

Практикум 14. Сборка de novo

Кушнарева Дарья 202

Подготовка чтений программой trimmomatic.

Скачивание соответствующего архива чтений:

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR424/009/SRR4240379/SRR4240379.fastq.gz
```

Данные чтения являются одиночными.

Теперь удаление адаптеров, сначала сложим их в файл adapters.fa

```
cat /mnt/scratch/NGS/adapters/*.fa > adapters.fasta
```

И удалим их:

```
TrimmomaticSE -phred33 SRR4240379.fastq.gz cleanSRR4240379.fastq.gz
```

```
ILLUMINACLIP:adapters.fasta:2:7:7
```

Input Reads: 7400155 Surviving: 7269852 (98.24%) Dropped: 130303 (1.76%)

TrimmomaticSE: Completed successfully

Выдача программы показала что 1.76% чтений оказались с частями адаптеров.

Размер исходного файла: 167M

Размер полученного файла: 165M

Теперь удалим нуклеотиды с качеством <20 и чтения с длиной <32

```
TrimmomaticSE -phred33 cleanSRR4240379.fastq.gz cleantrimSRR4240379.fasta.gz
```

```
TRAILING:20 MINLEN:32
```

Input Reads: 7269852 Surviving: 6974267 (95.93%) Dropped: 295585 (4.07%)

TrimmomaticSE: Completed successfully

Было удалено 4.07%.

Размер полученного файла 156M

Velveth

Для начала подготовим k-меры длины 31

```
velveth kmer31 31 -short -fastq.gz cleantrimSRR4240379.fasta.gz
```

В директории kmer31 находятся такие файлы:

Log Roadmaps Sequences

Сборка на основе k-меров

Затем с помощью программы была осуществлена сборка de novo на основе k-меров

