

FastQCParse

FastQCParse User Guide:

It is recommended to open this via browser: [User Guide](#)

Introduction

A Python program to parse FastQtext files, and generate reports and plots.

1. Clone directory: `git clone https://github.com/ms2206/FastQCParse.git`
 2. Make a new python environment based from requirments.yaml `conda env create -f requirements.yaml --<NAME>`
 3. Load environment env. `conda activate <NAME>`
 4. Change directory into FastQCParse
 5. Example Usage: `python3 src/main.py data/raw/fastqc_data2.txt fastqc_2 -a`
-

Set up

Example Usage: With python3, run executable found at `src/main.py` .

Pass `data/raw/fastqc_data2.txt` (or any fastqc file) - as input file, and `fastqc_2` as output directory (or use a customer directory name).

Use optional argument `-a` .

```
python3 src/main.py <FASTQ FILE> <DIR_NAME> [<OPTIONAL_ARGS>]
```

```
python3 src/main.py data/raw/fastqc_data2.txt fastqc_2 -a
```

Entry Point:

```
python3 src/main.py data/raw/fastqc_data2.txt fastqc_2 -a
```

Args Help

```
python src/main.py -h
```

usage: main.py [-h] [-b] [-t] [-s] [-c] [-g] [-n] [-l] [-d] [-o] [-p] [-k] [-a] input_file output_dir

Parse and plot FASTQC data.

positional arguments:

input_file Path to the FASTQC file.
output_dir Directory to save plots.

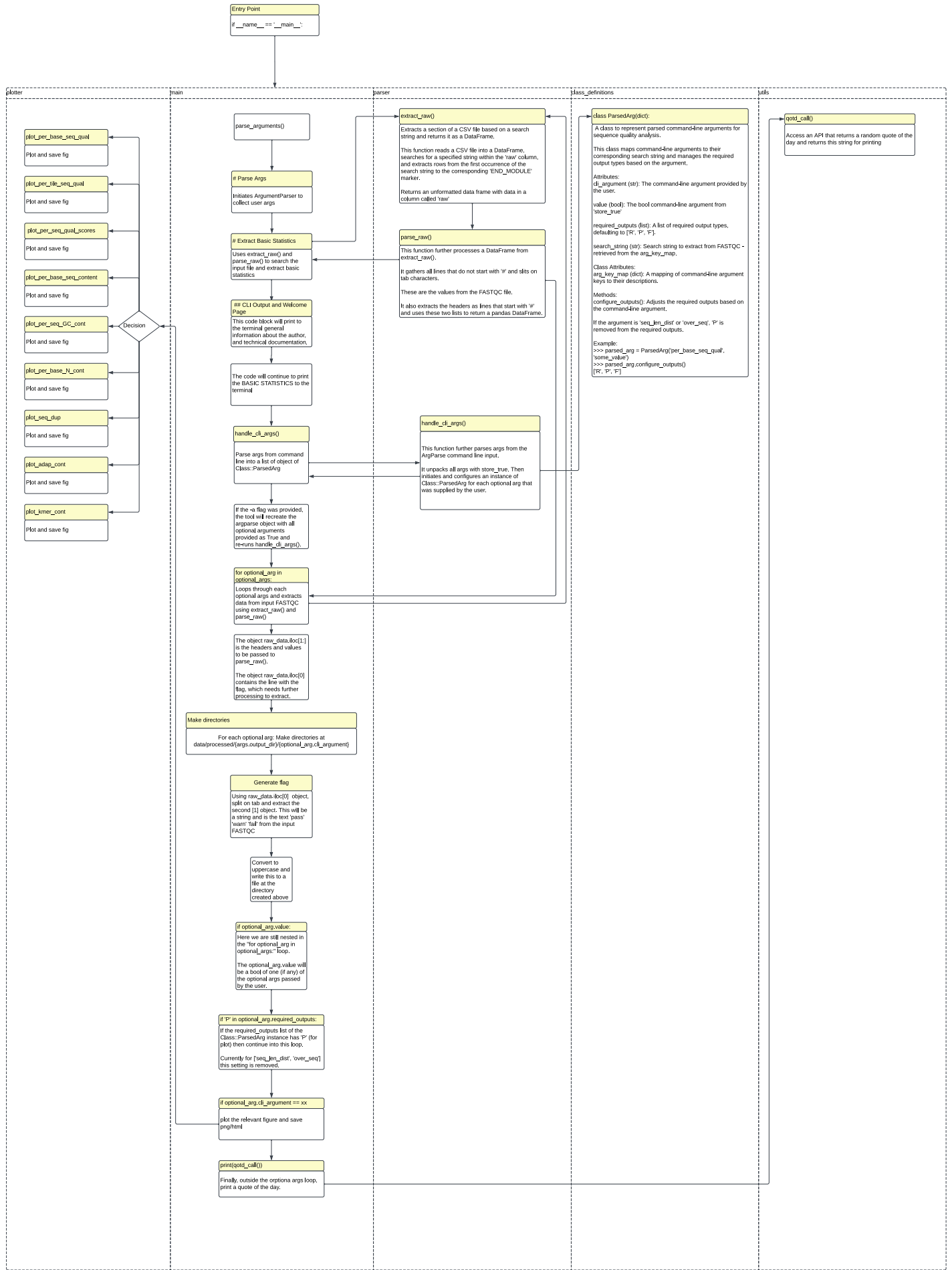
optional arguments:

-h, --help show this help message and exit
-b, --per_base_seq_qual Extract and plot per base sequence quality.
-t, --per_tile_seq_qual Extract and plot per tile sequence quality.
-s, --per_seq_qual_scores Extract and plot per sequence quality scores
-c, --per_base_seq_content Extract and plot per base sequence content
-g, --per_seq_GC_cont Extract and plot per sequence GC content
-n, --per_base_N_cont Extract and plot per base N content
-l, --seq_len_dist Extract sequence length distribution
-d, --seq_dup Extract and plot sequence duplication levels
-o, --over_seq Extract overrepresented sequences
-p, --adap_cont Extract and plot adapter content
-k, --kmer_cont Extract and plot K-mer Content
-a, --all Extract and plot all metrics

Optional Args

Help and misc information provided by ArgeParse for optional arguments.

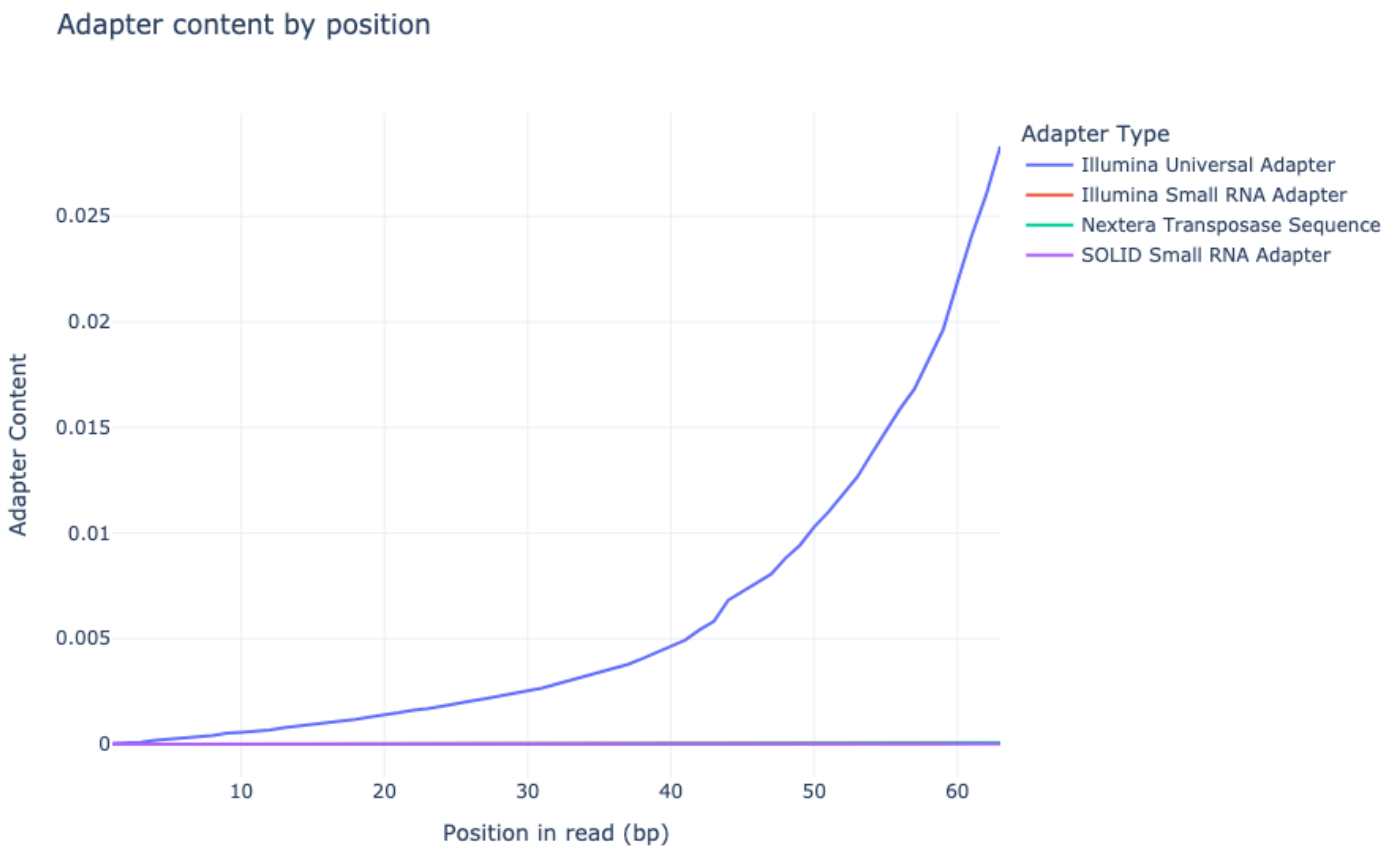
```
python3 src/main.py -h
```



Example Plots

Adapter Content

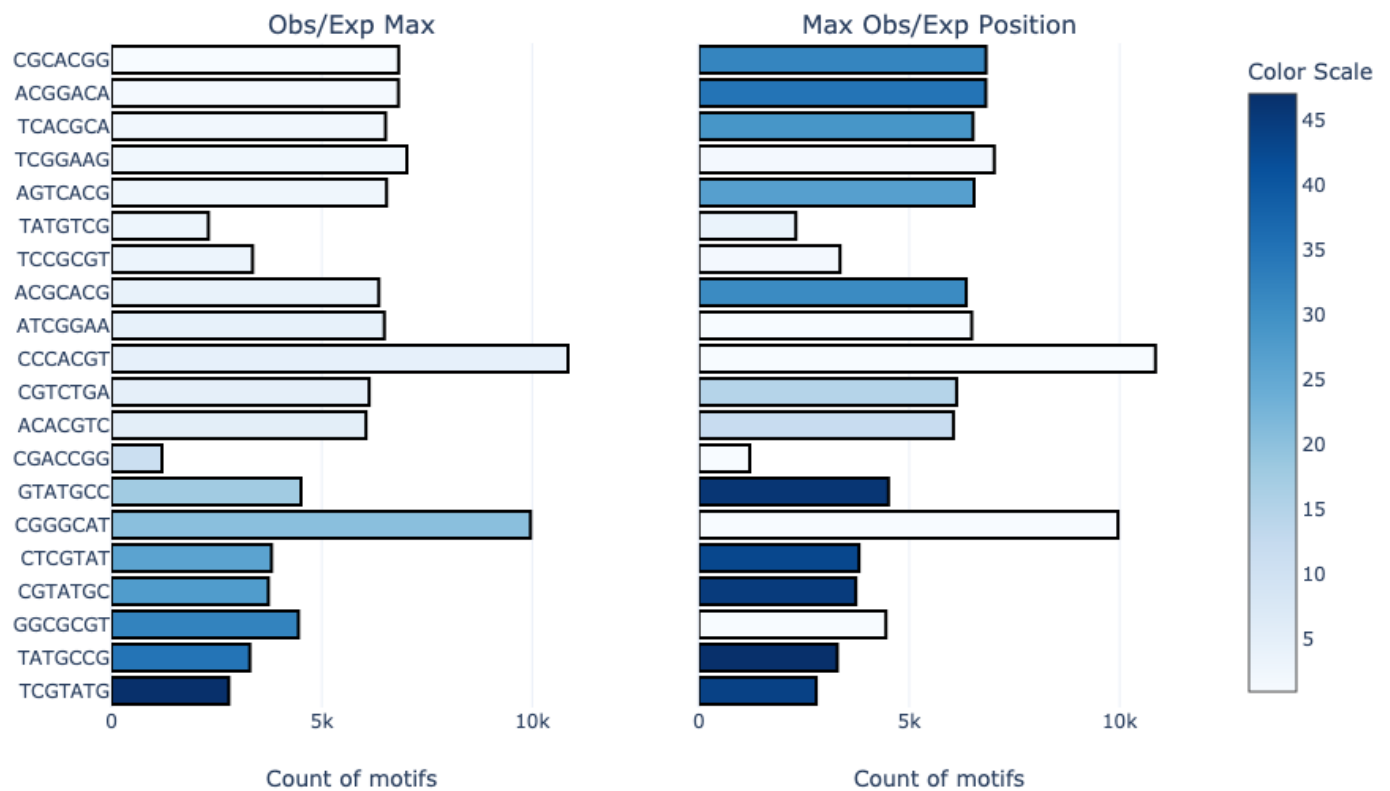
Plot's adapter content by position.



Kmer Content

Plot's kmer content by position.

