BV-BRC Test Report

A24. Service – FASTQ Utilities

Item to test	FASTQ Utilities Service using read files and SRA accessions
URL	https://www.bv-brc.org/app/FastqUtil
Prerequisites	Sample read files in the workspace
References	https://www.bv-brc.org/docs/quick_references/services/fastq_utilities_service.html https://www.bv-brc.org/docs/tutorial/fastq_utilities/fastq_utilities.html
Tester(s)	Rebecca Wattam, Maulik Shukla
Test date	10-May-2022 (follow-up from original test)
Test result	**Passed**

Overview

- Test FASTQ Utilities service using exemplar reads sets.
- Test input options, i.e. read files and SRA accession as input.
- Test different processing options, i.e. trim, fastqc, and real alignment to a reference genome.
- For each job submitted, verify successful completion of the job, presence of output files, and quality of the results from various processing steps.

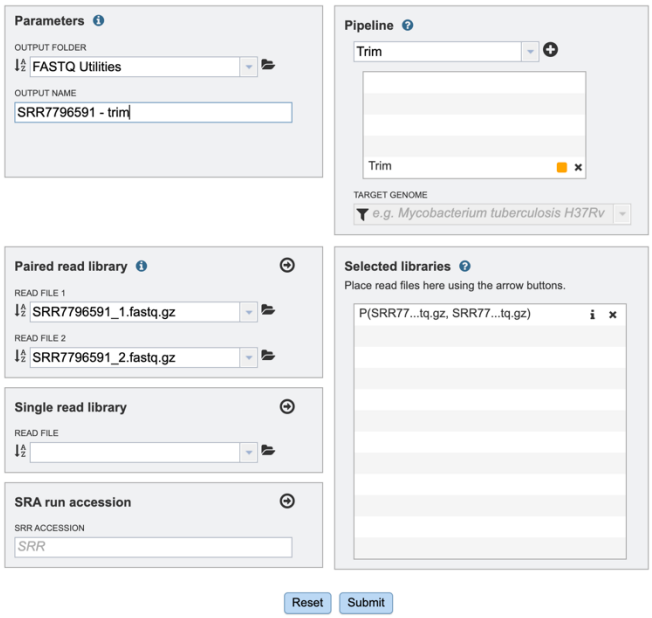
Test Data

Dataset	Rational	Input Format	Input
Buchnera aphidicola - SRR7796591	Workshop example	Read files	SRR7796591_1.fastq.gz SRR7796591_2.fastq.gz

- All test datasets and corresponding job results are available in the following public workspace: https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/FASTQ%20Utilities

Test Results

- All assembly jobs completed successfully, without any errors.
- All jobs resulted in expected output files in corresponding job output directory in the expected formats.
- All test datasets and corresponding job results are available in the following public workspace: https://www.bv-brc.org/workspace/BVBRC@patricbrc.org/BVBRC%20Tests/FASTQ%20Utilities
- Below are a series of screenshots showing successful completion of the jobs availability of the result files in the workspace.



