NORWEGIAN SEQUENCING CENTRE

Quality Control Report

March 12, 2015

This report aims to give a brief overview of the quality of the sequence produced from your sample and should be self-explanatory. However, you may want to refer to our website (http://www.sequencing.uio.no/services/data-delivery) for further details on sequence and quality control files.

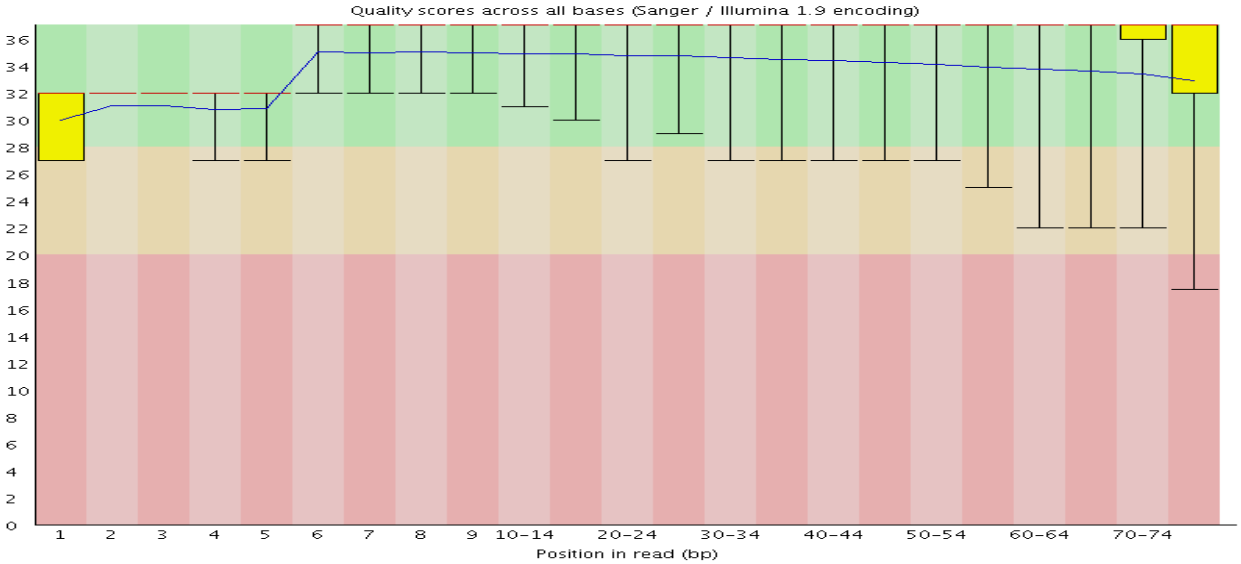
The graphs in the report were generated using fastqc (http://www.bioinformatics.bbsrc.ac.uk/). Fastqc generates many different reports and only a subset of them are included here. If you wish to see the full set, please download the software and run it on your sequence files.

1 Run summary

Measure	Value
Run name	150310_NS500336_0020_AH537YBGXX
RTA Version	2.1.3
Sample name	Anmarkrud-gDNA1-20-67809_S19_L00X_R1_001
Read	1
Total number of sequences	20867816

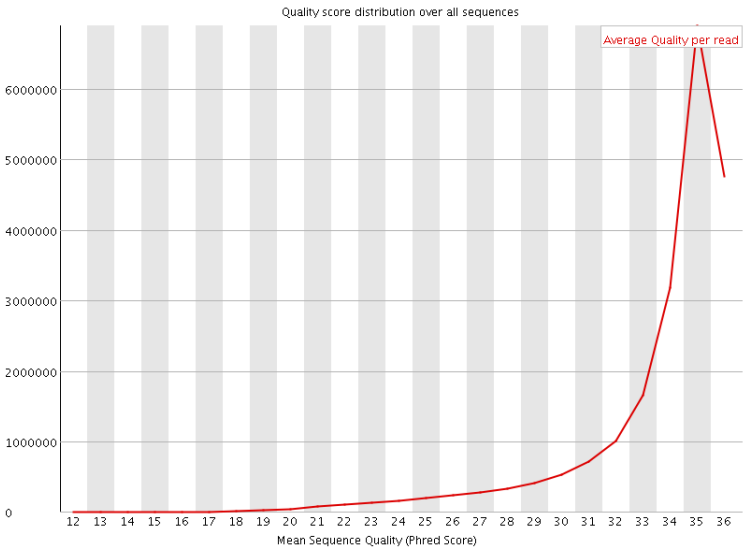
2 Per Base Sequence Quality

The graph shows an overview of the range of quality values across all bases at each position in the FastQ file. The central red line shows the median of the quality values and the blue line shows the mean of the quality values.



