наследования у 137 организмов. Данные организованы в таблице следующим образом: указывается имя объекта, его ID и ранг согласно NCBI Taxonomy, Inherited blast name, Rank, даётся указание на тип наследования, тип генетического кода согласно [16], и ссылка на источник. Собранный материал необходим для проверки получаемых экспериментальных данных литературными подтверждениями.
2. По данным таблицы 1 (см. Приложения Таблица 1) с помощью сервиса [17] было построено и дополнено изображениями дерево найденных видов (см. Приложения Дерево 1). Сервис использует информацию из базы данных NCBI Taxonomy, что даёт возможность строить согласно современным представлениям максимально правдоподобные деревья. По информации о родстве организмов с использованием скрипта, написанного на языке R, было построено дерево, отражающее распределение наследования митохондрий. В него были включены 104 вида из 137, за исключением близкородственных. Дерево (см. Приложения Дерево 2) и скрипт (см. Приложения Скрипт 1) доступны из репозитория GitHub (<https://github.com/Pavel-Kravchenko/R-script-to-draw-trees>).
3. Был написан **пайплайн** - последовательность программ, в которой данные, получаемые на выходе из одной программы, передаются на вход другой программы, выполняющей анализ последовательностей и визуализацию их параметров (Рисунок 2).

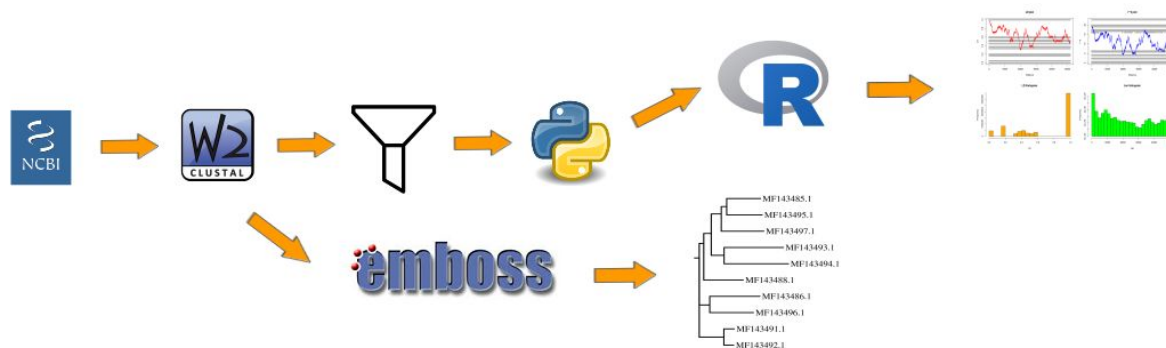


Рисунок 2. Схема пайплайна. Последовательность программ, выполняющих задачу по анализу данных. Записи полных митохондриальных геномов скачиваются из базы данных NCBI Nucleotide и группируются на подготовительном этапе с помощью скрипта 3 (см. Приложения Скрипт 3). Подготовленные последовательности выравниваются в программе clustalw. После выравнивания они проходят фильтрацию на качество блоков, из них удаляются столбцы с гэпами. Подготовленные отфильтрованные выравнивания подаются на вход анализирующему скрипту, написанному на Python. Скрипт вычисляет значения Linkage Disequilibrium (LD) для всех комбинаций пар колонок с полиморфизмами. Полученные в ходе работы пайплайна данные визуализируются средствами языка R, а также используются пакетом EMBOSS для построения деревьев.

На вход программе подаются скачанные из базы данных NCBI Nucleotide последовательности. На выходе пользователь получает:

- a. набор файлов, содержащих выравнивание последовательностей,
- b. скобочную нотацию дерева,
- c. деревья, построенные на основании нотации,
- d. файлы, содержащие срезы с полиморфизмами и Linkage Disequilibrium (LD), r^2 ,
- e. визуализацию значений LD, r^2 , гистограммы,
- f. результат параметрического теста Кендалла.

Основная функция пайплайна - попытка предсказания митохондриального типа наследования организма по множественным выравниваниям и сбор дополнительной

информации, облегчающей анализ и дальнейшее принятие решения о выборе наиболее вероятного типа наследования.

Пайплайн и инструкции по отладке доступны из репозитория GitHub

(<https://github.com/Pavel-Kravchenko/Evolution-of-mitochondrial-DNA-inheritance-patterns>).

4. Была оценена зависимость Linkage Disequilibrium (LD) от расстояния для имеющихся данных. Для установления зависимости двух векторов использовался пакет программы R-studio 'Kendall' версии 2.2 [18]. Данный тест можно использовать в качестве условного критерия, позволяющего определить тип наследования митохондрий. Был написан скрипт 2 (см. Приложения Скрипт 2), максимизирующий значения коэффициента Кендалла с целью выявления значения длины, при которой происходит оптимальная оценка зависимости. В дальнейшем все значения вычислялись до этого порога.

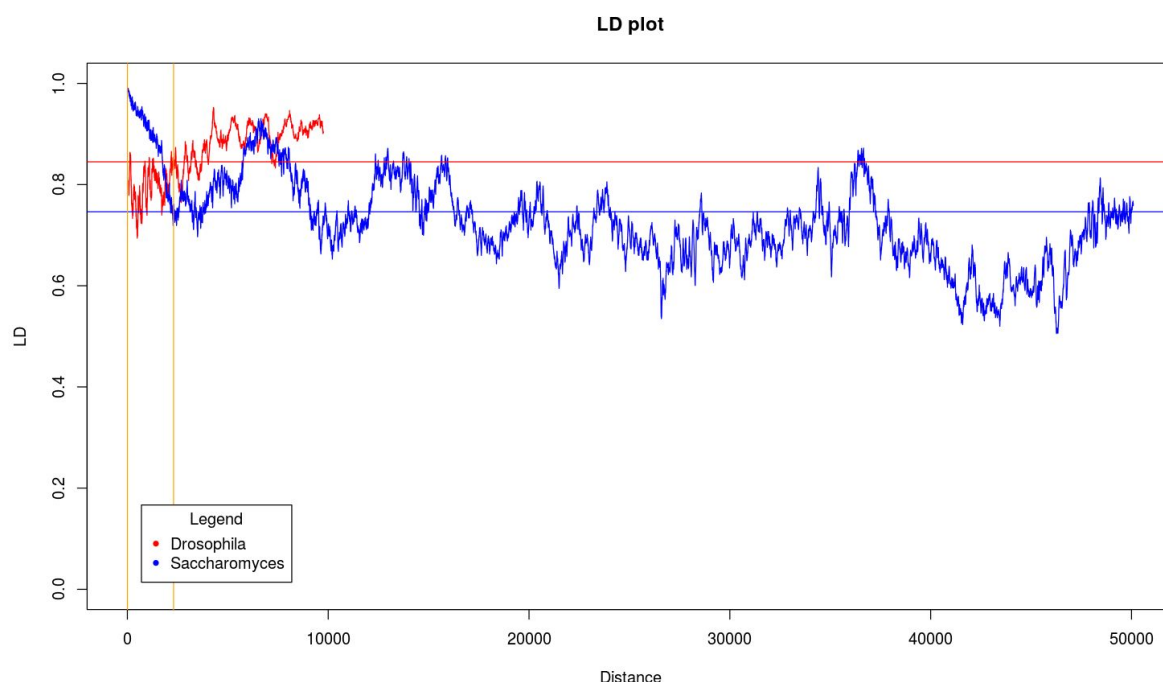


Рисунок 3. График Linkage Disequilibrium - Distance для Drosophila и Saccharomyces. По оси абсцисс отложены расстояния между двумя полиморфизмами в геноме. По оси ординат - значения LD для этих пар. Тау Кендалла равно 0.1406409 и -0.2385678, соответственно для Drosophila и Saccharomyces ($p\text{-value} < 2.2e-16$). На рисунке указаны средние значения LD, а также участок кривых, к которому применялся тест.

