MultiQC

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

> This report has been generated by the CBIO bacterial variant calling pipeline

Report generated on 2022-11-17, 10:01 based on data in: `/cbio/projects/019/bacterial_variant_calling/work/34/7e3440ddb70a33509f72deb5fae938`

General Statistics

Copy table | Configure Columns | Plot | Showing 3/3 rows and 6/9 columns.

Sample Name	Change rate	Ts/Tv	M Variants	% Dups	% GC	M Seqs
sample_1	4 262	1.587	0.00	8.7%	65%	3.6
sample_2	4 501	1.647	0.00	82.1%	64%	2.6
sample_3	4 107	1.592	0.00	10.8%	65%	4.9

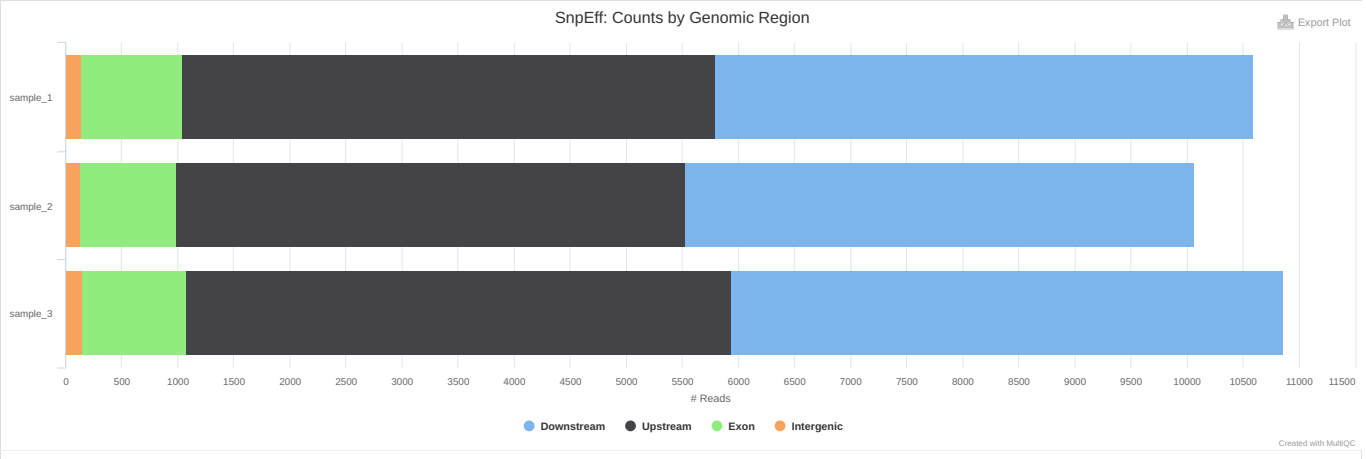
SnpEff

SnpEff is a genetic variant annotation and effect prediction toolbox. It annotates and predicts the effects of variants on genes (such as amino acid changes). . *DOI: 10.4161/fly.19695.*

Variants by Genomic Region

The stacked bar plot shows locations of detected variants in the genome and the number of variants for each location.