	Name	Size	Owner	Members
↑	Parent folder			-
	SRR7796591_1.fastq_trimming_report.txt	4.5 kB	me	Only me
	SRR7796591_1_fastqc.html	215.7 kB	me	Only me
	SRR7796591_1_ptrim.fq.gz	1.1 GB	me	Only me
	SRR7796591_1_ptrim.fq_SRR7796591_2_ptrim.fq.aligned.2.fq.gz	158.3 kB	me	Only me
	SRR7796591_1_ptrim.fq_SRR7796591_2_ptrim.fq.aligned.bam	541.0 kB	me	Only me
	SRR7796591_1_ptrim.fq_SRR7796591_2_ptrim.fq.aligned.bam.bai	1.9 kB	me	Only me
	SRR7796591_1_ptrim.fq_SRR7796591_2_ptrim.fq.aligned.fq.gz	134.0 kB	me	Only me
	SRR7796591_1_ptrim.fq_SRR7796591_2_ptrim.fq.all.bam.samstat.html	279.2 kB	me	Only me
	SRR7796591_1_ptrim.fq_SRR7796591_2_ptrim.fq.unmapped.1.fq.gz	1.1 GB	me	Only me
	SRR7796591_1_ptrim.fq_SRR7796591_2_ptrim.fq.unmapped.2.fq.gz	1.2 GB	me	Only me
	SRR7796591_2.fastq_trimming_report.txt	4.8 kB	me	Only me
	SRR7796591_2_fastqc.html	217.8 kB	me	Only me
	SRR7796591_2_ptrim.fq.gz	1.2 GB	me	Only me
	SRR7796591_meta.txt	1.4 kB	me	Only me
	bedtools.log.txt	347.9 kB	me	Only me

```
SUMMARISING RUN PARAMETERS
==========================
Input filename: /tmp/work/SRR7796591_1.fastq
Trimming mode: paired-end
Trim Galore version: 0.6.5dev
Cutadapt version: 2.2
Python version: 3.7.10
Number of cores used for trimming: 8
Quality Phred score cutoff: 20
Quality encoding type selected: ASCII+33
Using Illumina adapter for trimming (count: 993). Second best hit was smallRNA (count: 3)
Adapter sequence: 'AGATCGGAAGAGC' (Illumina TruSeq, Sanger iPCR; auto-detected)
Maximum trimming error rate: 0.1 (default)
Minimum required adapter overlap (stringency): 1 bp
Minimum required sequence length for both reads before a sequence pair gets removed: 20 bp
Output file will be GZIP compressed


This is cutadapt 2.2 with Python 3.7.10
Command line parameters: -j 8 -e 0.1 -q 20 -O 1 -a AGATCGGAAGAGC /tmp/work/SRR7796591_1.fastq
Processing reads on 8 cores in single-end mode ...
Finished in 685.12 s (58 us/read; 1.03 M reads/minute).

=== Summary ===

Total reads processed:              11,730,613
Reads with adapters:                 4,658,944 (39.7%)
Reads written (passing filters):    11,730,613 (100.0%)

Total basepairs processed: 1,771,322,563 bp
Quality-trimmed:              14,826,041 bp (0.8%)
Total written (filtered):  1,749,006,102 bp (98.7%)

=== Adapter 1 ===

Sequence: AGATCGGAAGAGC; Type: regular 3'; Length: 13; Trimmed: 4658944 times.

No. of allowed errors:
0-9 bp: 0; 10-13 bp: 1

Bases preceding removed adapters:
  A: 39.2%
  C: 15.6%
  G: 10.1%
  T: 34.8%
```

