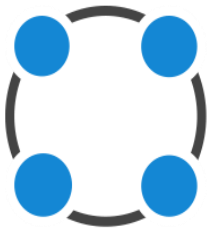


# De novo сборка бактериального генома и оценка скаффолдеров

Васильева Александра Андреевна

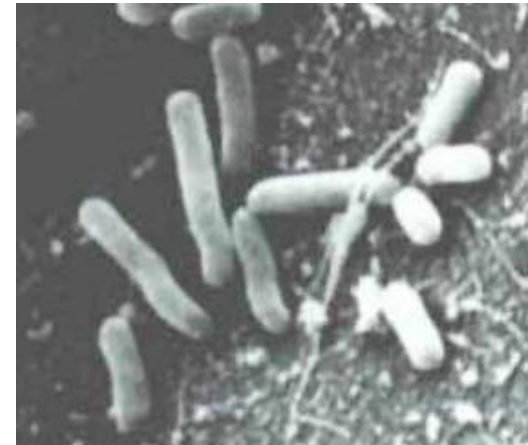
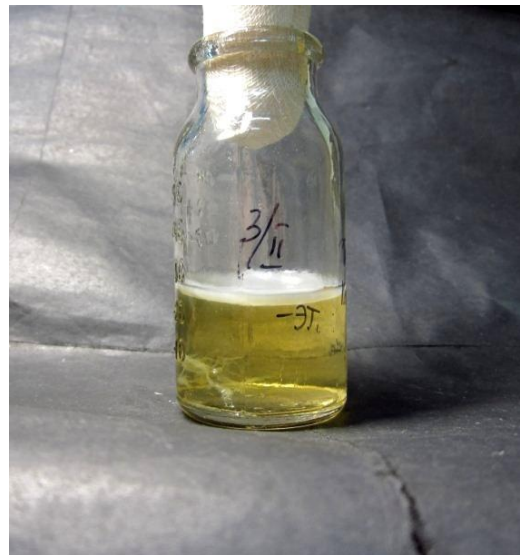
Руководитель:  
Михаил Райко  
Центр геномной биоинформатики  
им. Ф.Г. Добржанского



**Институт  
Биоинформатики**

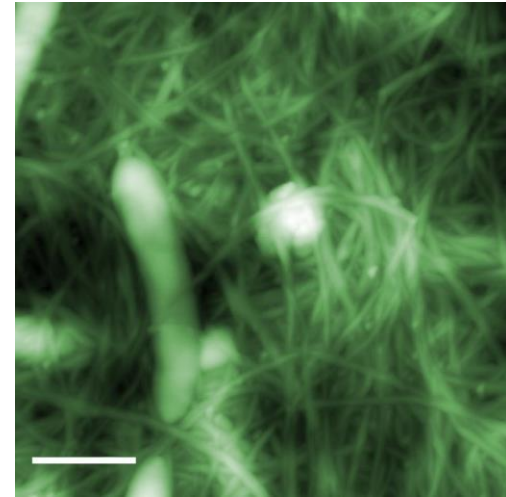
# *Gluconacetobacter* sp.

- Представители уксуснокислых бактерий:
  - Этанол → уксусная кислота, ацетат и лактат →  $\text{CO}_2$  и  $\text{H}_2\text{O}$



# *Gluconacetobacter* sp.

- Представители уксуснокислых бактерий:
  - Этанол → уксусная кислота, ацетат и лактат → CO<sub>2</sub> и H<sub>2</sub>O
- Развивающиеся культуры способны к синтезу целлюлозы



# Исследуемые штаммы

- *G. rhaeticus* AF1 (Corrêa dos Santos et al., 2014)
- *G. rhaeticus* CALU 1629

# Цель работы:

Пересборка бактериального генома *G. rhaeticus* AF1, сравнение со штаммом CALU 1629.

