# multiqc

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

Report generated on 2024-01-28, 22:12 CET based on data in: `/work_beegfs/sunam239/genomics`

## General Statistics

Copy table | Configure Columns | Plot   Showing ⁵/₅ rows and ¹³/₁₉ columns.

| Sample Name | N50 (Kbp) | Assembly Length (Mbp) | Organism | Contigs | CDS | % Duplication | M Reads After Filtering | GC content | % PF | % Adapter | % Dups | % GC | M Seqs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1_short_reads_qc \| 1 | | | | | | | | | | | 33.2% | 45% | 1.6 |
| 1_short_reads_qc \| 241155E_R1 | | | | | | 4.1% | 3.2 | 45.1% | 98.4% | 0.1% | | | |
| 1_short_reads_qc \| 3 | | | | | | | | | | | 33.6% | 45% | 1.6 |
| 3_hybrid_assembly \| quast \| 3_hybrid_assembly | 4 332.3Kbp | 4.5Mbp | | | | | | | | | | | |
| 4_annotated_genome \| strain | | | Genus species | 7.0 | 3883 | | | | | | | | |

## QUAST

QUAST is a quality assessment tool for genome assemblies, written by the Center for Algorithmic Biotechnology. *DOI: 10.1093/bioinformatics/btt086.*

### Assembly Statistics

Copy table | Configure Columns | Plot   Showing ¹/₁ rows and ⁹/₉ columns.

| Sample Name | N50 (Kbp) | L50 (K) | Largest contig (Kbp) | Length (Mbp) | Misassemblies | Mismatches/100kbp | Indels/100kbp | Genes | Genome Fraction |
|---|---|---|---|---|---|---|---|---|---|
| 3_hybrid_assembly \| quast \| 3_hybrid_assembly | 4 332.3Kbp | 0.0K | 4 332.3Kbp | 4.5Mbp | - | - | - | 2 495 | -% |

## Number of Contigs

This plot shows the number of contigs found for each assembly, broken down by length.

Counts | Percentages

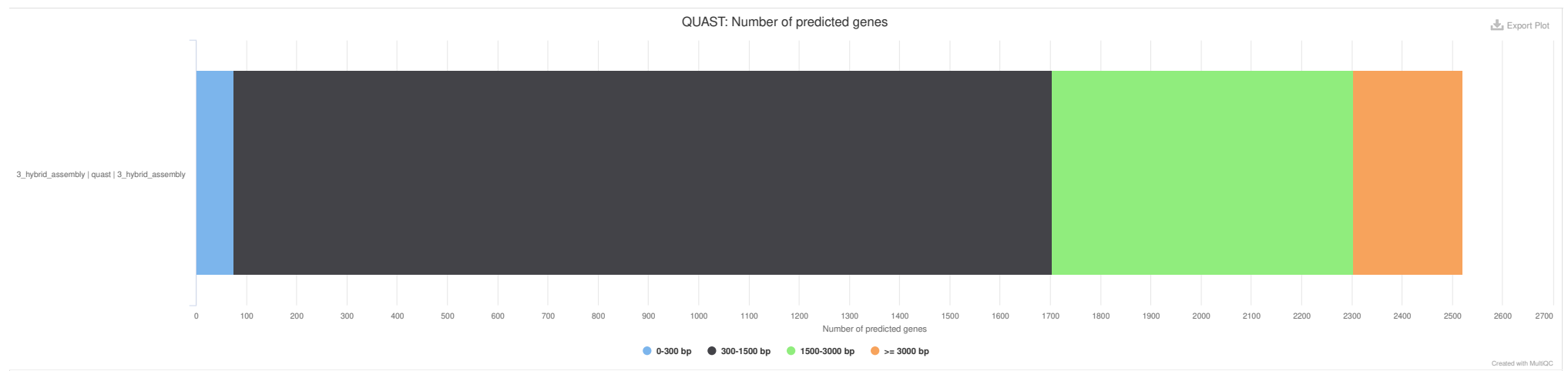QUAST: Number of Contigs



● >= 50000 bp    ● 5000-10000 bp    ● 0-1000 bp
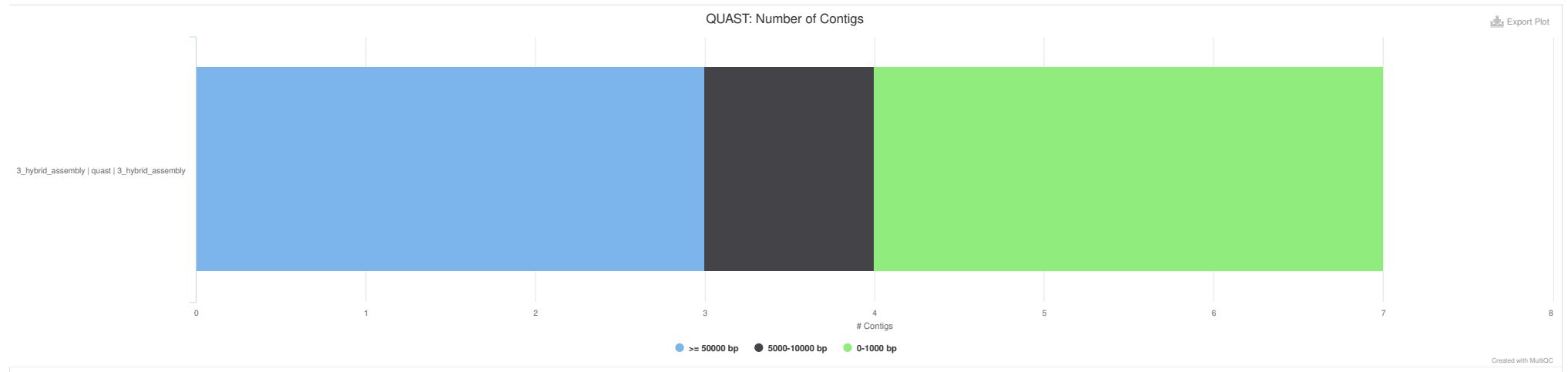
## Number of Predicted Genes

This plot shows the number of predicted genes found for each assembly, broken down by length.

