SEQUENCHER®

Tutorial for Windows and Macintosh

FastQ Quality Control Reports

Gene Codes Corporation

TCA **GENE**
AGT **CODES**

FastQ Quality Control Reports

FastQ Quality Control Reports

Before running any alignment or assembly, it is a good idea to check the quality of your data. We have added the FastQC program[1] so that you can now do this from within **Sequencher**. The analysis tests a number of different metrics and produces a consolidated report that you can view at the end of the analysis run or at a later time. Results are saved to a **FastQC Reports** folder within your **Documents** folder.

The FastQC report uses a 'traffic lights' system to indicate whether your results are good ✅ or bad ❌, there is also an intermediate warning ⚠ state.

GETTING STARTED

You will need to have installed FastQC using the FastQC installer that can be downloaded from the Free Download page on the Gene Codes website. FastQC requires Java so you may need to install that as well. Running the analysis will display the following dialog if FastQC is not installed. A similar dialog will display if Java is not installed.
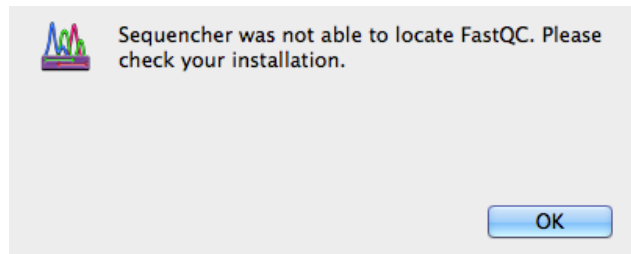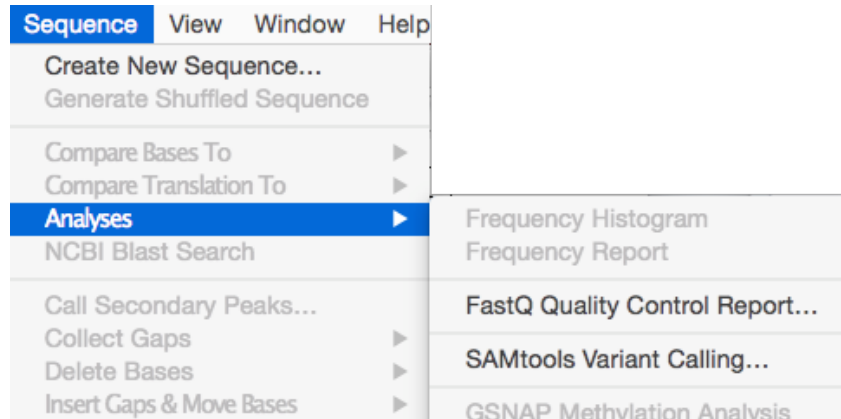

RUNNING A FASTQC REPORT

You can use FastQC on Sanger or NGS reads files as long as they are in FastQ format.

- Launch **Sequencher**.
- Choose the **Sequence > Analyses > FastQ Quality Control Report...** menu item.

---

[1] www.bioinformatics.babraham.ac.uk/projects/ © Simon Andrews, Pierre Lindenbaum, Brian Howard, Phil Ewels 2010-1015

- Use the **Add File** button to browse to the FastQ files you want to analyse. You will see the files listed in the FastQ reads files list. Note that there are no Advanced Options for this type of analysis.
- If you don't have any FastQ files of your own, then navigate to **Sequencher's Sample Data** folder and locate the **NGS Data** folder. You will find two FastQ reads files there.
- If you want to see the results immediately after analysis, then ensure that the **View Results, Yes** radio button is selected.
- Click on the **Analyze** button.

Note that if you have analysed large FastQ files, such as those from an RNA-seq run, or many FastQ files, then it may take some time for the analysis to complete. Bear this in mind when planning your analyses.

Once the analysis has completed, your FastQ reports will open automatically. Each report will open in its own window so that you may review reports side by side if you wish. If you have analysed greater than 4 files, the **Report Selection** dialog will appear.

- Select the report(s) you want to open. If you want to open more than one file to review, then use **Ctrl/Cmd+click** or **Shift+click** and click on the **Open** button.