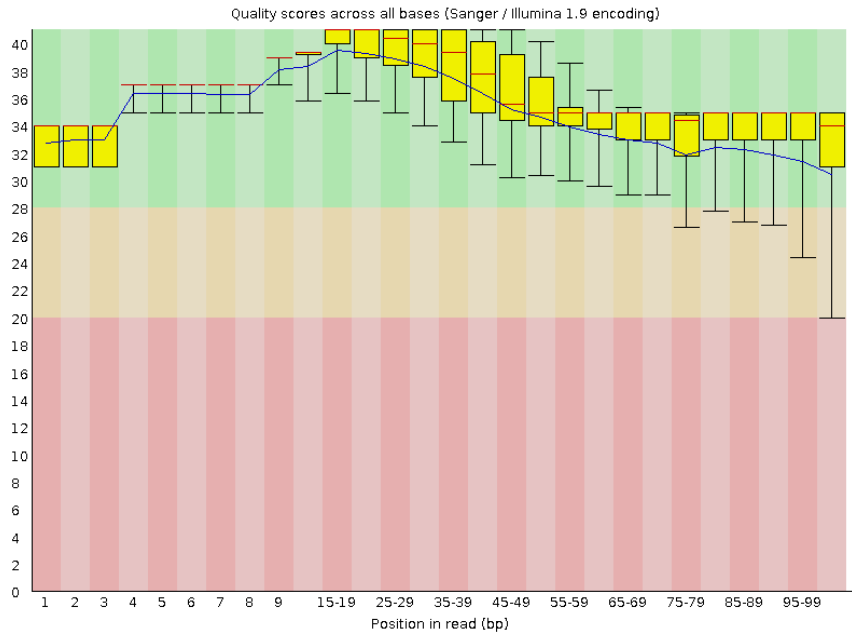
# Задачи:

1. Воспроизвести анализ из статьи (штамм AF1);
2. Пересобрать с помощью Platanus;
3. Сделать альтернативный скаффолдинг с SPAdes, BESST, OPERA, SORPA;
4. Сравнить со штаммом CALU 1629.

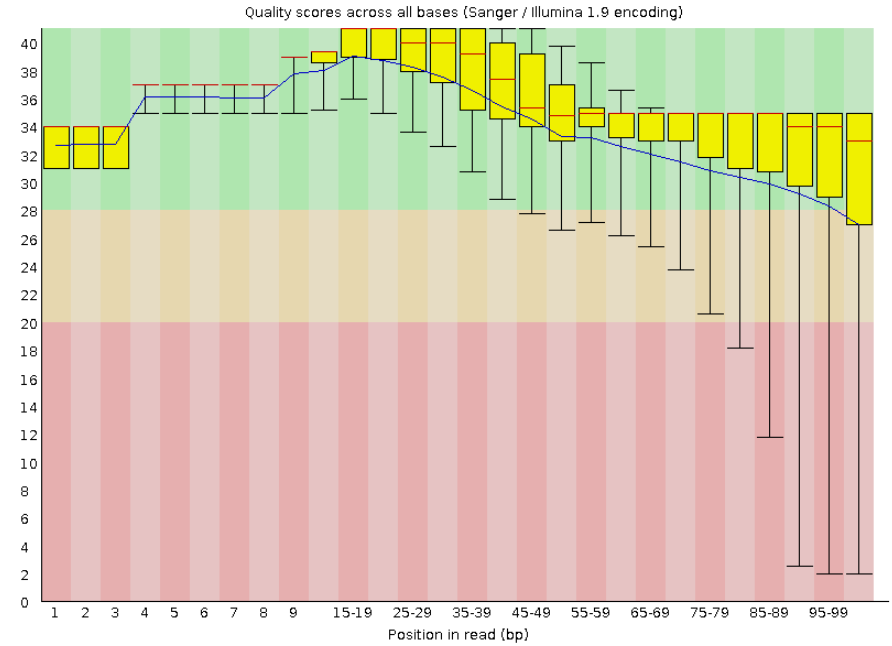
# Pipeline

1. Проверка качества исходных ридов (FastQC);
2. Очистка ридов (Trimmomatic);
3. Поиск и удаление последовательностей плазмид (Bowtie);
4. Сборка контигов (SPAdes);
5. Проверка качества сборки (Quast);
6. Проверка наличия характерных генов-маркеров бактерий (Amphora2);
7. Сборка скаффолда (SSPACE);
8. Проверка качества сборки (Quast).

