

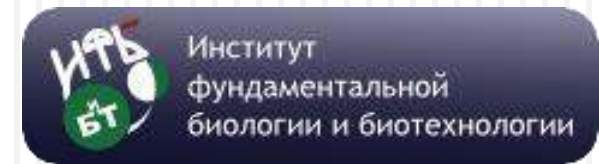
Федеральное государственное автономное образовательное учреждение  
высшего профессионального образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт фундаментальной биологии и биотехнологии  
Кафедра биофизики  
Лаборатория лесной геномики СФУ

# СБОРКА И АННОТИРОВАНИЕ ХЛОРОПЛАСТНОГО ГЕНОМА ЛИСТВЕННИЦЫ СИБИРСКОЙ

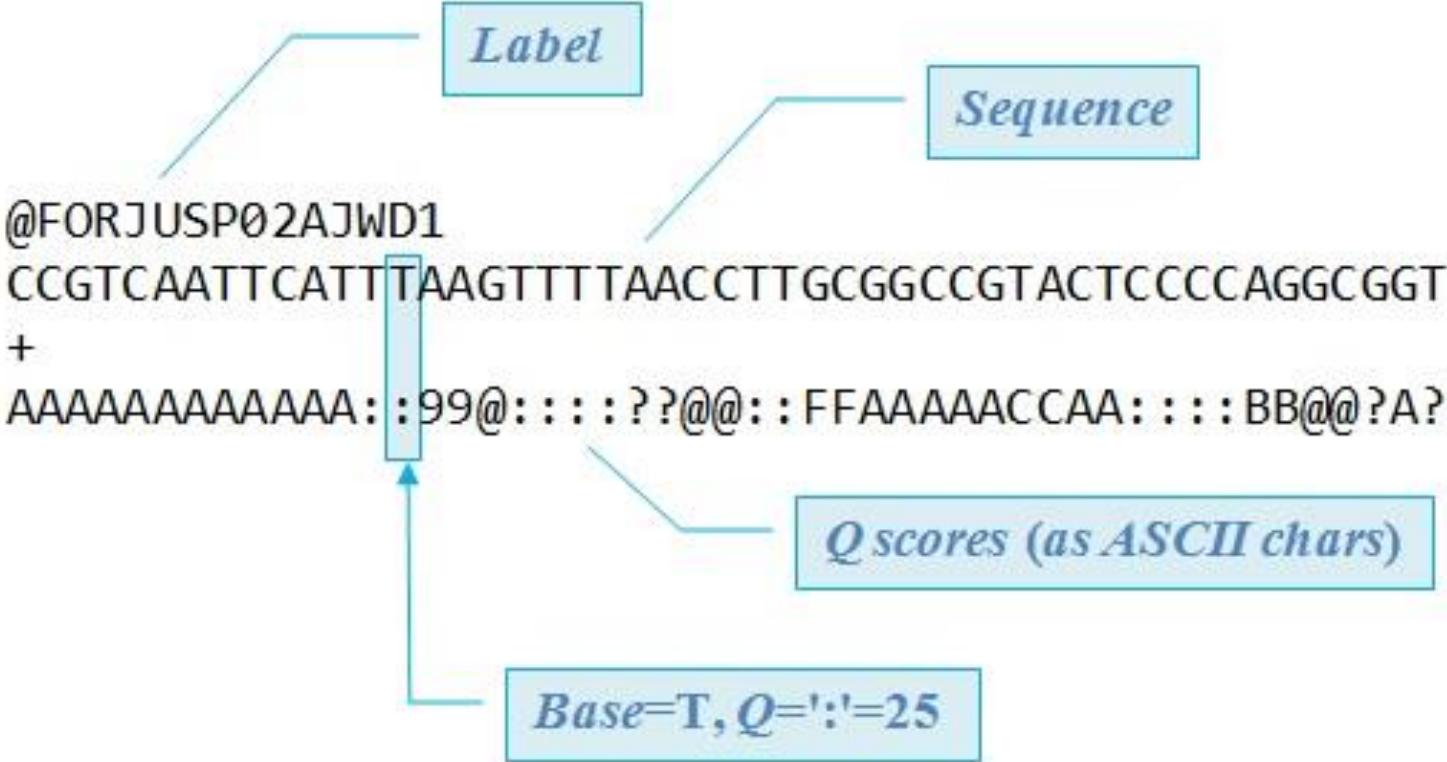
Выполнила: Е. И. Бондар

Научные руководители: д-р физ.-мат.  
наук Садовский М. Г., Ю. А. Путинцева

[bone-post@yandex.ru](mailto:bone-post@yandex.ru)