Counts | Percentages

QUAST: Number of predicted genes



● 0-300 bp    ● 300-1500 bp    ● 1500-3000 bp    ● >= 3000 bp

## Number of Partially Predicted Genes

This plot shows the number of partially predicted genes found for each assembly, broken down by length.

Error - was not able to plot data.

## Prokka

Prokka is a software tool for the rapid annotation of prokaryotic genomes.   *DOI: 10.1093/bioinformatics/btu153.*

This barplot shows the distribution of different types of features found in each contig.

Features   Percentages



Prokka: Feature Types

**CDS**   **rRNA**   **tRNA**   **CRISPR array**
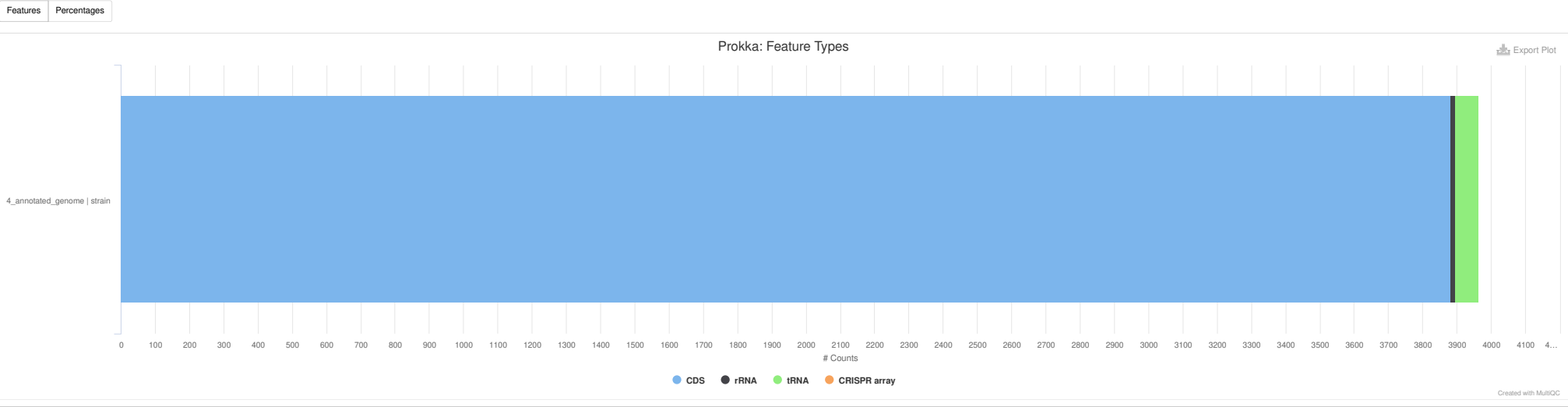
## NanoStat

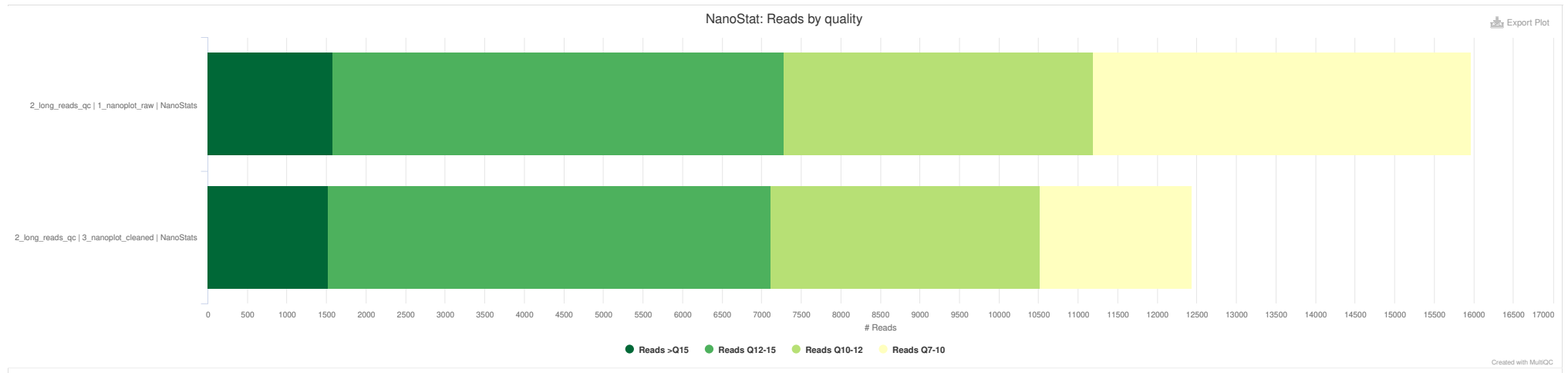NanoStat reports various statistics from a long read sequencing dataset in FASTQ, BAM or sequencing summary format.   *DOI: 10.1093/bioinformatics/bty149.*