# Проверка качества ридов

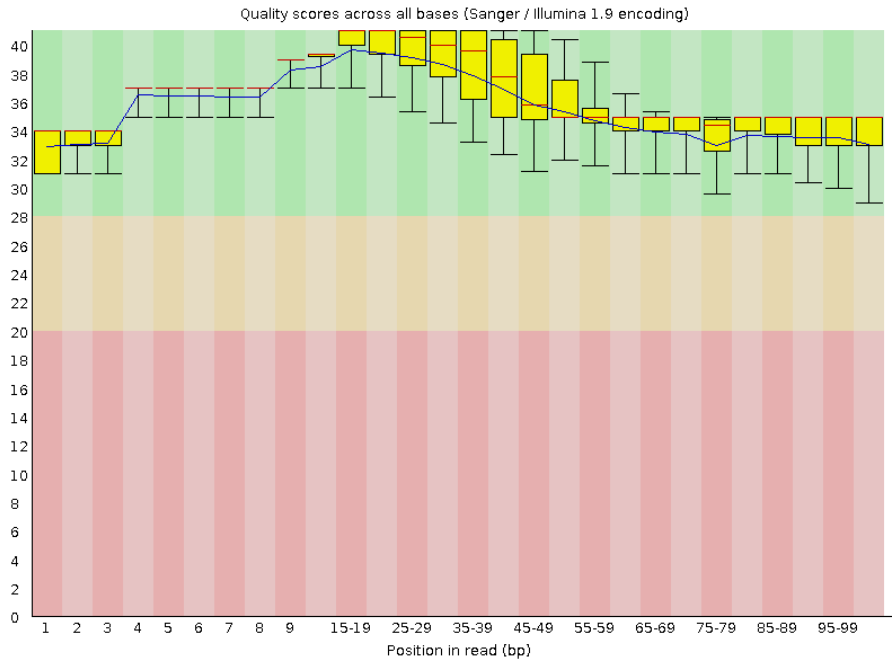


Forward reads

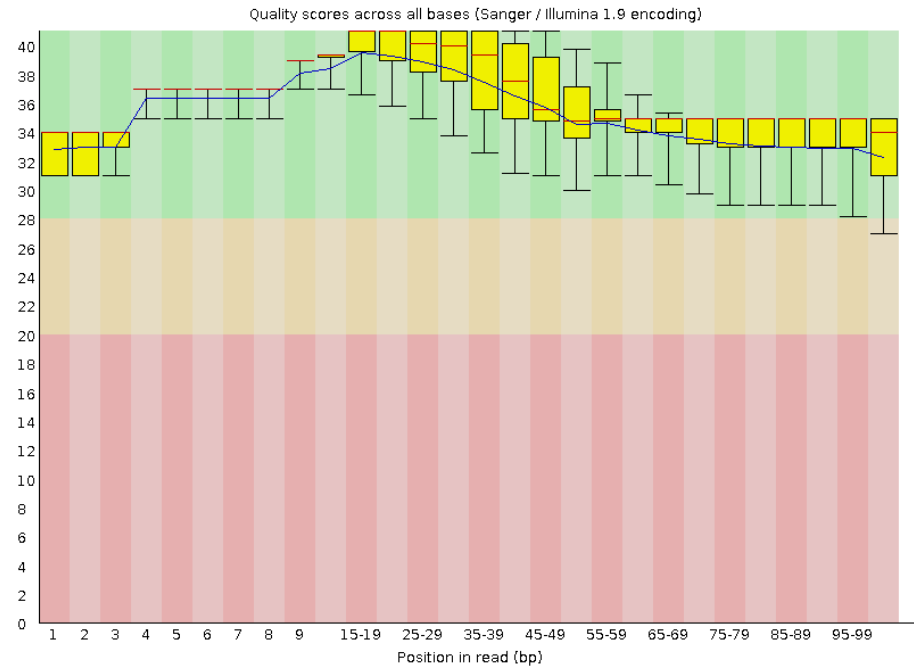


Reverse reads

# После очистки



Forward reads



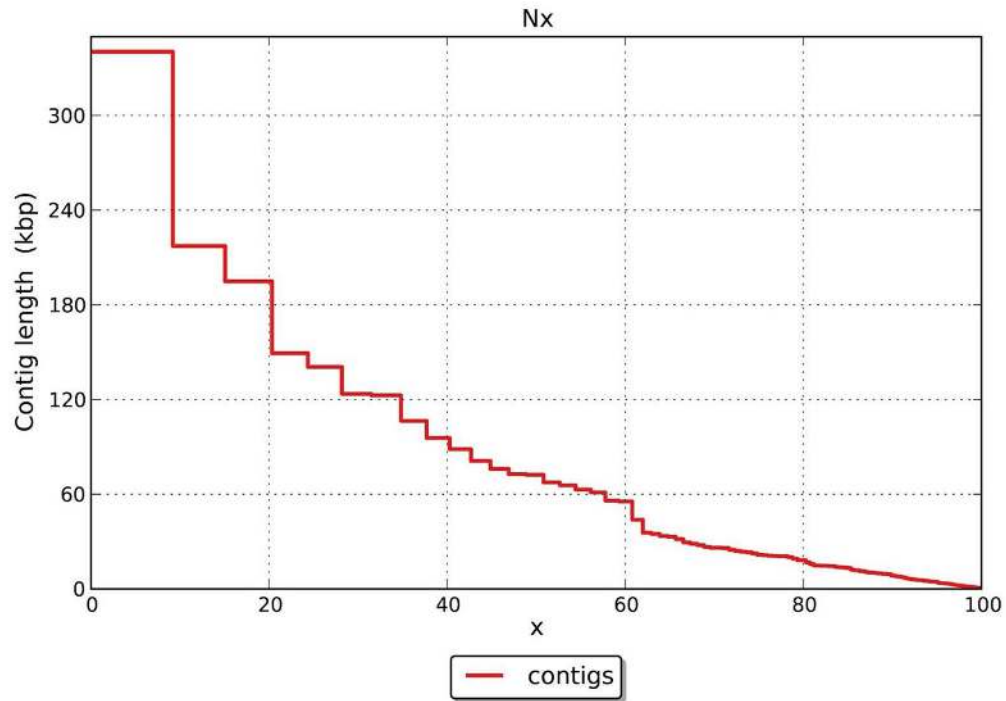
Reverse reads



# Pipeline

1. Проверка качества исходных ридов (FastQC);
2. Очистка ридов (Trimmomatic);
- 3. Поиск и удаление последовательностей плазмид (Bowtie);**