5. Дерево 2 было дополнено информацией о наследовании мтДНК, полученной из секвенированных последовательностей организмов.

§1.2. Сбор информации

Для определения типа наследования митохондрий обычно используют несколько основных методов, описанных ниже.

1. Ставят ПЦР, стараясь поднять участки, наличие рекомбинации в которых может свидетельствовать о типе наследования.
2. Скрещивают организмы с характерными признаками, стараясь по расщеплению понять тип наследования.

После проведённых операций статистически оценивают собранные в природе данные.

Большинство литературных данных, содержащихся в таблице 1 (см. Приложения Таблица 1), были получены одним из этих двух методов. В последнее время классическая генетика отходит на второй план, давая преимущество молекулярной биологии в лице первого метода.

Подготовительный этап

В базе данных NCBI Nucleotide с помощью библиотеки Biopython Entrez по запросу “db=“genome”, term=“mitochondrion, complete genome”” удалось найти 78445 полных митохондриальных геномов. С помощью написанного на Python скрипта 3 (см.

Приложения Скрипт 3) записи были скачаны и проанализированы. Всего нашлось 179 организмов, для которых было известно строго больше трёх последовательностей - необходимое условие в рамках нашего эксперимента. По данным организмам было распределено 8061 последовательностей согласно описанию.

§1.3. Алгоритм пайплайна

Алгоритм выглядит следующим образом:

1. Запускается управляющий bash скрипт.

- a. Наборы последовательностей, полученных на подготовительном этапе, выравниваются с помощью программы clustalw с параметрами
-TYPE=DNA -OUTPUT=FASTA
- b. По выравниваниям запускается вычисляющий LD Python скрипт с порогом 0.25 (описано далее). Для каждого выравнивания создаётся директория с промежуточной информацией и результатами. Значения LD рассчитываются по алгоритму, описанному в методическом пособии Summer Institute in Statistical Genetics [19]. В качестве примера рассмотрим выравнивание десяти последовательностей. Рассмотрим вычисление LD на примере: возьмём два столбца выравнивания, содержащих полиморфизмы (Рисунок 4). Для вычисления LD необходимо рассчитать содержание каждого из нуклеотидов в каждом столбце. При вычислении LD принимают во внимание лишь столбцы с двумя типами нуклеотидов, постольку наличие нескольких вариантов, в целом, маловероятно, а объём вычислений сильно возрастает. В данном случае мы имеем первый столбец с содержанием тимина 0.7 и цитозина 0.3, а второй - с содержанием тимина 0.3 и цитозина 0.7. Минимальное значение выше порога в 0.25, поэтому продолжаем вычисления. Рассчитываем D как разность произведений частот аллельных пар, и, для получения LD, нормируем D на максимальное из произведений частот одноимённых вариантов, взятых с отрицательным знаком (если $D < 0$) или на минимальное из произведений разноимённых вариантов (если $D > 0$). Получаем значения LD от 0 до 1. Значения, близкие к 0, говорят о независимости распределения полиморфизмов в столбцах выравниваний. Значения близкие к 1 указывают на обратное - зависимость полиморфизмов в двух столбцах.

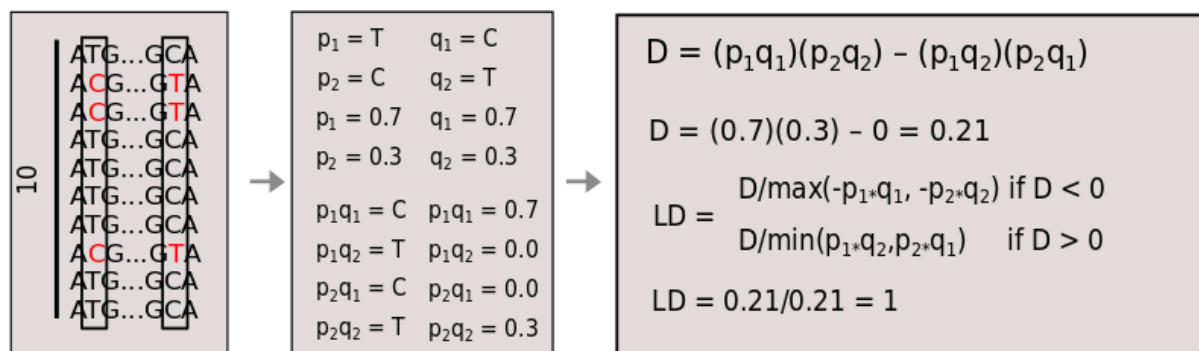


Рисунок 4. Алгоритм расчёта Linkage Disequilibrium (LD) по выравниванию полных митохондриальных геномов. Наравне с методом расчёта LD могут использоваться подходы, описанные в [20], [21]. Пояснения в тексте.

- с. В ключевом файле записывается таблица из значений LD и расстояний между столбцами в выравнивании, между которыми были рассчитаны значения LD.
2. По значениям LD и расстояний в каждой директории строятся графики (LD-расстояние между полиморфизмами) с помощью R скрипта.
3. Рассчитывается значение тау Кендалла для значений LD и расстояния между полиморфизмами.
4. Графики и гистограммы LD анализируются вручную с привлечением тау Кендалла и построенных по выравниванию деревьев.

§1.4. Материалы и методы

При выполнении работы использовались следующие базовые материалы и методы.

§1.4.1. Программное обеспечение

- Командная строка Unix.
- Jupyter 5.4.1
- Anaconda 5.1
- Пакет EMBOSS

- Основные библиотеки Python и Biopython - обработка последовательностей и парсинг данных
 - re
 - Bio.SeqRecord
 - Bio
 - Bio.Seq
 - numpy
 - os
 - random
 - datetime
 - sys
 - time
- R + пакеты к нему- визуализация и обработка вторичных данных. Графический и консольный варианты.
 - cor.test
 - plot
 - hist
 - png
- Jalview - построение выравниваний.
- Bash - создание скриптов и пайплайнов.
- Excel - отработка алгоритма взаимодействия скриптов и первичный анализ данных. Внепотоковая визуализация данных.
- Infogram - сбалансированная визуализация избранных данных для постеров и статей.
- GitHub - репозиторий и контроль версий.
- Atom - среда разработки.
- MEGA- программа для построения деревьев по выравниваниям.
- Подключаемые библиотеки

§1.4.2. Базы данных

1. NCBI - архив
 - a. Taxonomy
 - b. PopSet
 - c. Nucleotide
2. Refseq
3. GenBank

§1.5. Результаты экспериментальной работы

В ходе работы по литературным данным были найдены организмы с известными типами наследования митохондриальной ДНК. Была создана таблица 1 (см. Приложение Таблица 1), в которой была собрана информация о типах наследования мтДНК у 137 организмов.

