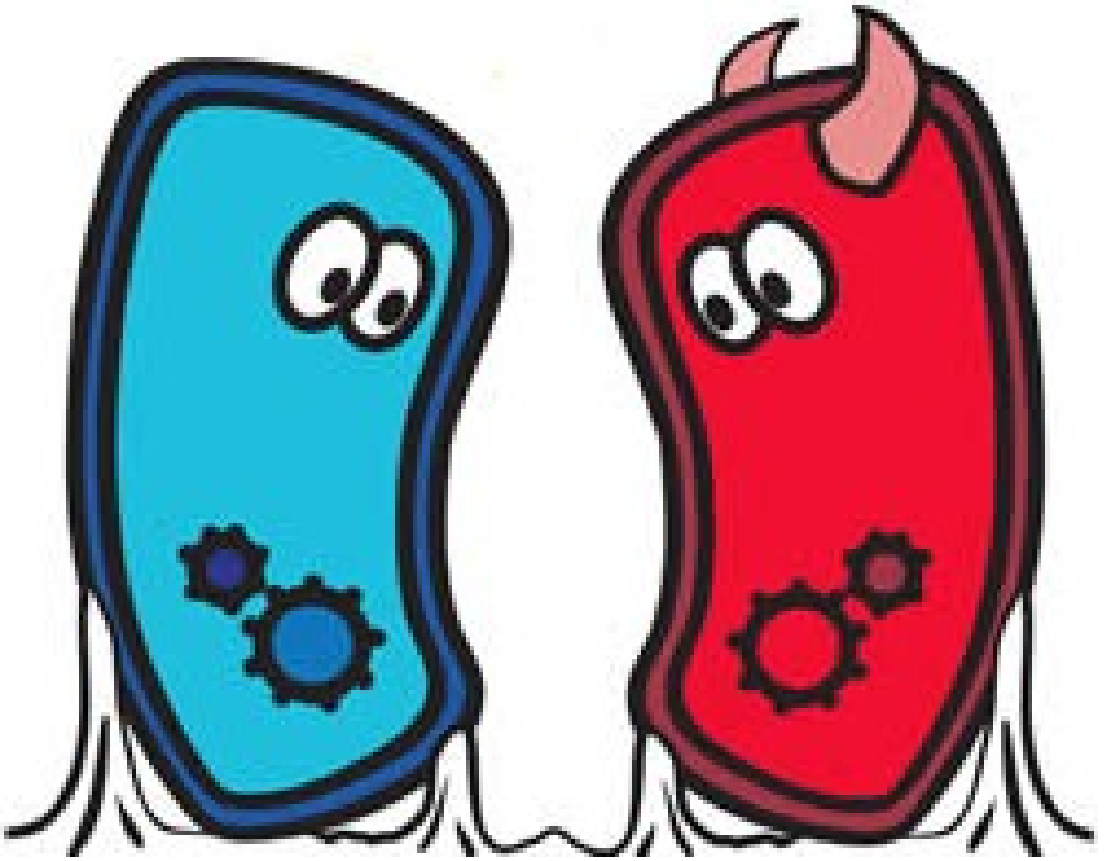
Поиск участков патогенности в геноме штамма *E.coli* 0104:H4

Руководитель проекта: Михаил Райко

В 2011 году в Германии произошла страшная эпидемия острой кишечной инфекции. Пострадало более 3'000 человек.

Почему новый штамм вызывает такие тяжёлые симптомы?



Задачи

1. Собрать геном нового штамма
2. Сравнить с родственным штаммом
3. Найти причину аномальной патогенности

Этапы геномного проекта

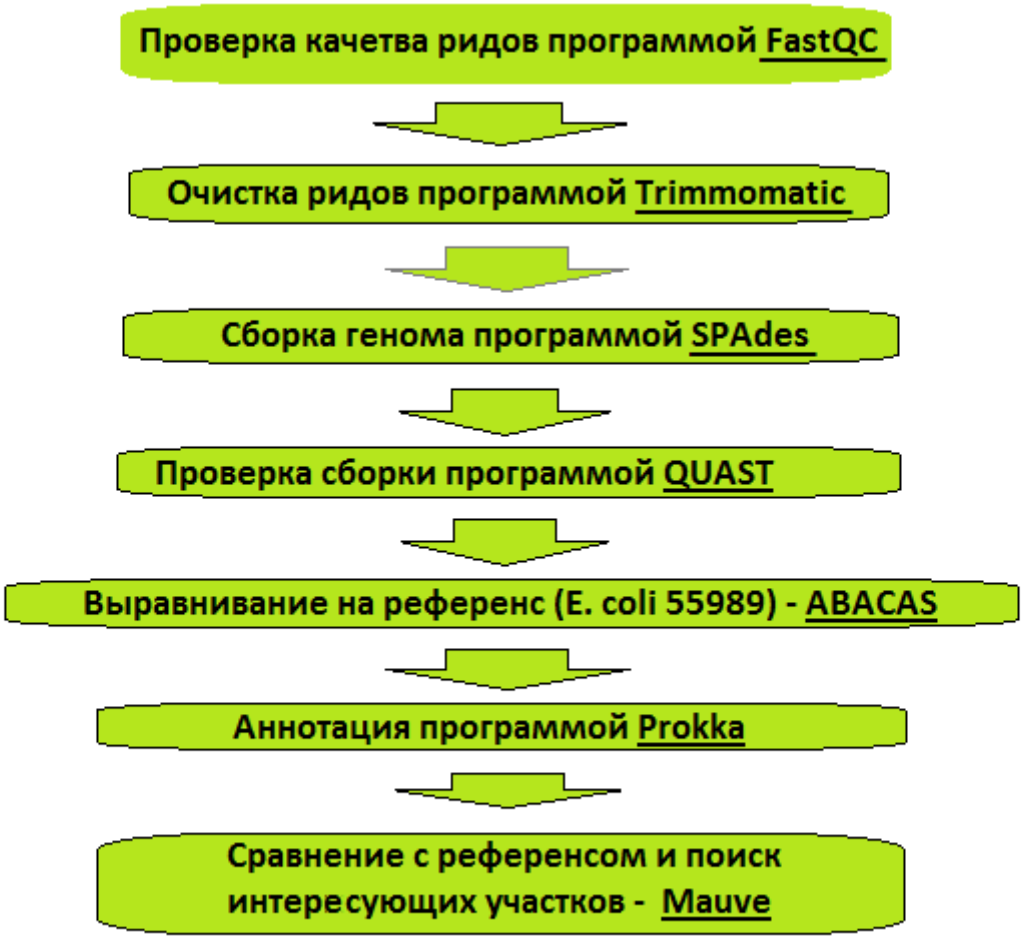
1. Убедиться в качестве исходных ридов (FastQC)
2. Собрать контиги (SPAdes)
3. Проверить качество сборки (Quast)
4. Собрать скаффолд с помощью референсного генома (Abacas, Contiguator)
5. Проаннотировать геном (Prokka, RAST)
6. Сравнить с референсным геномом (Mauve)