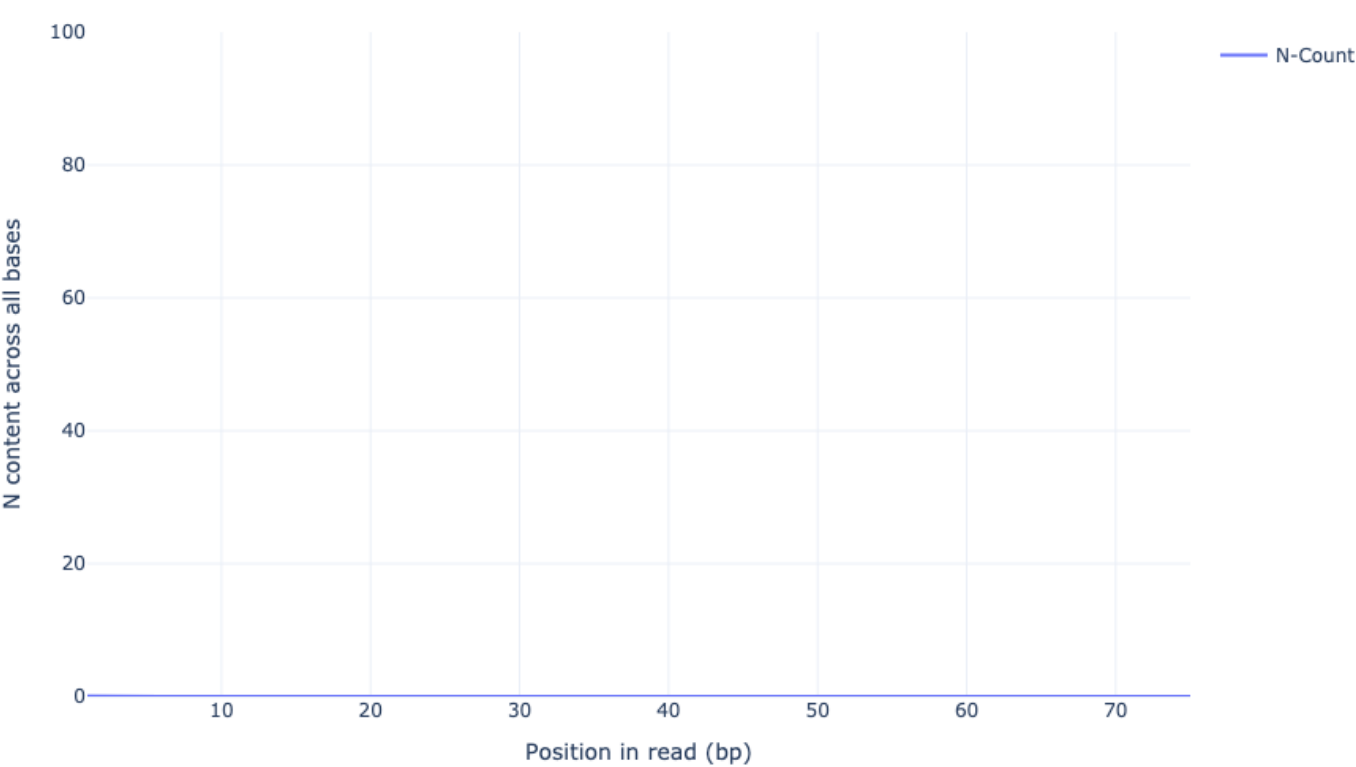
Sequence Counts



Overrepresented sequences

Plot's Per base N content.