FastQC Report

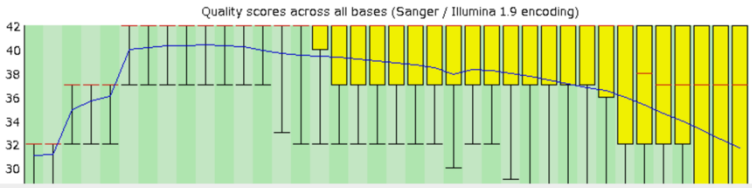
Tue 21 Dec 2021
SRR7796591_1.fastq

Summary

Basic Statistics

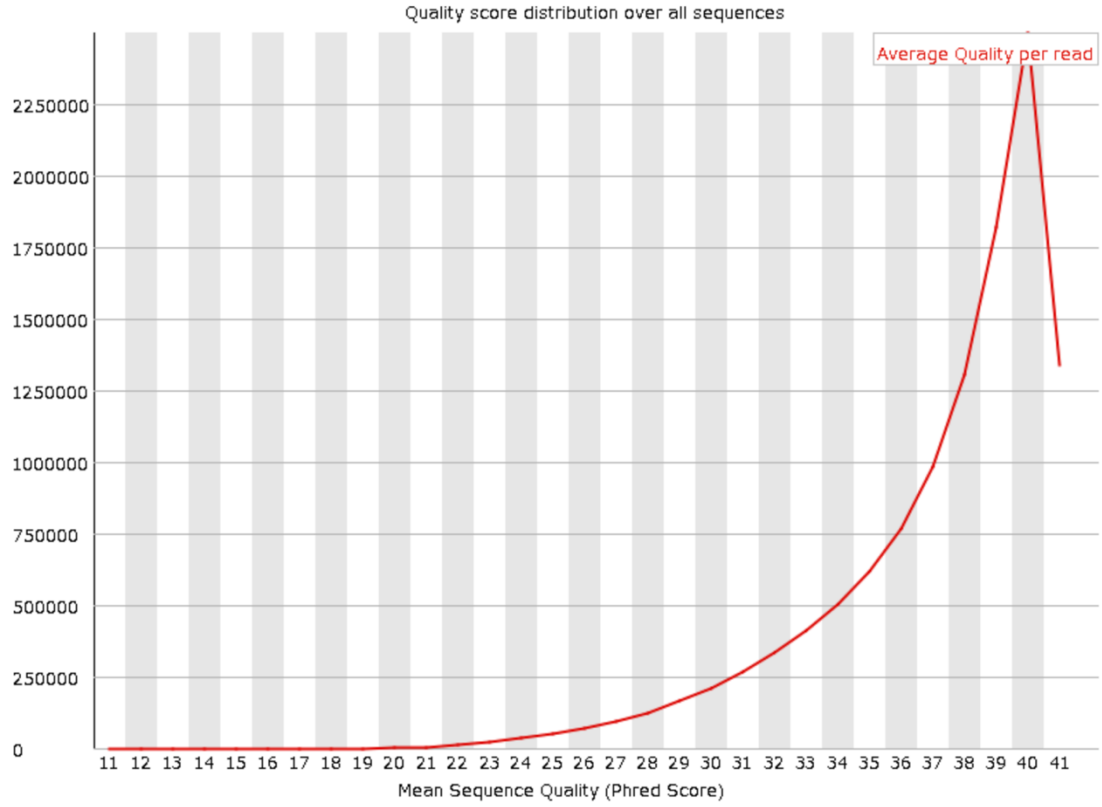
Measure	Value
Filename	SRR7796591_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	11730613
Sequences flagged as poor quality	0
Sequence length	151
%GC	28