Группа №1

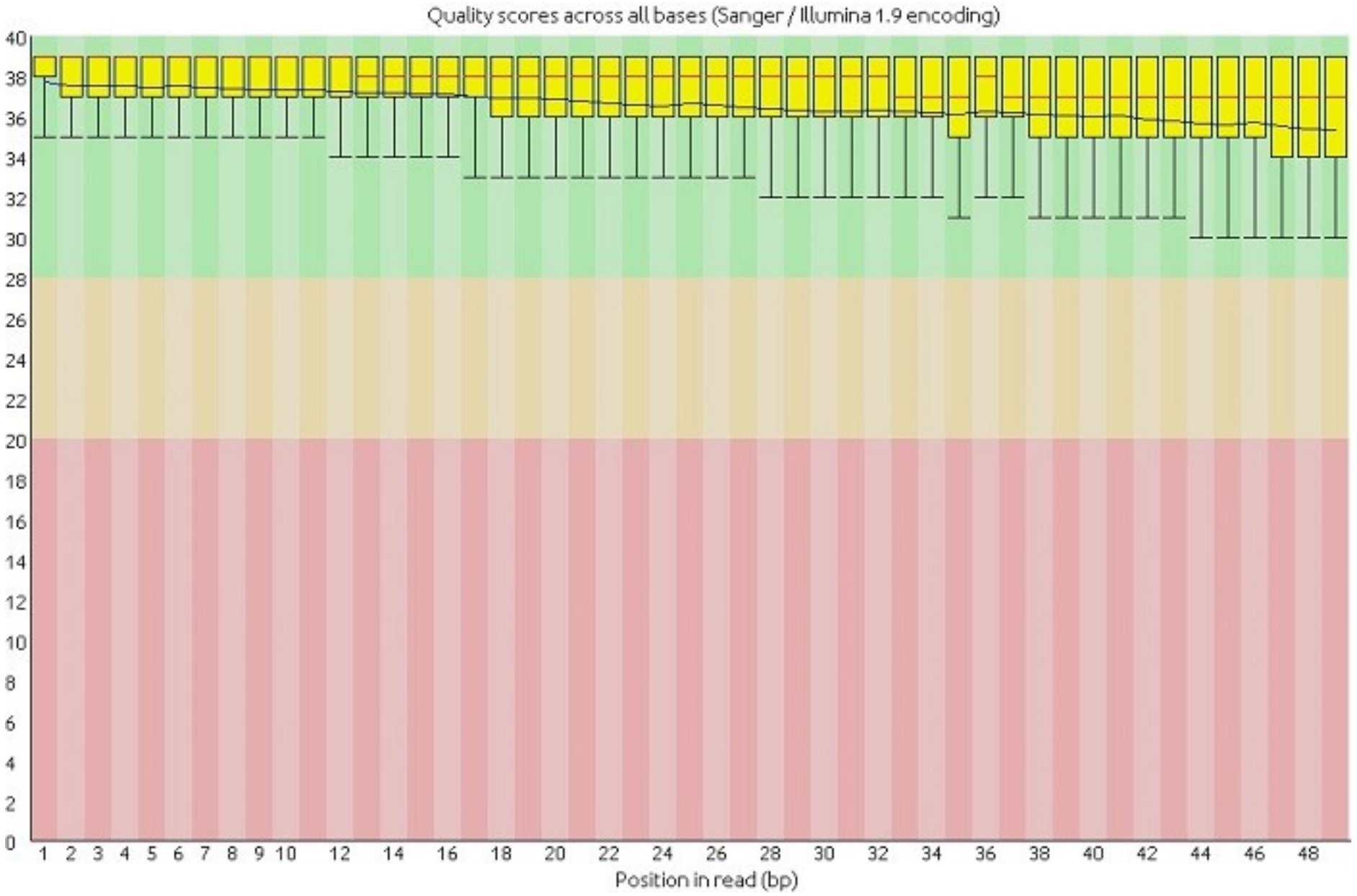
Доменикан Александр
Шишкина Элина
Бутенко Николай
Зайнагтдинова Галия