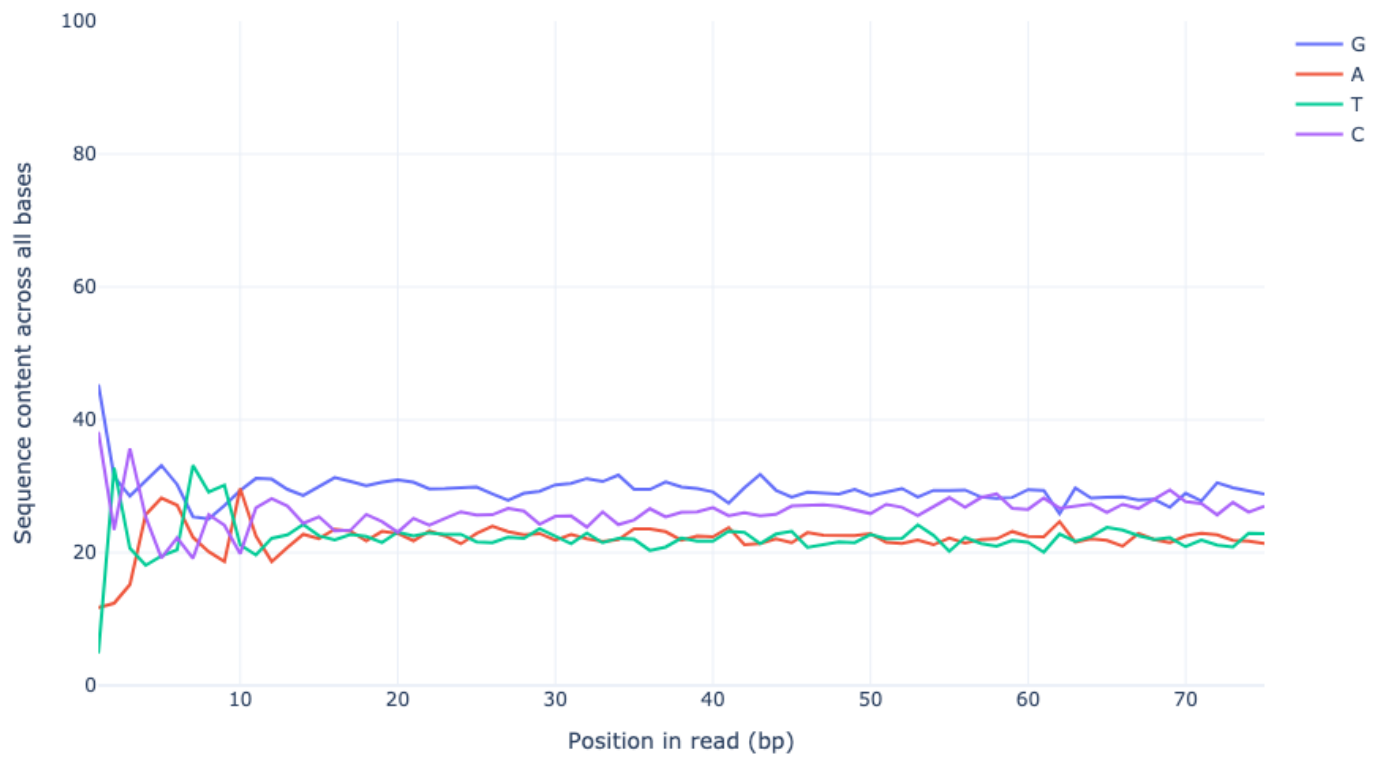
Adapter content by position



Per base sequence content

Plot's Per base sequence content.

