

Отчет по практикуму 14

Абзалимов Амир

Задание 1 Получение и обработка чтений

Загрузка чтений:

```
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR424/006/SRR4240356/  
SRR4240356.fastq.gz
```

Объединяем адаптеры:

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

Удаляем остатки адаптеров:

```
TrimmomaticSE -phred33 SRR4240356.fastq.gz SRR4240356_trash.fastq.gz  
ILLUMINACLIP:adapters.fa:2:7:7
```

Вывод программы:

```
Input Reads: 7511529 Surviving: 7358438 (97.96%) Dropped: 153091 (2.04%)
```

```
TrimmomaticSE: Completed successfully
```

В результате работы программы было удалено 2 процента адаптеров.

Триммируем второй раз с необходимыми параметрами:

```
TrimmomaticSE -phred33 SRR4240356_trash.fastq.gz  
SRR4240356_double.fastq.gz TRAILING:20 MINLEN:32
```

Вывод программы:

```
Input Reads: 7358438 Surviving: 7053346 (95.85%) Dropped: 305092 (4.15%)
```

```
TrimmomaticSE: Completed successfully
```

В результате работы программы было удалено еще 4 процента адаптеров.

Задание 2 К-меры

Делаем k-меры длины 31 с помощью velveth:

```
velveth velv 31 -short -fastq.gz SRR4240356_double.fastq.gz
```

Задание 3 Сборка *de novo*

Далее запускаем velvetg:

```
velvetg velv
```

Вывод работы программы:

```
Final graph has 286 nodes and n50 of 65554, max 111962, total 659837, using 0/7053346 reads
```

Из результатов работы программы можно определить следующие параметры:

N50: 6555

Самая большая длина контига - 111962

Далее путем обработки различными конвейерами с использованием команд: `grep`, `sort`, `head`, `tail`, получаем статистику по контигам:

3 самых длинных континга и их покрытие:

```
>NODE 8 length 111962 cov 38.660198
```

```
>NODE 6 length 107488 cov 34.174030
```

```
>NODE 10 length 80939 cov 37.524174
```

Находим медианную величину:

```
Медиана: >NODE 78 length174 cov 16.798851
```

Находим контиги с аномальными покрытиями:

Аномально низкие значения покрытия:

```
1. >NODE 123 length91 cov 2.362637
```

```
2. >NODE 74 length31 cov 3.064516
```

```
3. >NODE 165 length78 cov 3.666667
```

4.	>NODE	59	length31	cov	4.129032
5.	>NODE	65	length125	cov	4.624000
6.	>NODE	146	length61	cov	5.196721
7.	>NODE	145	length74	cov	5.486486
8.	>NODE	149	length31	cov	5.838710
9.	>NODE	82	length135	cov	5.940741
10.	>NODE	73	length143	cov	5.951049

Аномально высокие значения покрытия:

1.	>NODE	27	length282	cov	458.429077
2.	>NODE	17	length950	cov	447.494751
3.	>NODE	14	length934	cov	444.608124
4.	>NODE	29	length308	cov	436.269470
5.	>NODE	2	length524	cov	419.526703
6.	>NODE	20	length501	cov	309.155701
7.	>NODE	55	length42	cov	211.428574
8.	>NODE	36	length45	cov	166.199997
9.	>NODE	39	length84	cov	43.488094
10.	>NODE	12	length20510	cov	40.211361

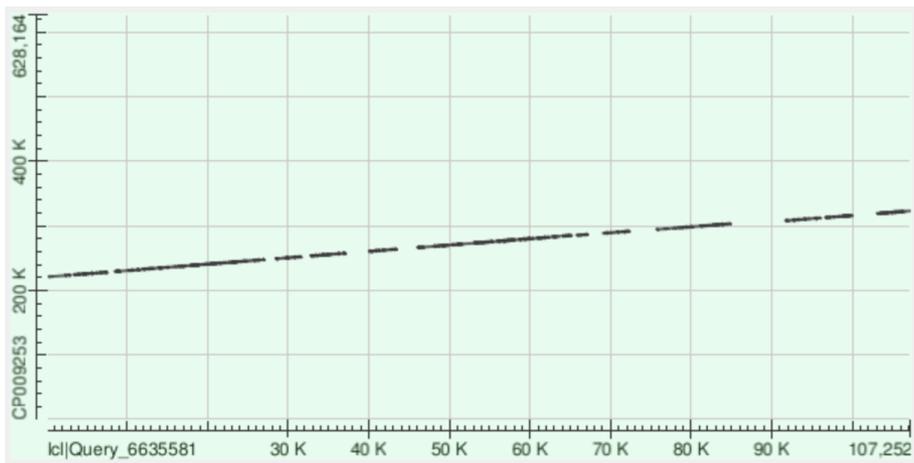
Выводы:

- Было найдено 10 аномально больших и аномально малых покрытий:
- Максимальное превышает медиану в 30 раз, минимальное из аномальных в 250 раз
- Длина аномально больших контигов варьирует от 42 до 20510
- Минимальное покрытие - 2.36
- Максимальное аномальное среди минимальных - 5,95
- Длина аномально малых контингов варьирует от 31 до 143

Задание 4 BLAST

Сравнение самых длинных контигов с хромосомой *Buchnera aphidicola* (GenBank/EMBL AC – CP009253) проводим в программе megaBLAST со стандартными параметрами, предварительно выделив последовательности необходимых контигов:

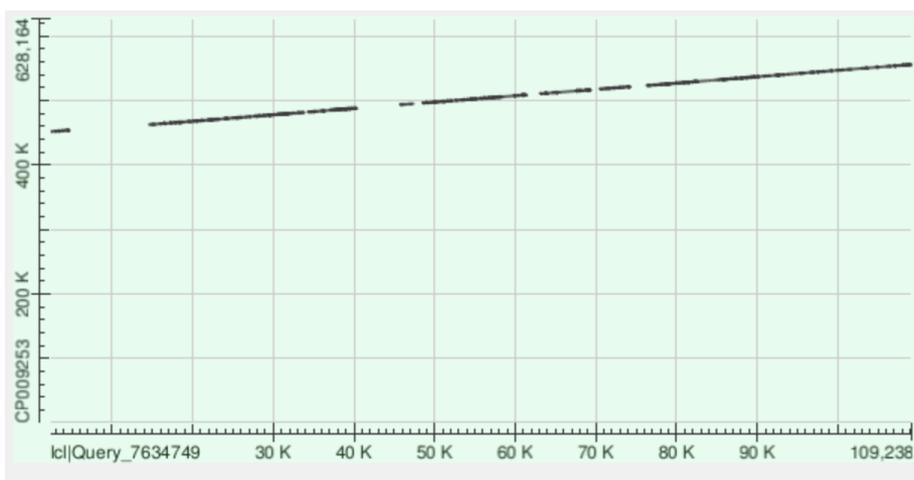
Контиг 6.



6 контиг выровнялся на участок 220-380к на хромосоме.

Процент сходства 78,76%.

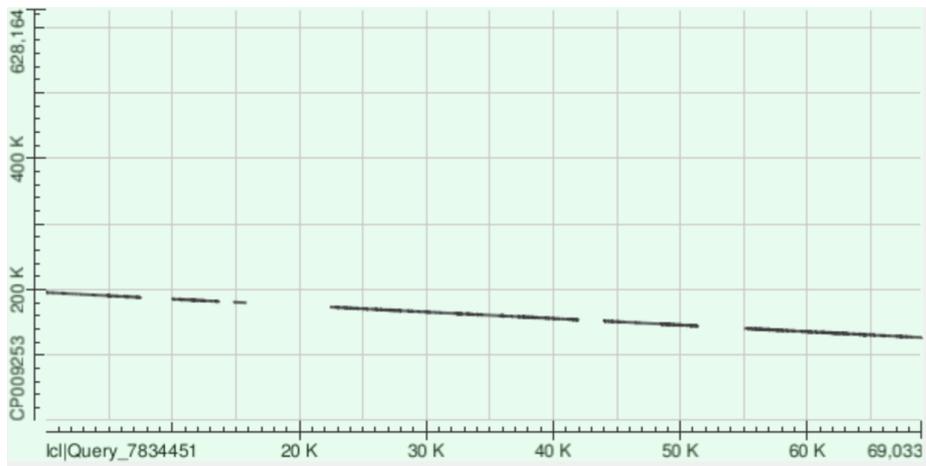
Контиг 8.



8 контиг выровнялся на участок 445-540к на хромосоме.

Процент сходства 81,46%

Контиг 10



10 контиг выровнялся на участок 120-199к на хромосоме.

Процент сходства: 74,88%