Собственно Pipeline

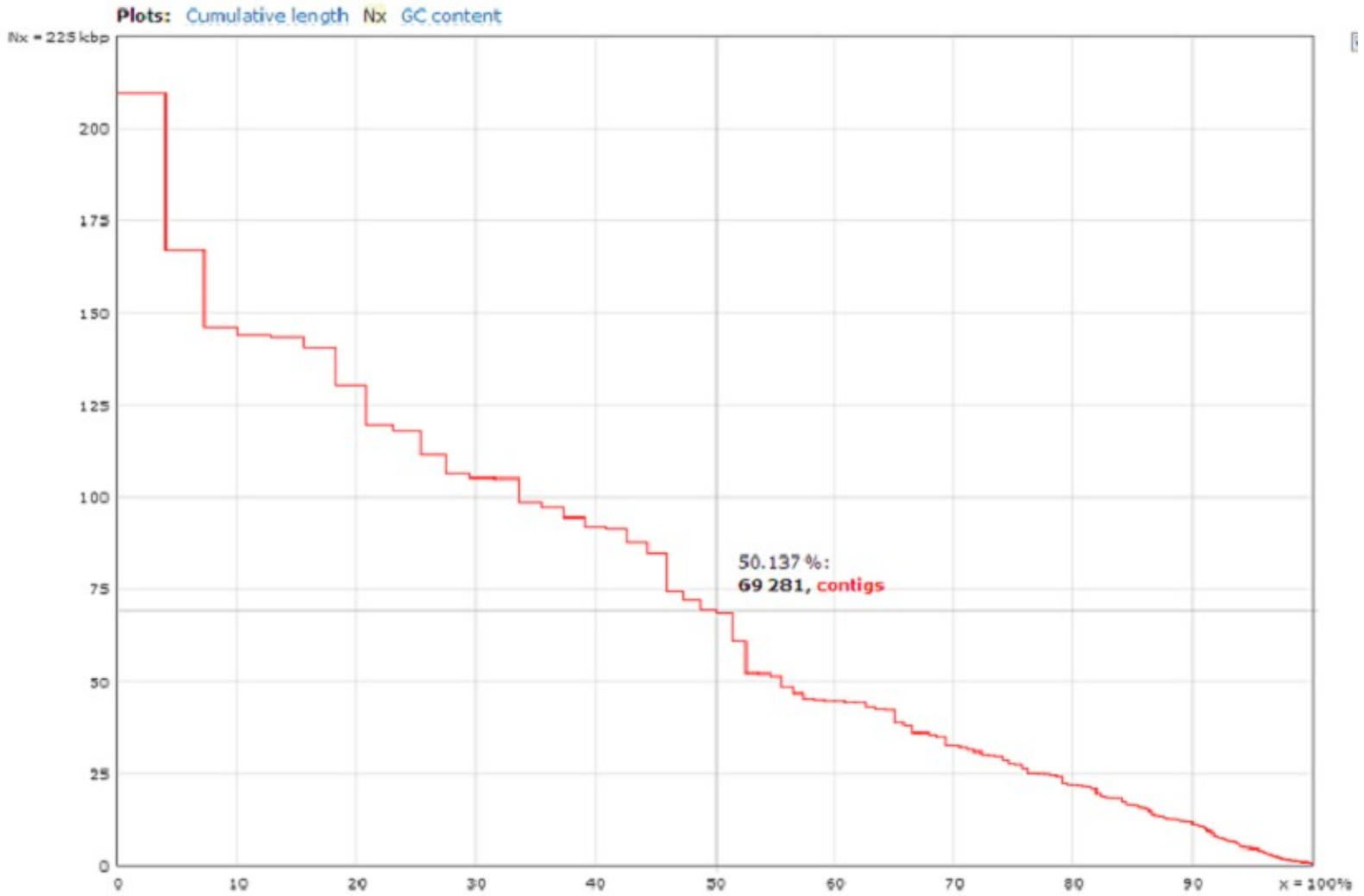
Pipeline


Проверка качетва ридов программой FastQC
Очистка ридов программой Trimmomatic
Сборка генома программой SPAdes
Проверка сборки программой QUAST
Выравнивание на референс (E. coli 55989) - ABACAS
Аннотация программой Prokka
Сравнение с референсом и поиск интересующих участков - Mauve