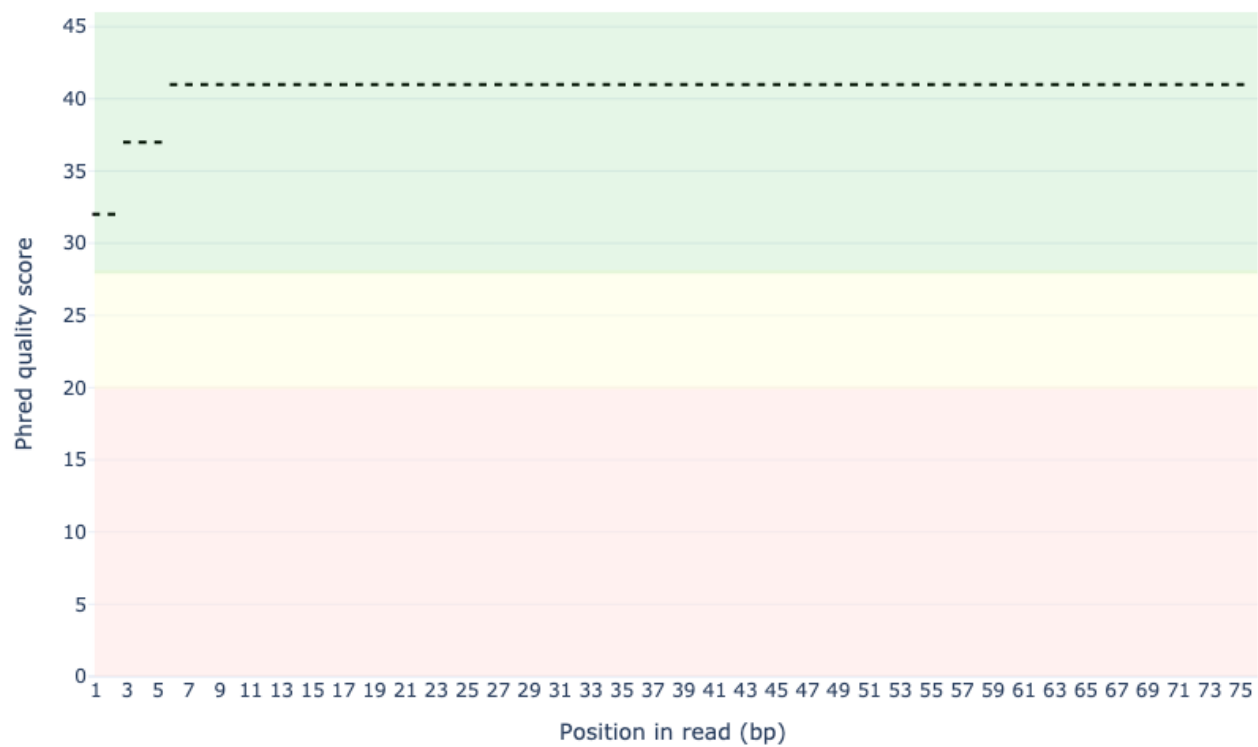
Per base sequence content



Per sequence quality scores

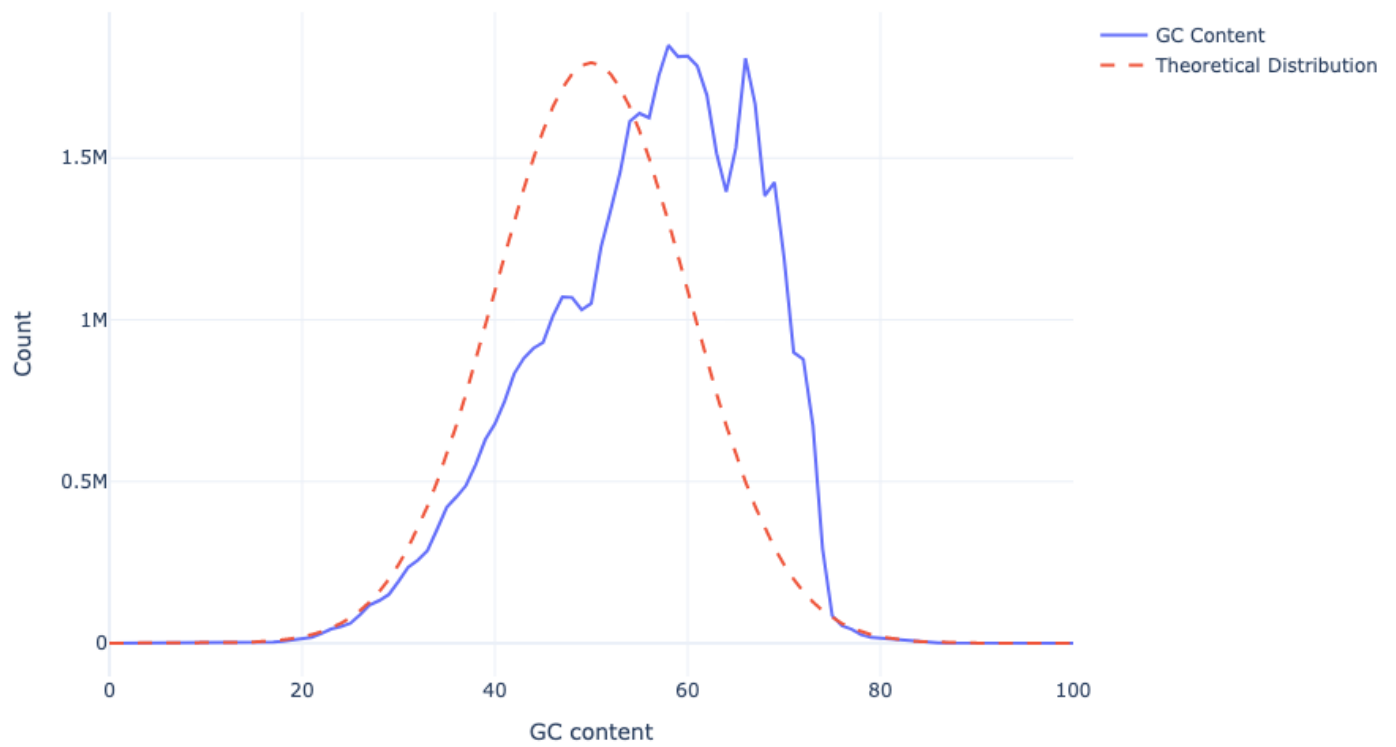
Plot's Per sequence quality scores.

Per base sequence quality



Per sequence GC content
Plot's Per sequence GC content.