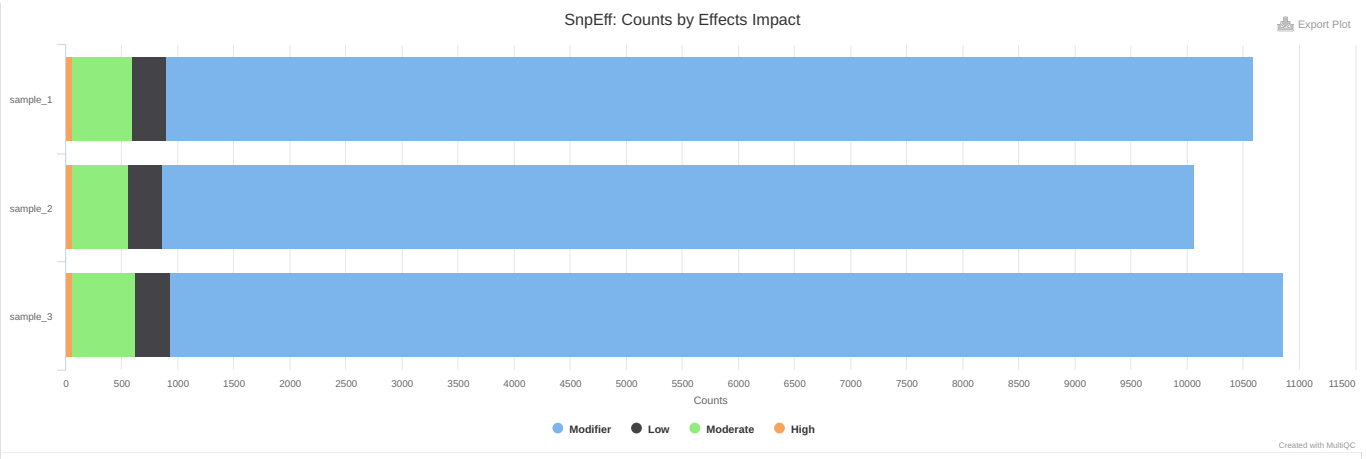
Counts | Percentages | Log10



Variant Effects by Impact

The stacked bar plot shows the putative impact of detected variants and the number of variants for each impact.

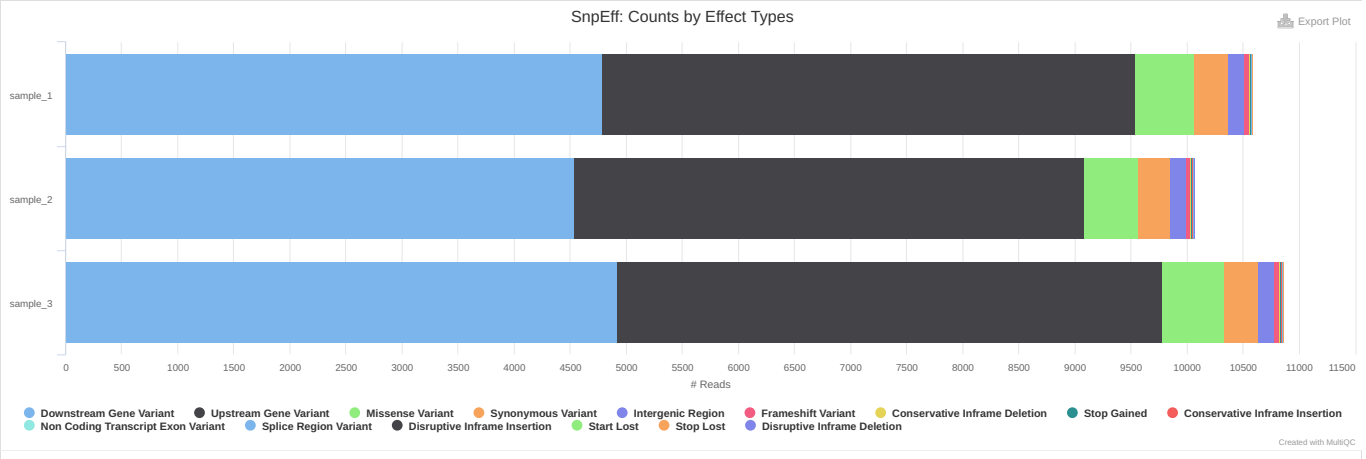
Counts | Percentages | Log10



Variants by Effect Types

The stacked bar plot shows the effect of variants at protein level and the number of variants for each effect type.