4. Сборка контигов (SPAdes);
5. Проверка качества сборки (Quast);
6. Проверка наличия характерных генов-маркеров бактерий (Amphora2);
7. Сборка скаффолда (SSPACE);
8. Проверка качества сборки (Quast).

# Сборка контигов (SPAdes)

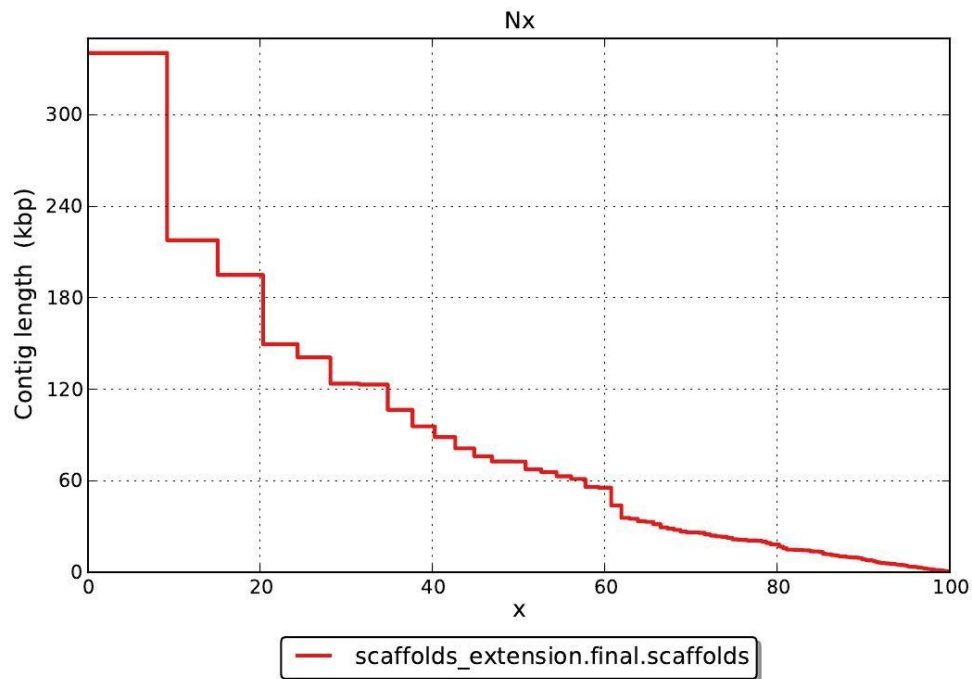


Statistics without reference	contigs
# contigs	222
Largest contig	340 395
Total length	3 700 623
N50	72 344