### Summary Statistics

Copy table   Configure Columns   Plot   Showing $^2/_2$ rows and $^5/_7$ columns.

| Sample Name | Median length | Read N50 | Median Qual | # Reads (K) | Total Bases (Mb) |
|---|---|---|---|---|---|
| **2_long_reads_qc | 1_nanoplot_raw | NanoStats** | 3 270 bp | 21 971 bp | 11.7 | 16.0 | 146.3 |
| **2_long_reads_qc | 3_nanoplot_cleaned | NanoStats** | 4 477 bp | 22 747 bp | 12.5 | 12.4 | 131.7 |

## Reads by quality

Read counts categorised by read quality (Phred score).

| Number of Reads | Percentages |

NanoStat: Reads by quality



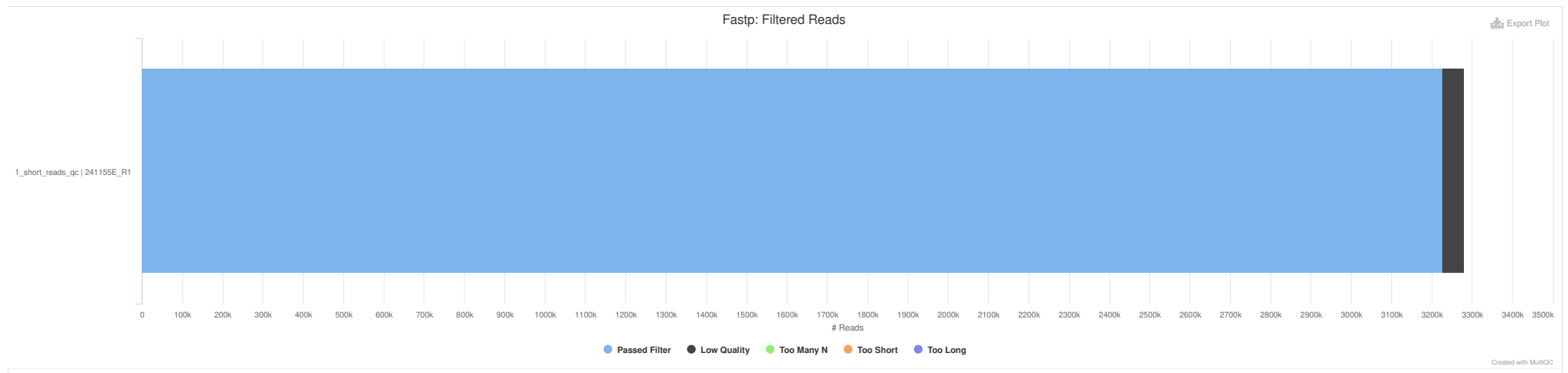Reads >Q15 ● Reads Q12-15 ● Reads Q10-12 ● Reads Q7-10

# fastp *Version:* `0.23.4`

fastp An ultra-fast all-in-one FASTQ preprocessor (QC, adapters, trimming, filtering, splitting...).   *DOI: 10.1093/bioinformatics/bty560.*

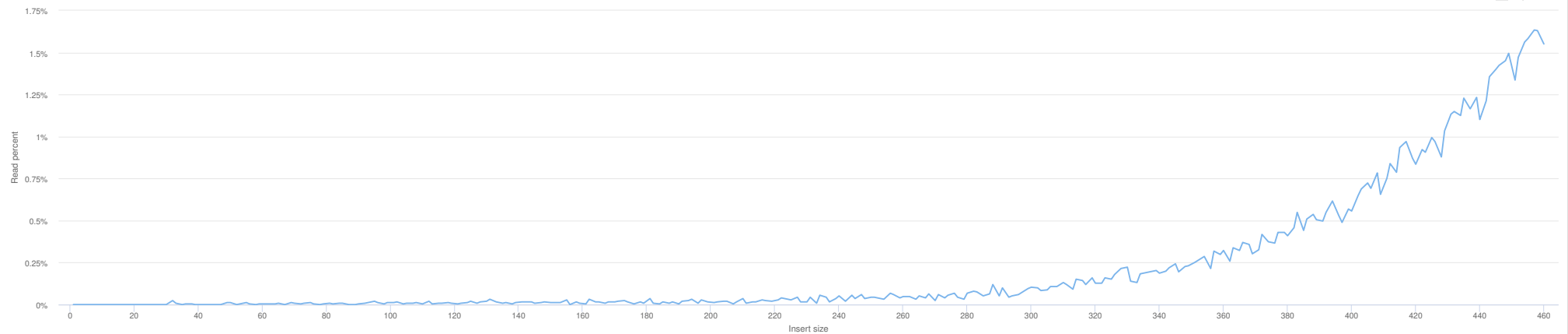## Filtered Reads

Filtering statistics of sampled reads.

| Number of Reads | Percentages |

Fastp: Filtered Reads



● Passed Filter  ● Low Quality  ● Too Many N  ● Too Short  ● Too Long

## Insert Sizes

Insert size estimation of sampled reads.

