

Сборка de novo

Задание 1

Командой был получен архив с чтениями:

[wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR424/009/SRR4240379/SRR4240379.fastq.gz](ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR424/009/SRR4240379/SRR4240379.fastq.gz)

Был собран файл с адаптерами:

[Cat TruSeq* NexteraPE-PE.fa > adapters.fa](#)

Input: 7400155

Output: 7269852

Итого 130 303 чтения оказались с остатками адаптеров(1,8%)

[TrimmomaticSE SRR4240379.fastq.gz SRR4240379_output_trim_filter.fq.gz ILLUMINACLIP:adapters.fa :2:7:7 TRAILING:20 MINLEN:32](#)

Далее проводила фильтрацию чтений с качеством выше 20 и длиной более 32

После фильтрации осталось 6974267 чтений, то есть очистилось 7269852 – 6974267=295 585(4,1%)

Задание 2

[velveth ./vel 31 -short -fastq SRR4240379_output_trim_filter.fq.gz](#)

./vel – папка в которой сохраняются результаты

31 – k-меры

Short – короткие и непарные чтения

Задание 3

Строит граф де Брёйна из k-меров предыдущего задания

[velvetg ./vel](#)

```
[29.756784] Writing into stats file ./vel/stats.txt...
[29.757735] Writing into graph file ./vel/LastGraph...
Final graph has 440 nodes and n50 of 25646, max 49912, total 664650, using 0/6974267 reads
liyaskina.irina@kodomomo:/mnt/scratch/NGS/irina.liyaskina/pr15$ ls -l
```

Получилось N50 = 25646

Самые длинные контиги:

- 33085(ID:5, coverage: 36.259030)
- 49262(ID:9, cov: 34.772177)
- 49912(ID:6, cov:35.907237)

Также есть контиги с аномально малым покрытием:

- Cov: 1.000000 ID: 429 len:1
- Cov: 1.000000 ID:437 len:1
- 1.000000 ID:348 len:3

С аномально большим покрытием:

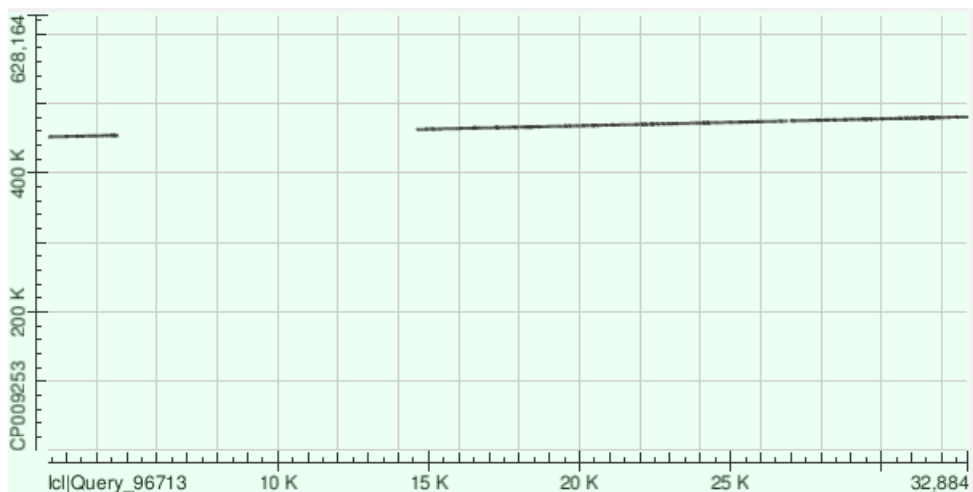
- Cov: 610.000000 ID:174 len:1
- Cov: 2694.000000 ID:105 len:1
- Cov: 474299.000000 ID:133 Len:1