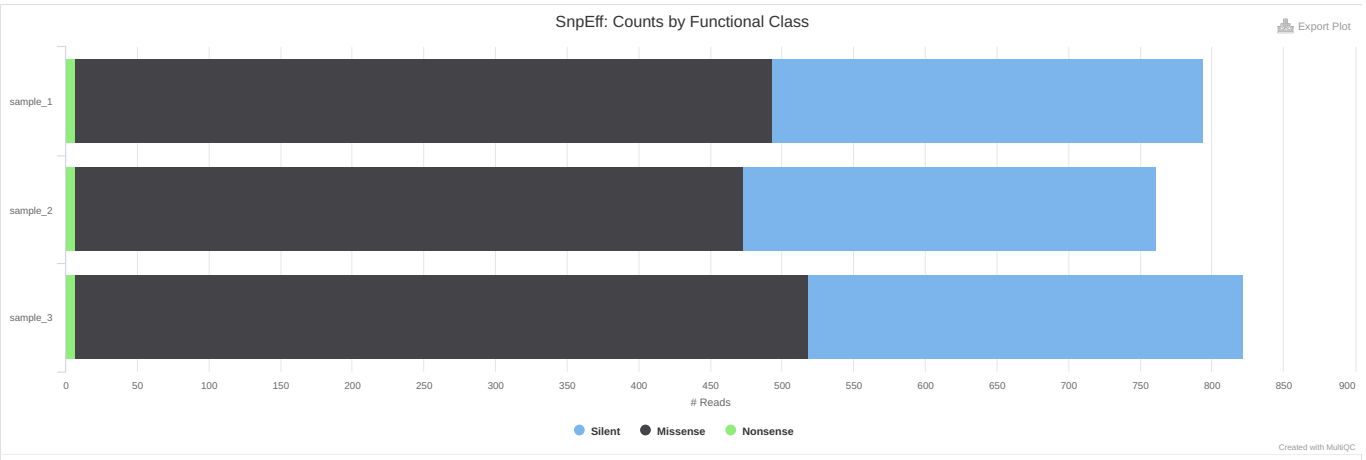
Counts | Percentages



Variants by Functional Class

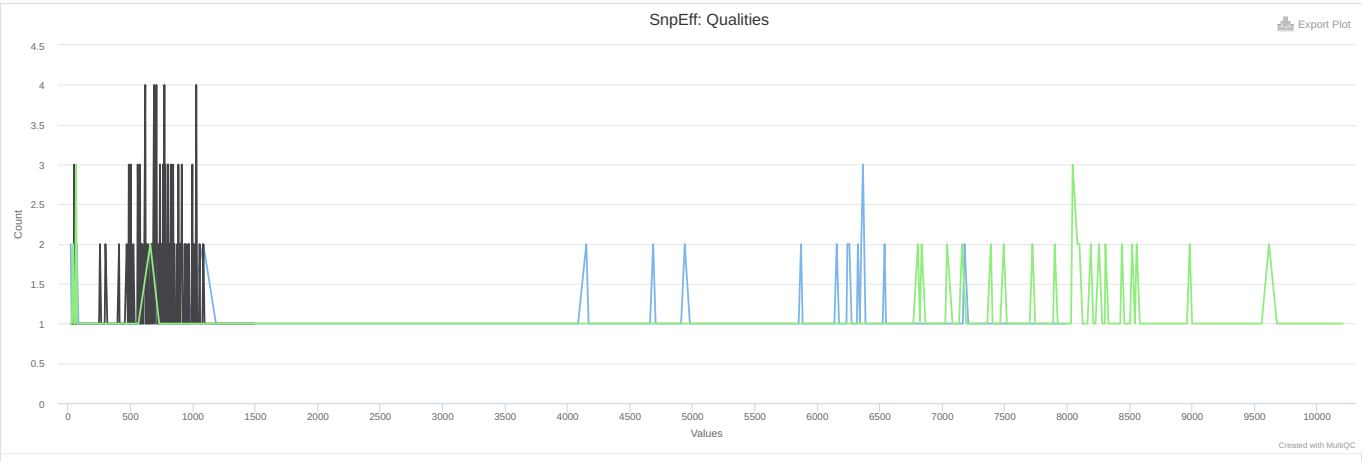
The stacked bar plot shows the effect of variants and the number of variants for each effect type.

Counts | Percentages | Log10



Variant Qualities

The line plot shows the quantity as function of the variant quality score.



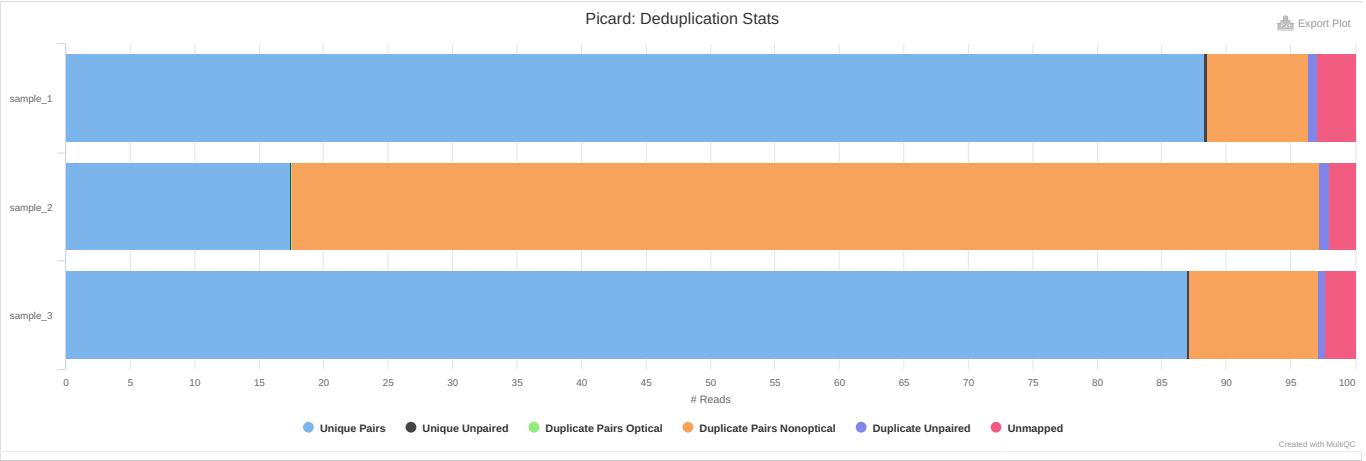
Picard

Picard is a set of Java command line tools for manipulating high-throughput sequencing data.

