

# Построение графа связей геномных последовательностей

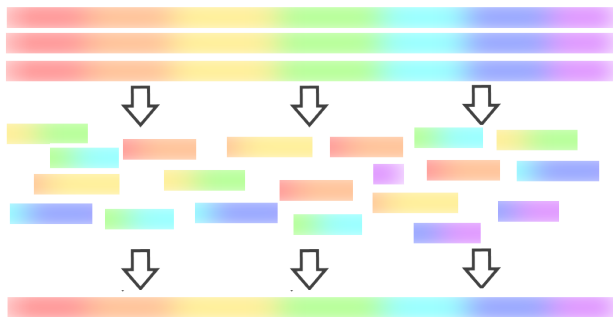
Черникова Ольга

Руководитель: Пржибельский Андрей Дмитриевич

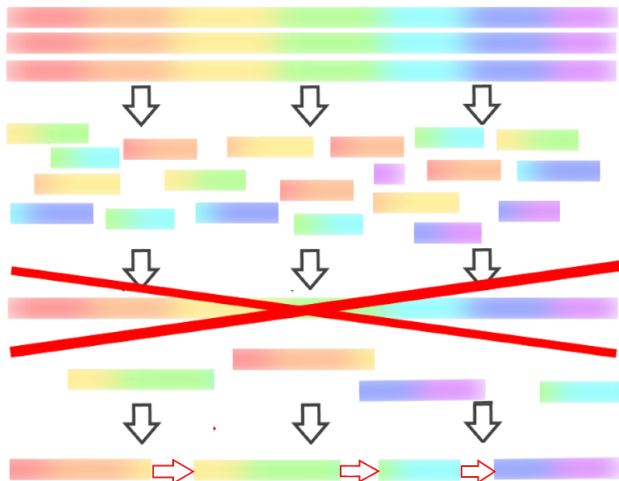
СПб АУ РАН

Осень 2016

# Задача сборки генома



# Задача сборки генома

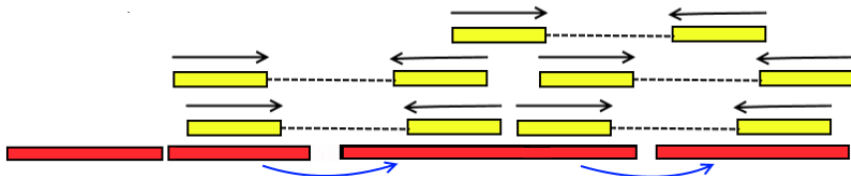


# По парным рядам ДНК

## ■ Парные ряды:

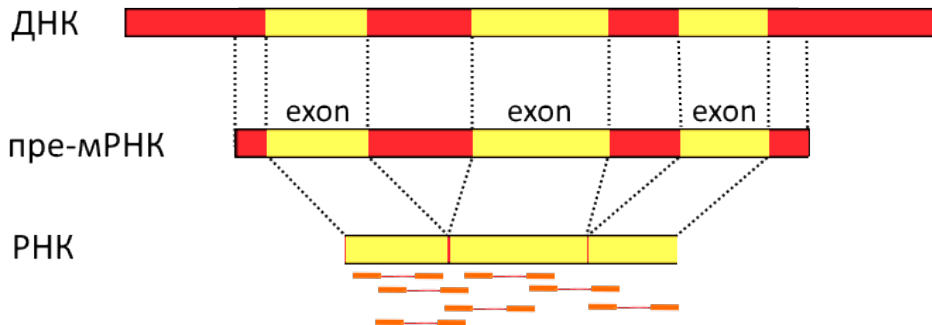


## ■ Нахождение связей с помощью парных рядов:



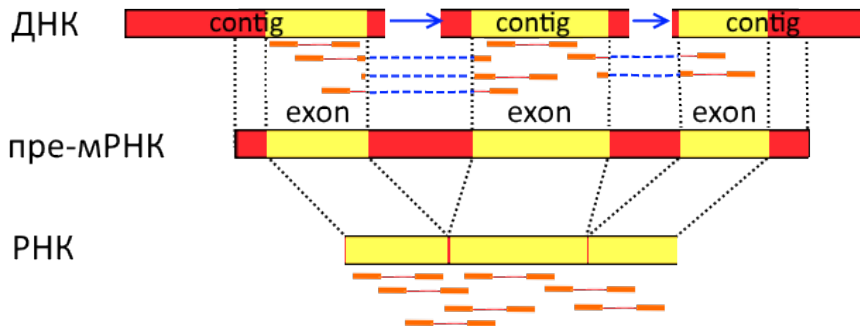
## По ридам РНК

Связь между ДНК и РНК:

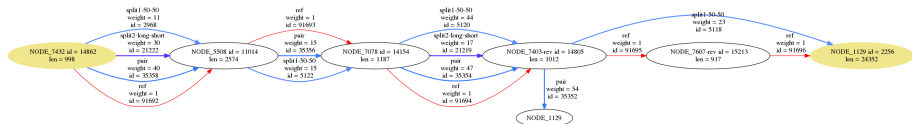
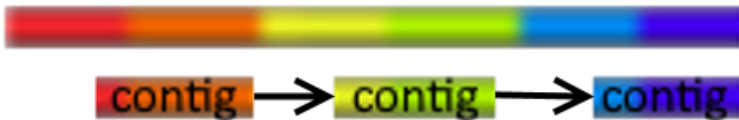


## По ридам РНК

Выравнивание ридов РНК на контиги:



## По эталлонной сборке



# Задачи

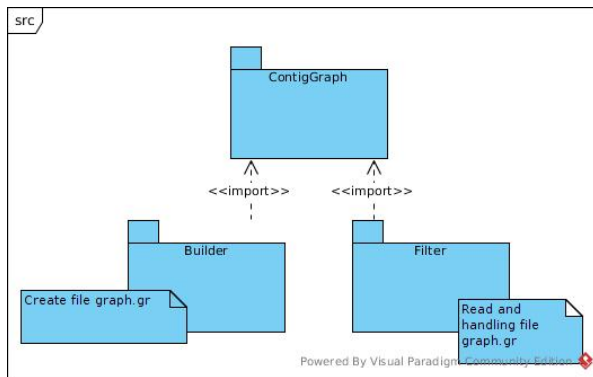
- Исследовать возможные варианты построения графа по ридам РНК.
- Объединить разные виды связей между контигами в один граф.
- Решить проблему с подбором параметров для фильтрации графа.
- Декомпозиция графа.



# Архитектура

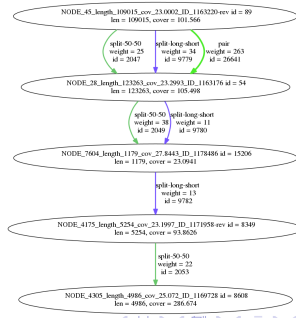
Задача делится на две части:

- Построение графа связей
- Фильтрация



# Результаты

- Реализация модуля для нахождения связей между контигами различными способами.
- Реализация модуля с различными возможностями для визуализации и фильтрации графа.
- Удалось улучшить сборку *C.elegans*.



# Используемые инструменты

- Язык разработки - **C++**
- **SeqAn** - библиотека для работы с файлами в SAM/BAM и fasta/fastq форматах.
- **gtest** - библиотека для тестирования.
- Программы для выравнивания - **STAR, nucmer, bowtie2**.
- **QUAST** - для анализа качества сборки.
- **Tablet** - для визуализации выровненных рядов.

## Дальнейшее развитие

- Подбор и тестирование приложения на различных данных. Например на данных *A.thaliana*.



- Ускорение работы приложения.
- Новые способы фильтрации связей и декомпозиции графа.
- Нахождение путей в графе и улучшение сборки.

# Спасибо за внимание

Репозиторий: [https://github.com/olga24912/bio\\_scaffolder](https://github.com/olga24912/bio_scaffolder)