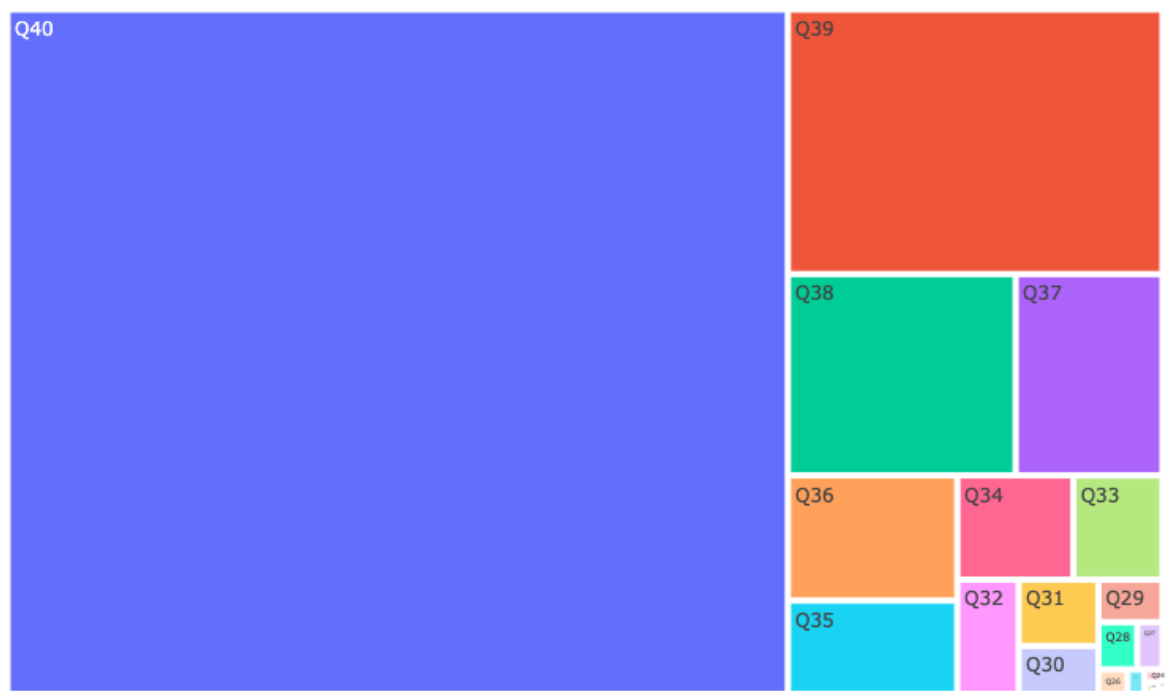
Per sequence GC content



Per sequence quality scores

Plot's Per sequence quality scores.

