

Лабораторная работа 2

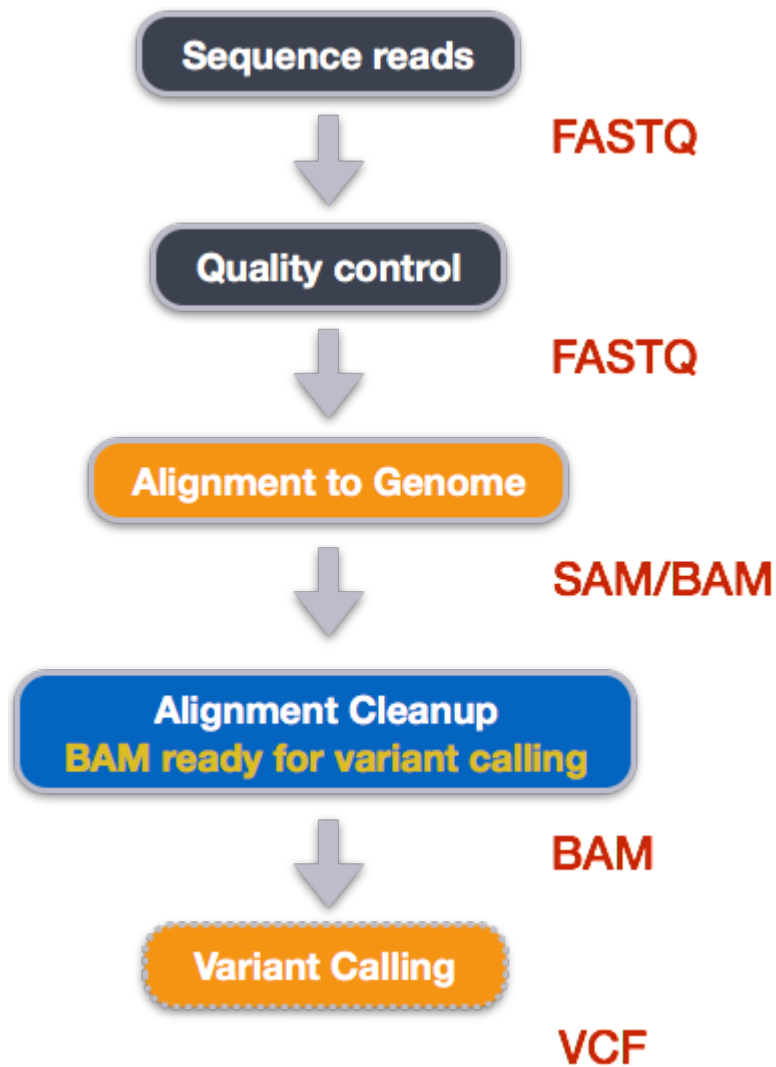
Целью эксперимента является нахождение мутаций в течении 15000 и 30000 поколений *E. coli* в определенных условиях.

Численный эксперимент проводится на ПК 16Gb RAM, Apple M2 Pro, с предустановленной MacOS Ventura 13.6.2.

Для проведения эксперимента были установлены программы:

- FastQC, Trimmomatic – при помощи пакетного менеджера Conda
- BWA, SAMtools, BCFtools – при помощи пакетного менеджера brew
- IGV – с официального сайта
(<https://igv.org/doc/desktop/#DownloadPage/>)

План работы отображён на схеме:



1. Сырые данные были загружены командами

```
curl -O ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR258/004/SRR2589  
curl -O ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR258/004/SRR2589  
curl -O ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR258/003/SRR2584  
curl -O ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR258/003/SRR2584  
curl -O ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR258/006/SRR2584  
curl -O ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR258/006/SRR2584
```

2. Архивы с данными были распакованы командой

```
gunzip *.gz
```

3. Контроль качества ридов был произведён командой

```
fastqc *.fastq
```

Все данные показали хорошие результаты качества, работа с ними была продолжена.

4. Тримминг производился командой

```
trimmomatic PE -threads 12 SRR_1056_1.fastq SRR_1056_2.fastq  
SRR_1056_1.trimmed.fastq SRR_1056_1un.trimmed.f  
SRR_1056_2.trimmed.fastq SRR_1056_2un.trimmed.f  
ILLUMINACLIP:SRR_adapters.fa SLIDINGWINDOW:4:20
```

для каждой пары сырых ридов.