VIEWING OLDER REPORTS

Sometimes you may wish to review an older report at a later time.

- Choose the **View > Display FastQC Report…** menu item.
- Choose the reports you want to review by selecting them. Use **Ctrl/Cmd+click** or **Shift+click** if you want to review more than one file at a time.

Once you are looking at the report, you will see that it is comprised of 11 or more separate analyses. Since the report is in HTML format, you can jump down the report by clicking on one of the links on the left-hand side of the report under the Summary text.

FastQC Report

Summary

- Basic Statistics
- Per base 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
- Kmer Content

Basic Statistics

Measure	Value
Filename	read2.fq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	75000
Sequences flagged as poor quality	0
Sequence length	35
%GC	33

Per base sequence quality

Quality scores across all bases (Sanger / Illumina 1.9

The report starts with some basic statistics, then follows with more detailed analyses. While a single result may give the impression that your results are poor, it is wise to take into account the results from all of the analyses before deciding whether to proceed with further analysis.

Lets look at some of the analyses.

- Click on the '**Per base sequence quality'** link.

Look at the distribution of the bars in this plot. The x-axis represents the per-base confidence scores and the y-axis represents the position in the read. Notice how the longer the read gets, the lower the quality of the base.

✅ Per base sequence quality

Quality scores across all bases (Sanger / Illumina 1.9 encoding)
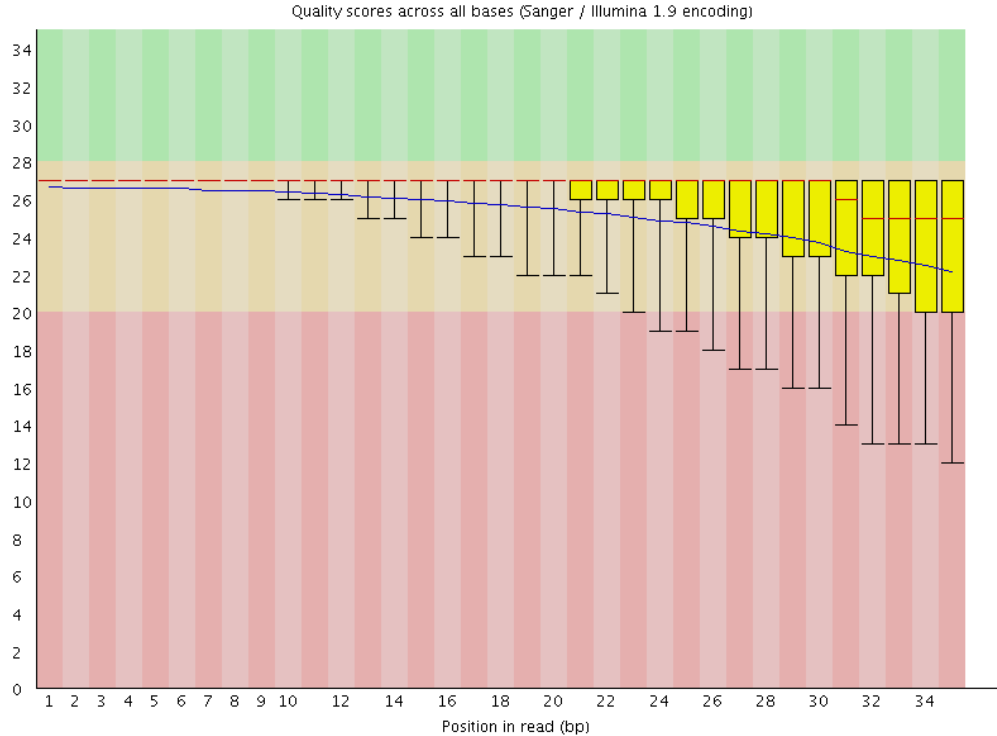
In this example, the overall quality of the reads is good enough that trimming probably wouldn't be required unless you had very stringent requirements. The blue line indicates the mean quality, the red lines indicate the median value.

CONCLUSION

In this tutorial, you have learned how to run a quality control report on your FastQ data using FastQC. You have also learned how to review reports prepared at an earlier time.

For detailed information on individual modules in FastQC, refer to the online help at
http://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/