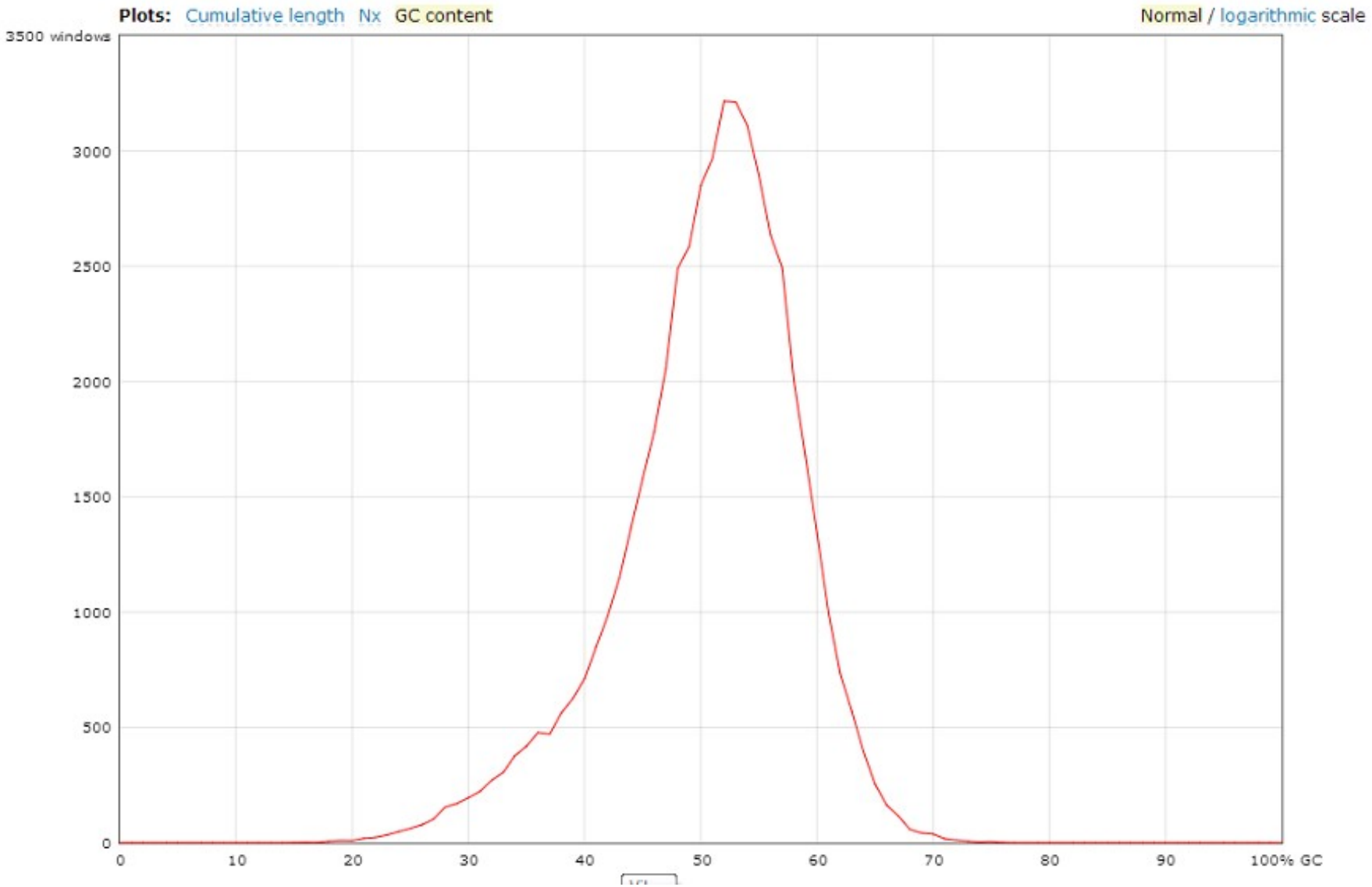
Проверка качества ридов



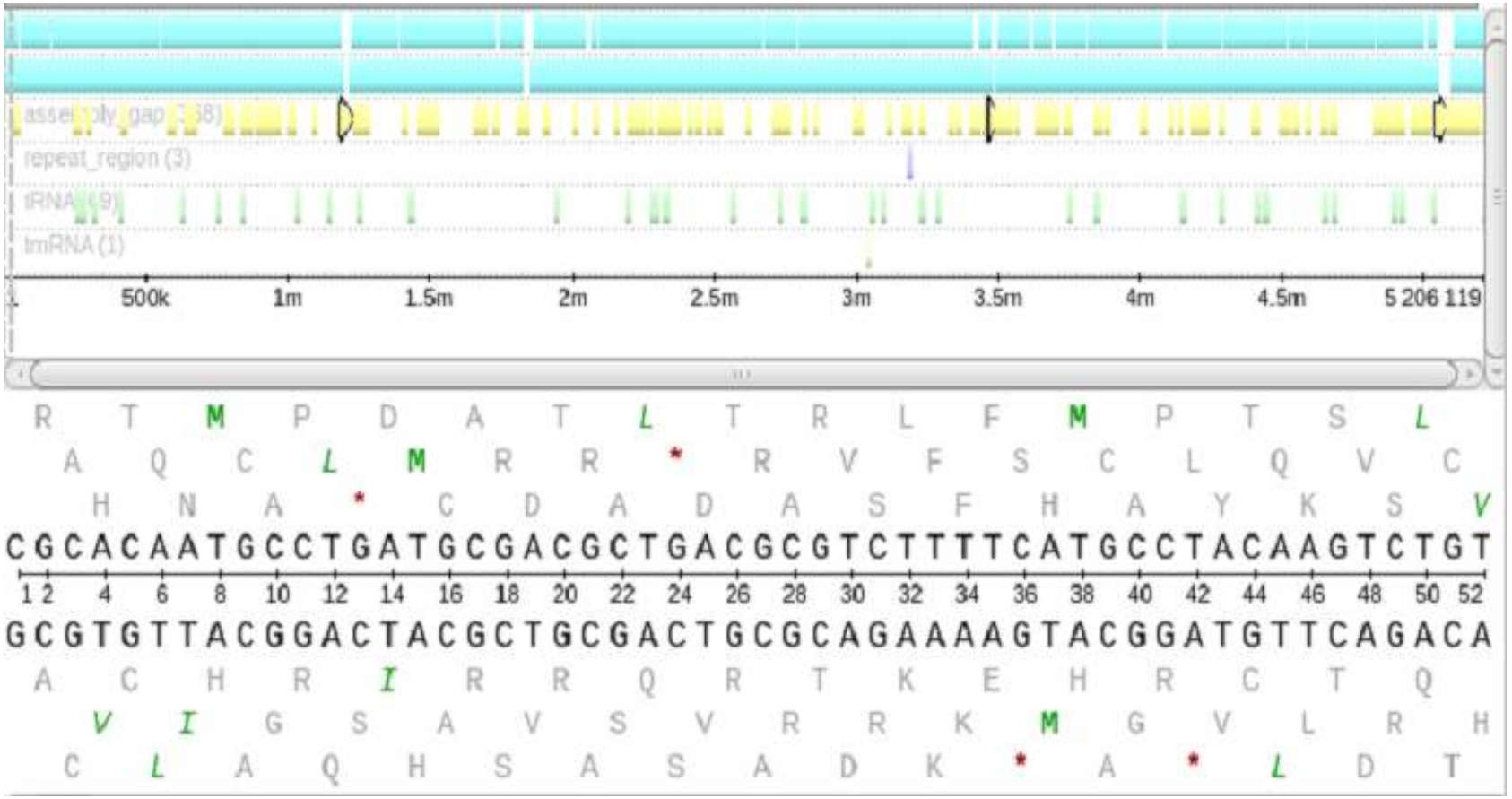
Проверка качества сборки (N50)



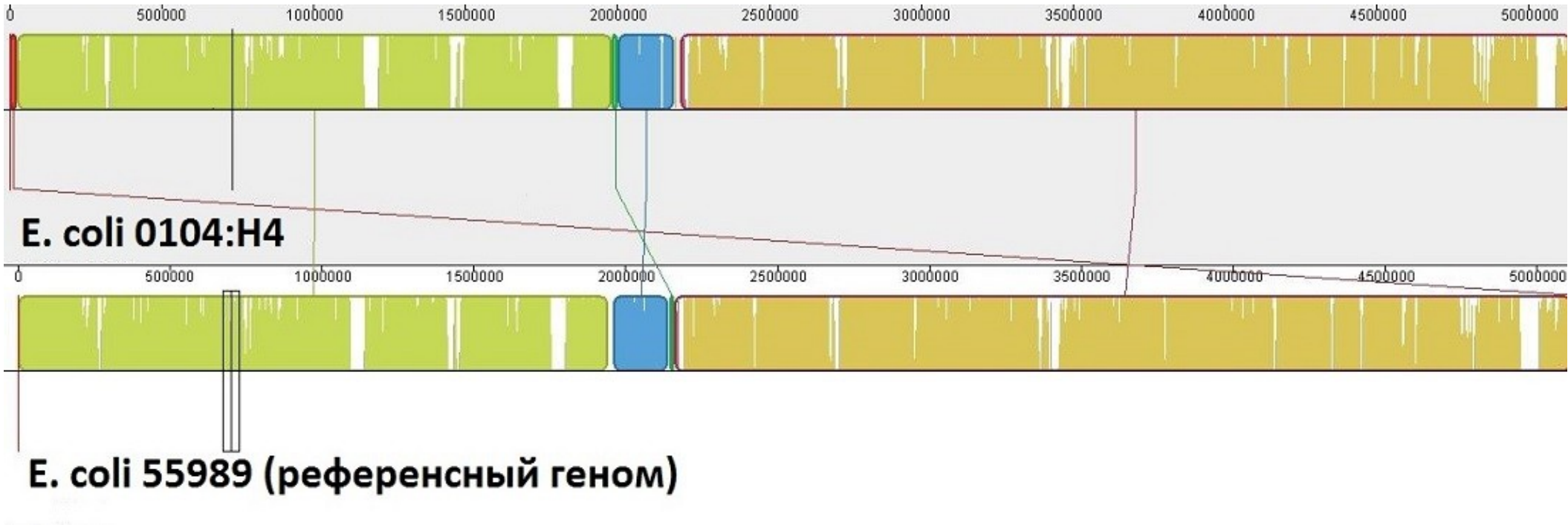
Проверка качества сборки (GC контент)



Визуализация аннотации (Ugene)