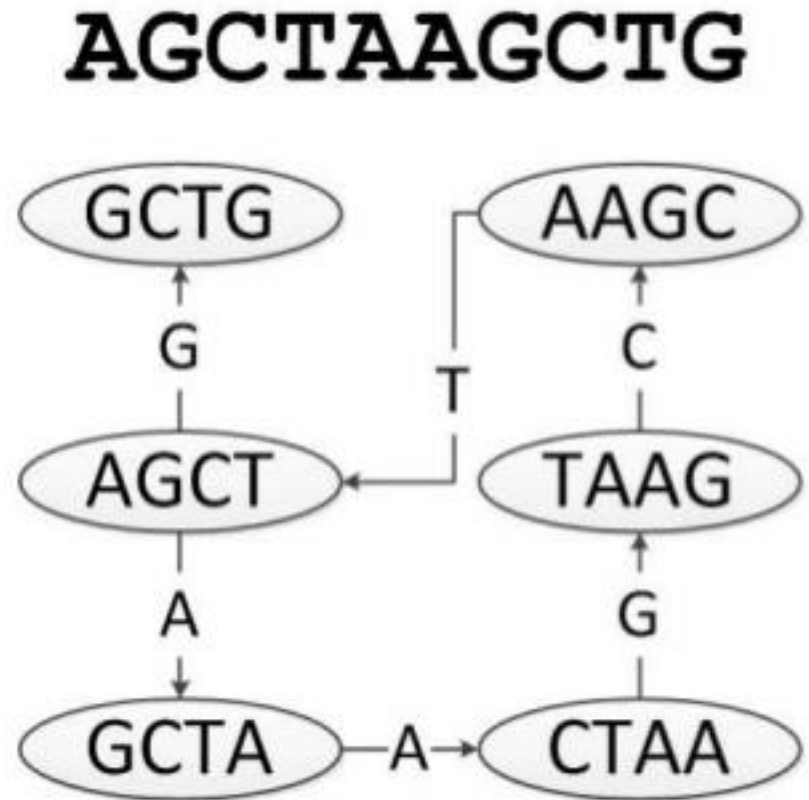
# FASTQ format



# Граф де Брёйна

3

- Ориентированный граф
- Вершины (k-1)
- Ребра k
- Эйлеров путь =  
геномная  
последовательность



# BASECLEAR

4

FOR 100% DNA RESULTS

SSPACE STANDARD

## Лаборатория алгоритмической биологии

Санкт-Петербургского Академического университета РАН

[О лаборатории](#) ▾ [Геномика](#) ▾ [Proteomics](#) [Transcriptomics](#) ▾ [Immunoproteogenomics](#) [Публикации](#) ▾ [Коллектив](#) ▾ [Образование](#) ▾

Геномный ассемблер SPAdes

### SPAdes 3.5.0

Изменения.

[Руководство по ассемблеру \(версия 3.5.0, англ.\)](#)

[Руководство по dipSPAdes \(для геномов с высоким уровнем полиморфизма, англ.\)](#)

[Скачать ассемблер](#)

[Сборка длинных парных ридов Иллюмины \(2x150 и 2x250\)](#)

[SPAdes на данных GAGE-B](#)

[SPAdes на других данных](#)

Техническая поддержка: [spades.support@bioinf.spbau.ru](mailto:spades.support@bioinf.spbau.ru)



# Ситуация на сегодняшний день

5

По состоянию на 20 апреля 2015 года в базе данных Genbank:

- опубликовано 117 хлоропластных геномов представителей семейства Сосновые.
- Для представителей рода *Larix* опубликованы две последовательности хлоропластных геномов для видов *L. decidua* Mill. и *L. occidentalis* Nutt.
- Для хлоропластного генома *L. sibirica* имеется 11 последовательностей некоторых генов и их частей.

