

Скачиваю с помощью wget:

```
wget -nc ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR424/008/SRR4240378/SRR4240378.fastq.gz
```

Получаю файл **SRR4240378.fastq.gz**

Организм: Buchnera aphidicola str. Tuc7 (Acyrthosiphon pisum)

AC: SRR4240378

Прибор: Illumina Genome Analyzer II

Тип чтений: одноконцевые

Скопировала все адаптеры, объединила `cat *fa > adapters.fa`

Удалила адапторы: `java -jar /usr/share/java/trimmomatic.jar SE -threads 10 SRR4240378.fastq.gz SRR4240378-ad.fastq.gz ILLUMINACLIP:adapters.fa:2:7:7 2> adapter.txt`

Input Reads: 4420587 (было чтений)

Surviving: 4338744 (98.15%) (осталось после удаления адапторов)

Dropped: 81843 (1.85%) (удалено=было с адаптерами)

Триммировала: `java -jar /usr/share/java/trimmomatic.jar SE -threads 10 SRR4240378-ad.fastq.gz trimmedSRR4240378.fastq.gz TRAILING:20 MINLEN:32 2> trim.txt`

Input Reads: 4338744 (было чтений после удаления адаптеров)

Surviving: 4154738 (95.76%) (осталось после триммирования)

Dropped: 184006 (4.24%) (удалено)

Размеры:

93МБ **SRR4240378-ad.fastq.gz**

94МБ **SRR4240378.fastq.gz**

87МБ **trimmedSRR4240378.fastq.gz**

Подготовка k-меров

```
velveth velveth/ 31 -fastq.gz -short trimmedSRR4240378.fastq.gz
```

velveth папка куда загрузится выход

31 длина k-меров, short короткие непарные

На вход файл с триммированными чтениями

Сборка k-меров

```
velvetg velveth/
```

Выход программы: n50 of 7028, max 36746

Ищем самые большие контиги:

```
cut -f2 stats.txt | sort -h |tail -3
```

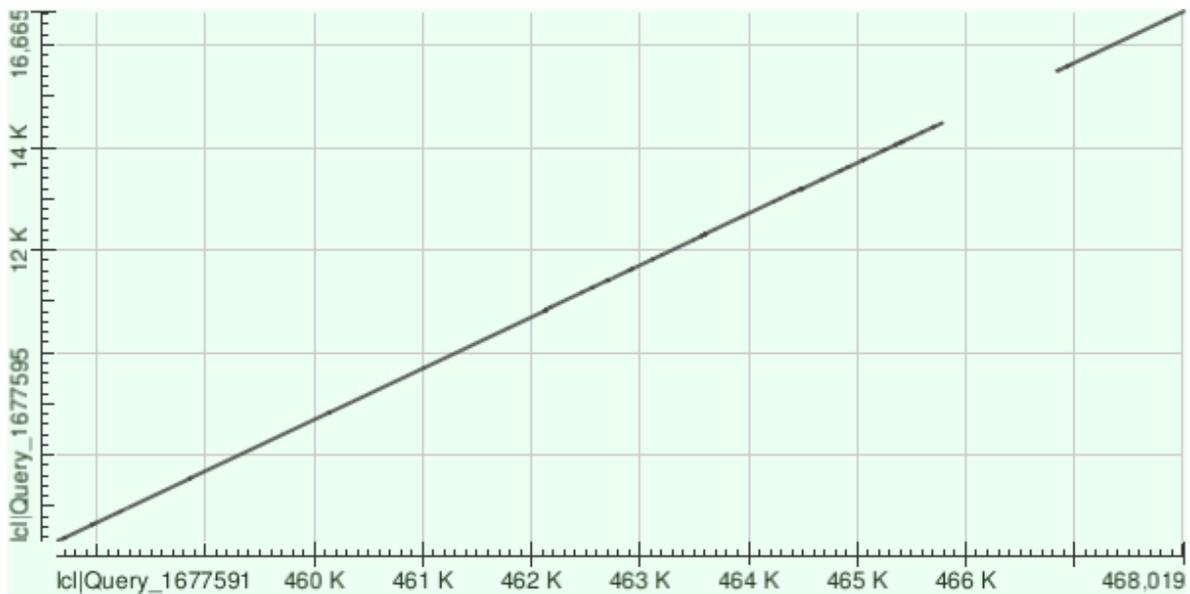
Контиг (ID)	Длина	Coverage
15	16745	9,96%
57	19371	11,5%
8	36746	21,86%

Аномально большие значения: 829.0, 923.0, 968.0, 148170.0, аномально маленьких нет (всех остальных по 15 штук примерно). Нашла с помощью сортировки шестого столбца.

