QC Tutorial

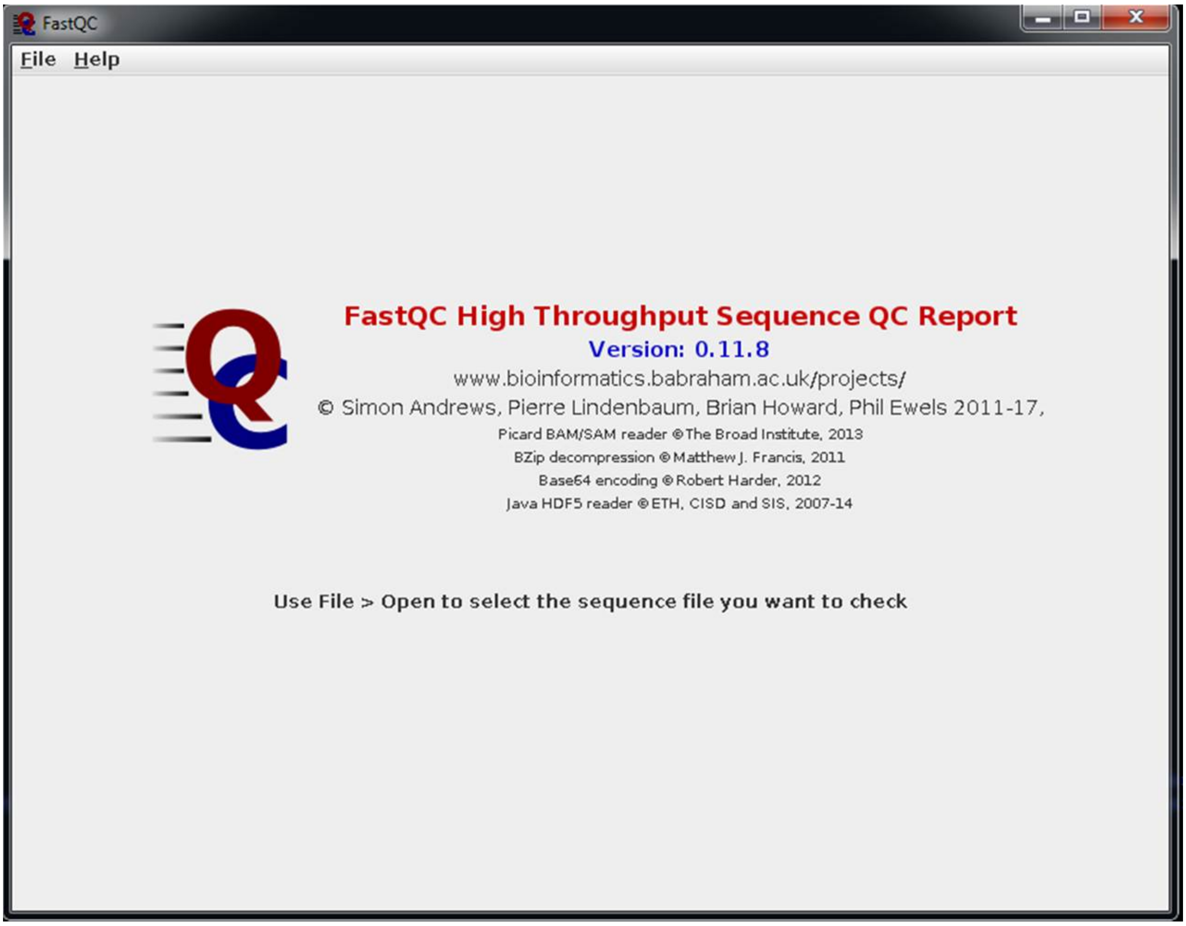
For this tutorial we will work in the folder called **qc_tutorial**

```
$ cd
$ mkdir qc_tutorial
$ cp ~/Course_Data/{PG15-BW001432.R1.fastq.gz,PG15-BW001432.R2.fastq.gz}
~/qc_tutorial
$ cd qc_tutorial
$ ls -lh
```