# Цель

6



- Целью данного исследования является сборка и аннотирование хлоропластного генома лиственницы сибирской (*Larix sibirica* Ledeb.)

# Материалы и методы

7

- Образцы ДНК были выделены из семян и гаплоидного каллуса трех деревьев, произрастающих в разных регионах – Уральский, Хакасский и Красноярский.
- Данные полногеномного секвенирования *L. sibirica*, получены на приборе Illumina HiSeq2000 в Лаборатории лесной геномики СФУ.
- Размер библиотек : 400-500 п.н.о. (урал) и 5000 п.н.о. (Черное озеро).
- Референсная последовательность: последовательности хлоропластных геномов *L. decidua* и *L. occidentalis* (AB501189.1 и FJ899578.1) из GenBank .

# Схема процесса сборки

8

Выравнивание коротких ридов на существующие последовательности хлоропластных геномов при помощи Bowtie2



Ассемблирование ридов с помощью SPAdes. Выравнивание полученных контигов на целевые геномы при помощи BLASTN



Ассемблирование с помощью SPAdes с указанием флага «--trusted contigs»



Формирование скаффолдов с использованием библиотек MP-ридов и программы SSPACE.



# Аннотация

9

- Проверка собранных геномов при помощи RepeatModeler.
- Rapid Annotation using Subsystem Technology (RAST)



**The SEED Viewer** SEED Viewer version 2.0  
Welcome to the SEED Viewer - a read-only browser of the curated SEED data.  
For more information about The SEED please visit [theSEED.org](http://theSEED.org).

»Navigate »Organism »Comparative Tools »Help find  Eugenya Bondar

### Export for *Larix sibirica* subsp. *altaica* (349908.3)

We support a large set of different information about our organisms. All this information is based on the manual curation effort of our annotators. The underlying data they created is freely available on our ftp site at: <ftp://ftp.theseed.org/>

This is a RAST organism, which is currently not in the SEED. If you would like to download this genome, please follow this link to get to the according RAST page:

- [RAST job details page for \*Larix sibirica\* subsp. \*altaica\*](#)



- 1) tRNAscan-SE and "search\_for\_rnas" by Niels Larsen – выполняют поиск tRNA и rRNA. В дальнейшем эти области считаются достоверными и не рассматриваются при поиске кодирующих участков.
- 2) GLIMMER-2 – ab initio предиктор. Использует интерполированные марковские модели для нахождения CDS.
- 3) Филогенетический контекст – в аннотируемом геноме проводится поиск небольшого количества универсальных генов (FIGfams). Эти гены используются для поиска ближайших соседей.
- 4) Целенаправленный поиск генов в FIGfams филогенетических соседей и именованье участков, определенных GLIMMER.
- 5) Проверка оставшихся участков – выравнивание против FIGfams для нахождения потенциальных функциональных семейств, затем проверка этих семейств.
- 6) Удаление дублей и регулировка стартовых позиций и перекрытий – blast для разрешения конфликтов.

