**Задание 10**

Трещева Мария

*1) Анализ качества прочтений до тримминга*:

conda activate trimming

*mkdir -p fastqc\_raw\_output*

fastqc /projects/mipt\_dbmp\_biotechnology/genome/illumina\_reads\_R1\_001.fastq /projects/mipt\_dbmp\_biotechnology/genome/illumina\_reads\_R2\_001.fastq -o fastqc\_raw\_output

*2) Тримминг: bash-script.*

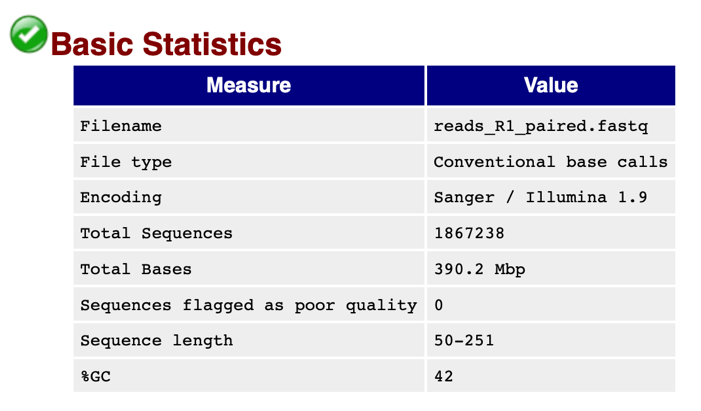
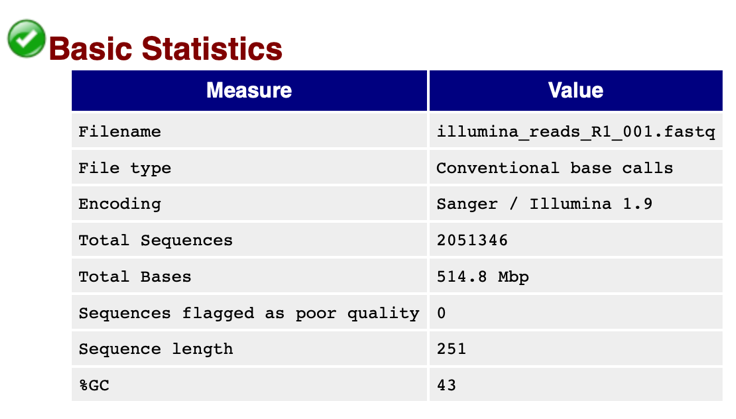
1. Удаление адаптеров с помощью опции ILLUMINACLIP и файла adapters.fa.
2. Обрезка низкокачественных концов ридов: LEADING:20 и TRAILING:20.
3. Обрезка коротких ридов: MINLEN:50.

*3) Анализ качества прочтений после тримминга:*

fastqc trimmed\_output/reads\_R1\_paired.fastq trimmed\_output/reads\_R2\_paired.fastq -o fastqc\_trimmed\_output

*4) Результаты: до и после тримминга, соответственно.*

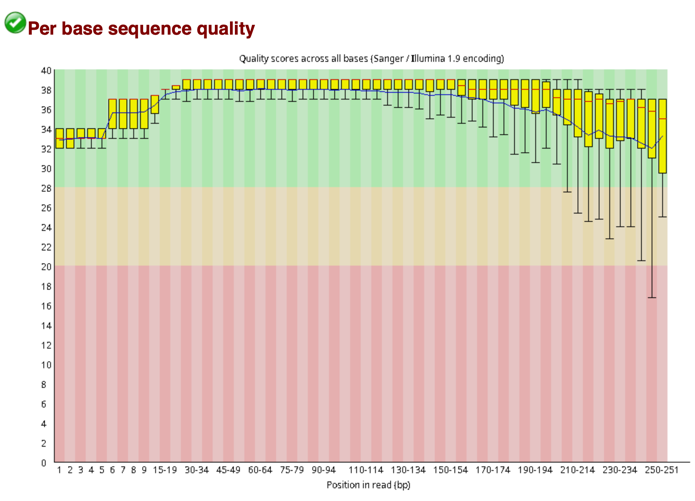
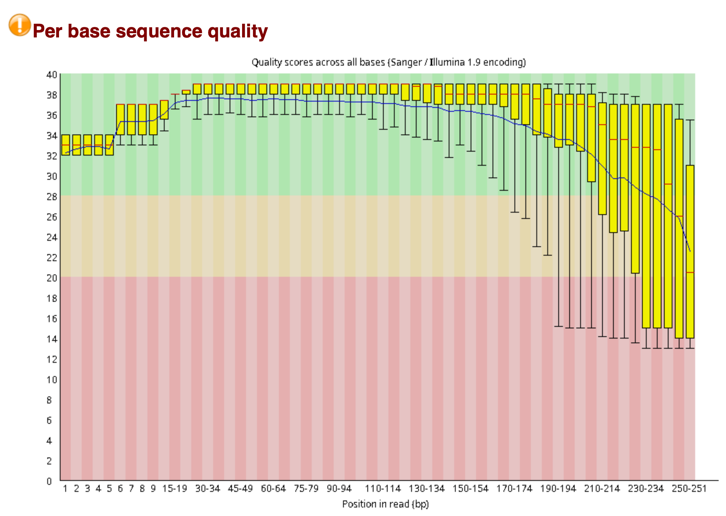
**Basic statistics для R1 до и после:**