# Pipeline

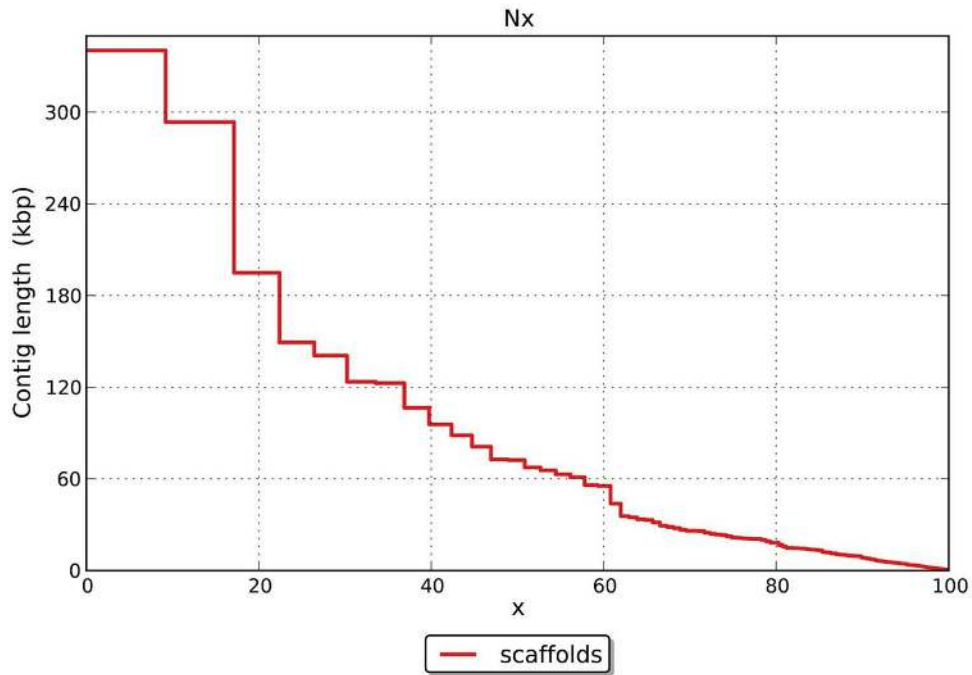
1. Проверка качества исходных ридов (FastQC);
2. Очистка ридов (Trimmomatic);
3. Поиск и удаление последовательностей плазмид (Bowtie);
4. Сборка контигов (SPAdes);
5. Проверка качества сборки (Quast);
- 6. Проверка наличия характерных генов-маркеров бактерий (Amphora2);**
7. Сборка скаффолда (SSPACE);
8. Проверка качества сборки (Quast).

# Сборка скаффолда (SSPACE)



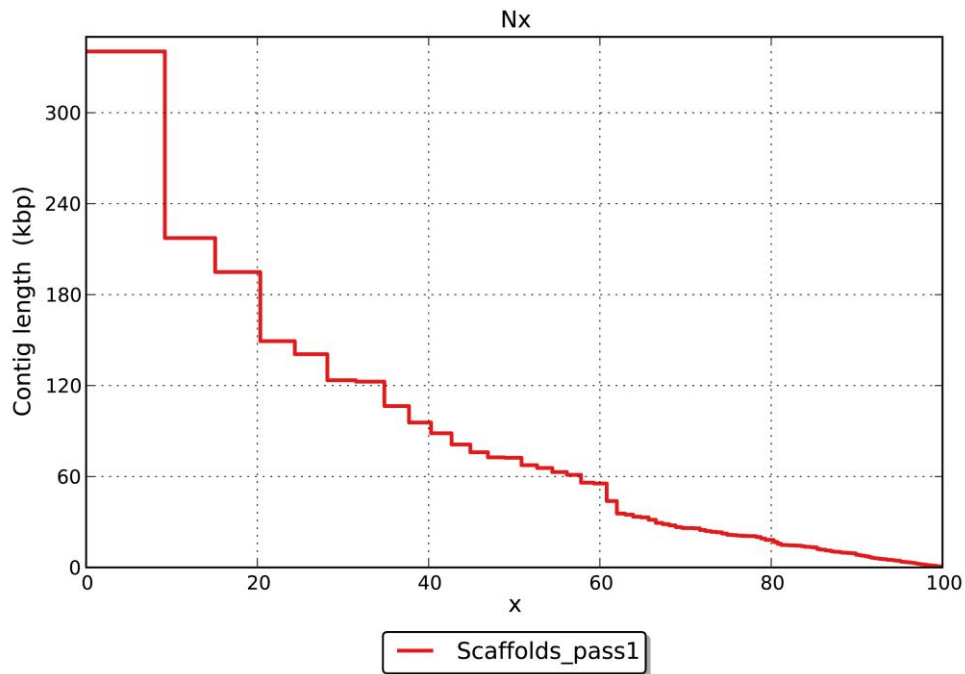
Statistics without reference	contigs
# contigs	224
Largest contig	340 441
Total length	3 708 566
N50	72 518

