

Сборка de novo

Подготовка чтений

Я составила файл adapters.fasta из всех файлов с адаптерами из папки /mnt/scratch/NGS/adapters с помощью nano. Далее я провела триммирование:

```
TrimmomaticSE -phred33 SRR4240360.fastq.gz trim.fastq.gz ILLUMINACLIP:adapters.fasta:2:7:7
```

Результаты:

Input Reads: 8254632 Surviving: 8212774 (99.49%) Dropped: 41858 (0.51%)

По итогам фильтрования удалилось 41858 прочтений, они и содержали в себе адаптерные последовательности.

Я ввела команду для удаления с правых концов чтений нуклеотиды с качеством ниже 20 и чтений, длина которых меньше 32 нуклеотидов:

```
TrimmomaticSE -phred33 trim.fastq.gz trim_1.fastq.gz TRAILING:20 MINLEN:32
```

Результаты:

Input Reads: 8212774 Surviving: 7915474 (96.38%) Dropped: 297300 (3.62%)

На вход было дано 8212774 чтений, удалено 297300 (3.62%)

Размеры файлов:

- Скачанный файл – 194 Мб
- Чтения без адаптеров – 193 Мб
- Чтения после очистки – 184 Мб

Подготовка k-меров

Для создания k-меров длиной k=31 я написала команду:

```
velveth . 31 -short -fastq.gz trim_1.fastq.gz
```

Сборка de novo

Для сборки de novo из получившихся k-меров я использовала команду:

```
velvetg .
```

N50=43070

С помощью команд grep, cut и sort я пришла к следующим выводам:

Табл. 1. Характеристика трех самых длинных контигов из сборки

| ID | Длина | Покрытие |
|----|--------|-----------|
| 1 | 113474 | 33.525459 |
| 5 | 83603 | 33.646065 |
| 4 | 64155 | 35.847324 |

Табл. 2. Характеристика контига с аномально большим покрытием

| ID | Длина | Покрытие |
|----|-------|------------|
| 40 | 69 | 109.391304 |

Табл. 3. Характеристика одного из контигов с аномально маленьким покрытием

| ID | Длина | Покрытие |
|-----|-------|----------|
| 565 | 31 | 1.612903 |

Анализ

С помощью программы megablast на сайте NCBI я сделала выравнивания каждого из 3 самых длинных контигов с хромосомой *Buchnera aphidicola* (GenBank/EMBL AC — CP009253).

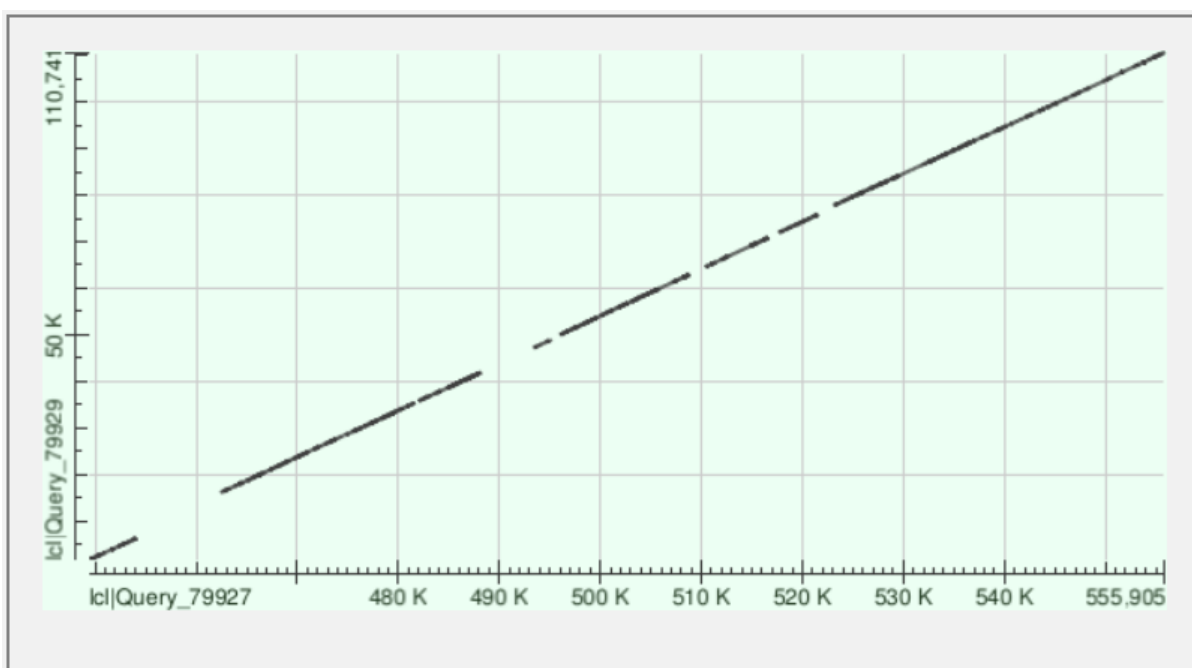


Рис. 1. Выравнивание с контигом 1

1 контиг выровнялся на участок 449-551к на хромосоме.