# Результаты ассемблирования

11



***L. sibirica*** –  
122561 bp



***L. decidua*** -  
122474 bp

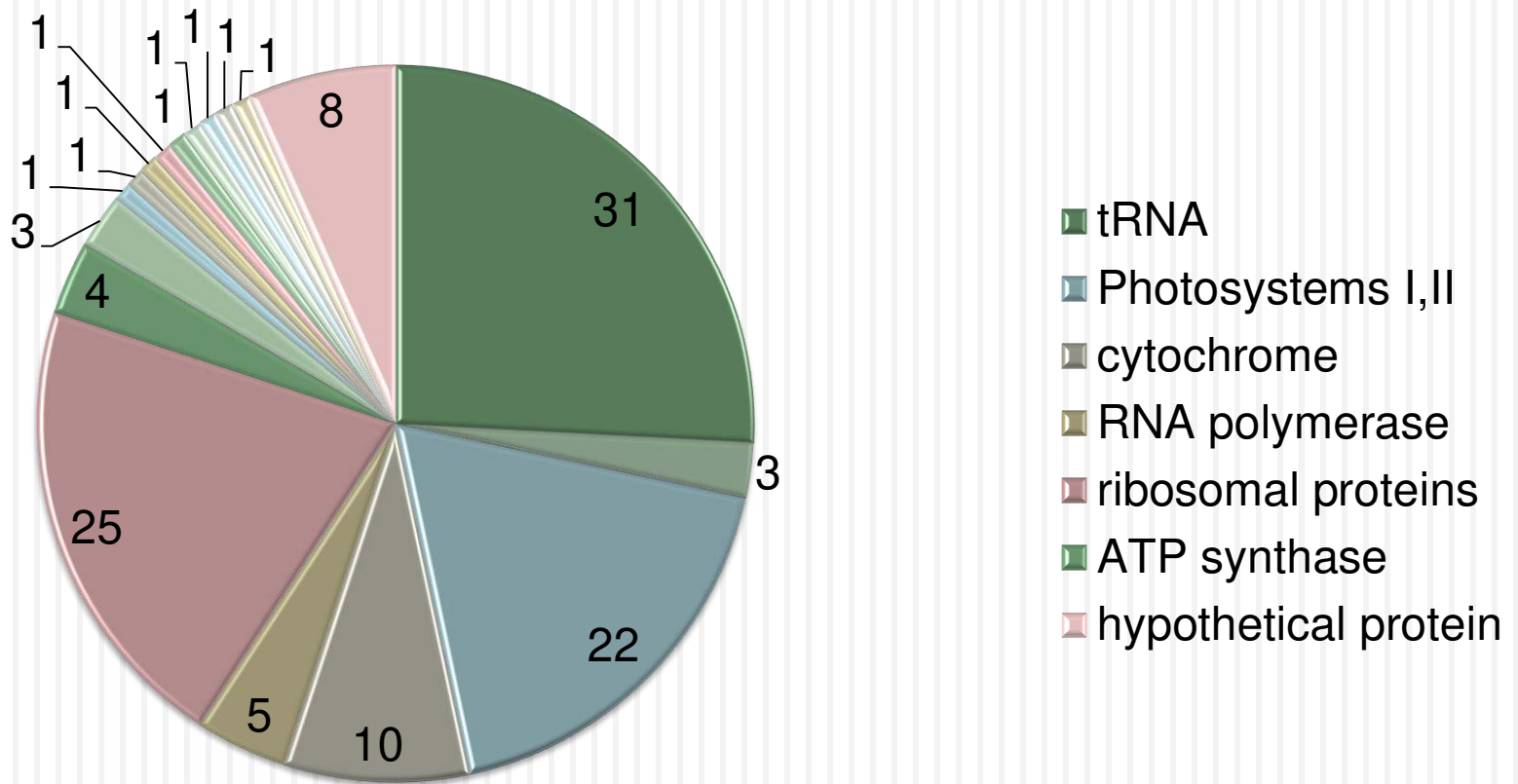


***L. occidentalis*** -  
119680 bp

<b>Protein coding sites</b>	<b>Abundance</b>
tRNA	31
rRNA	3
Ribosomal proteins	25
Photosystems I,II	22
Cytochrome (b6/f, b559)	10
RNA polymerase	5
ATP synthase	4
Light-independent protochlorophyllide reductase	3
Translation initiation factor 1	1
NAD(P)H-quinone oxidoreductase	1
Cell-surface adhesin2C putative	1
Cell division protein FtsH	1
Gene trnK, product="maturase K"	1
Ribulose biphosphate carboxylase	1
Acetyl-coenzyme A carboxyl transferase	1
Membrane protein PxcA2C	1
ATP-dependent Clp protease proteolytic	1
Hypothetical protein	8