5. Референсный геном *E. coli* был скачан и распакован командами

```
curl -L -o ref_genome/ecoli_rel606.fasta.gz \  
ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/017/985/GCA_00  
  
gunzip ref_genome/ecoli_rel606.fasta.gz
```

6. Для упрощения variant calling были скачаны и распакованы некоторые подмножества ридов (после тримминга)

```
curl -L -o sub.tar.gz https://ndownloader.figshare.com/files/  
$ tar xvf sub.tar.gz
```

7. Референсный геном был проиндексирован командой

```
bwa index ref_genome/ecoli_rel606.fasta
```

Индексирование генома позволяет программе для выравнивания быстро находить потенциальные сайты выравнивания для искомым последовательностей в геноме, тем самым ускоряя процесс выравнивания.

8. Выравнивание было произведено командой

```
$ bwa mem ref_genome/ecoli_rel606.fasta \  
trimmed_fastq_small/SRR2584866_1.trim.sub.fastq \  
trimmed_fastq_small/SRR2584866_2.trim.sub.fastq \  
> results/sam/SRR2584866.aligned.sam
```

для каждой пары ридов.

9. Перевод из формата выравнивания .sam в его бинарный аналог .bam для ускорения работы и уменьшения размера файлов выравнивания был произведён командой

```
samtools view -S -b results/sam/SRR2584866.aligned.sam \  
> results/bam/SRR2584866.aligned.bam
```

для каждого поколения бактерии.

10. Сортировка по координатам была выполнена командой

```
samtools sort -o results/bam/SRR2584866.aligned.sorted.bam \  
results/bam/SRR2584866.aligned.bam
```

11. Для выявления мутаций, покрытия и SNV был произведён variant calling.

Покрытие было посчитано командой

```
bcftools mpileup -O b -o results/bcf/SRR2584866_raw.bcf \  
-f data/ref_genome/ecoli_rel606.fasta results/bam/SRR2584866..
```

Для выявления SNV была использована команда

```
bcftools call --ploidy 1 -m -v -o results/vcf/SRR2584866_vari.  
results/bcf/SRR2584866_raw.bcf
```