Per base sequence quality

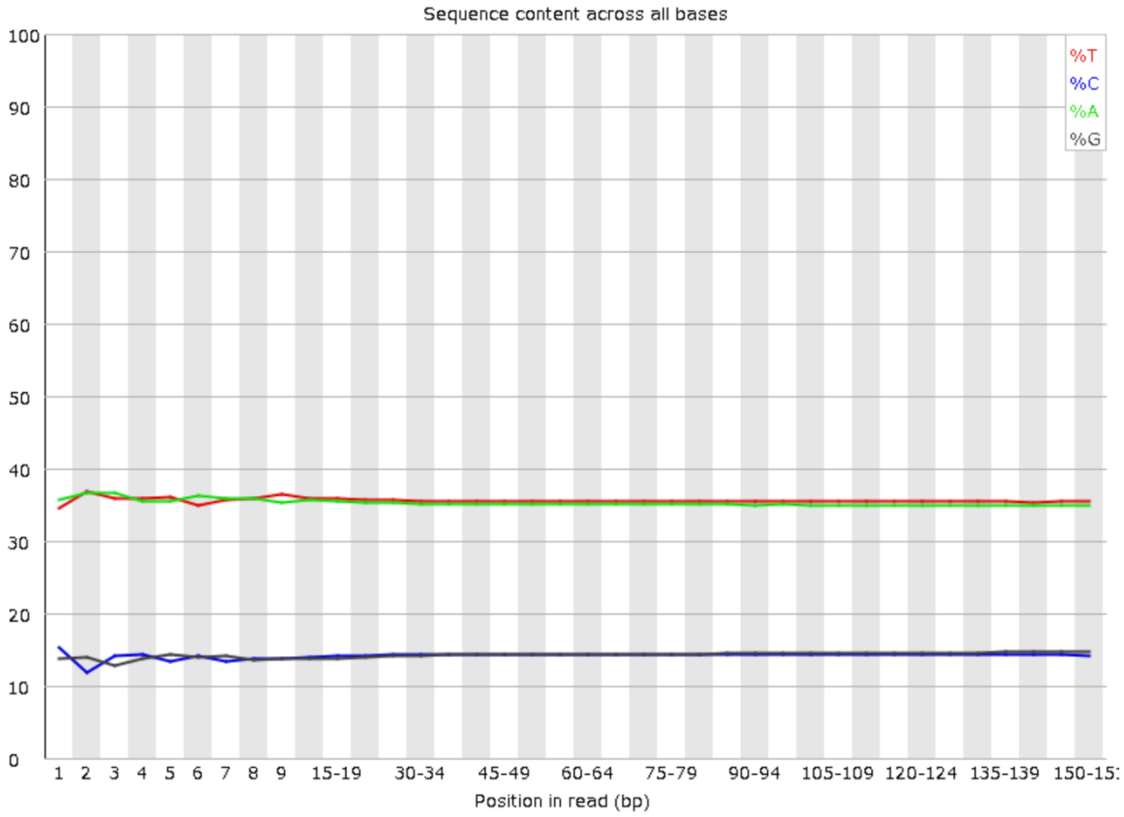


Produced by FastQC (version 0.11.9)