По данным таблицы 1 было построено дерево найденных видов, отражающее распределение и возможную эволюцию типов наследования митохондрий (Рисунок 5). За исключением близкородственных в него были включены 104 вида из 137.

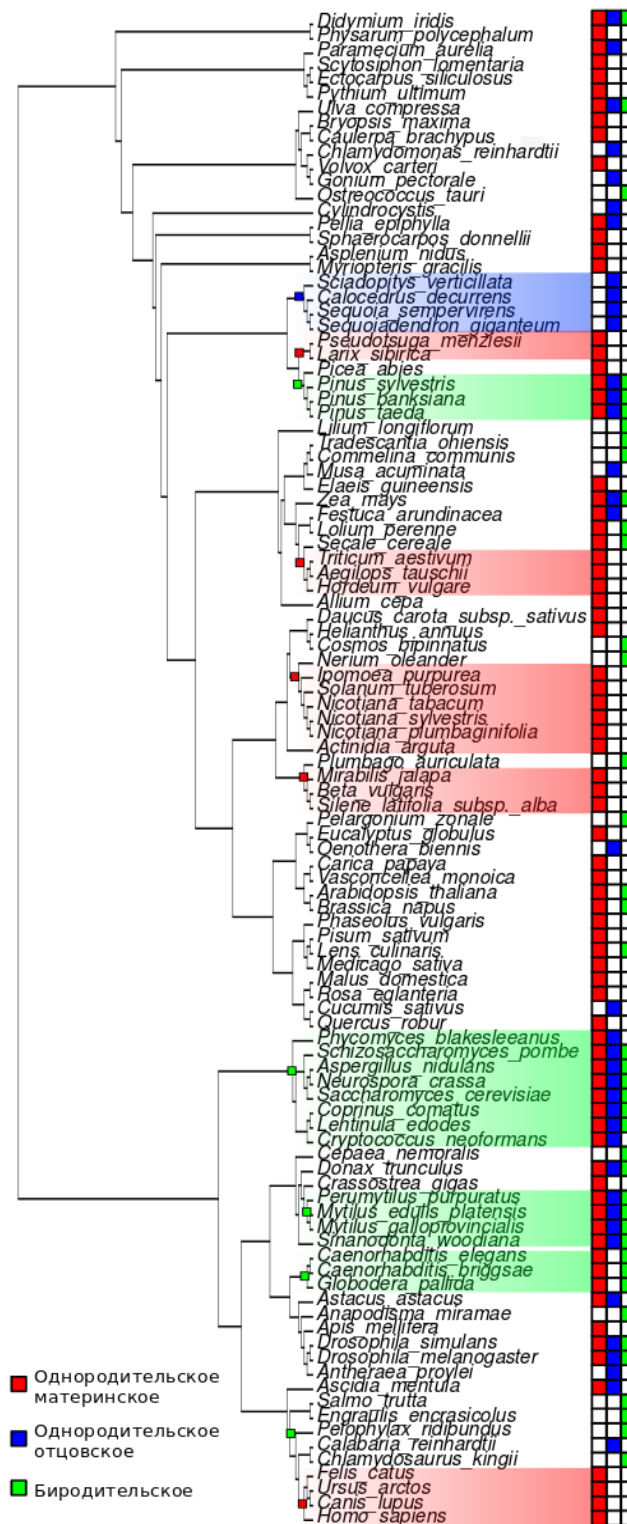


Рисунок 5. Дерево найденных видов с известной информацией о типах наследования мтДНК. Красным цветом обозначены организмы с материнским типом наследования, синим - с отцовским, а зелёным - с биродительским. Наиболее интересные на данном этапе монофилетические группы выделены прямоугольными областями одного цвета с мажорным типом наследования.

Был написан пайплайн, осуществляющий анализ эволюционно связанных последовательностей и способный полуколичественно предсказывать тип наследования мтДНК по значениям LD. Значения близкие к 0, как ожидается, должны говорить о независимом распределении полиморфизмов - биродительском наследовании и активной рекомбинации, а близкие к 1 - о монородительском или отсутствие рекомбинации при биродительском наследовании. Однако построение эксперимента не позволяет установить, какой из типов монородительского наследования был реализован в популяции.

Из базы данных NCBI нам удалось собрать последовательности для 179 организмов. Только для 92 из них удалось предположить тип наследования мтДНК. На вход пайплайна подавалось 1867 подходящих последовательностей из базы данных NCBI Nucleotide, кластеризованных по названию вида из описания записи. На выходе были получены графики и дополнительная информация для каждого организма. До сих пор у нас нет решающего правила, согласно которому можно было бы в автоматическом режиме с той или иной степенью уверенности определить тип наследования митохондрий. Для этих целей мы используем комбинацию визуализации LD и теста Кендалла.

В качестве примера работы алгоритма рассмотрим два организма, для которых хорошо изучены типы и механизмы наследования митохондриальной ДНК: *Drosophila melanogaster* и *Saccharomyces cerevisiae*. Для сравнения использовались выравнивания 10 последовательностей с ограничением на содержание полиморфных нуклеотидов 0.25 в столбце.

Drosophila melanogaster

Видно, что график LD-Distance (Рисунок 6) *Drosophila melanogaster* прижимается к 1, что может говорить о монородительском типе передачи митохондриальной ДНК. Полиморфизмы в мтДНК дрозофилы не показывают отрицательной корреляции, а скорее наоборот обнаруживают достоверную положительную корреляцию.

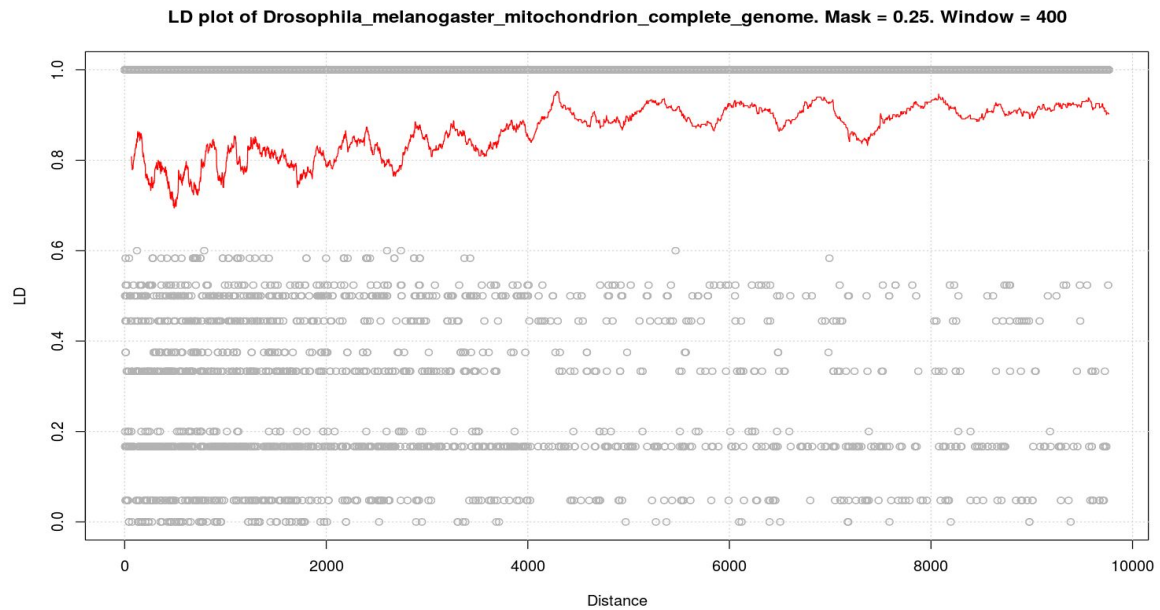


Рисунок 6. Зависимость LD от расстояния между полиморфными сайтами (distance) для *Drosophila melanogaster*. Красная линия скользящего среднего (moving window average) вычислялась как среднее от значений вектора LD, взятых в определённом окне. В данном случае размер окна равен 400.