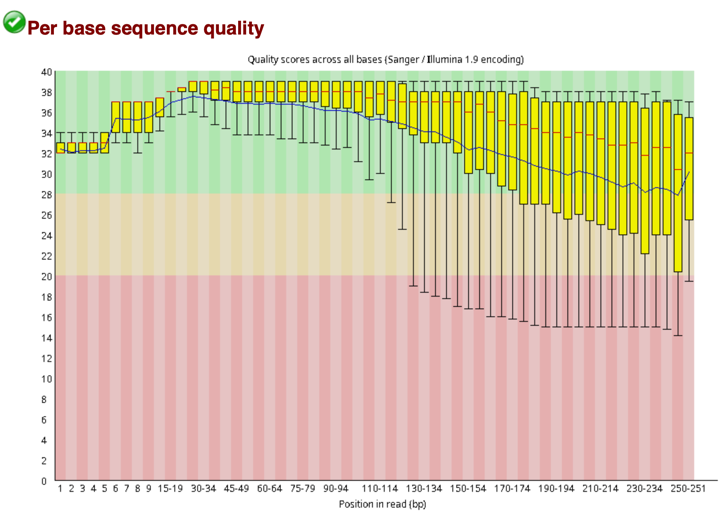
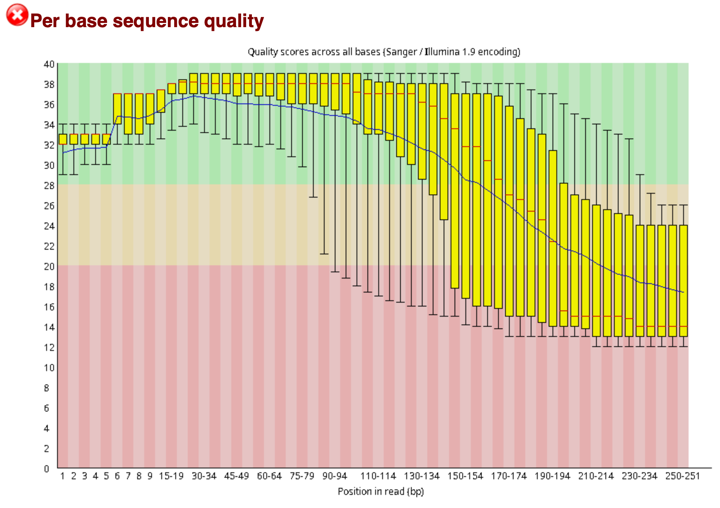
Длина ридов стала варьироаться от 50 до 251, GC-состав почти тот же.

**Качество прочтений:**

R1, до/после:

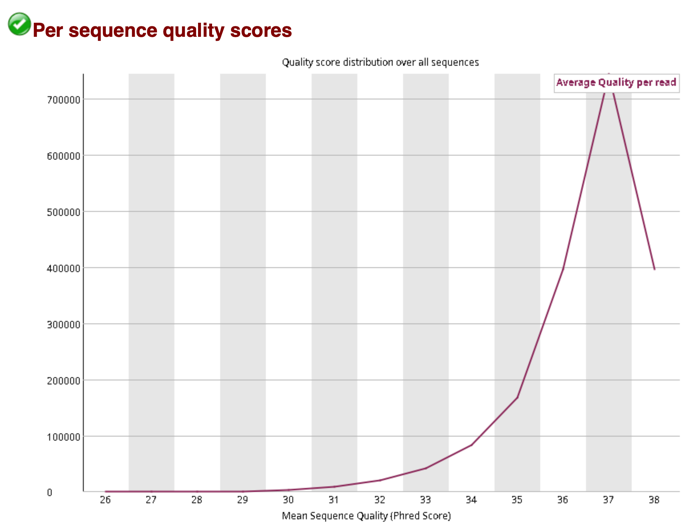
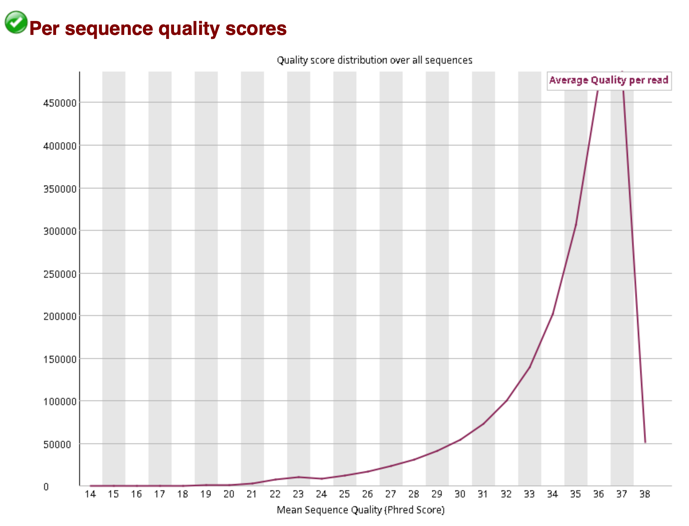


R2, до/после:

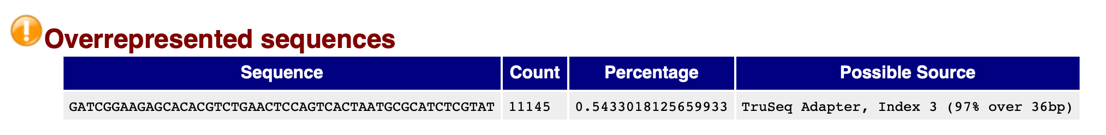


Качество прочтений улучшилось после тримминга, поскольку мы избавились от ошибок на концах прочтений.

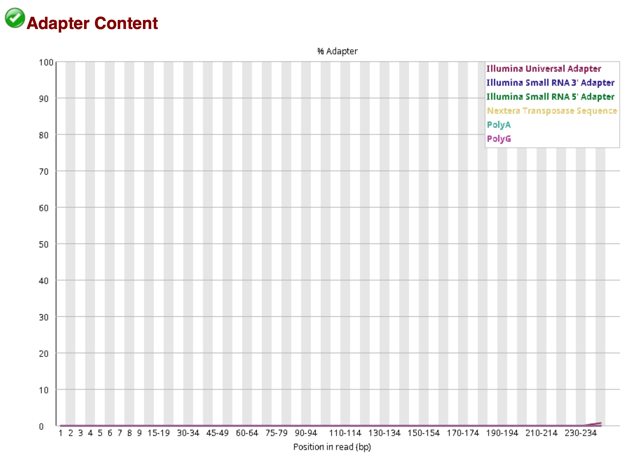
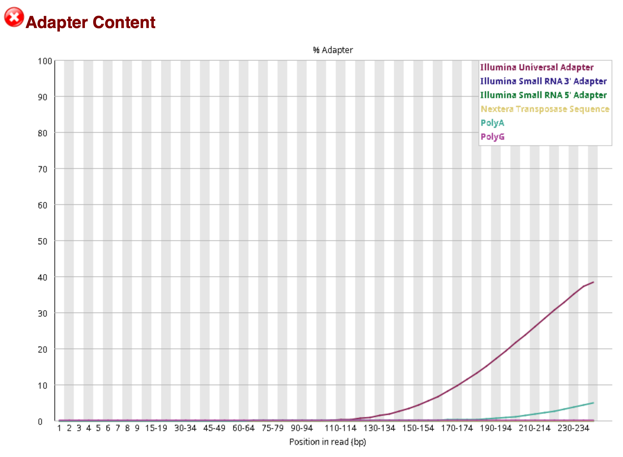
Далее данные для R1, для R2 аналогично. До/после:



Ось х представлена меньшим диапазоном значений – совсем низкокачественные прочтения исправлены с помощью тримминга. Они сместились из хвоста распределения к более высокому качеству. В итоге получилось вместо 500к хороших прочтений – 700к.



Адаптер, конечно, был перепредставлен в последовательностях, не подвергнутых триммингу. Ниже видно, что адаптер исключен из прочтений после тримминга:



Итак, в результате тримминга качество ридов повысилось:

удалили адаптеры, риды с низким качеством и улучшили качество на концах прочтений.