Y-Limits: on



## Sequence Quality

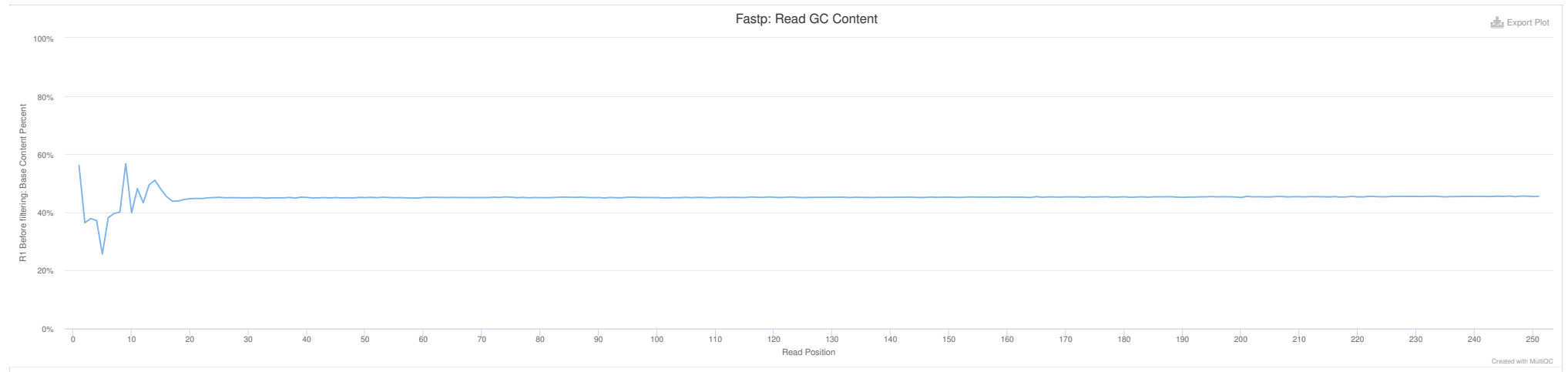Average sequencing quality over each base of all reads.

Y-Limits: on

| Read 1: Before filtering | Read 1: After filtering | Read 2: Before filtering | Read 2: After filtering |

MultiQC Report

file:///run/user/767620/gvfs/sftp:host=caucluster.rz.uni-kiel.de/work_beegfs/s...
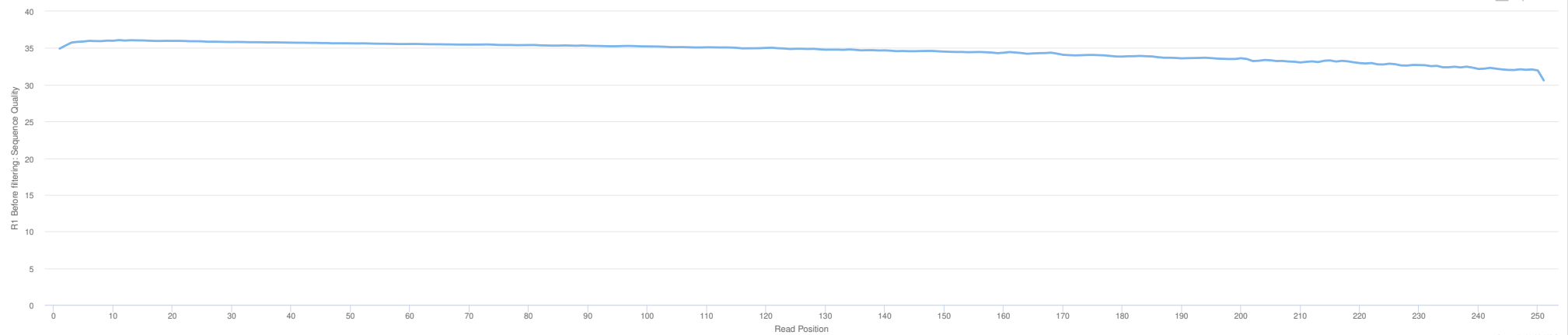
## GC Content

Average GC content over each base of all reads.

Y-Limits: on

| Read 1: Before filtering | Read 1: After filtering | Read 2: Before filtering | Read 2: After filtering |



Fastp: Read GC Content

## N content

Average N content over each base of all reads.

Y-Limits: on

| Read 1: Before filtering | Read 1: After filtering | Read 2: Before filtering | Read 2: After filtering |



Fastp: Read N Content
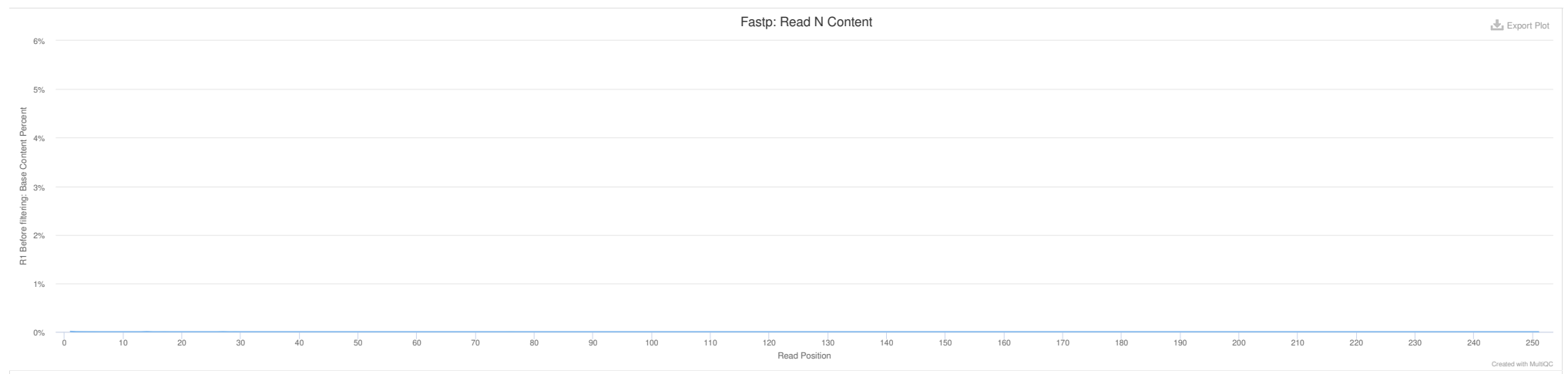
# FastQC  *Version:* `0.12.1`