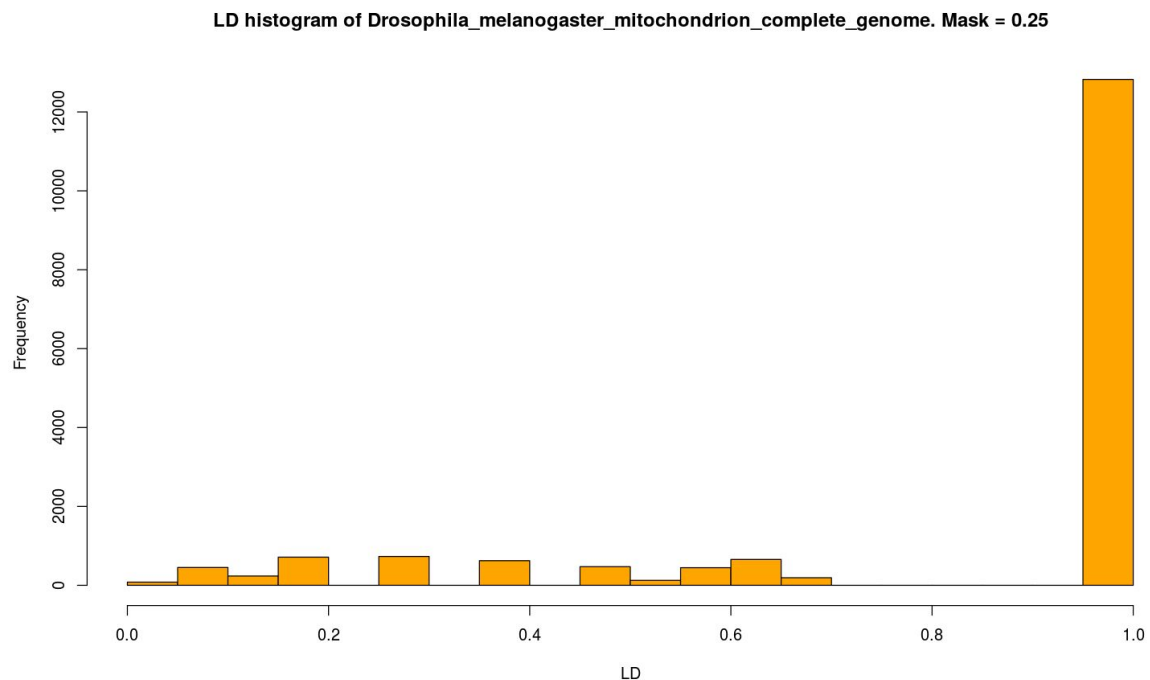


Рисунок 7. Гистограмма значений LD всех возможных пар полиморфных сайтов для *Drosophila melanogaster*. Наблюдается большое количество сайтов со значением LD 1.

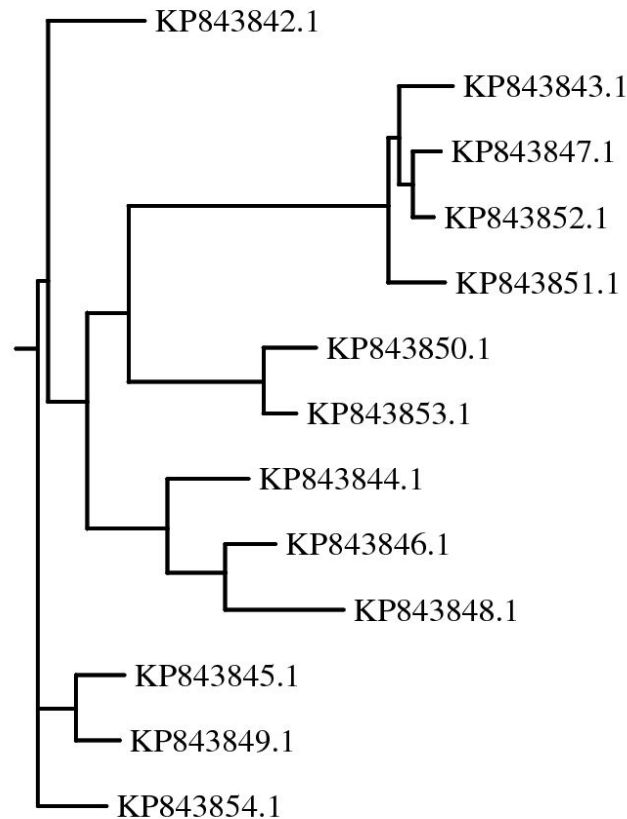


Рисунок 8. Реконструкция филогении для *Drosophila melanogaster*

Saccharomyces cerevisiae

График LD-Distance для *Saccharomyces cerevisiae*

(Рисунок 9) прижимается резко падает на расстоянии до 50000 bp; что может говорить о биродительском типе передачи митохондриальной ДНК.

Мы обнаружили значительную отрицательную корреляцию между полиморфизмами и расстоянием между ними в пределах от 1 до 50000 bp. Этот результат согласуется с литературными данными о типе наследования и скорости мутирования мтДНК в дрожжах.