Mark Duplicates

Number of reads, categorised by duplication state. **Pair counts are doubled** - see help text for details.

Number of Reads | Percentages



uct-cbio/bacterial_variant_calling Workflow Summary

- this information is collected when the pipeline is started.

Run Name	`deadly_babbage`
Reads	`/cbio/projects/019/bacterial_variant_calling/sample_sheet.csv`
Data Type	`Paired-End`

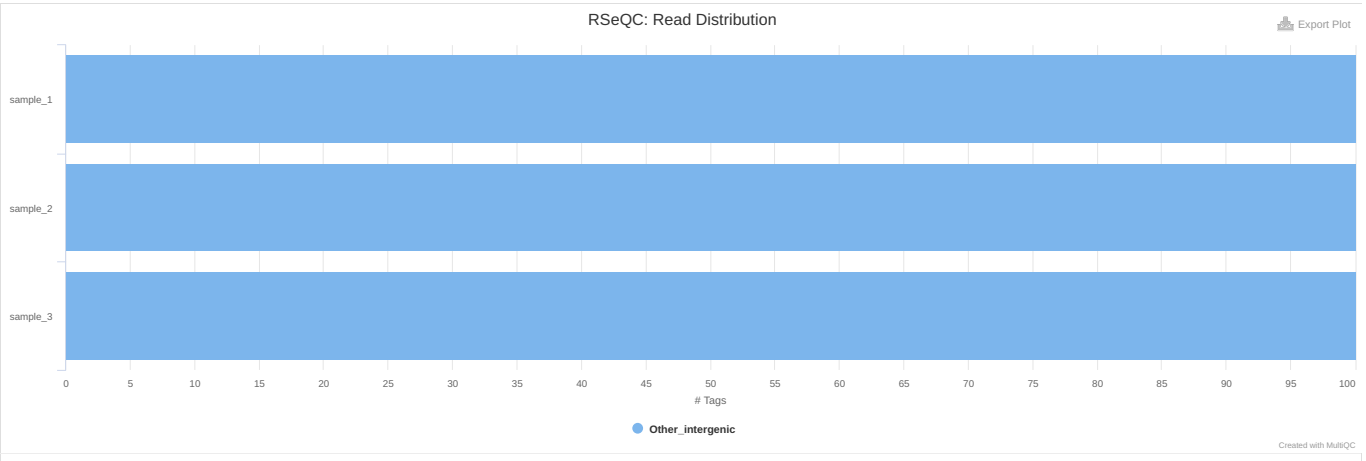
RSeQC

RSeQC package provides a number of useful modules that can comprehensively evaluate high throughput RNA-seq data. *DOI: 10.1093/bioinformatics/bts356.*

Read Distribution

Read Distribution calculates how mapped reads are distributed over genome features.