Задание 4

Seqretsplit разделила файл contigs.fa на несколько

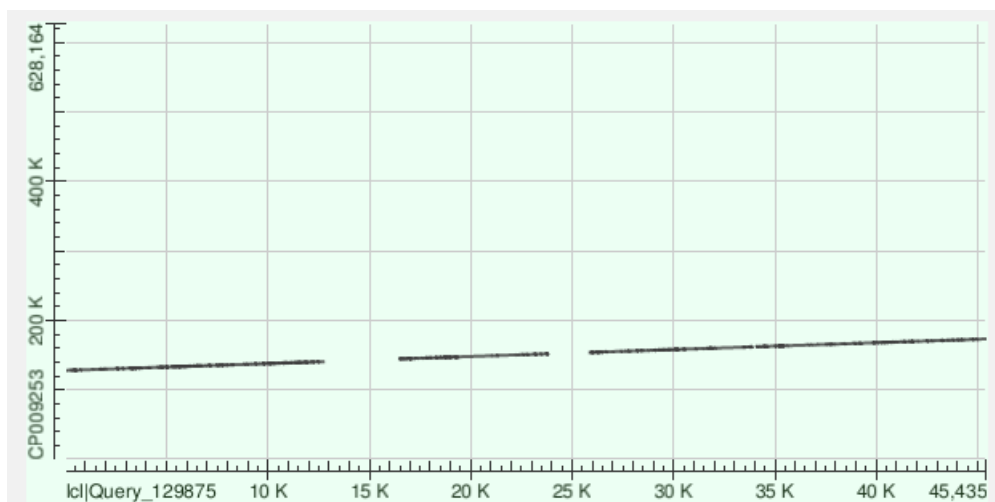
Далее BLAST с контигом ID:5 дал такой результат:

№ Находки	Регион выравнивания	Совпадающих позиций	Гэпы
1	467412 - 474667	5691/7388 (77%)	208/7388 (2%)
2	462496 - 467421	3861/5015 (77%)	162/5015 (3%)
3	474844 - 480660	4431/5974 (74%)	255/5974 (4%)
4	451729 - 454069	1827/2370 (77%)	55/2370 (2%)



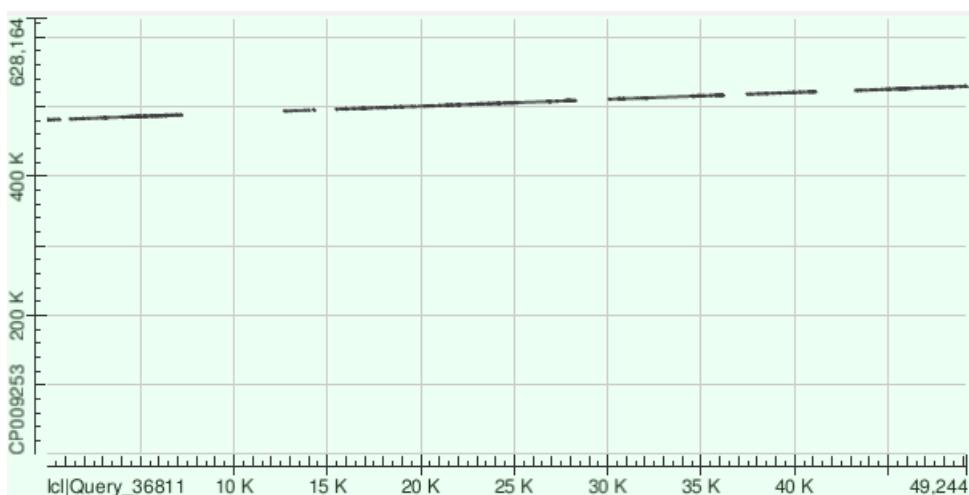
Для контига 6:

№ Находки	Регион выравнивания	Совпадающих позиций	Гэпы
1	127825 - 140555	9741/13008 (75%)	544/13008 (4%)
2	153752 - 161738	6347/8169 (78%)	266/8169 (3%)
3	144368 - 151796	5863/7536 (78%)	243/7536 (3%)
4	161898 - 166752	3910/4912 (80%)	108/4912 (2%)
5	166750 - 173180	4965/6517 (76%)	159/6517 (2%)



Для контига 9:

№Находки	Регион выравнивания	Совпадающих позиций	Гэпы
1	500370 - 508806	6516/8617 (76%)	351/8617 (4%)
2	510438 - 516539	4897/6234 (79%)	187/6234 (2%)
3	523105 - 528679	4369/5685 (77%)	207/5685 (3%)
4	481997 - 488106	4621/6238 (74%)	308/6238 (4%)
5	517766 - 521500	2922/3783 (77%)	101/3783 (2%)
6	496111 - 500325	3255/4324 (75%)	154/4324 (3%)
7	493487 - 494864	1109/1384 (80%)	13/1384 (0%)
8	480874 - 481545	564/686 (82%)	20/686 (2%)
9	528794 - 529211	357/425 (84%)	26/425 (6%)
10	495033 - 495148	108/120 (90%)	5/120 (4%)



Во всех 3х случаях отчётливо прослеживаются делеции в хромосоме по сравнению с контигами, в первом случае это одна большая делеция, во втором две поменьше, в третьем 5 маленьких.