FASTQ Metrics

Opening FastQC

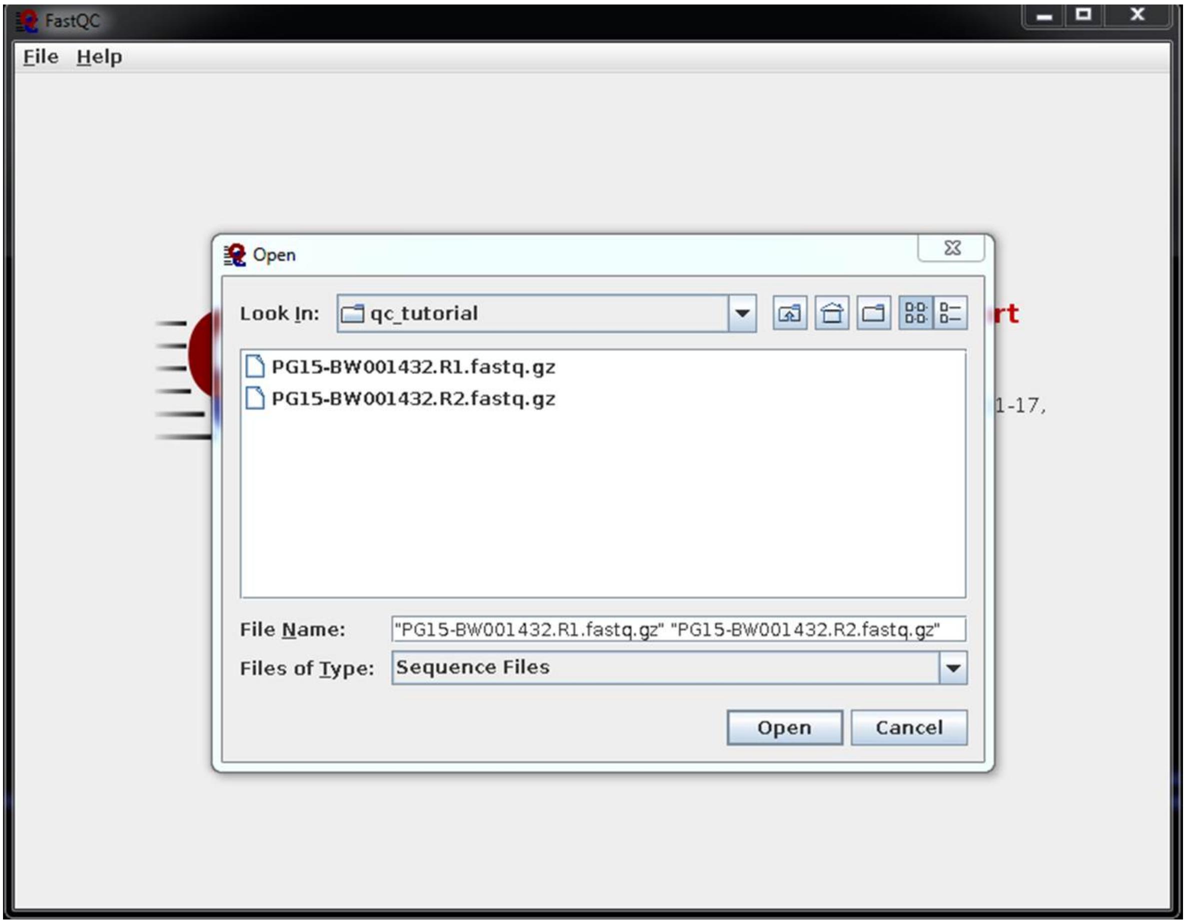
```
$ ~/software/fastqc_v0.11.8/FastQC/fastqc
or
$ fastqc
```



FastQC High Throughput Sequence QC Report
Version: 0.11.8
www.bioinformatics.babraham.ac.uk/projects/
© Simon Andrews, Pierre Lindenbaum, Brian Howard, Phil Ewels 2011-17,
Picard BAM/SAM reader ©The Broad Institute, 2013
BZip decompression ©Matthew J. Francis, 2011
Base64 encoding ©Robert Harder, 2012
Java HDF5 reader ©ETH, CISD and SIS, 2007-14