Number of Tags | Percentages



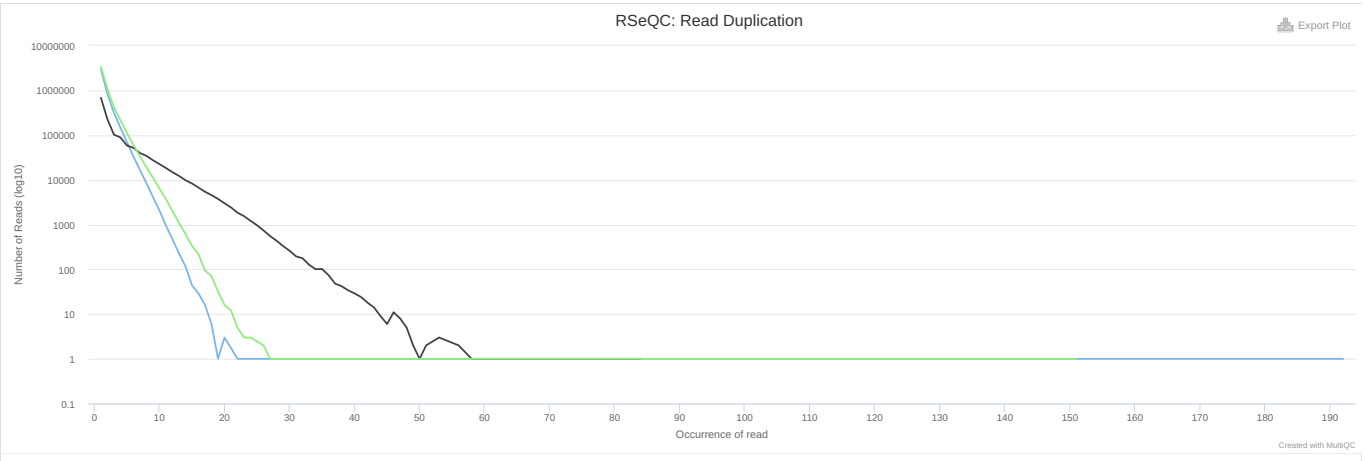
Inner Distance

Inner Distance calculates the inner distance (or insert size) between two paired RNA reads. Note that this can be negative if fragments overlap.

Counts | Percentages

Read Duplication

read_duplication.py calculates how many alignment positions have a certain number of exact duplicates. Note - plot truncated at 500 occurrences and binned.



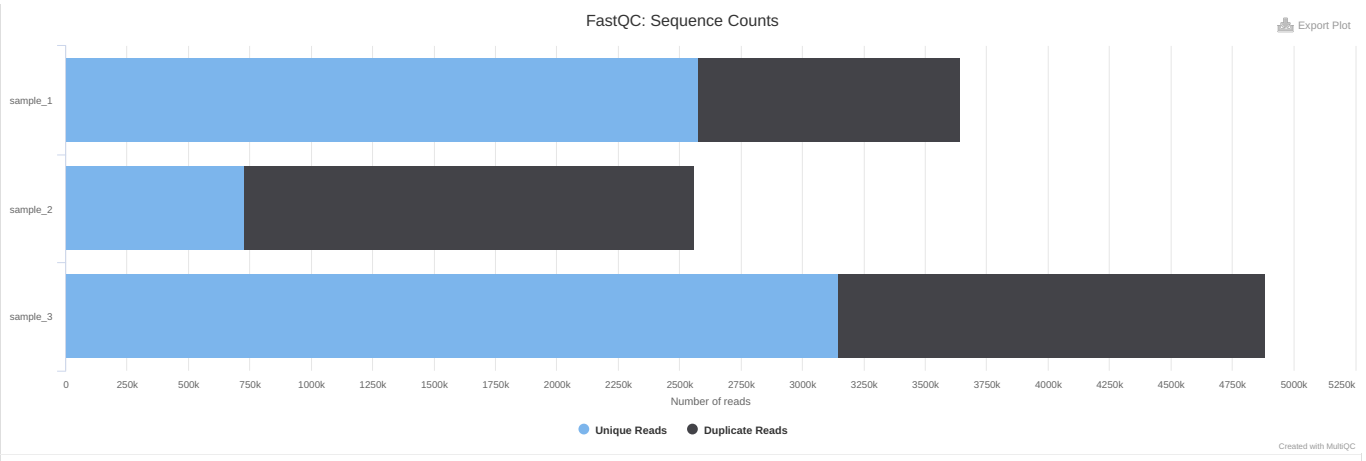
FastQC

FastQC is a quality control tool for high throughput sequence data, written by Simon Andrews at the Babraham Institute in Cambridge.

Sequence Counts

Sequence counts for each sample. Duplicate read counts are an estimate only.