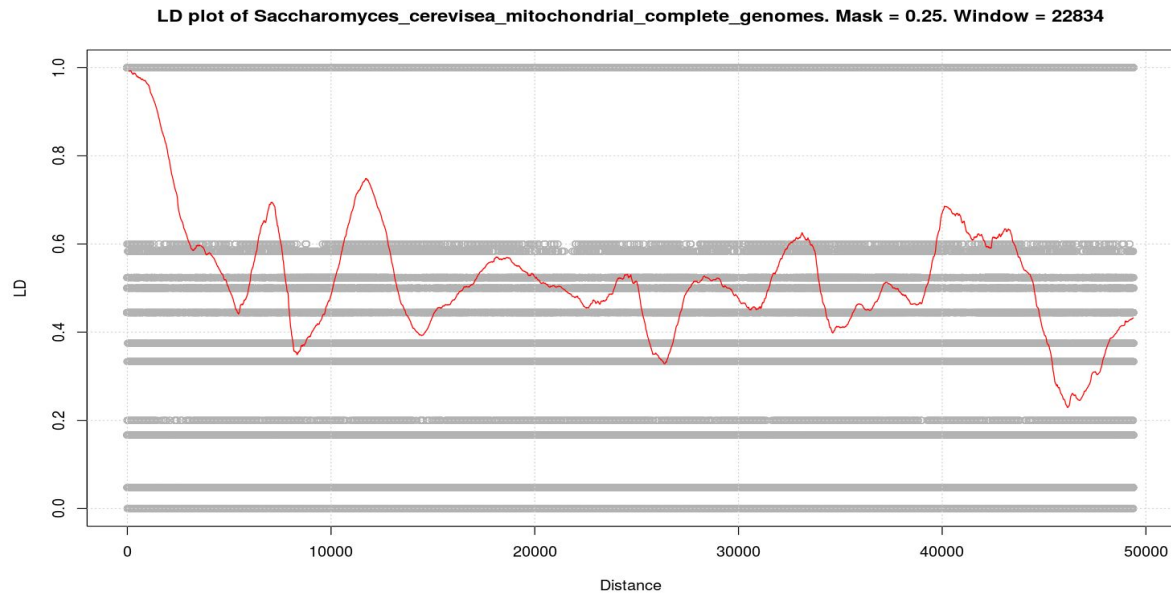


Рисунок 9. Зависимость LD от расстояния между полиморфными сайтами (distance). для *Saccharomyces cerevisiae*. Красная линия скользящего среднего (moving window average) вычислялась как среднее от значений вектора LD, взятых в определённом окне. В данном случае размер окна равен 22834.

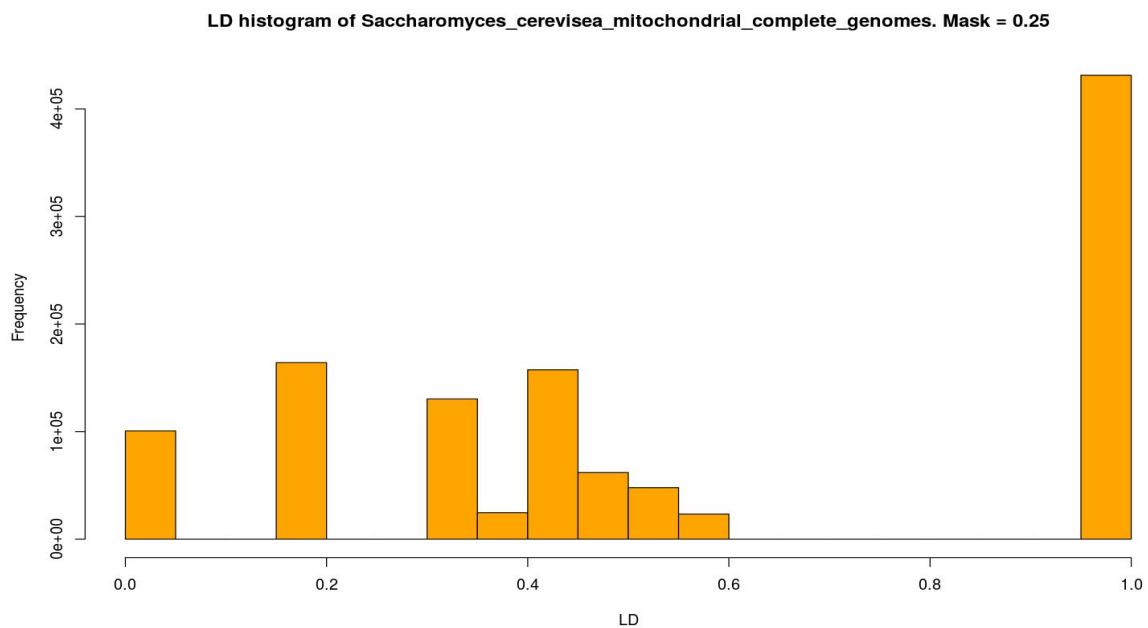


Рисунок 10. Гистограмма значений LD всех возможных пар полиморфных сайтов для *Saccharomyces cerevisiae*. Наблюдается большое количество сайтов со значением LD 1 и около значения 0,4.

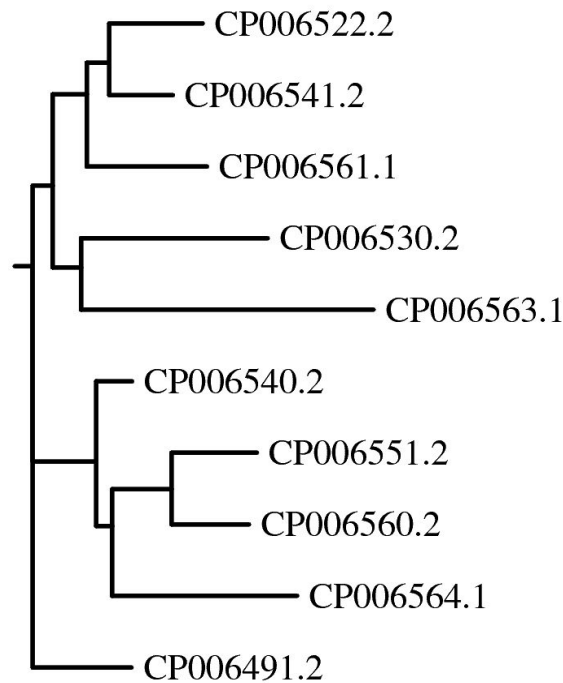
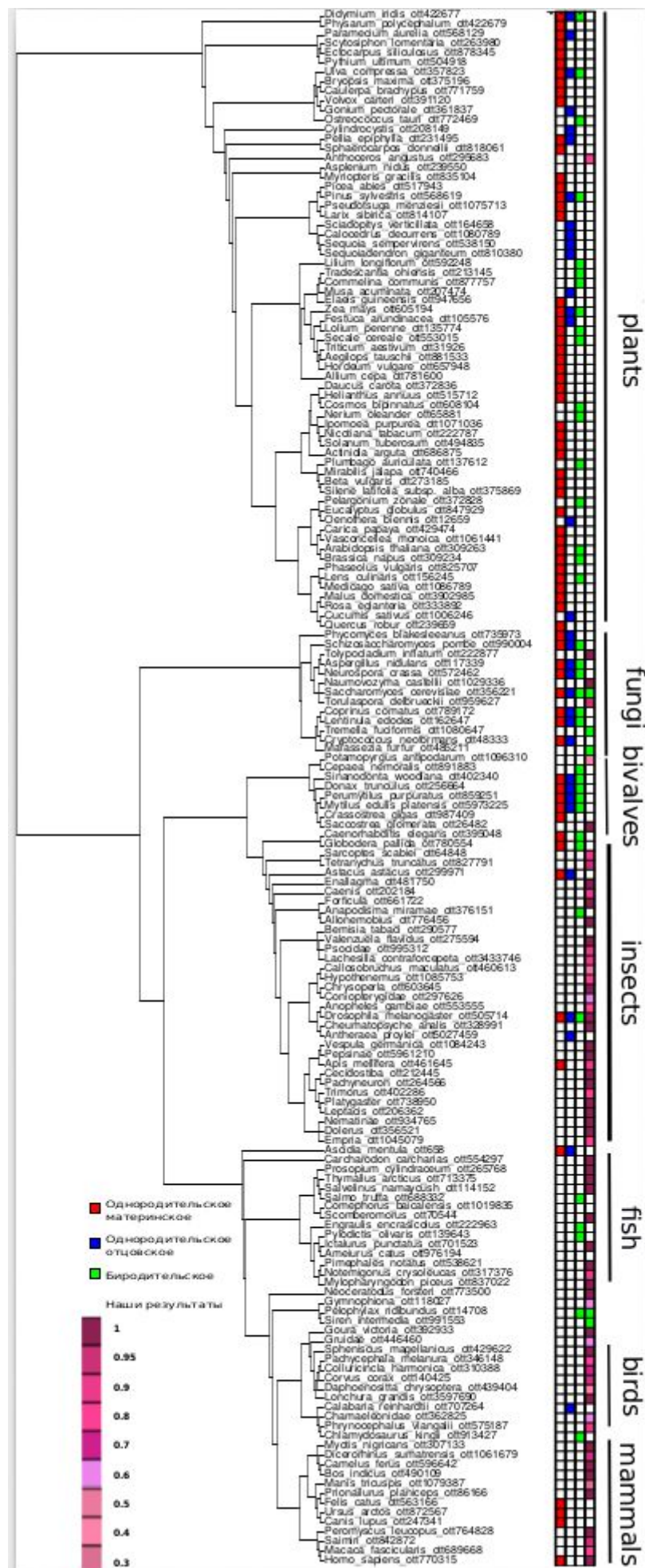