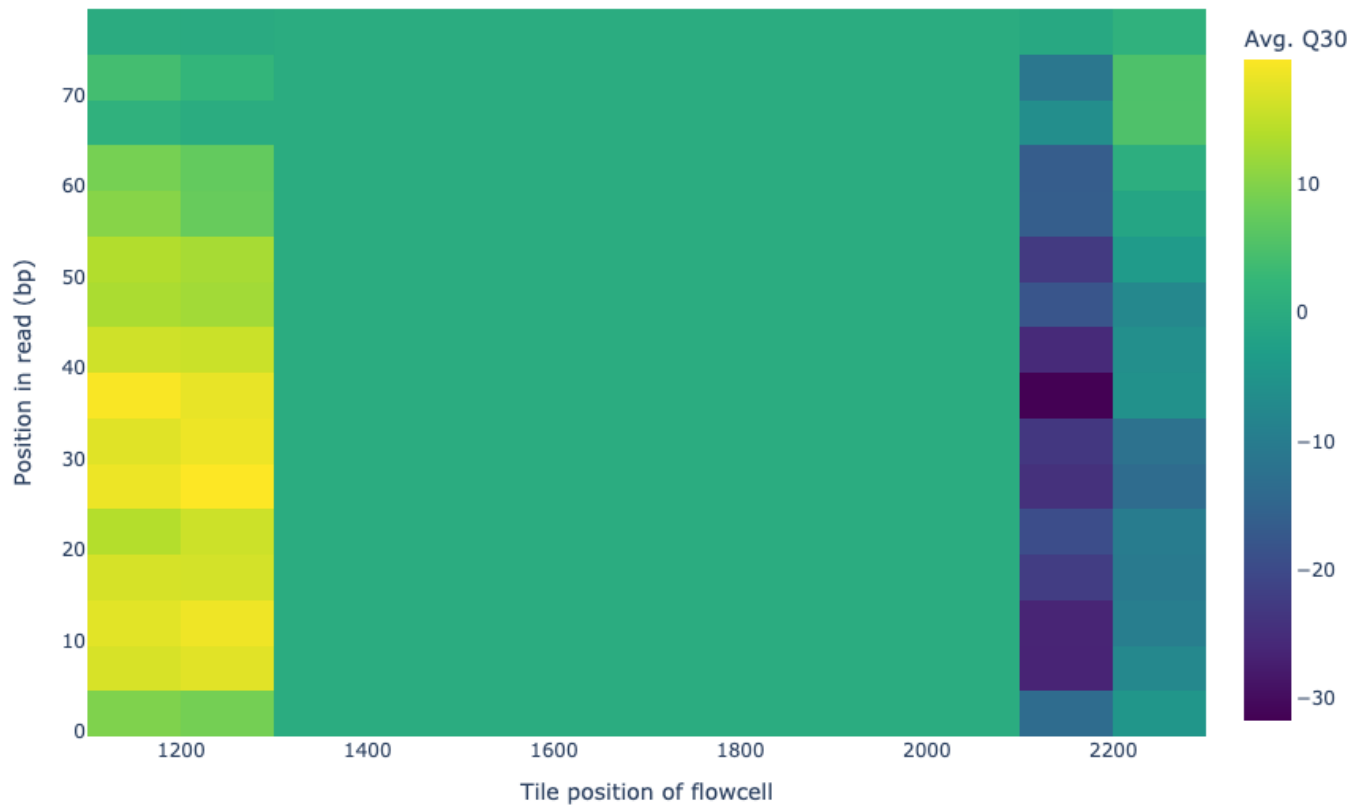
Count of Phred score



Per tile sequence quality

Plot's Per tile sequence quality.

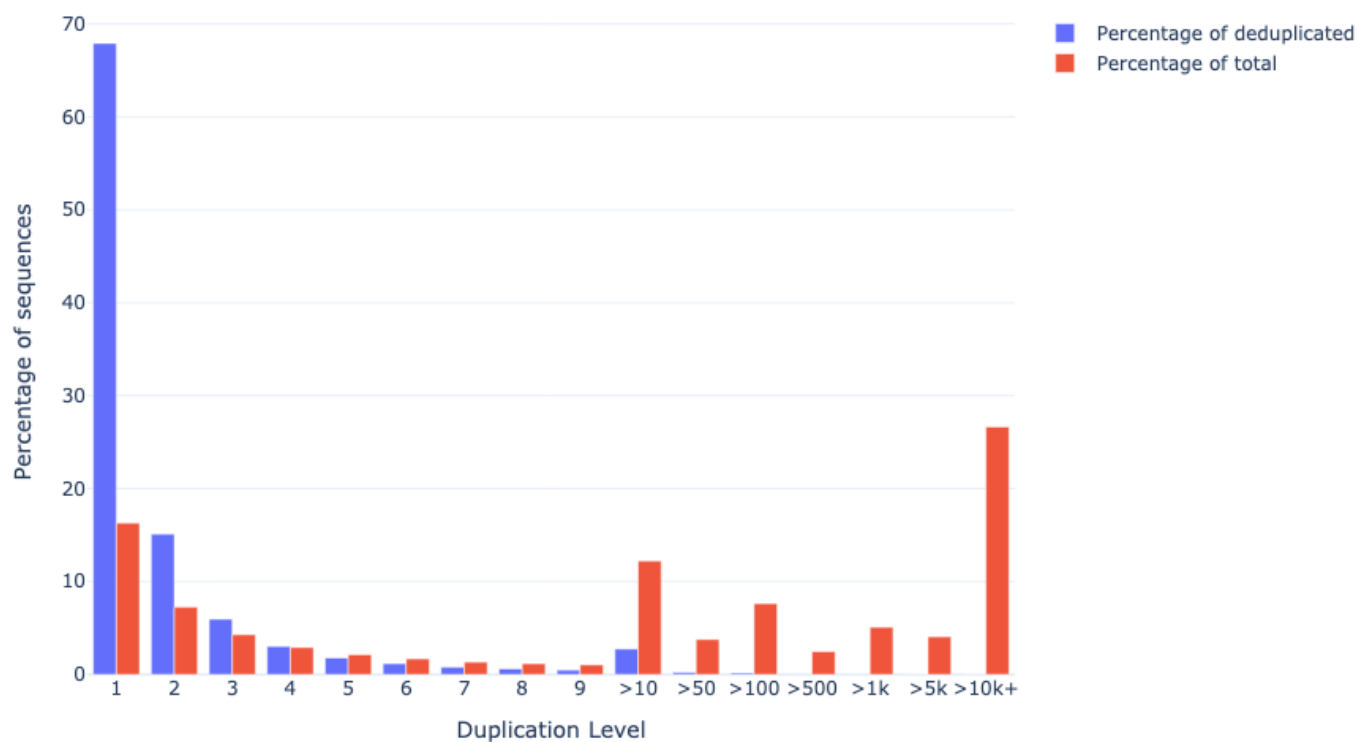
Per aggregated tile sequence quality



Sequence Duplication Levels

Plot Sequence Duplication Levels.

Sequence Duplication Levels



GitHub

<https://github.com/ms2206/FastQCParser.git>

Documentation built with [MkDocs](#).