```
velvetg kmer31
```

Final graph has 440 nodes and n50 of 25646, max 49912, total 664650, using 0/6974267 reads

N50 = 25646

Теперь в директории kmer31 находятся файлы:

contigs.fa Graph LastGraph Log PreGraph Roadmaps Sequences stats.txt

Final graph has 440 nodes and n50 of 25646, max 49912, total 664650, using 0/6974267 reads

N50 = 25646

Три самых длинных контига:

1. 6, 49 912 bp, 35.907237
2. 9, 49 262 bp, 34.772177
3. 5, 33 085 bp, 36.259030

Контиги с аномально большим показателем покрытия:

- 2 контига длины 1 bp и покрытиями 474299 и 2694
1 контиг длины 2083 bp и покрытием 172,52

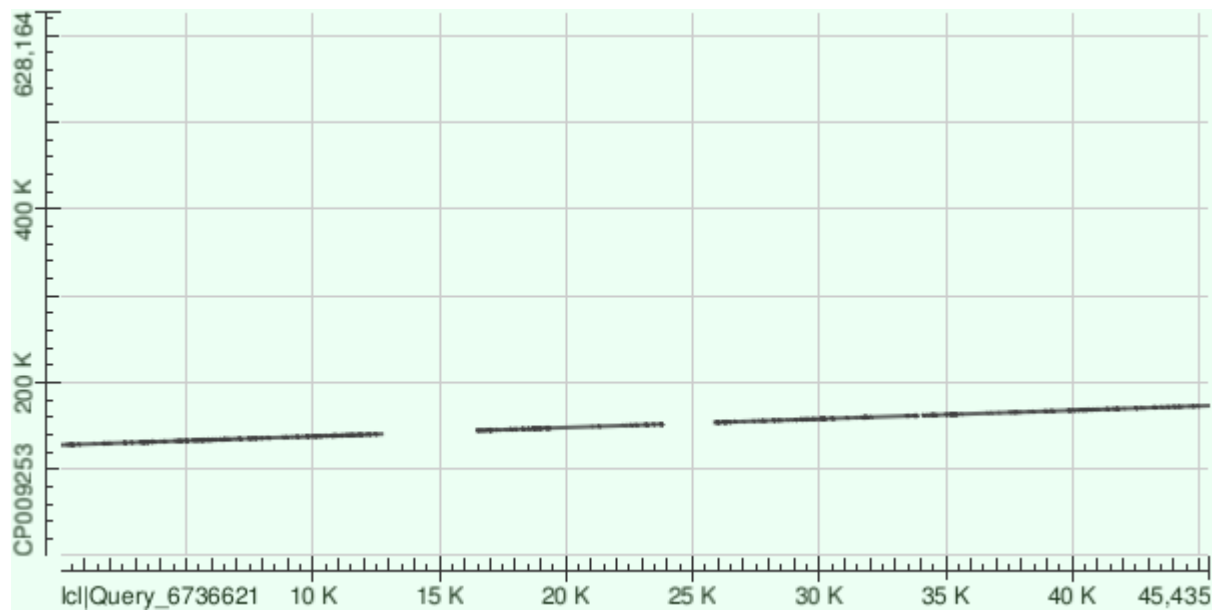
Анализ

Для дальнейшего анализа данной командой были получены .fasta контигов

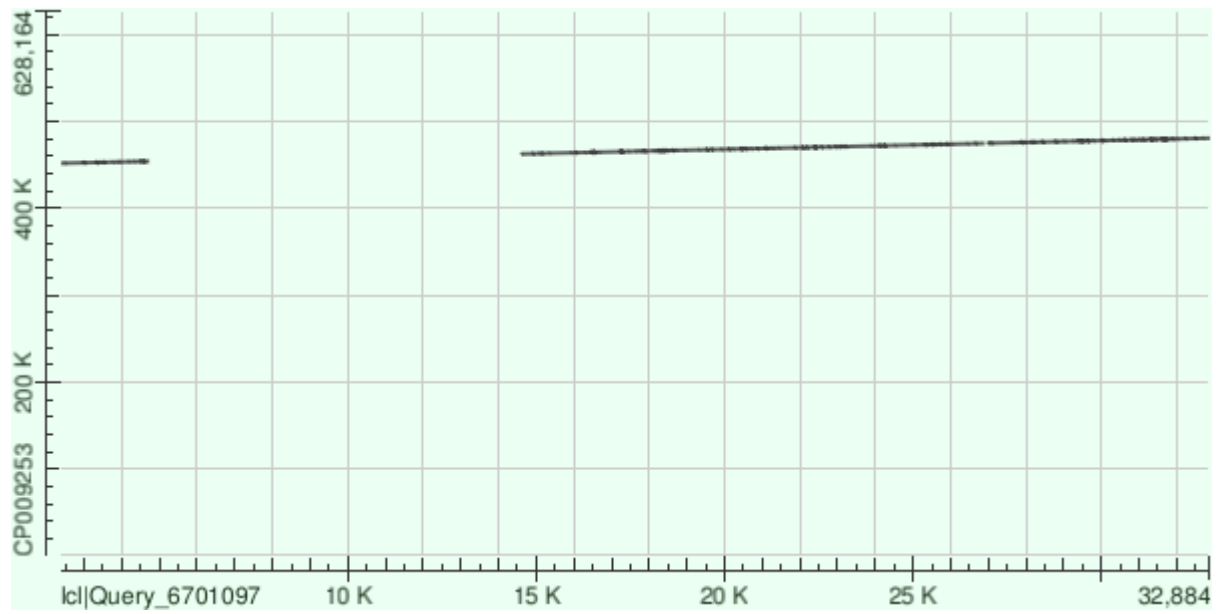
```
seqretsplit -filter kmer31/contigs.fa dir/name.format
```

Сравнение программой megablast трёх самых длинных контигов с хромосомой *Buchnera aphidicola* (GenBank/EMBL AC — CP009253):

1. Шестой контиг, 49 912 bp, 35.907237



2. Девятый континг, 49 262 bp, 34.772177



3. Пятый континг, 33 085 bp, 36.259030