Use File > Open to select the sequence file you want to check

Open both the read files in **qc_tutorial**

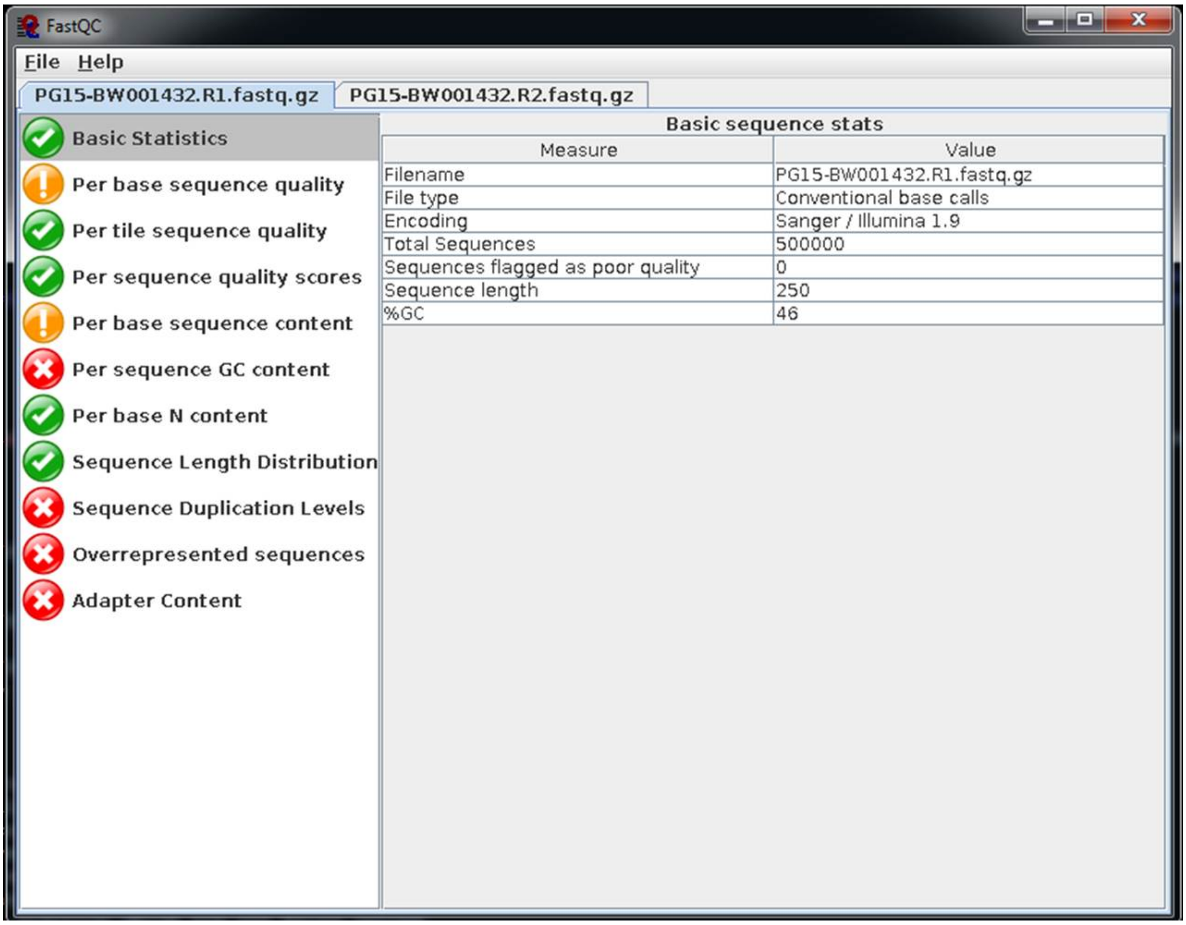


Look In: qc_tutorial

PG15-BW001432.R1.fastq.gz
PG15-BW001432.R2.fastq.gz

File Name: "PG15-BW001432.R1.fastq.gz" "PG15-BW001432.R2.fastq.gz"