# Альтернативный скаффолдинг (SPAdes)



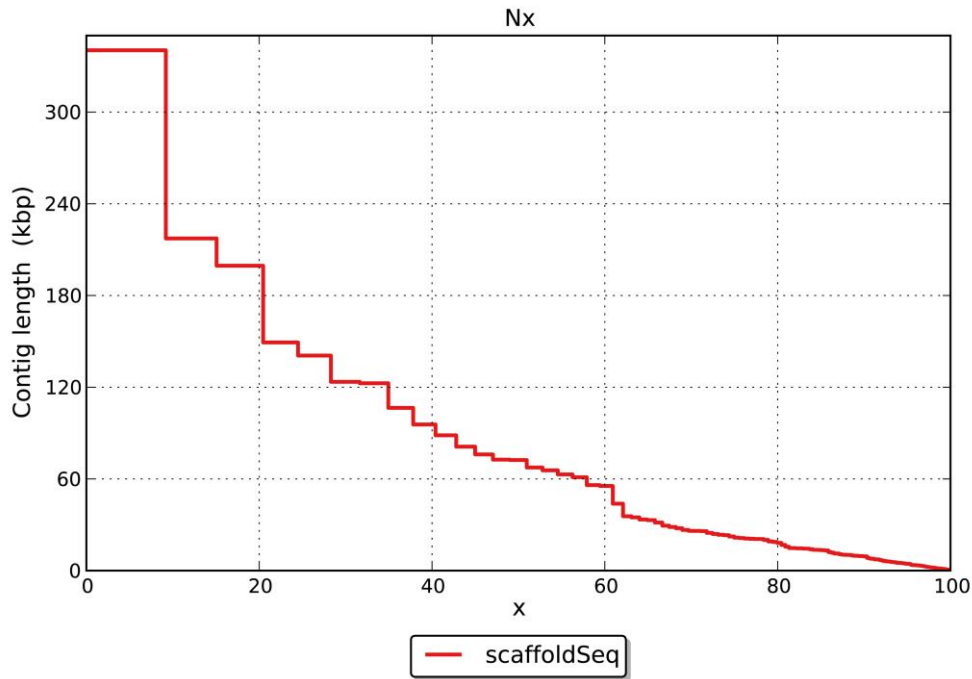
Statistics without reference	contigs
# contigs	220
Largest contig	340 395
Total length	3 700 900
N50	72 344

# Альтернативный скаффолдинг (BESST)



Statistics without reference	contigs
# contigs	222
Largest contig	340 395
Total length	3 700 623
N50	72 344

# Альтернативный скаффолдинг (OPERA)

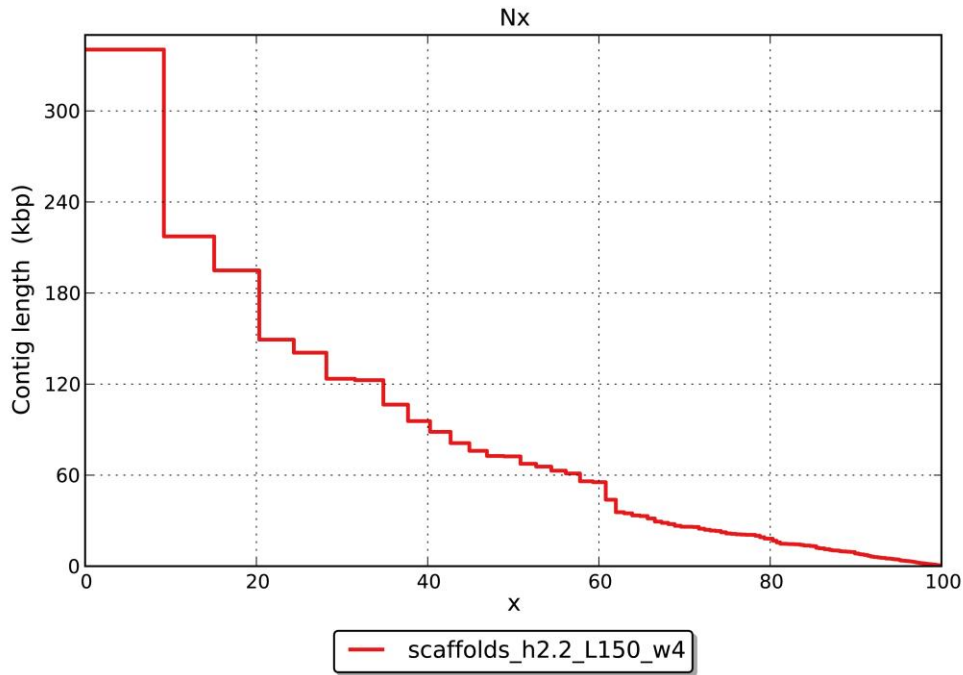


---

Statistics without reference	contigs
# contigs	213
Largest contig	340 395
Total length	3 700 713
N50	72 344