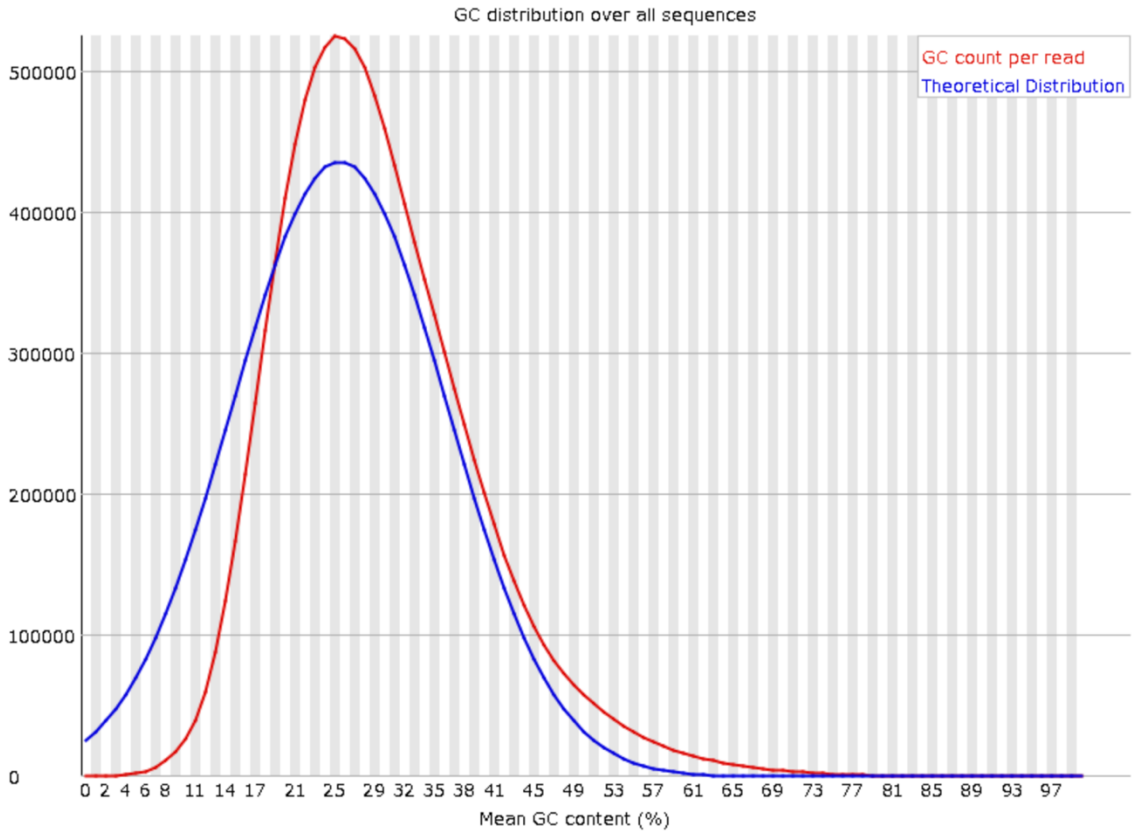
Per sequence quality scores



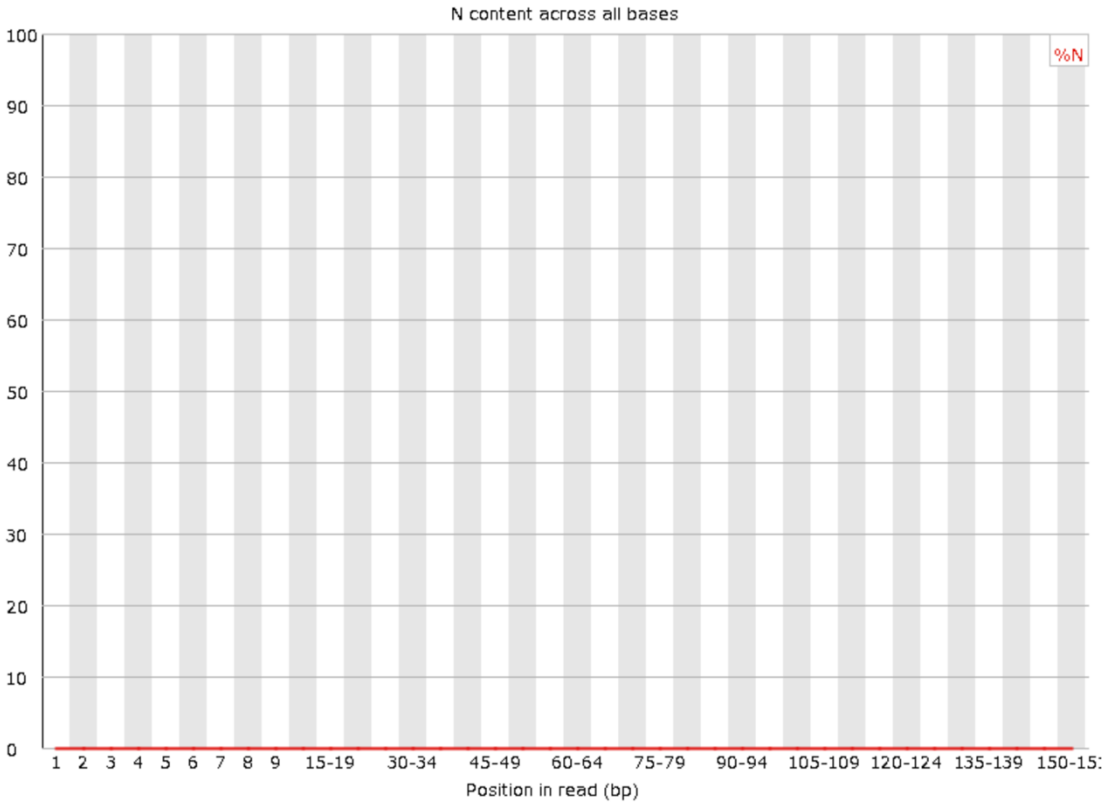
Per base sequence content



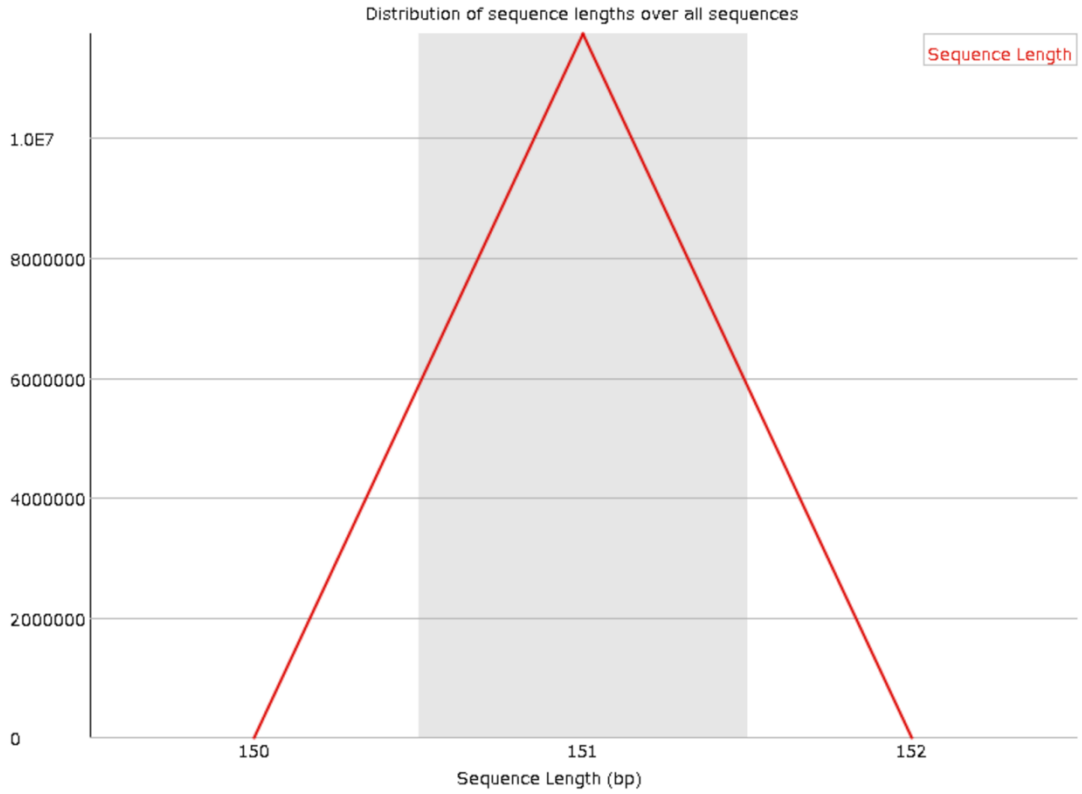