Анализ

С помощью seqretsplit разделяю последовательности в файле contings.fa и оставляю только самые длинные
Сравниваю их с хромосомой *Buchnera aphidicola* (CP009253) в blastn

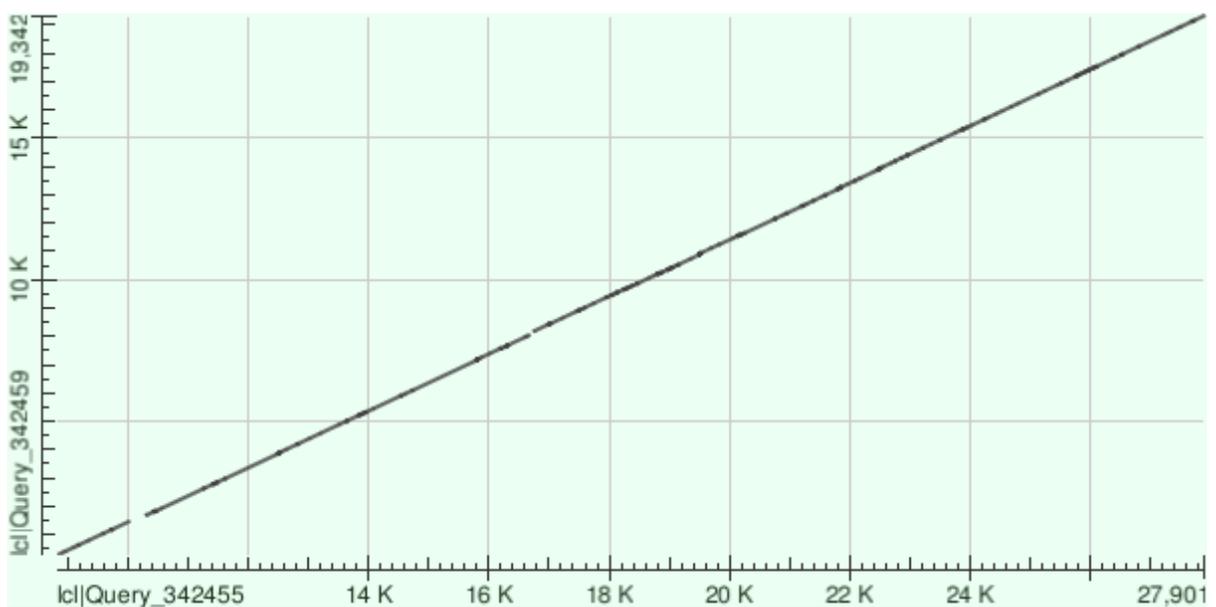
Для контига 15:



Есть делеция

Контиг	Организм	Гэпы	Identity
6315-14480	457626-465787	184(2%)	79%
15488-16665	466839-468019	17(1%)	78%

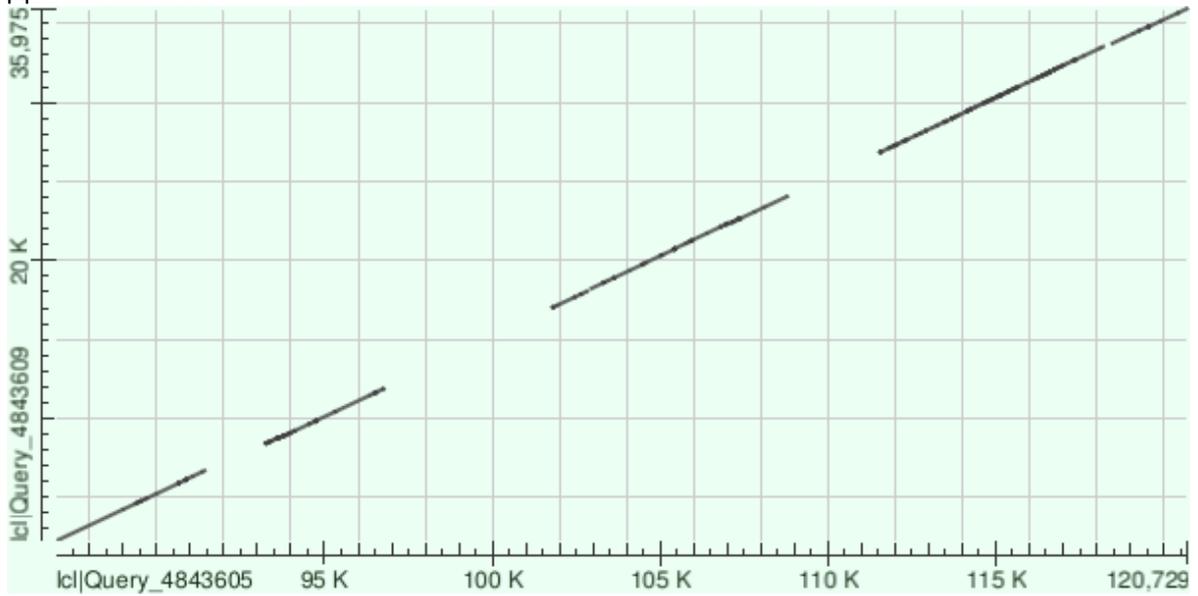
Для контига 57:



Контиг	Организм	Гэпы	Identity
8183-19342	16750-27901	484(4%)	76%
1671-8062	10298-16690	189(2%)	78%

272-1480	8827-10041	20(1%)	77%
----------	------------	--------	-----

Для контига 8:



Контиг	Организм	Гэпы	Identity
2248-6737	87013-91478	76(1%)	82%
18179-22359	102904-107026	110(2%)	78%
8396-11888	93232-96769	155(4%)	79%
26830-33590	111533-118252	347(5%)	73%
33704-35975	118428-120729	44(2%)	79%
22392-24064	107118-108795	43(2%)	82%
16992-18070	101766-102843	19(1%)	80%