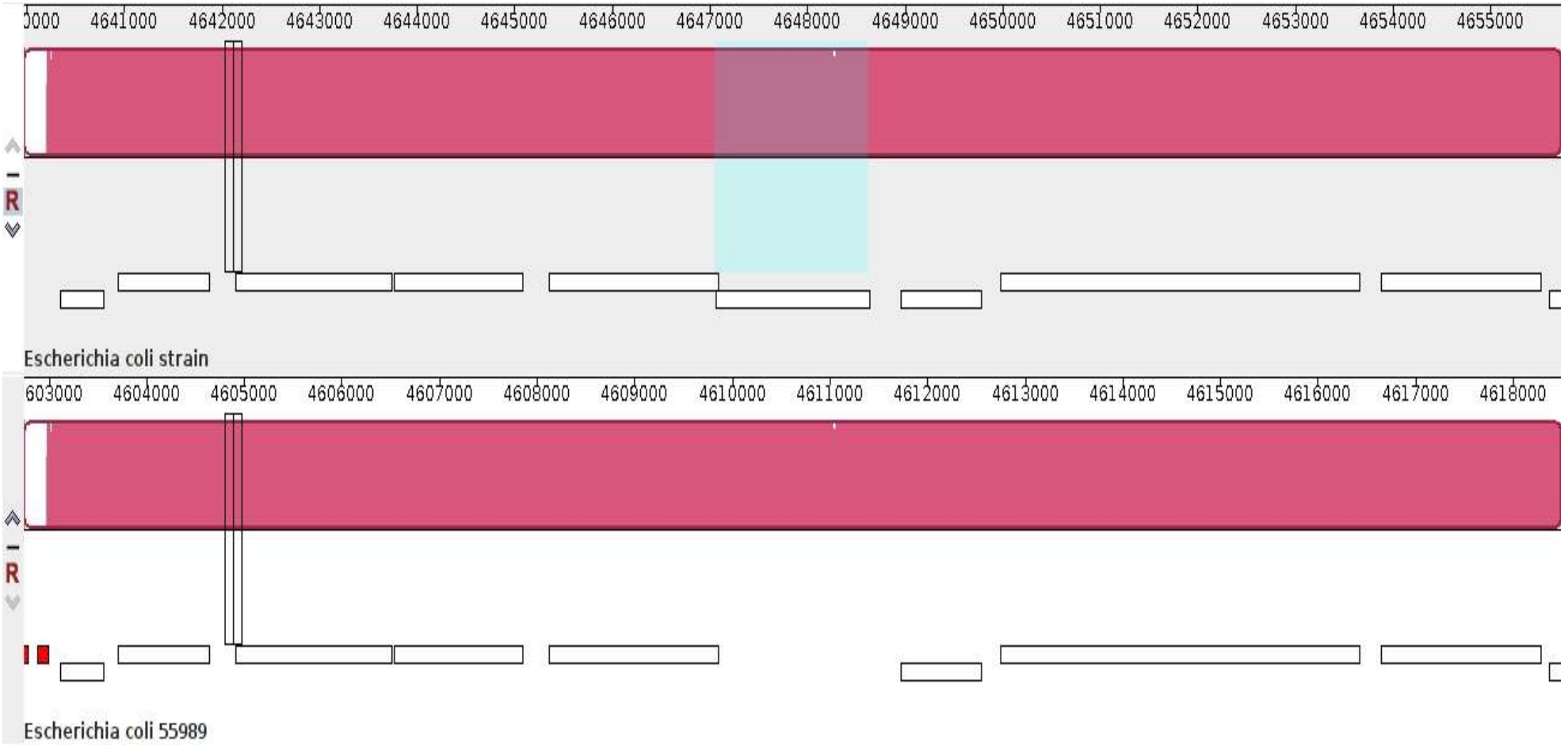


Сравнение с геномом E. coli 55989

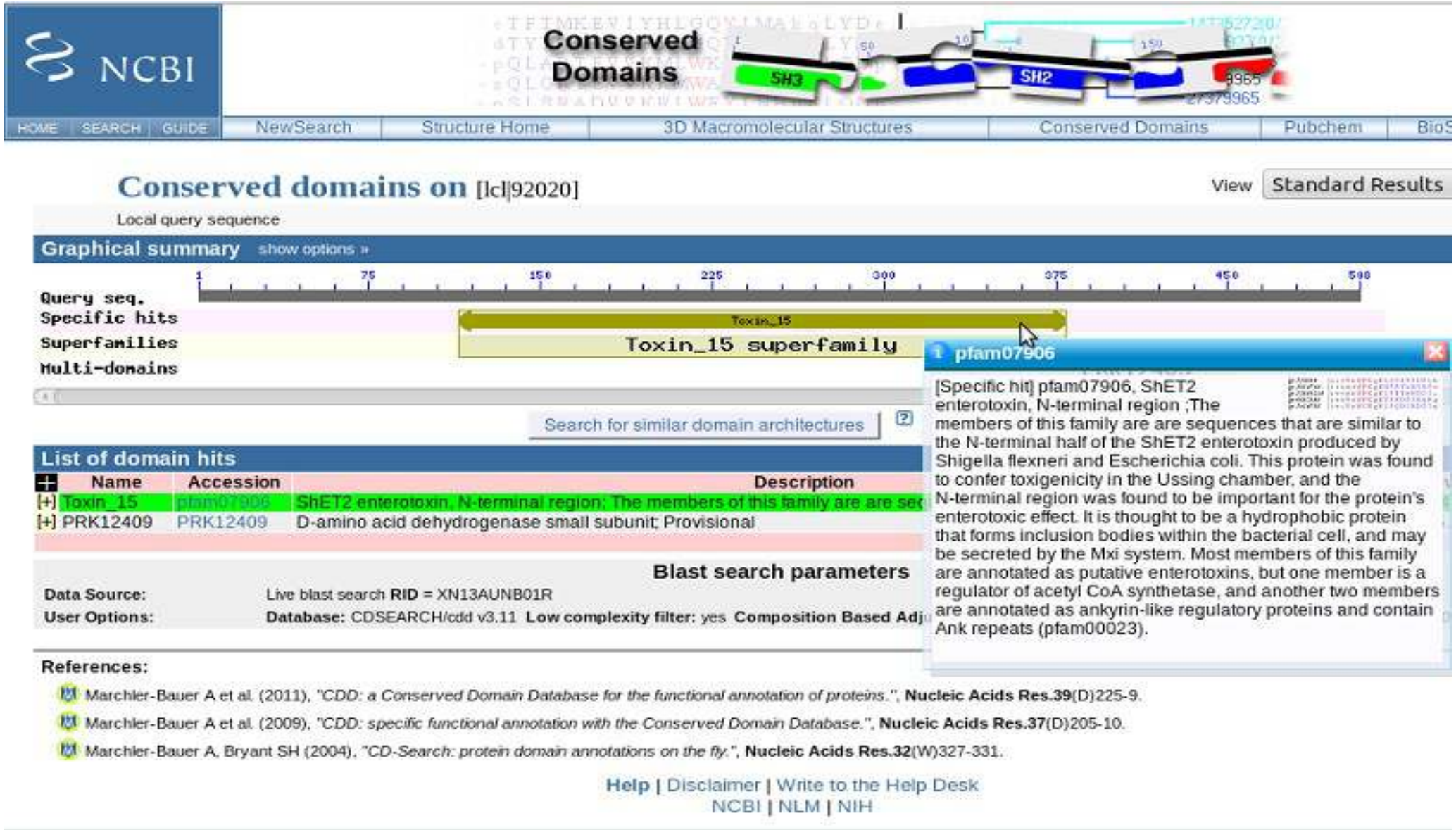


Как мы искали нужный участок? Ручками.