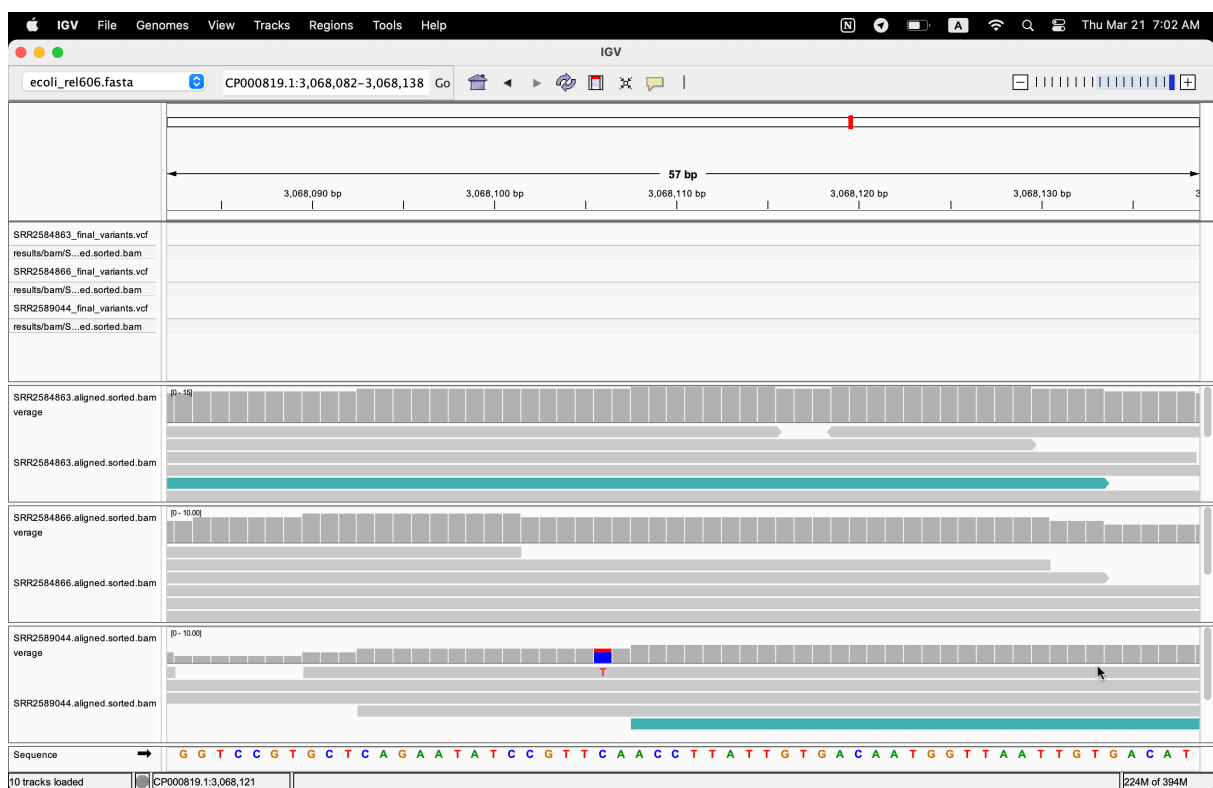
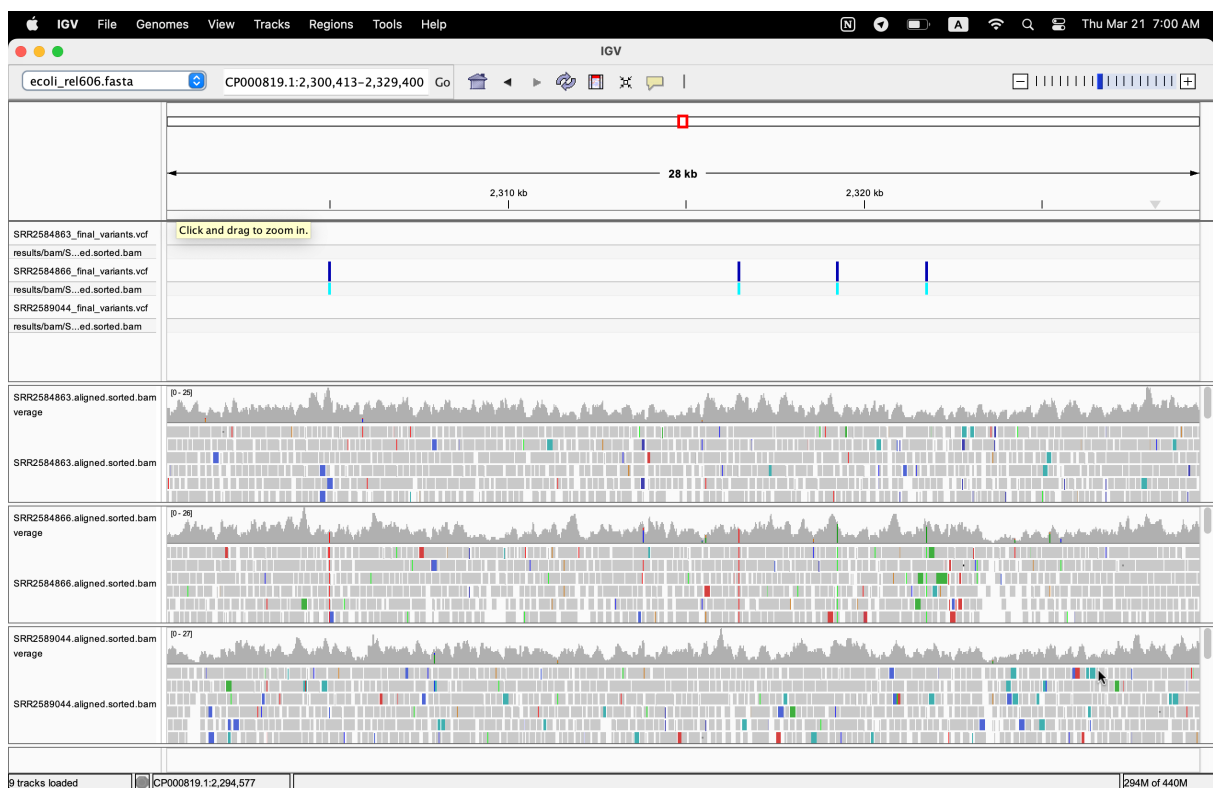
Фильтрация и получение итогового отчёта были произведены командой

```
vcfutils.pl varFilter results/vcf/SRR2584866_variants.vcf \  
> results/vcf/SRR2584866_final_variants.vcf
```

(все команды были использованы с каждым поколением бактерии)

12. При помощи геномного браузера IGV была получена визуализация результатов исследования. Использовались референсный геном, файлы .bam и .vcf.





Изображения показывают результат исследования при разном увеличении. Подсветка выбрана по размеру вставки и парной ориентации.