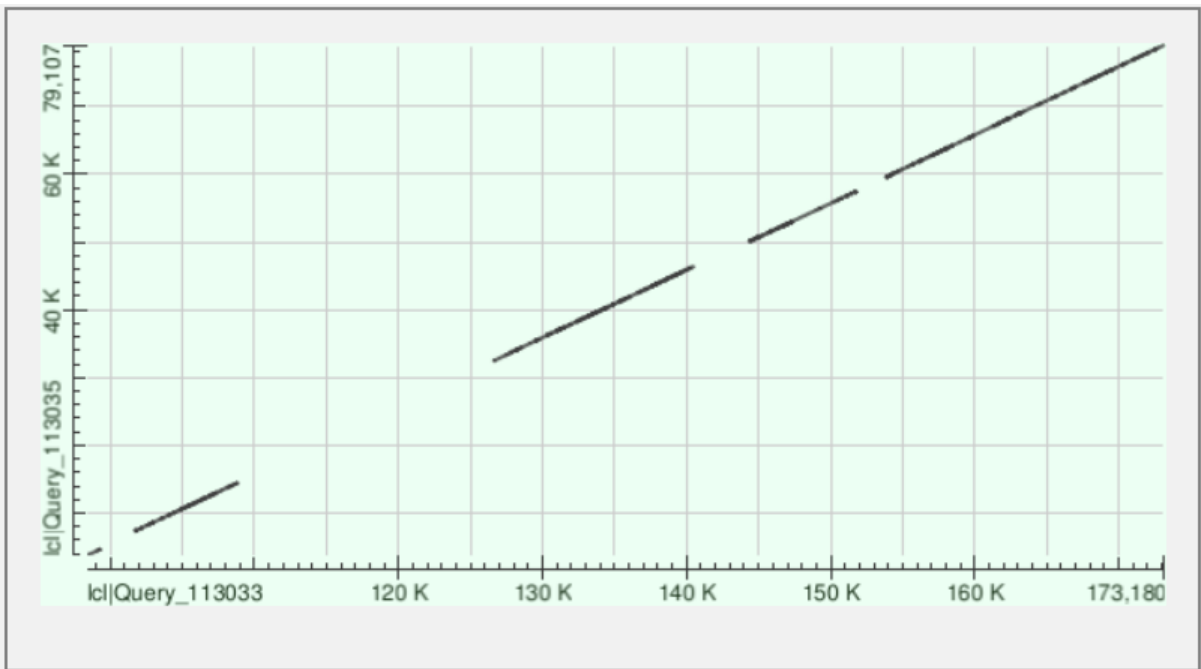


Рис. 2. Выравнивание с контигом 5

5 контиг выровнялся на участок 98-174к на хромосоме.

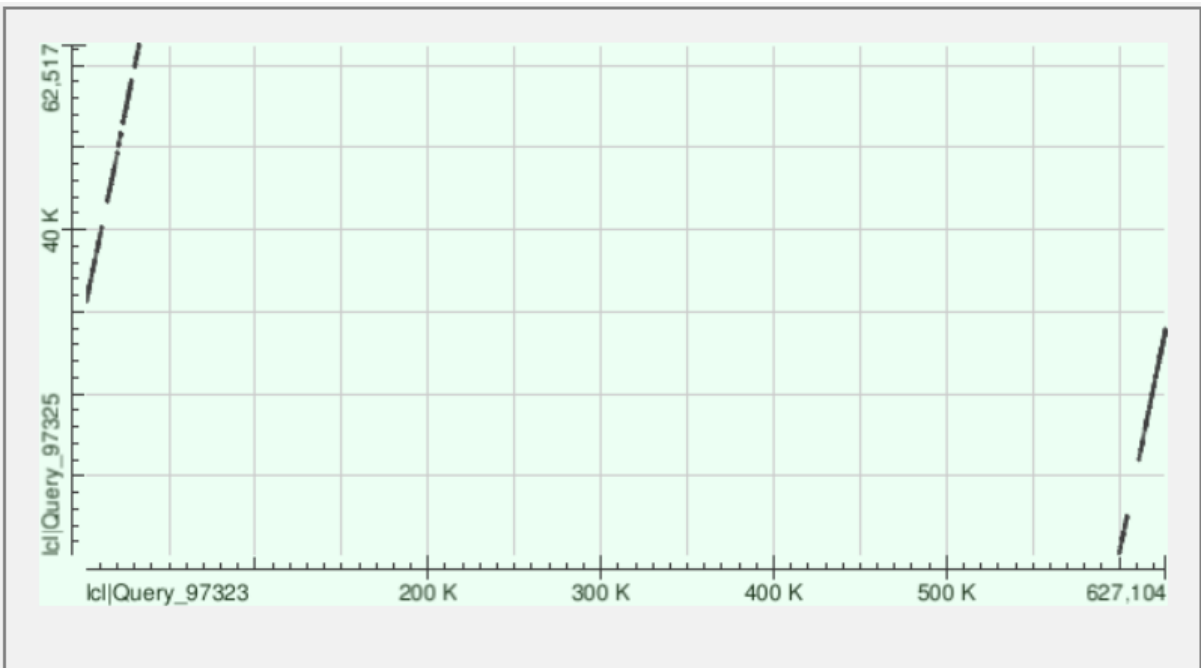


Рис. 3. Выравнивание с контигом 4

4 контиг выровнялся на участок 599-33к на хромосоме.