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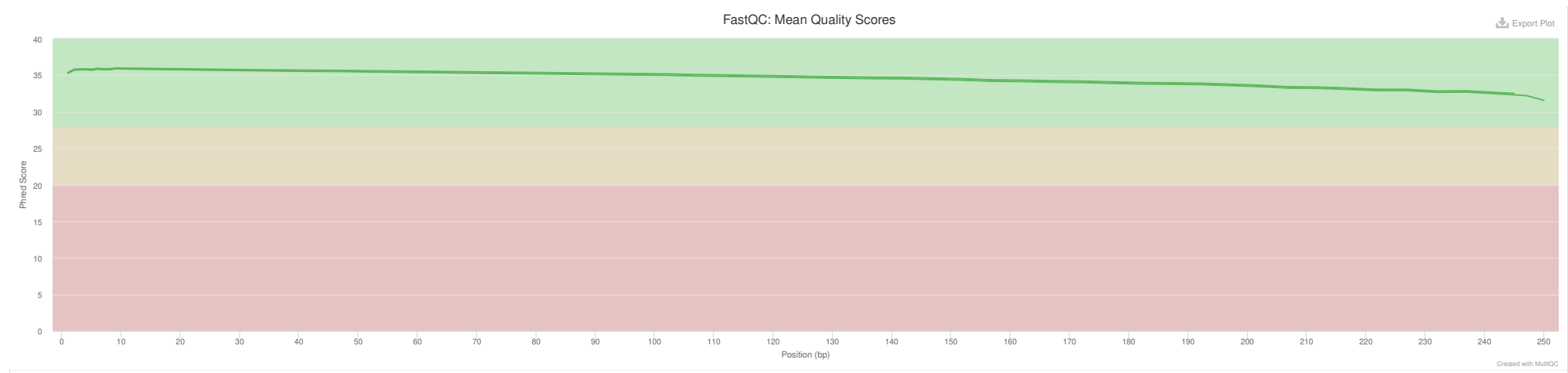
## Sequence Counts