Рисунок 11. Реконструкция филогении для *Saccharomyces cerevisiae*

Рисунок 12. Дерево найденных видов с известной информацией о типах наследования мтДНК, дополненное результатами работы пайплайна. Обозначения и цветовая схема аналогичны таковым для рисунка 5, за исключением дополнительного столбца справа, отражающего средние значения LD в диапазоне от 0.3 до 1, а также предсказания для скачанных последовательностей. Тип наследования определялся вручную по графикам зависимости LD от расстояния между полиморфными сайтами, а также по результатам теста Кендалла для первых 50000 bp с учётом дополнительной информации в виде гистограмм и реконструкции филогении. Если тест Кендалла не давал значимой корреляции, то в качестве значения в правом столбце бралось среднее значение LD.

Оригинал доступен в репозитории GitHub

(https://github.com/Pavel-Kravchenko/R-script-to-draw-trees/blob/master/Rplot_tree_with_findings.pdf).



Для проверки достоверности получаемых результатов нами было разработано два скрипта, позволяющих симулировать данные для пайплайна. В первом скрипте (см.

Приложения Скрипт 4) варьируются эффективная численность популяции (N_e), количество поколений и число мутаций на поколение. Во втором можно задавать интервалы для таких величин как число рекомбинаций между последовательностями, количество поколений и число мутаций на поколение (см. Приложения Скрипт 5).

Элементарный цикл симуляции с помощью первого скрипта выглядит следующим образом. Генерируется случайная последовательность заданной длины, она размножается до эффективной численности популяции (N_e) с внесением указанного пользователем числа мутаций. Далее каждая из последовательностей размножается до N_e , претерпевая мутирование. Получается множество последовательностей, из которого отбирается новая популяция численности N_e . Между отобранными последовательностями осуществляется рекомбинация: указанное в процентном соотношении число случайно выбранных особей к N_e . Получается новая популяция, цикл повторяется.

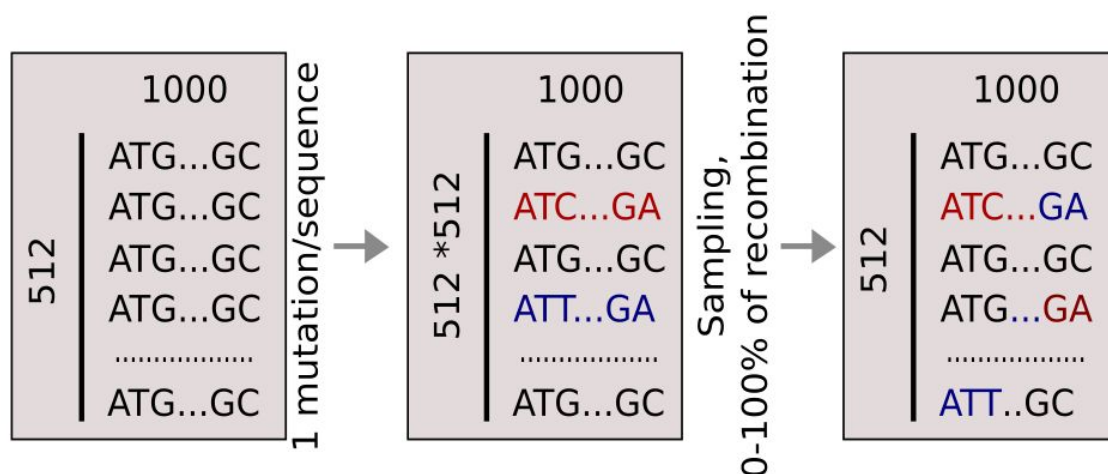


Рисунок 13. Элементарный цикл симуляции популяционных данных для пайплайна.

Пояснения в тексте.

В результате получают следующие результаты визуализации LD.

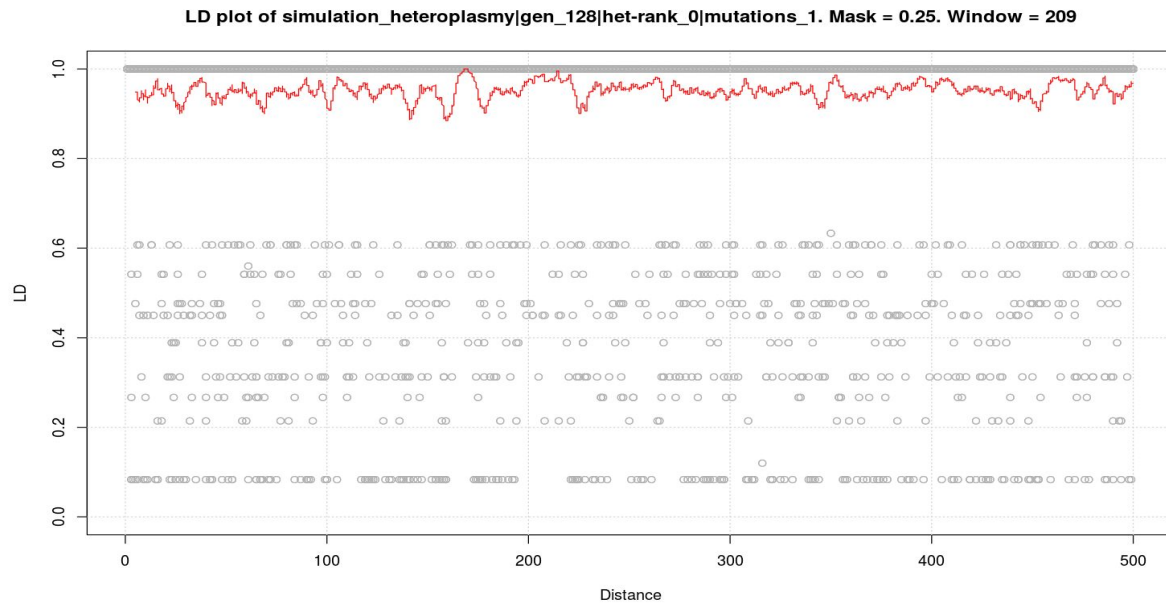


Рисунок 14. График LD-Distance для последовательности длины 1000 bp. Частота рекомбинации 1%. Число поколений - 128. Число мутаций на поколение - 1. ограничение на содержание минорного варианта - 0.25.