Number of reads | Percentages

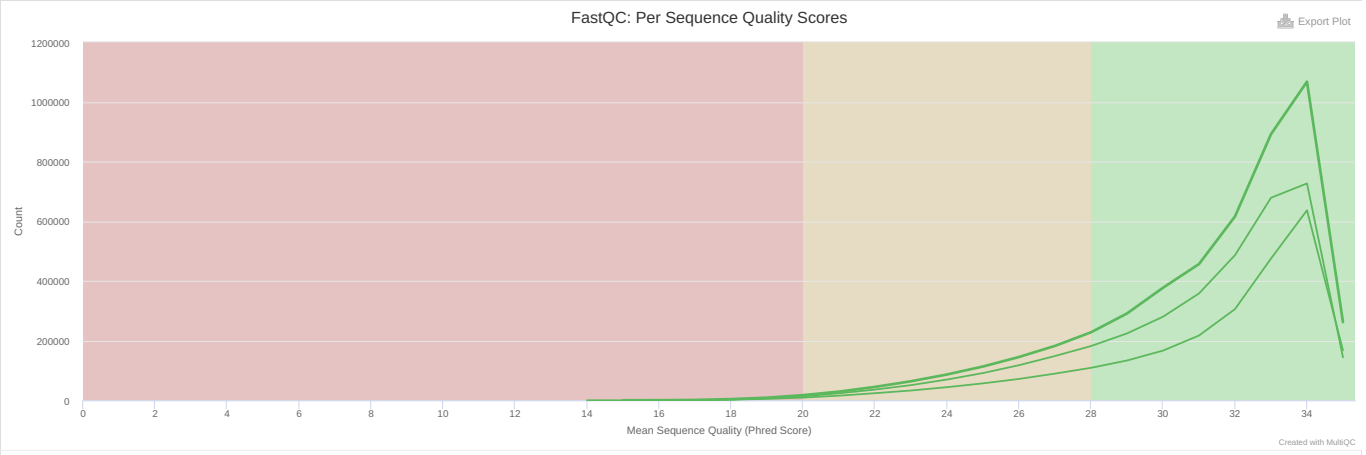


Sequence Quality Histograms 3

The mean quality value across each base position in the read.

Per Sequence Quality Scores 3

The number of reads with average quality scores. Shows if a subset of reads has poor quality.



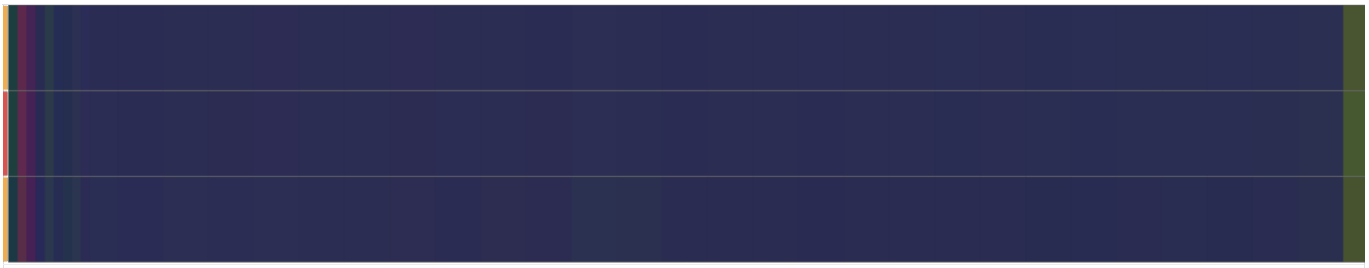
Per Base Sequence Content 2

The proportion of each base position for which each of the four normal DNA bases has been called.

Click a sample row to see a line plot for that dataset.

Rollover for sample name

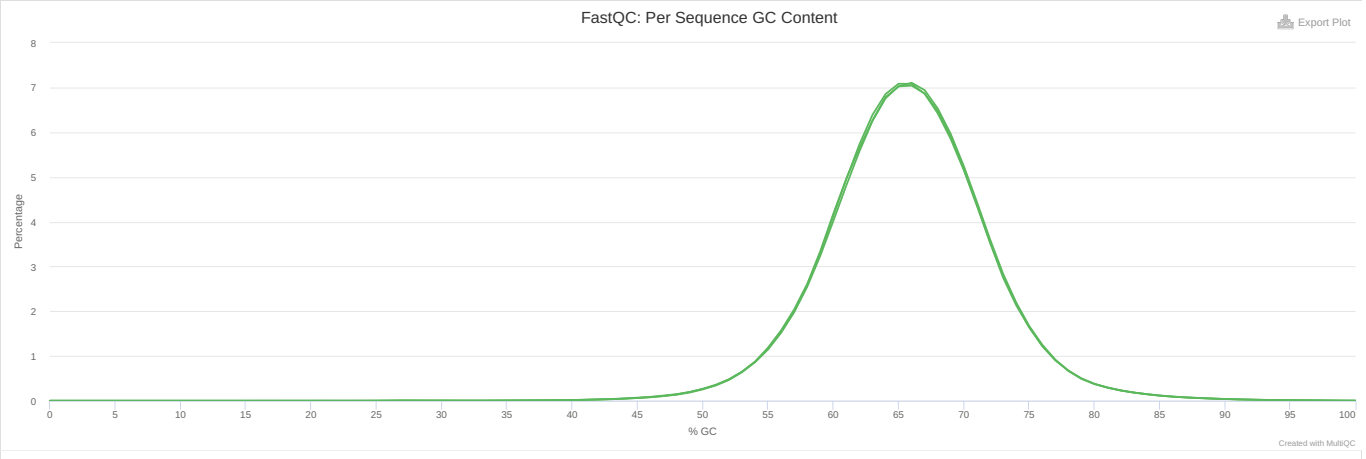
Position: - %T: - %C: - %A: - %G: - Export Plot