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

| Number of reads | Percentages |
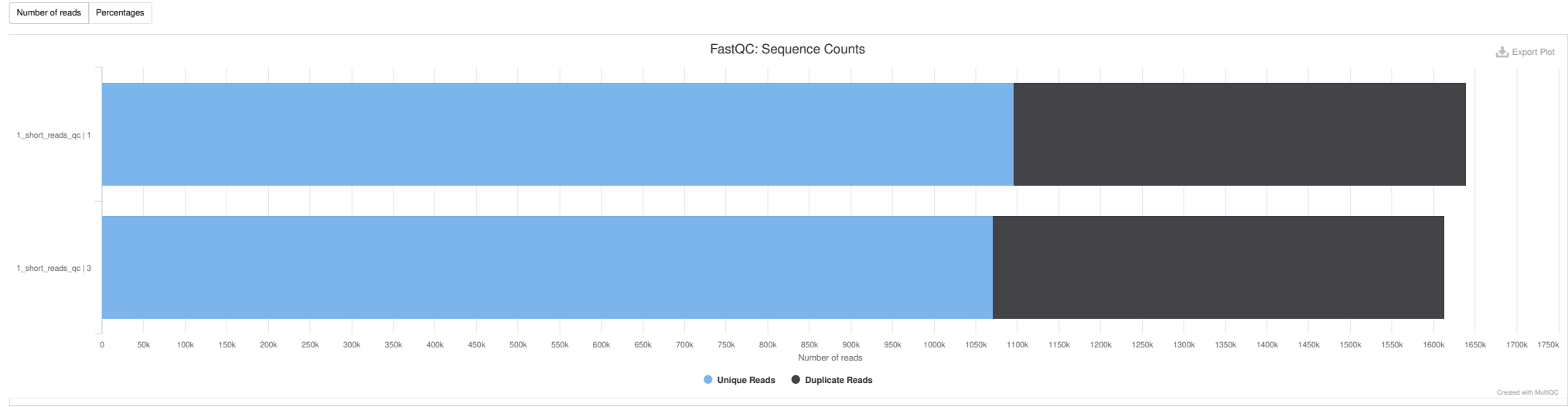


## Sequence Quality Histograms  `2`

The mean quality value across each base position in the read.
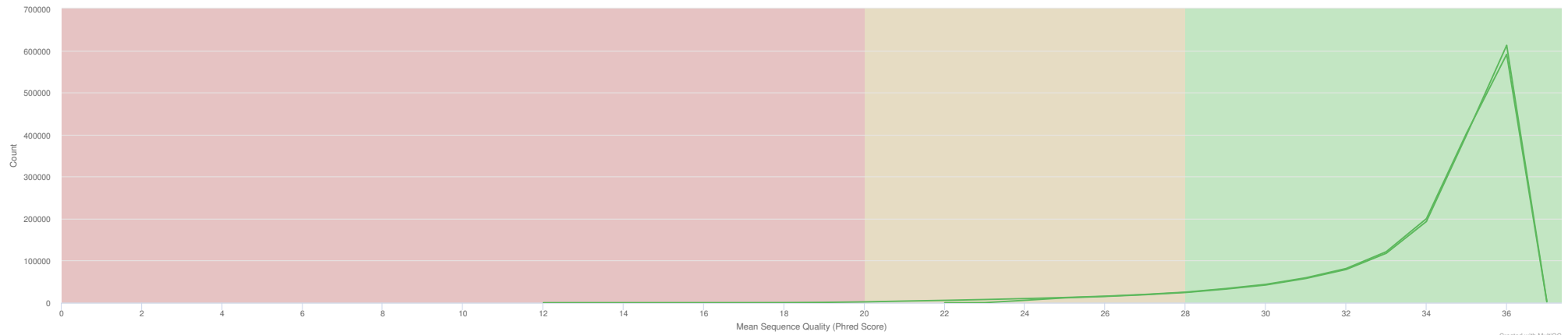
Y-Limits: `on`

## Per Sequence Quality Scores [2]

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

Y-Limits: on

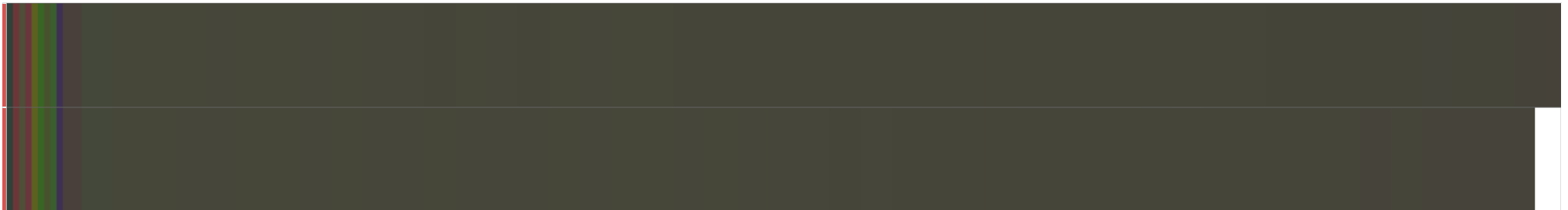

FastQC: Per Sequence Quality Scores

## Per Base Sequence Content

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.

ℹ Rollover for sample name

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

MultiQC Report

file:///run/user/767620/gvfs/sftp:host=caucluster.rz.uni-kiel.de/work_beegfs/s...
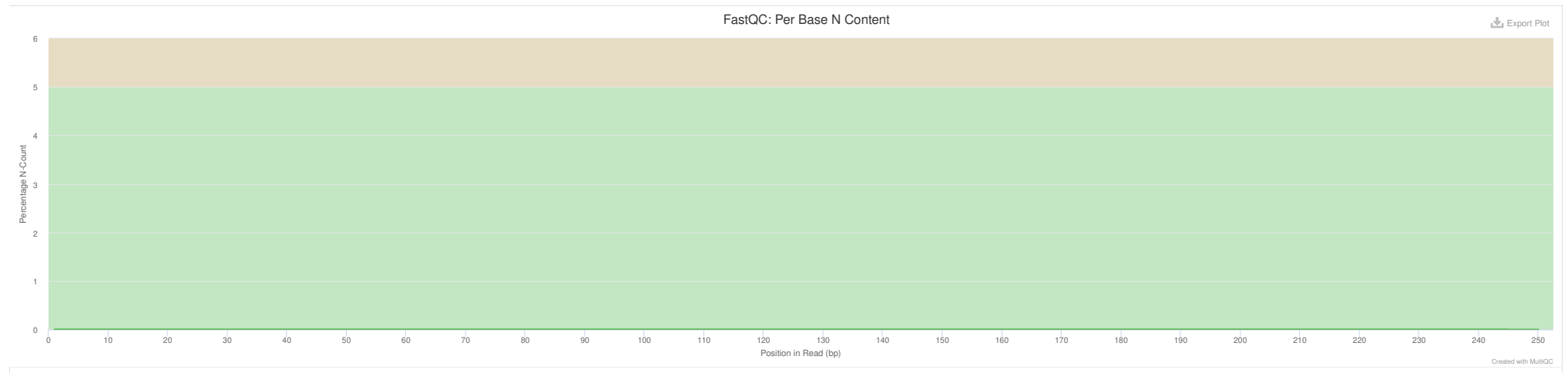
## Per Sequence GC Content  2