# Результаты аннотирования

13

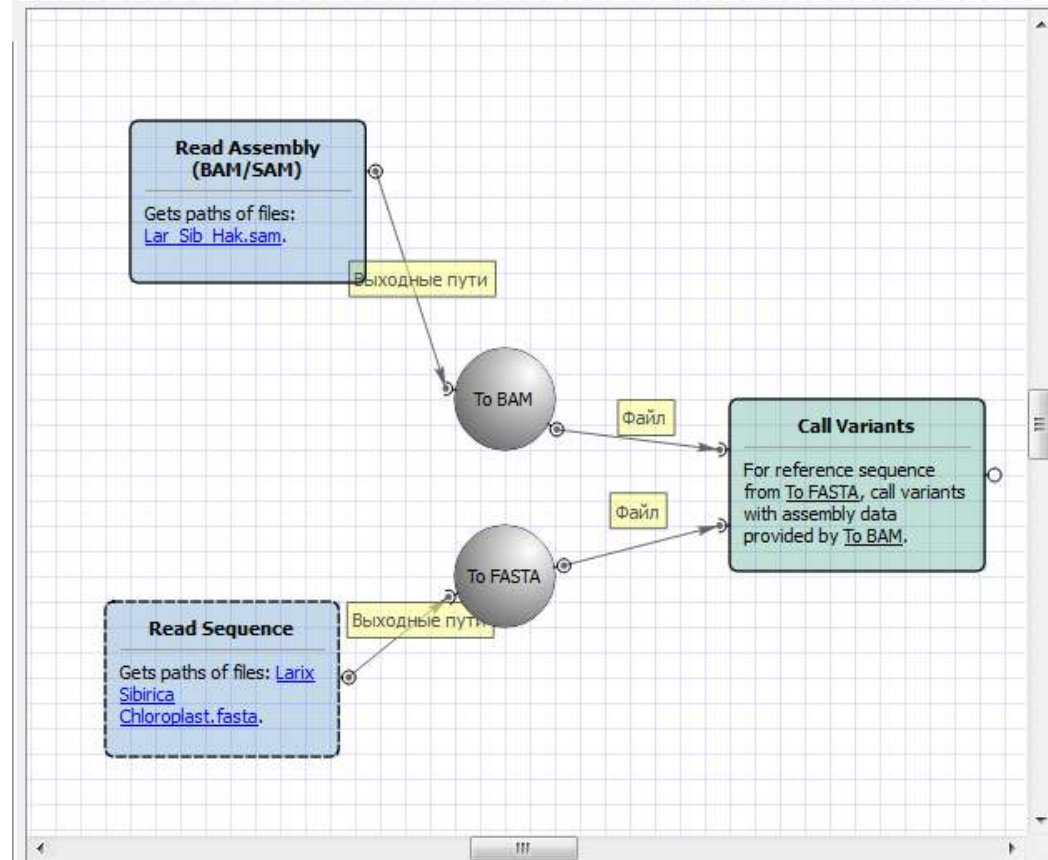


Кодирующие области, найденные после сравнения аннотации с уже имеющимися данными по *L. decidua* и *L. occidentalis*.

# Поиск SNPs

14

- Поиск между уральским, хакасским, и красноярским деревьями
- Для трех деревьев найдено 13 позиций с SNP, 2 из них находятся в кодирующих участках генома: *tRNA-Arg* и *Cell division protein FtsH*.



# Заключение

- Длина хлоропластного генома *L. sibirica* составила 122561 bp и близка к 122474 bp у близкородственной *Larix decidua* Mill.
- В результате аннотирования и сравнения полученных данных с уже имеющимися близкородственными видами, был выявлен 121 кодирующий участок. Из них 87 соответствуют белок-кодирующим генам и 34 – генам.
- Для трех деревьев *L. sibirica* найдено 13 позиций с SNP, 2 из них находятся в кодирующих участках генома: *tRNA-Arg* и *Cell division protein FtsH*.

# Благодарности

16

- Мы благодарим М. Г. Садовского за обсуждение результатов и ценные комментарии, И. Н. Третьякову, М.Е. Пак, В. Л. Семерикова и К. Дейч за помощь с отбором образцов, Д.А. Кузьмина и С. В. Маколова за помощь в вычислениях.
- Работа выполнена в рамках проекта «Геномные исследования основных бореальных лесообразующих хвойных видов и их наиболее опасных патогенов в Российской Федерации», финансируемого Правительством РФ (договор № 14.У26.31.0004).



Спасибо за  
внимание!