Per Sequence GC Content 3

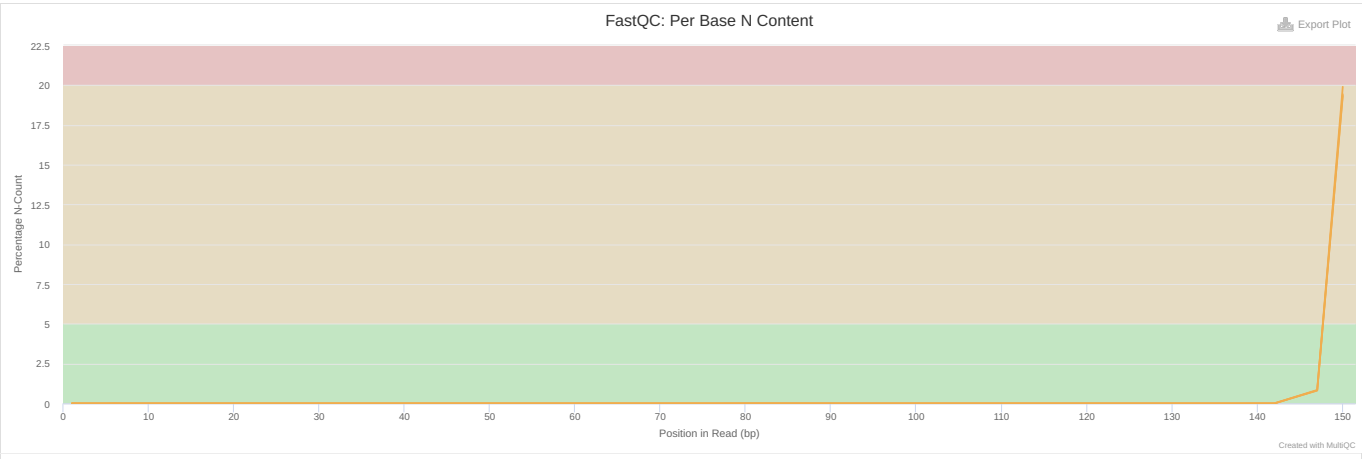
The average GC content of reads. Normal random library typically have a roughly normal distribution of GC content.

Percentages | Counts



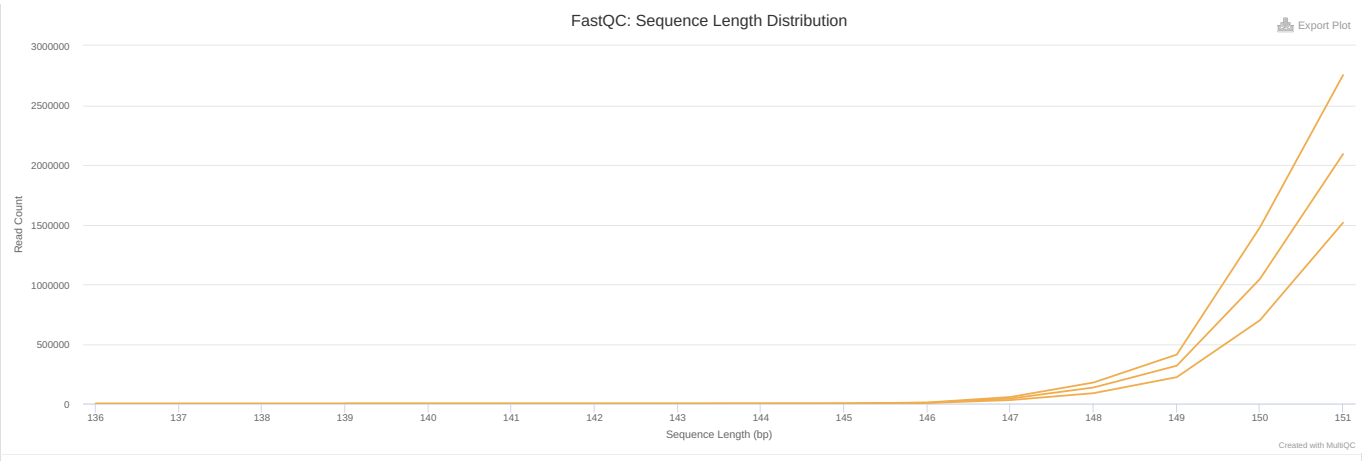
Per Base N Content 3

The percentage of base calls at each position for which an `N` was called.