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

Y-Limits: on

Percentages | Counts



FastQC: Per Sequence GC Content
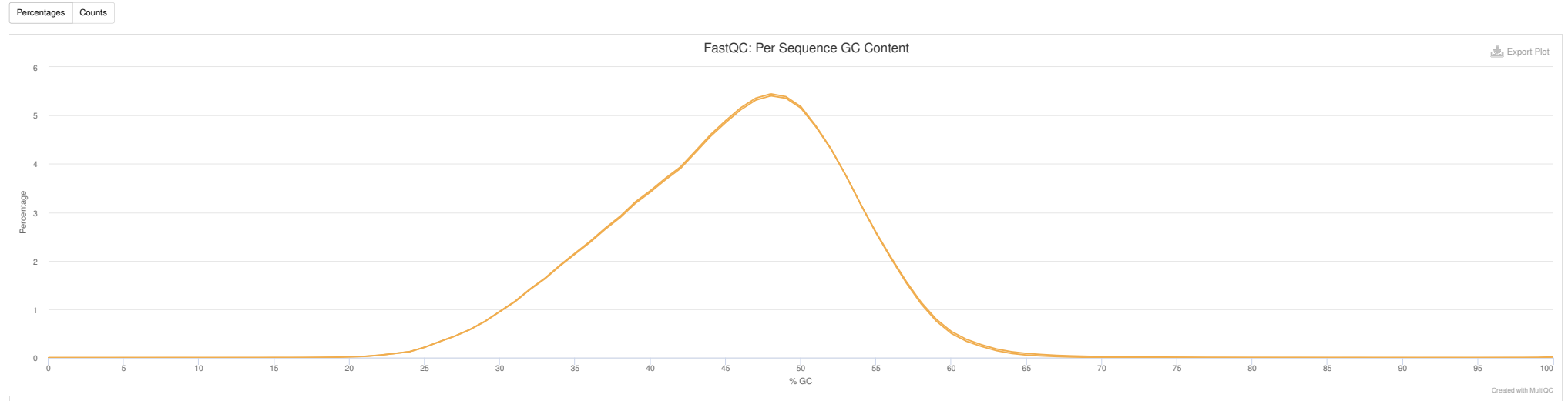
## Per Base N Content  2

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

Y-Limits: on



FastQC: Per Base N Content

MultiQC Report

file:///run/user/767620/gvfs/sftp:host=caucluster.rz.uni-kiel.de/work_beegfs/s...

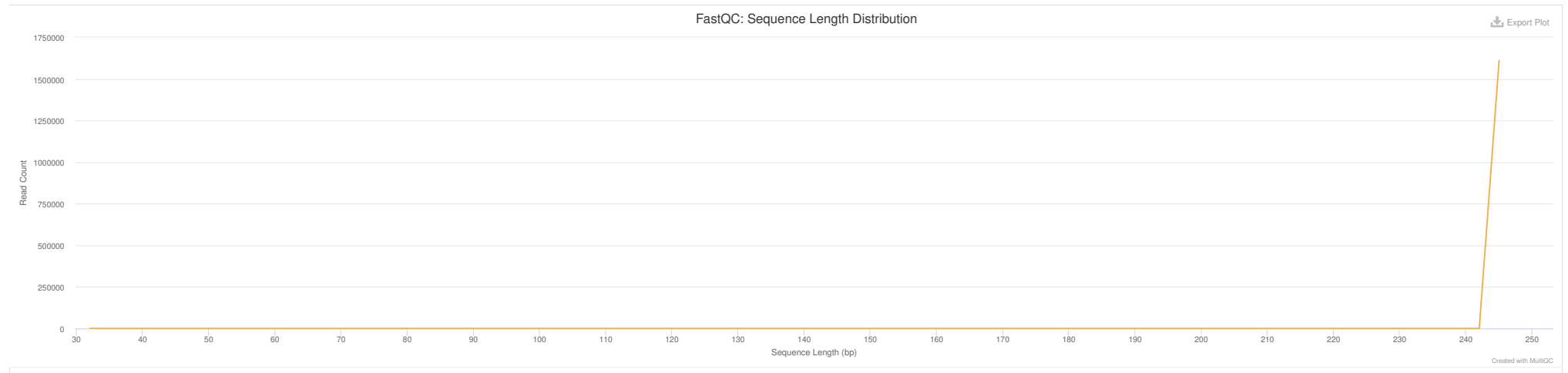## Sequence Length Distribution

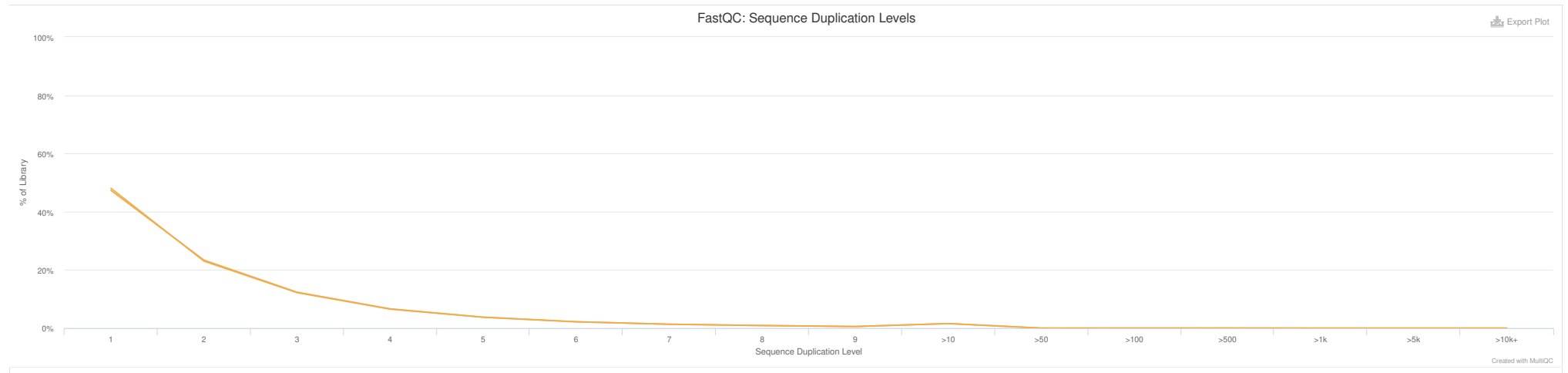The distribution of fragment sizes (read lengths) found. See the FastQC help

Y-Limits: on



FastQC: Sequence Length Distribution

## Sequence Duplication Levels

The relative level of duplication found for every sequence.