Files of Type: Sequence Files

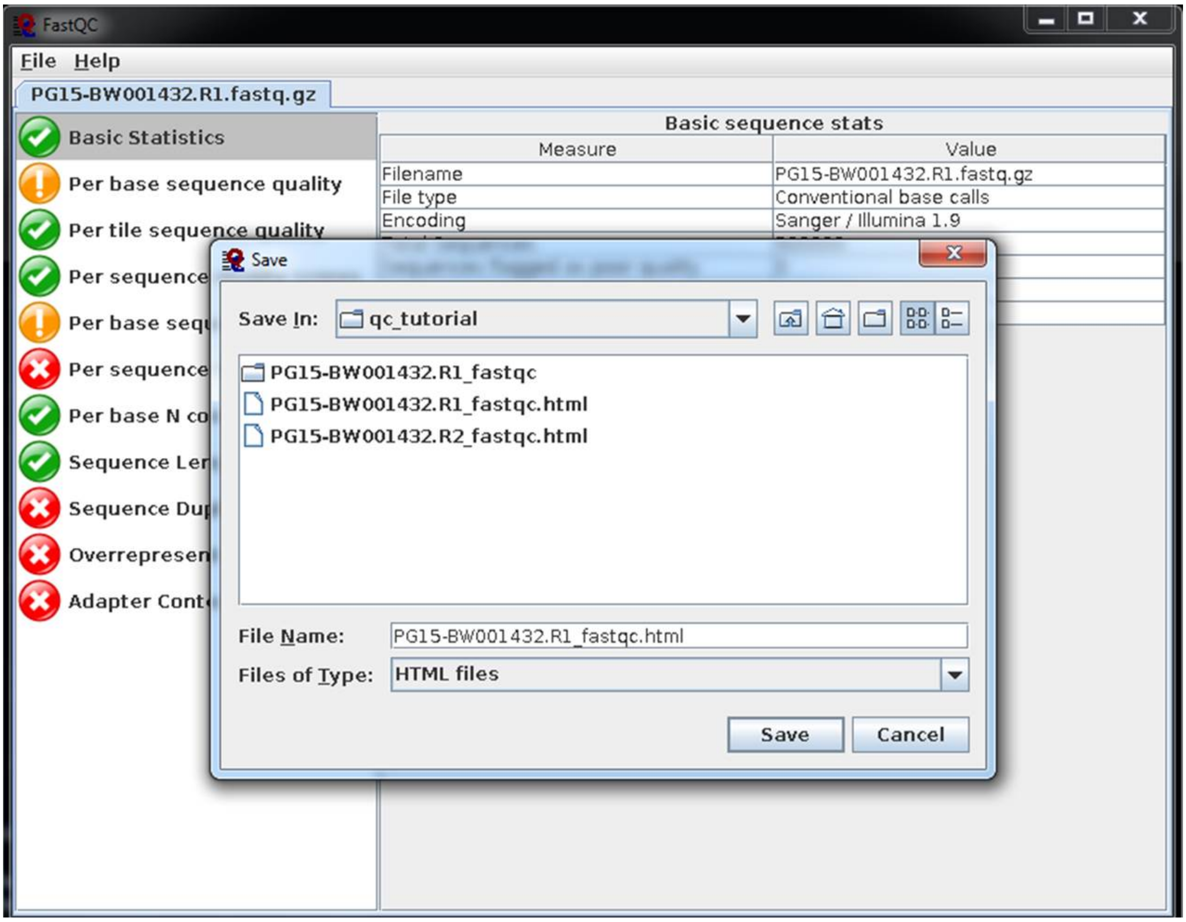


- ✓ Basic Statistics
- ! Per base sequence quality
- ✓ Per tile sequence quality
- ✓ Per sequence quality scores
- ! Per base sequence content
- ✗ Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✗ Adapter Content

Basic sequence stats

Measure	Value
Filename	PG15-BW001432.R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	500000
Sequences flagged as poor quality	0
Sequence length	250
%GC	46

Save the report in HTML format



Basic sequence stats

Measure	Value
Filename	PG15-BW001432.R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9

Save In: qc_tutorial

PG15-BW001432.R1_fastqc
PG15-BW001432.R1_fastqc.html
PG15-BW001432.R2_fastqc.html

File Name: PG15-BW001432.R1_fastqc.html

Files of Type: HTML files

FastQC without GUI

FastQC can be invoked from the command line to directly run on a batch of read files in any folder

```
$ mkdir output
$ fastqc -h
$ fastqc *.gz -o output/ --extract
```

-o : output directory

--extract : Unzips the files after creation

All the output files are now in **~/qc_tutorial/output/**

```
$ cd output
$ ls -lh
$ lynx PG15-BW001432.R1_fastqc.html
$ cd PG15-BW001432.R1_fastqc/Images
$ ls -lh
$ eog adapter_content.png
$ cd ~/qc_tutorial
```

Trimming Reads

Running Trim Galore

```
$ ~/software/TrimGalore-0.6.5/trim_galore --help
or
$ trim_galore --help
$ cd ~/qc_tutorial
$ trim_galore
```

Trimming adapters and Low quality bases

```
$ mkdir output_q30
$ trim_galore *.gz -q 30 --phred33 -o output_q30/ --illumina --max_n 2 --paired
$ cd output_q30
$ nano PG15-BW001432.R1.fastq.gz_trimming_report.txt
```

-q : Quality score to use

-o : Output folder

--phred33 : ASCII code used for quality scores

--illumina : Adapters to be trimmed (can specify using **-a** and -**a2**)

--max_n : Total no of 'N' in a read before it is removed

--paired : Reads are paired

- How many Reads contain adapters?
- How many reads passed?
- How many bases were quality trimmed?
- Open the trimmed files in FastQC (*val_1.fq.gz and *val_2.fq.gz)

Increasing and decreasing the stringency of Trimming

```
$ mkdir output_q40
$ trim_galore *.gz -q 40 --phred33 -o output_q40/ --illumina --max_n 2 --paired
$ cd output_q40
$ nano PG15-BW001432.R1.fastq.gz_trimming_report.txt
$ ls -lh
```

- How many Reads contain adapters?
- How many reads passed?
- How many bases were quality trimmed?

```
$ mkdir output_q10
$ trim_galore *.gz -q 10 --phred33 -o output_q10/ --illumina --max_n 2 --paired
$ cd output_q10
$ nano PG15-BW001432.R1.fastq.gz_trimming_report.txt
```

- How many Reads contain adapters?
- How many reads passed?
- How many bases were quality trimmed?
- Open the trimmed files in FastQC

Using wrong adapters for trimming

```
$ mkdir wrong_adapter
$ trim_galore *.gz -q 30 --phred33 -o wrong_adapter/ --nextera --max_n 2 --paired
$ cd wrong_adapter
$ nano PG15-BW001432.R1.fastq.gz_trimming_report.txt
```

- How many Reads contain adapters?
- Open the trimmed files in FastQC to check adapter content