---

# Альтернативный скаффолдинг (SORPA)



Statistics without reference	contigs
# contigs	222
Largest contig	340 395
Total length	3 700 623
N50	72 344



# Platanus

---

	Construct contigs	Construct scaffolds	Gap close
# contigs	314	189	189
Largest contig	140 718	339 693	339 602
Total length	3 667 401	3 637 585	3 638 490
N50	27 327	72 349	72 325

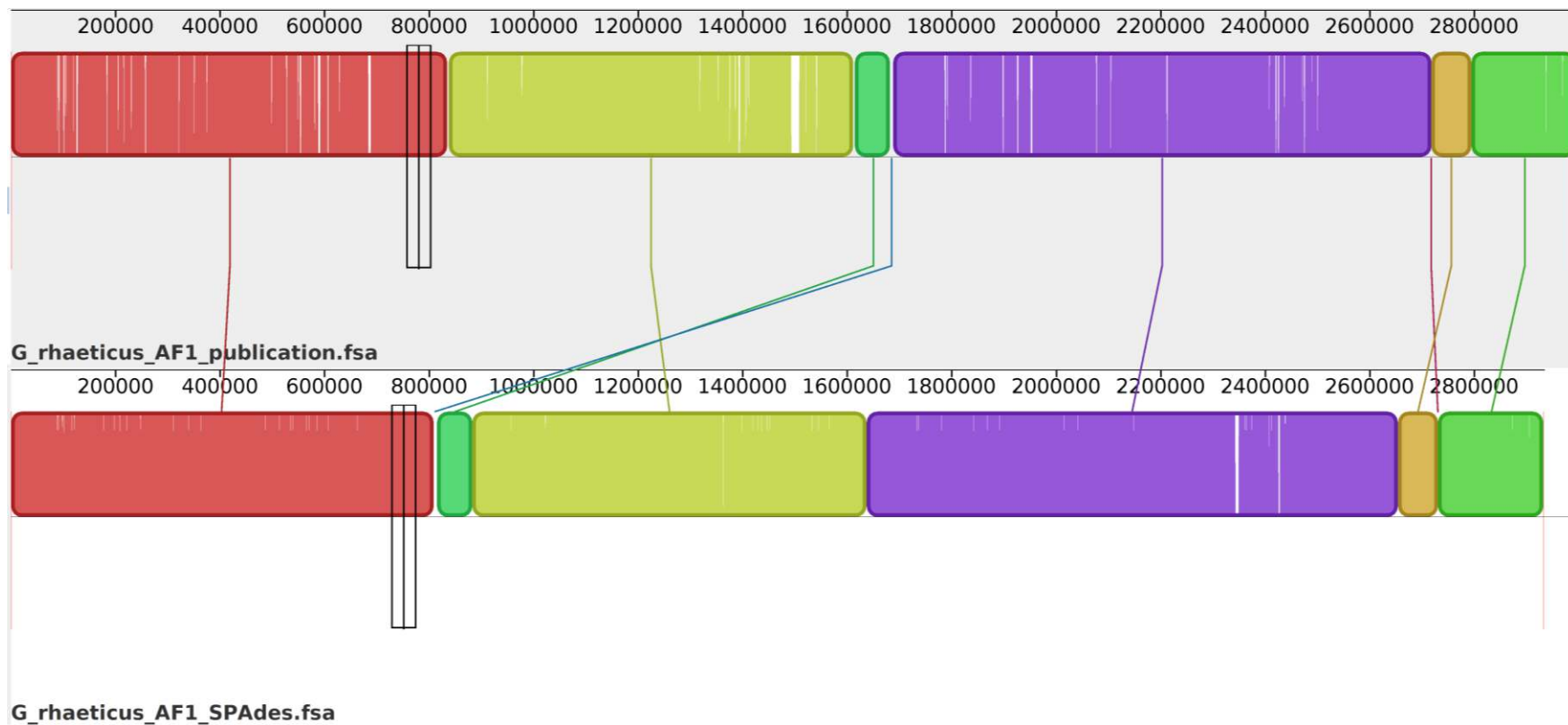
---

---

	<b>SSPACE</b>	<b>SPAdes</b>	<b>BESST</b>	<b>OPERA</b>	<b>SORPA</b>	<b>Platanus</b>	<b>Публикация</b>
# contigs	224	220	222	213	222	189	219
Total length	3 708 566	3 700 900	3 700 623	3 700 713	3 700 623	3 638 490	3 944 291
N50	72 518	72 344	72 344	72 344	72 344	72 325	73 183