Y-Limits: on



FastQC: Sequence Duplication Levels

## Overrepresented sequences by sample

The total amount of overrepresented sequences found in each library.

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

## Top overrepresented sequences

Top overrepresented sequences across all samples. The table shows 20 most overrepresented sequences across all samples, ranked by the number of samples they occur in.
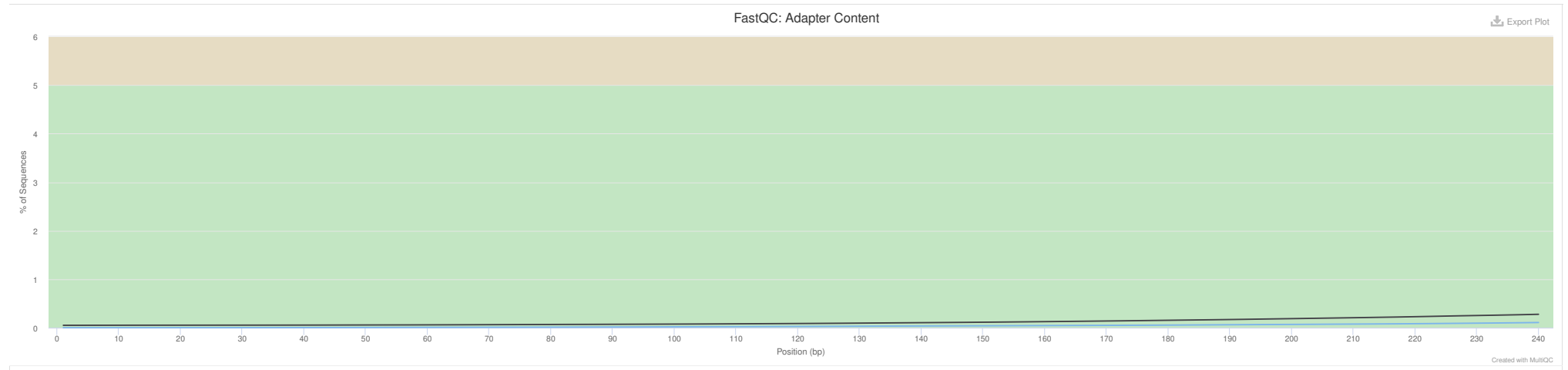
📋 Copy table    Showing $^0/_0$ rows.

**Overrepresented sequence**

## Adapter Content   2

The cumulative percentage count of the proportion of your library which has seen each of the adapter sequences at each position.
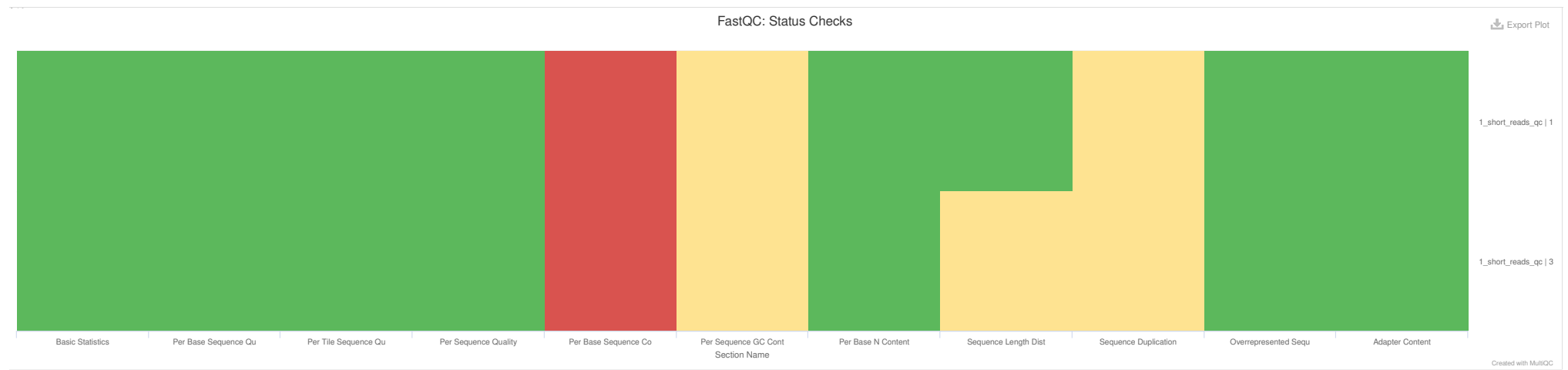
Y-Limits: on



FastQC: Adapter Content

## 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



FastQC: Status Checks

## Software Versions

Software Versions lists versions of software tools extracted from file contents.

📋 Copy table

| Software | Version |
| --- | --- |
| FastQC | 0.12.1 |
| fastp | 0.23.4 |

**MultiQC v1.19** - Written by Phil Ewels, available on GitHub.
This report uses HighCharts, jQuery, jQuery UI, Bootstrap, FileSaver.js and clipboard.js.

≋ seqera