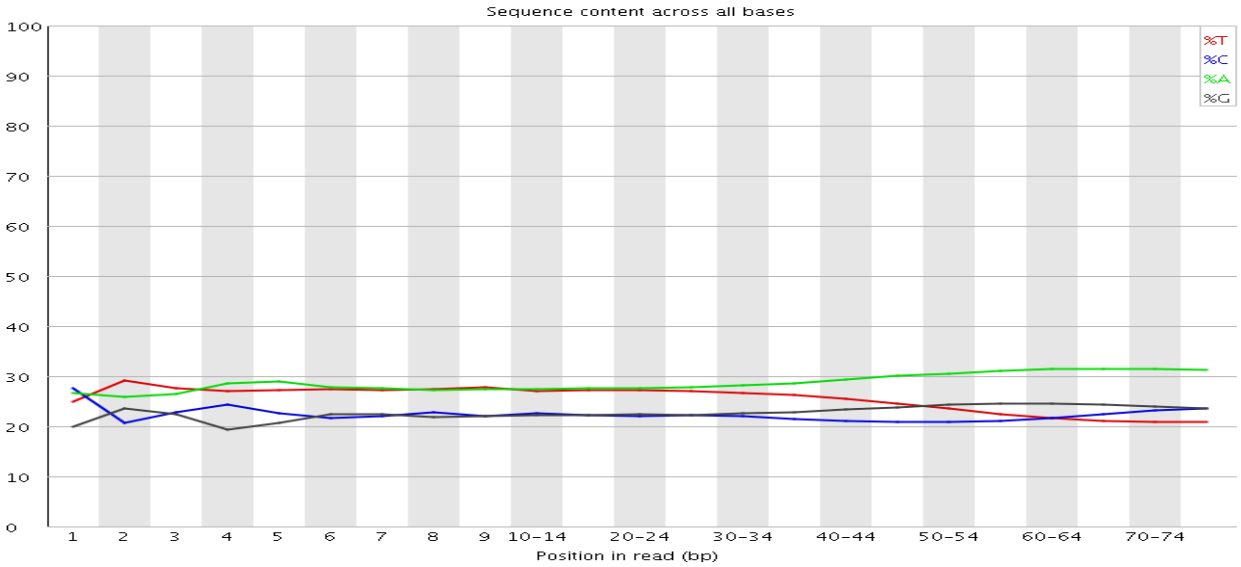
3 Per Sequence Quality Scores

The graph is generated by computing the average quality of a read (by averaging across read positions) and then plotting the distribution of this average quality. It thus enables you to tell whether low quality bases are located in a subset of the reads or distributed across all reads.



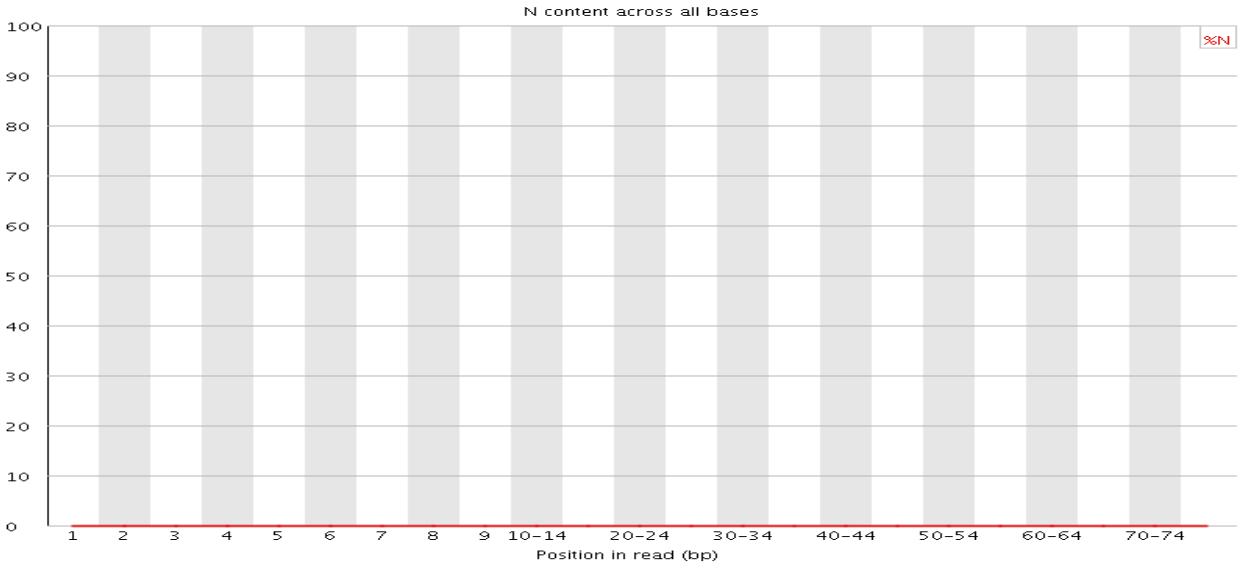
4 Per Base Sequence Content

The graph plots the proportion of each base type in each read position.



5 Per Base N Content

The graph shows the percentage of base calls at each position for which an N was called. If a sequencer is unable to make a base call with sufficient confidence then it will normally call an N rather than A, T, G or C.



6 Sequence Duplication levels

The graph shows the number of sequences with different degrees of duplication relative to the number of unique sequences (which is set to 100%). In a diverse library, most sequences will occur only once in the final set and the graph will show a peak in the unique category. However, some sequences may be present in more than one copy (for example, as the result of PCR amplification), in which the graph may show high numbers of sequences in the other categories (2 copies, 3 copies, etc). The last category is for 10 copies or more.