Note: 'enterotoxin'



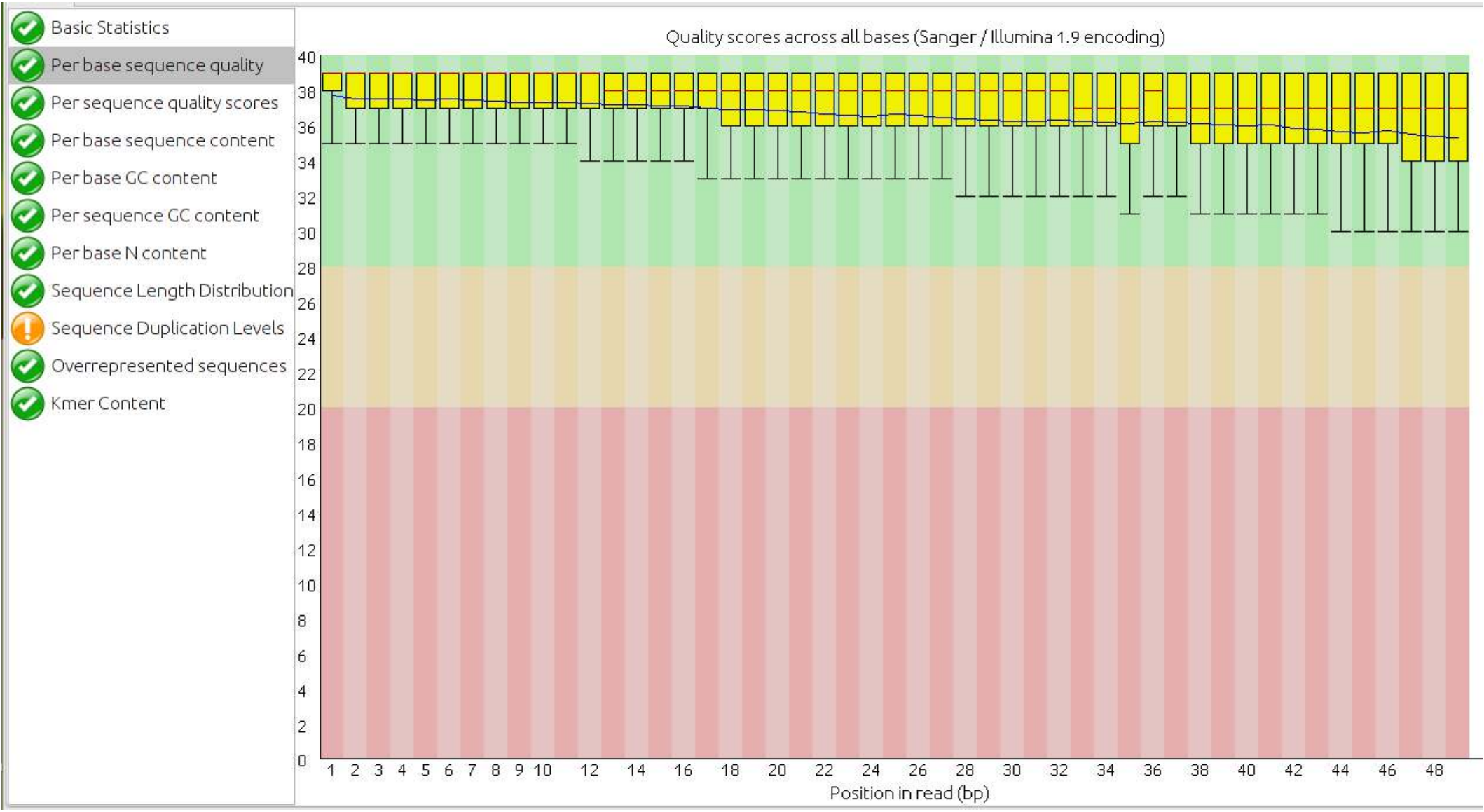
Результат.
Выравние BLASTом на гомологичные белки



Группа №2

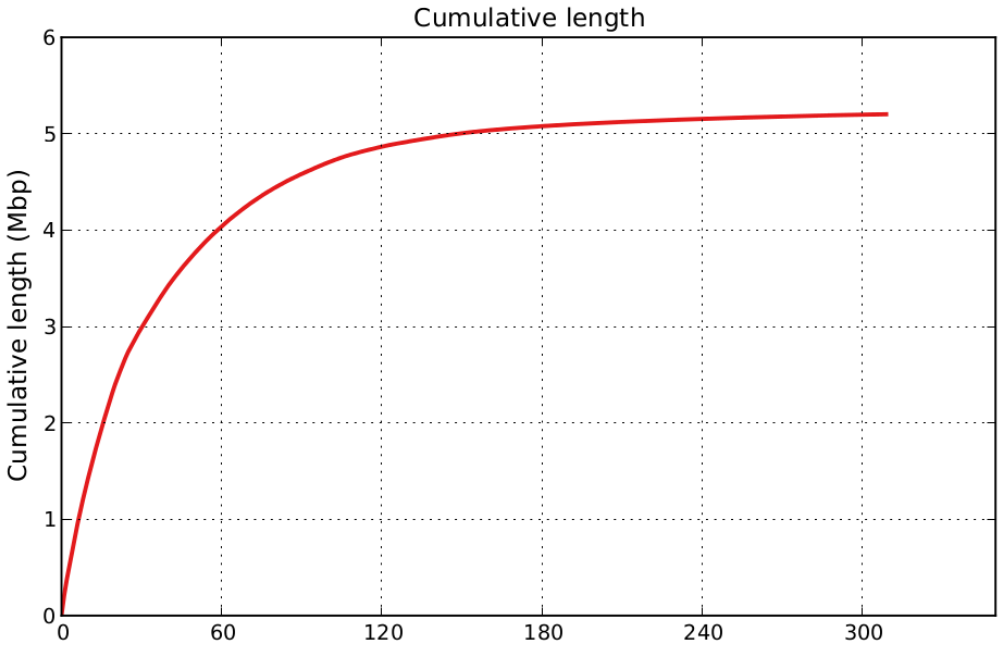
- Голокоз Александр
- Гусак Юлия
- Котенко Анастасия
- Малиновский Илья
- Ненарокова Анна
- Орлова Марина