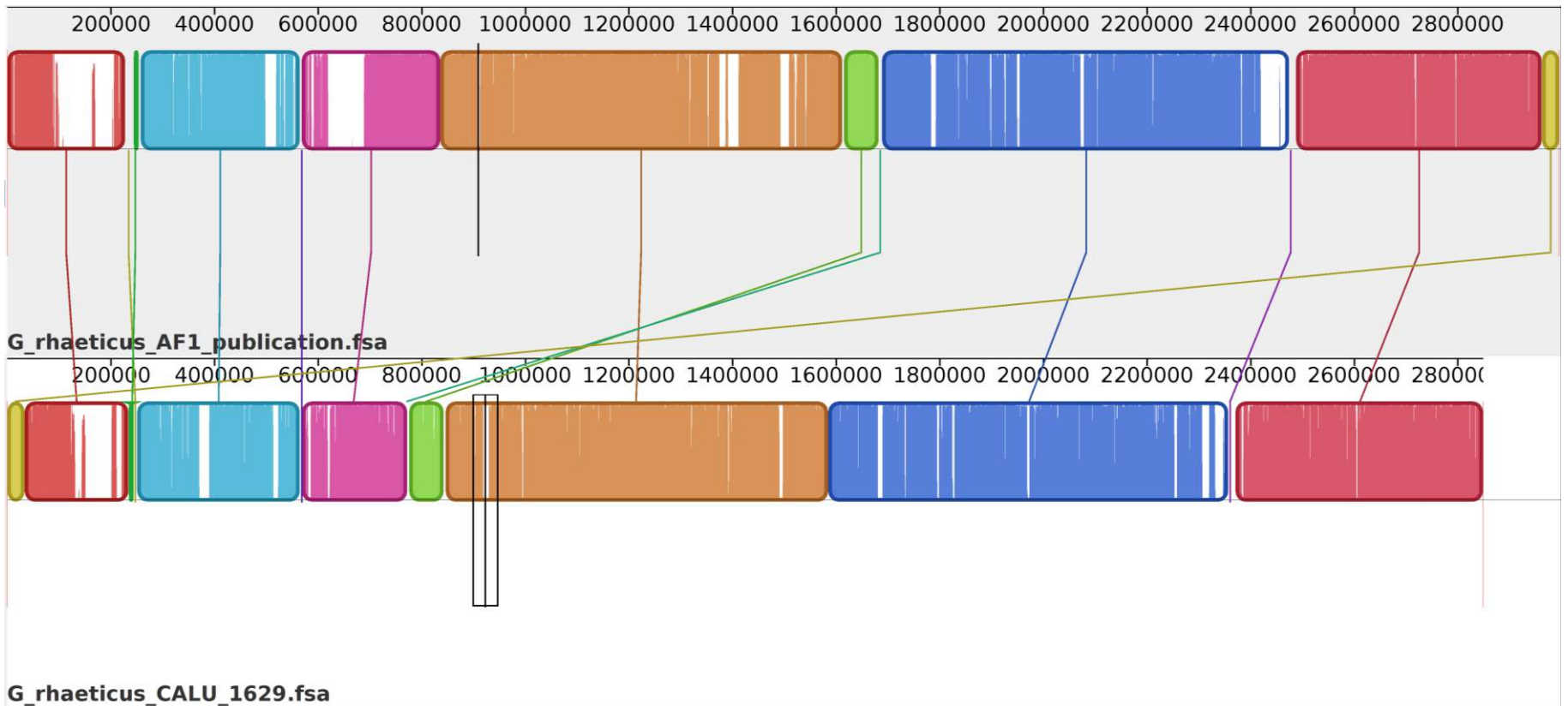
---

# Сравнение со сборкой из публикации



Выравнивание с использованием MAUVE (Multiple Genome Alignment)

# Сравнение со штаммом CALU 1629



Выравнивание с использованием MAUVE (Multiple Genome Alignment)

**Спасибо за внимание :)**