Sequence Length Distribution 3

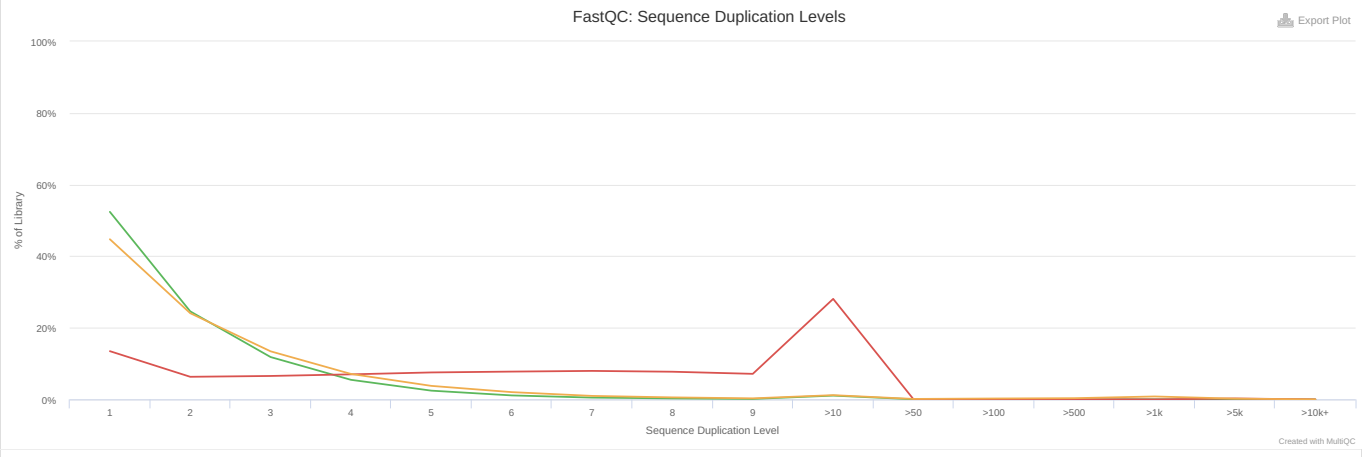
The distribution of fragment sizes (read lengths) found. See the FastQC help



Sequence Duplication Levels

1 1

The relative level of duplication found for every sequence.



Overrepresented sequences

1 2

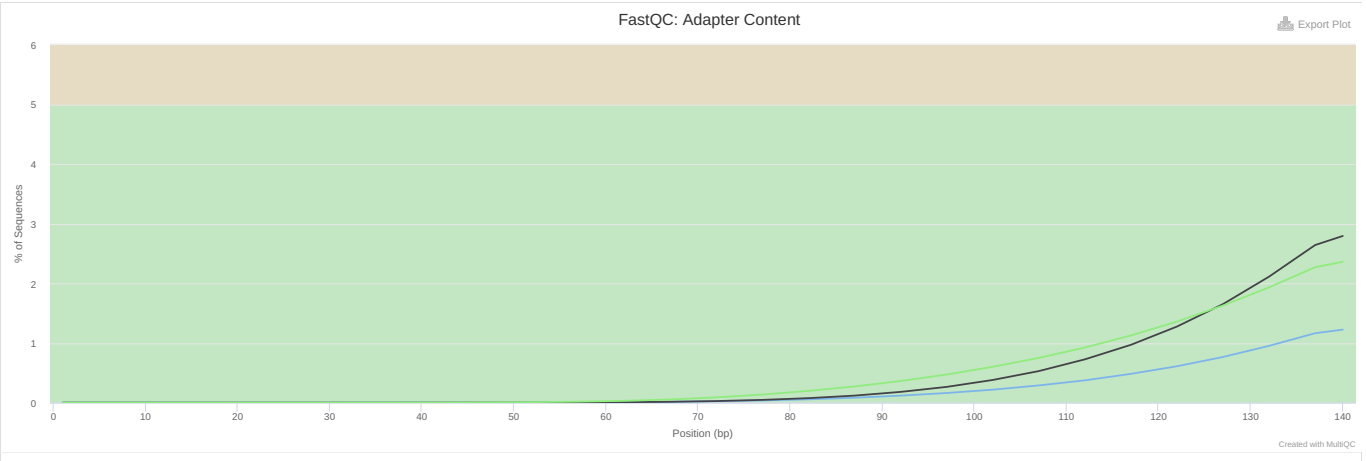
The total amount of overrepresented sequences found in each library.

3 samples had less than 1% of reads made up of overrepresented sequences

Adapter Content

3

The cumulative percentage count of the proportion of your library which has seen each of the adapter sequences at each position.



Status Checks

Status for each FastQC section showing whether results seem entirely normal (green), slightly abnormal (orange) or very unusual (red).

Sort by highlight **Min:** 0 **Max:** 1