Per sequence GC content



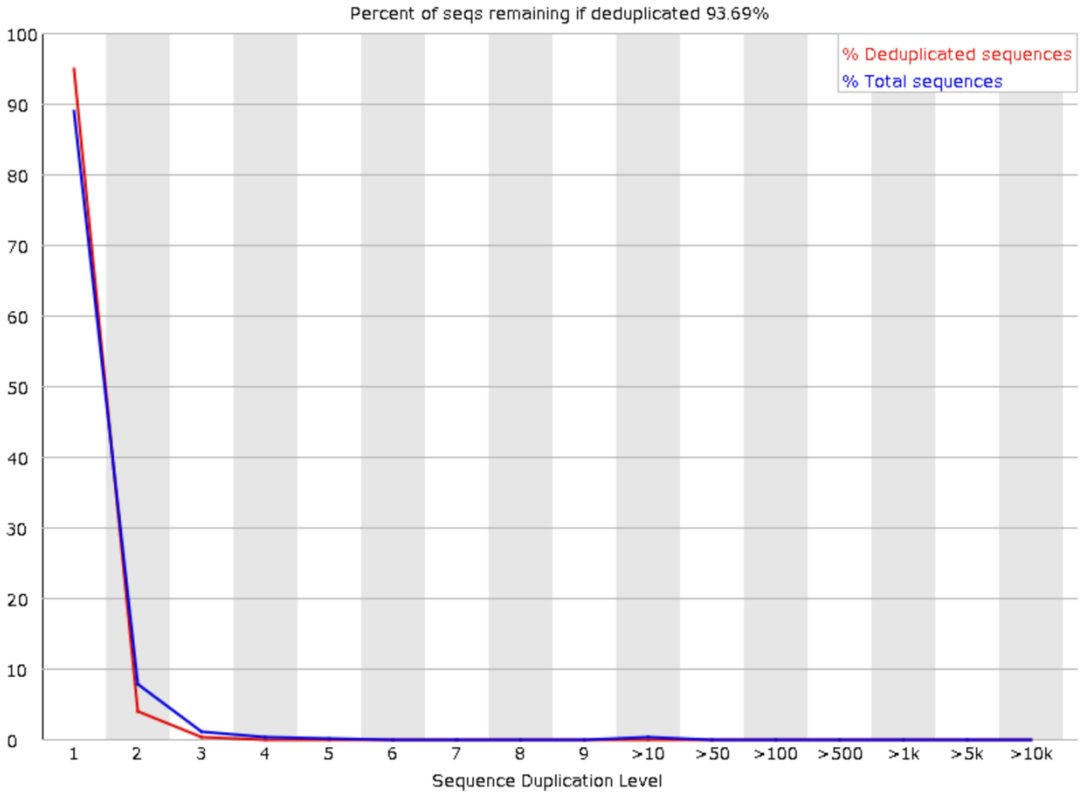
Per base N content



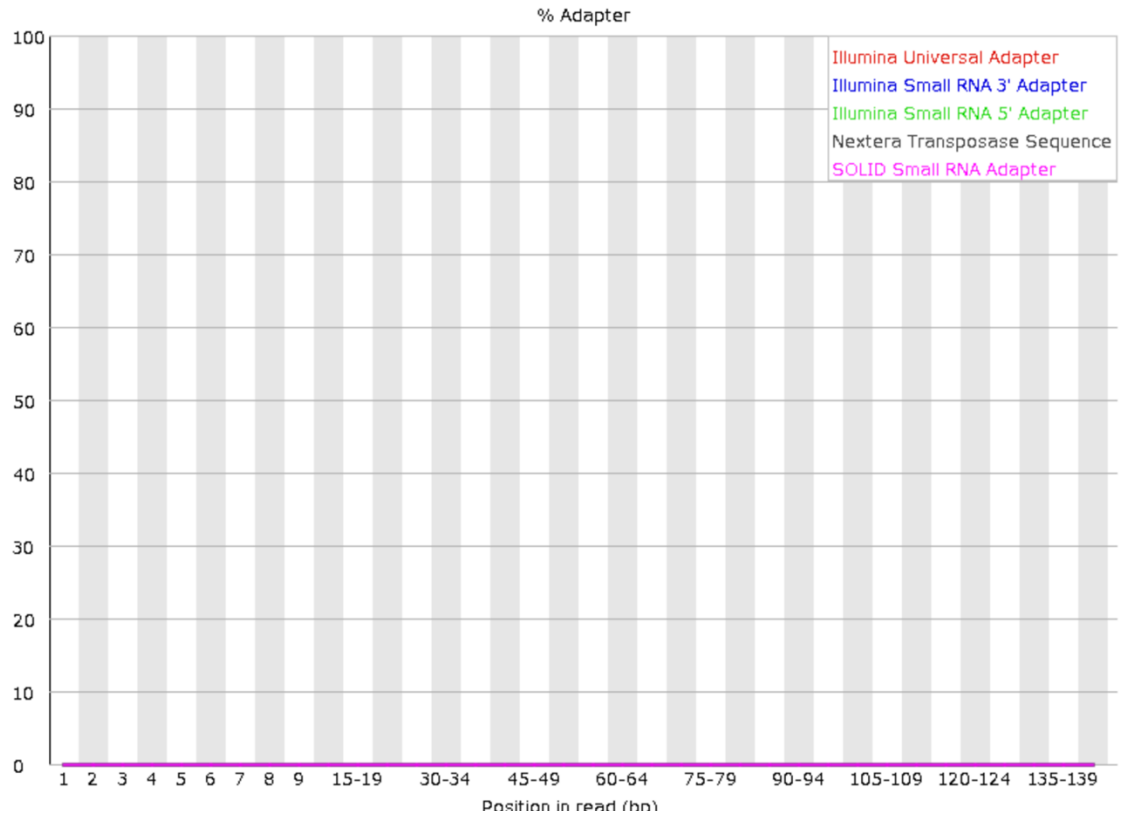
Sequence Length Distribution



Sequence Duplication Levels



Adapter Content



References

- FASTQ Utilities Service Quick Reference Guide
- FATSQ Utilities Service Tutorial