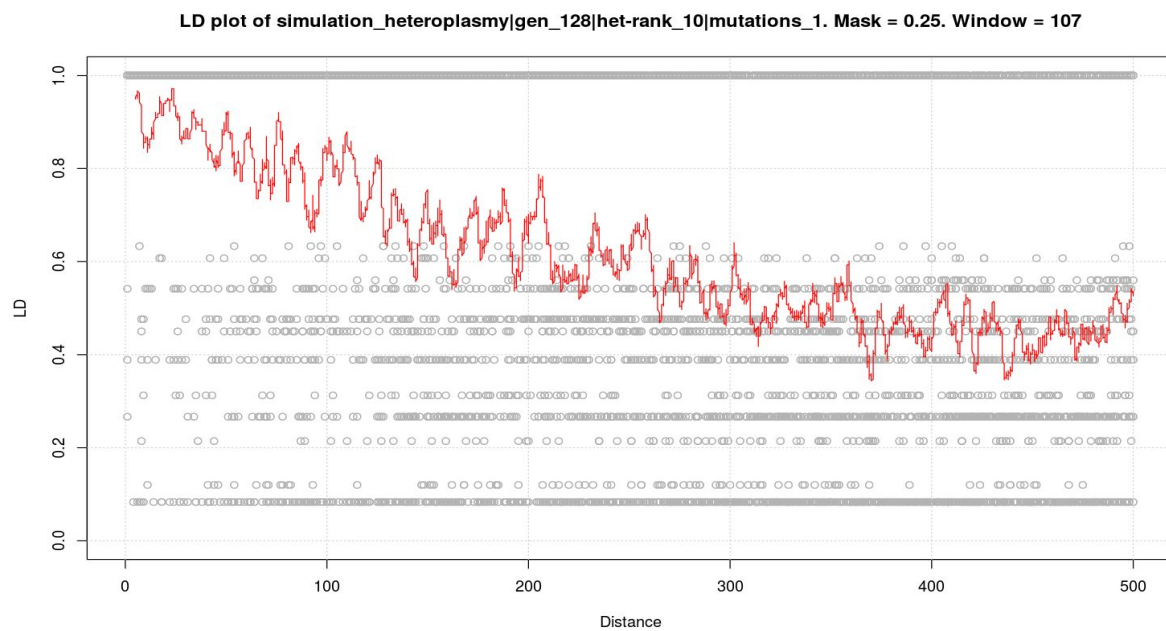


Рисунок 15. График LD-Distance для последовательности длины 1000 bp. Частота рекомбинации 10%. Число поколений - 128. Число мутаций на поколение - 1. ограничение на содержание минорного варианта - 0.25.

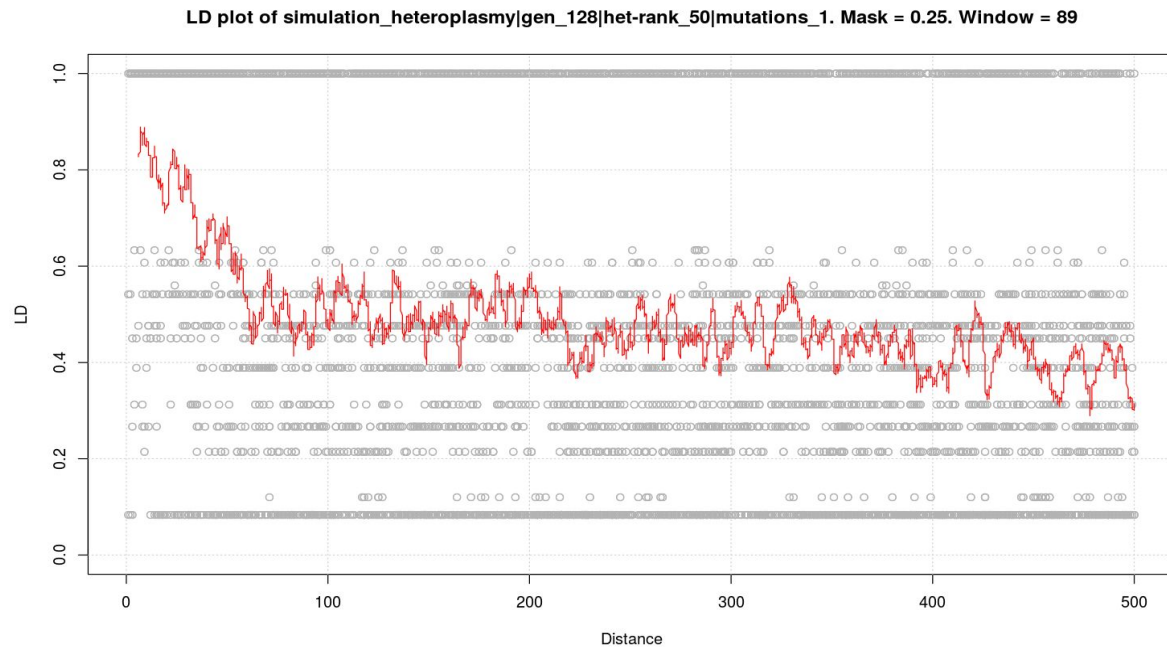


Рисунок 14. График LD-Distance для последовательности длины 1000 bp. Частота рекомбинации 50%. Число поколений - 128. Число мутаций на поколение - 1. ограничение на содержание минорного варианта - 0.25.

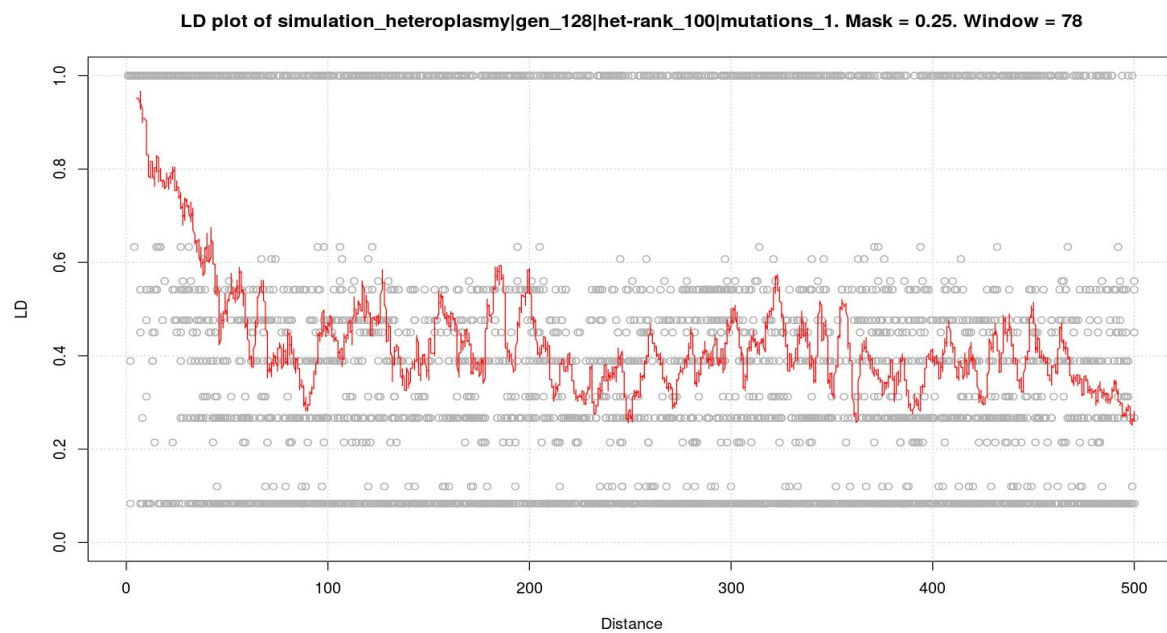


Рисунок 14. График LD-Distance для последовательности длины 1000 bp. Частота рекомбинации 100%. Число поколений - 128. Число мутаций на поколение - 1. ограничение на содержание минорного варианта - 0.25.