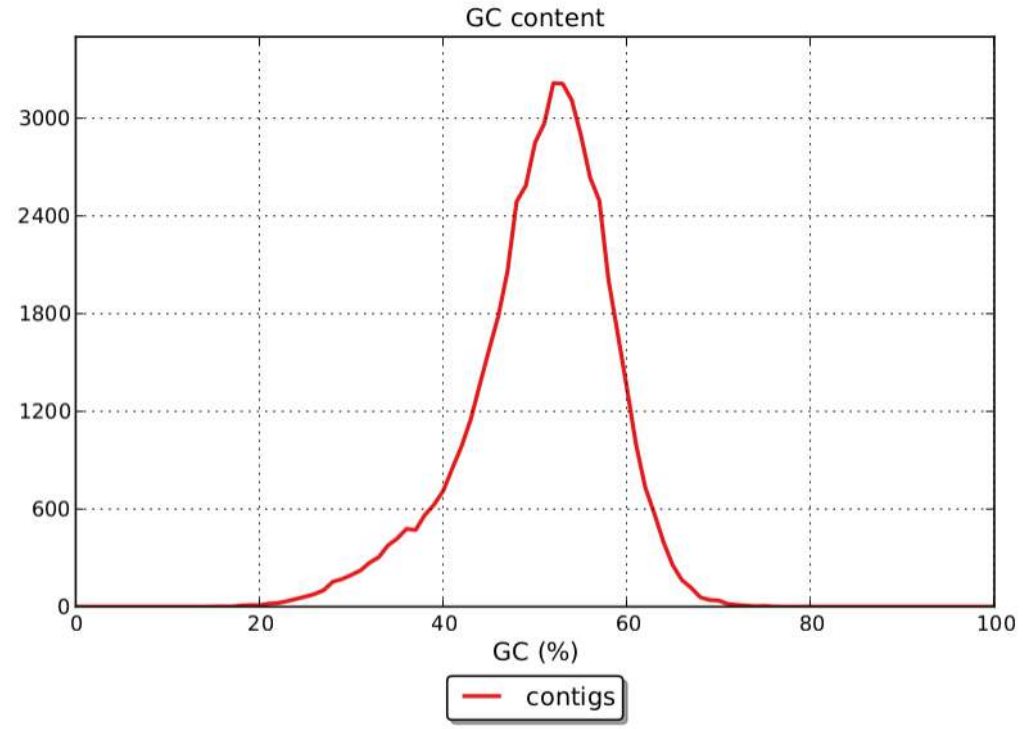
Качество исходных ридов

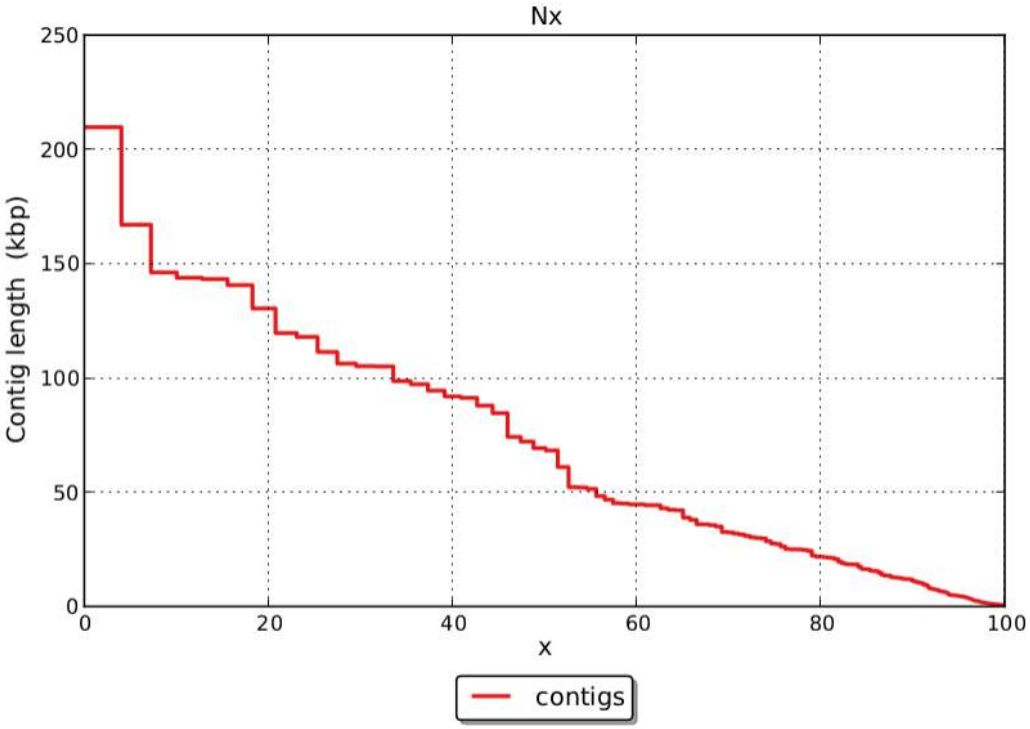


Отчёт о качестве сборки

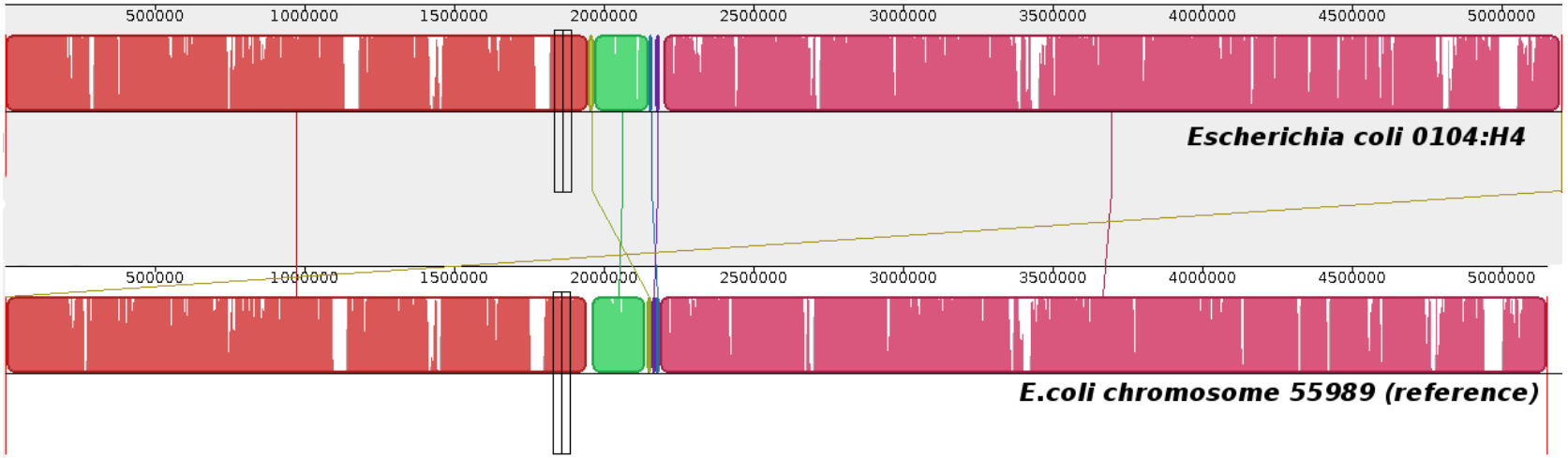
Statistics without reference	contigs
# contigs	309
Largest contig	209 682
Total length	5 201 829
N50	69 281
Mismatches	
# N's per 100 kbp	0