Также был написан скрипт, позволяющий визуализировать LD окружение единичного полиморфизма (см. Приложения Скрипт 6), но он не нашёл применения в данной работе.

§1.6. Обсуждение

С некоторой степенью точности можно утверждать, что у организмов митохондриальная ДНК наследуется только от матери, и что она имеет более высокую скорость мутирования, чем ядерная днк. Однако такое описание подходит не для всех организмов. Анализируя последние статьи, посвящённые выяснению типа и механизма наследования в той или иной группе, можно заметить, что данное правило материнского наследования часто нарушается, причём нарушается оно в разной степени.

Скорее всего, не существует четкого разграничения типов наследования мтДНК. Согласно нашим и литературным данным, стратегия передачи мтДНК многократно менялась на древе жизни. По-видимому, исходно существовал организм, передававший митохондрии бигермическим путём. В ходе эволюции происходила адаптация к различным условиям и стратегиям, специализация и, иногда, переход к чисто монородительскому отцовскому или монородительскому материнскому типу.

Из-за недостатка данных не представляется возможным подтвердить или опровергнуть правомерность использования предлагаемого алгоритма. Множество организмов, чьи типы наследования известны по литературным данным, и множество тех, для которых тип наследования предсказали мы, имеют пересечение всего по пяти организмам. Однако для всех пяти наш метод дает результат, подтверждаемый литературными данными.

ЗАКЛЮЧЕНИЕ

В заключение можно сказать, что полученные результаты нуждаются в подтверждении и более достоверном анализе. Необходимо получить более качественные популяционные данные и использовать альтернативные способы определения зависимости полиморфизмов. Это может быть сделано, например, с помощью методов, описанных в работах [20], [21]. Дальнейшая работа над инструментом, по предсказаниям которого станет возможным проследить изменение типов наследования мтДНК организмов, безусловно, сама по себе очень интересна.

ВЫВОДЫ

1. Было проанализировано 194 статьи и собрана база данных, включающая в себя информацию о типе наследования мтДНК для 137 видов. Больше всего видов с известным типом наследования мтДНК включало в себя надцарство Archaeplastida.
2. Был разработан пайплайн, позволяющий полуколичественно оценивать тип наследования митохондрий по полиморфизмам в последовательностях мтДНК различных особей одного вида. С помощью симуляции *in silico* показано, что данный подход позволяет различить типы наследования с рекомбинацией и без таковой. Для пяти видов с известным типом наследования мтДНК было показано, что результаты предсказания совпадают с ранее опубликованными экспериментальными результатами.
3. База данных дополнена информацией о типе наследования мтДНК для еще 103 видов.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

1. Baur, E. (1909) Zeit. Vererbungsl. 1, 330- 351.
2. <https://blog.opentreeoflife.org/2014/10/07/tree-for-all-hackathon-series-introduction/>
3. И.А. Захаров-Гезехус. ЦИТОПЛАЗМАТИЧЕСКАЯ НАСЛЕДСТВЕННОСТЬ. ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ 18, 2014 (2014).
4. L. Margulis, Origin of Eukaryotic Cells (Yale Univ. Press, New Haven, CT, 1970).
5. Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521, 173–179 (2015).
6. Pollard, T. D. & Cooper, J. A. Actin, a central player in cell shape and movement. Science 326, 1208–1212 (2009)
7. M. W. Gray and D. F. Spencer, in Evolution of Microbial Life, D. M. Roberts, P. Sharp, G. Alderson, M. Collins, Eds. (Cambridge Univ. Press, Cambridge, 1996), pp. 107–126.
8. Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. Science (80-.). 283, 1476–1481 (1999).
9. Diversity and reductive evolution of mitochondria among microbial eukaryotes
10. Sinha K, Das J, Pal PB, Sil PC. Oxidative stress: the mitochondria-dependent and mitochondria-independent pathways of apoptosis. Arch Toxicol. 2013;87:1157–80.
11. Chandel NS. Evolution of mitochondria as signaling organelles. Cell Metab. 2015;22:204–6.
12. Cheng Z, Ristow M. Mitochondria and metabolic homeostasis. Antioxid Redox Signal. 2013;19:240–2.
13. Avise JC. Molecular markers, natural history, and evolution. 2nd ed. Sunderland: Sinauer Associates; 2004.
14. Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker
15. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution
16. <https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi>

17. <http://www.phylotastic.org/>
18. Hipel, K.W. and McLeod, A.I., (2005). Time Series Modelling of Water Resources and Environmental Systems. Electronic reprint of our book originally published in 1994. <http://www.stats.uwo.ca/faculty/aim/1994Book/>.
19. http://faculty.washington.edu/tathornt/sisg2013/Kerr/3LD_Kerr.pdf
20. Posada, D. and Crandall, K.A. (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl Acad. Sci. USA, 98, 13757–13762.
21. Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. Genetics 172:2665–2681.

ПРИЛОЖЕНИЯ

Таблицы

Таблица 1

<https://docs.google.com/spreadsheets/d/1-tPHDElJOxV64XZ6Vo2ipHEN4AdnvloTfYKdSUNDeKQ/edit?usp=sharing>

Деревья

Дерево 1

https://github.com/Pavel-Kravchenko/R-script-to-draw-trees/blob/master/nice_tree.png

Дерево 2

<https://github.com/Pavel-Kravchenko/R-script-to-draw-trees/blob/master/Rplot.pdf>

Скрипты

Скрипт 1

https://github.com/Pavel-Kravchenko/R-script-to-draw-trees/blob/master/tree_maker_with_boxes.R

Скрипт 2

https://github.com/Pavel-Kravchenko/R-script-to-compare-two-LDs/blob/master/R_2_plots.R

Скрипт 3

<https://github.com/Pavel-Kravchenko/Python-script-for-parsing-complete-genome-sequences/blob/master/parser.ipynb>

Скрипт 4

https://github.com/Pavel-Kravchenko/evolution-simulation-pipeline_Ne_and_generations

Скрипт 5

https://github.com/Pavel-Kravchenko/evolution-simulation-pipeline_recombination_and_generations

Скрипт 6

<https://github.com/Pavel-Kravchenko/LD-plotter-package>

Сторонние ресурсы для подтверждения правильности работы алгоритма

<http://web.cbio.uct.ac.za/~darren/rdp.html>

<http://www.lifesci.sussex.ac.uk/CSE/test/ld.php>

Файл с последовательностями использованных митохондриальных геномов без фильтрации оказался слишком большим для загрузки на сервер. Он может быть получен выполнением скрипта

<https://github.com/Pavel-Kravchenko/Python-script-to-parsing-complete-genome-sequences/blob/master/parser.ipynb>.