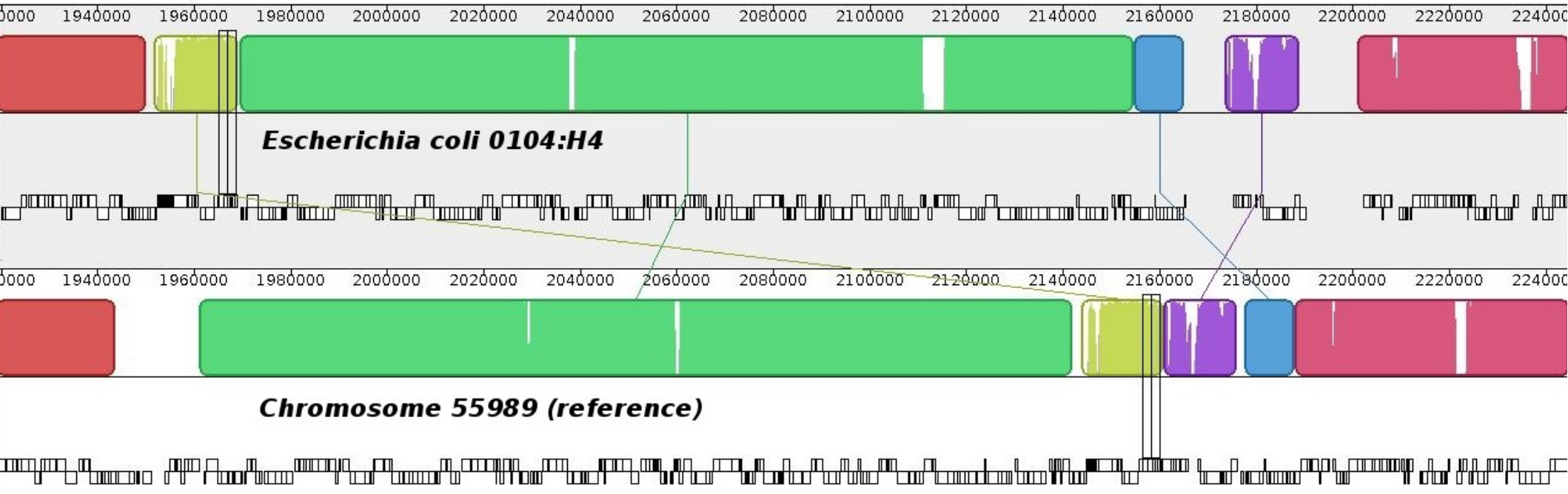






Сравнение изучаемого штамма с родственным





Скрипт для автоматической обработки ридов

Скрипт для сравнения gbk файлов

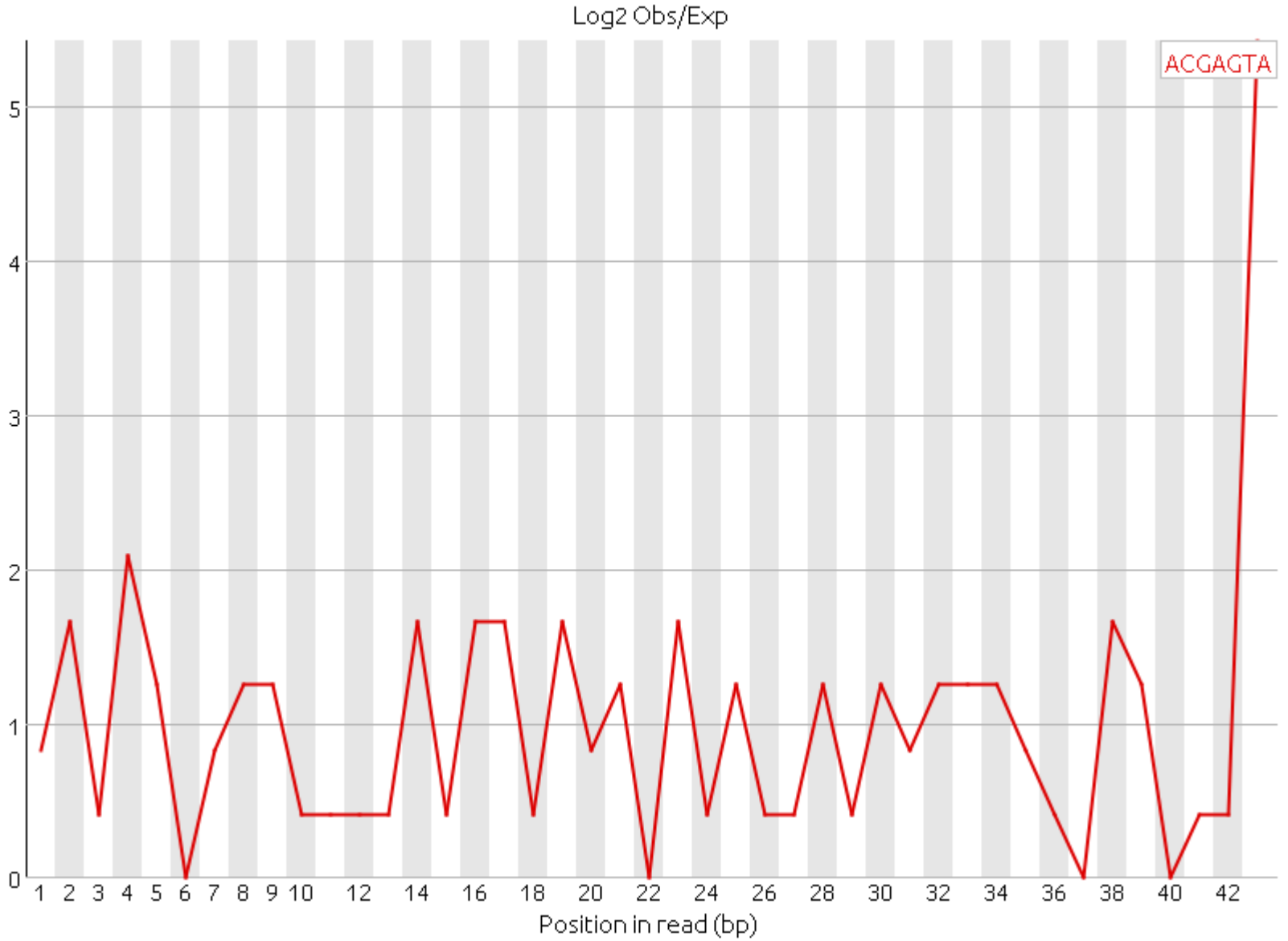
1. Сравнивает по названию генов и в output выдает названия новых генов и их белков

2. Сравнивает последовательности белков и выдает названия новых белков

Группа 3

Иванов Кирилл
Фурменков Александр
Митрофанова Ольга

Отчет в FastQC



Приобретенный опыт

- Навыки работы с ранее незнакомым биоинформатическим софтом
- Знакомство с новыми форматами представления данных и их структурой
- Практический скилл сборки генома с проверками качества, ранее незнакомый
- Улучшение мастерства в черной магии, необходимой для получения результатов там, где другие методы не подходят

Результаты

- Найден ген шига-токсина, который отсутствует у референсного штамма
- Найдены гены фимбрий I типа
- Найдены гены других токсинов, которых нет у референсного штамма
- В базе NCBI найдены гомологичные последовательности для несобранных контигов, они принадлежат *E.coli* и родственным видам

Спасибо за внимание!