

SF-seq

Jake VanCampen

Wed Sep 27 15:08:08 2017

SF-seq read quality score distributions

In this assignment I worked with the following read files from `/projects/bgmp/2017_sequencing/demultiplexed`:

```
2_2B_control_S2_L008_R1_001.fastq.gz
2_2B_control_S2_L008_R2_001.fastq.gz
29_4E_fox_S21_L008_R1_001.fastq.gz
29_4E_fox_S21_L008_R2_001.fastq.gz
```

To determine quality score distributions for each read the program FastQC was loaded on talapas along with its dependencies: `ml easybuild intel/2017a FastQC`

In a working directory called SF-seq, a new directory was created for the output of FastQC. FastQC was then run using the following options on all four files.

```
$ fastqc --noextract -o fastqc_out -t 4 \
2_2B_control_S2_L008_R1_001.fastq.gz 2_2B_control_S2_L008_R2_001.fastq.gz \
29_4E_fox_S21_L008_R1_001.fastq.gz 29_4E_fox_S21_L008_R2_001.fastq.gz
```

Results

Quality score distributions for the forward and reverse reads as well as per-base N content is shown below:

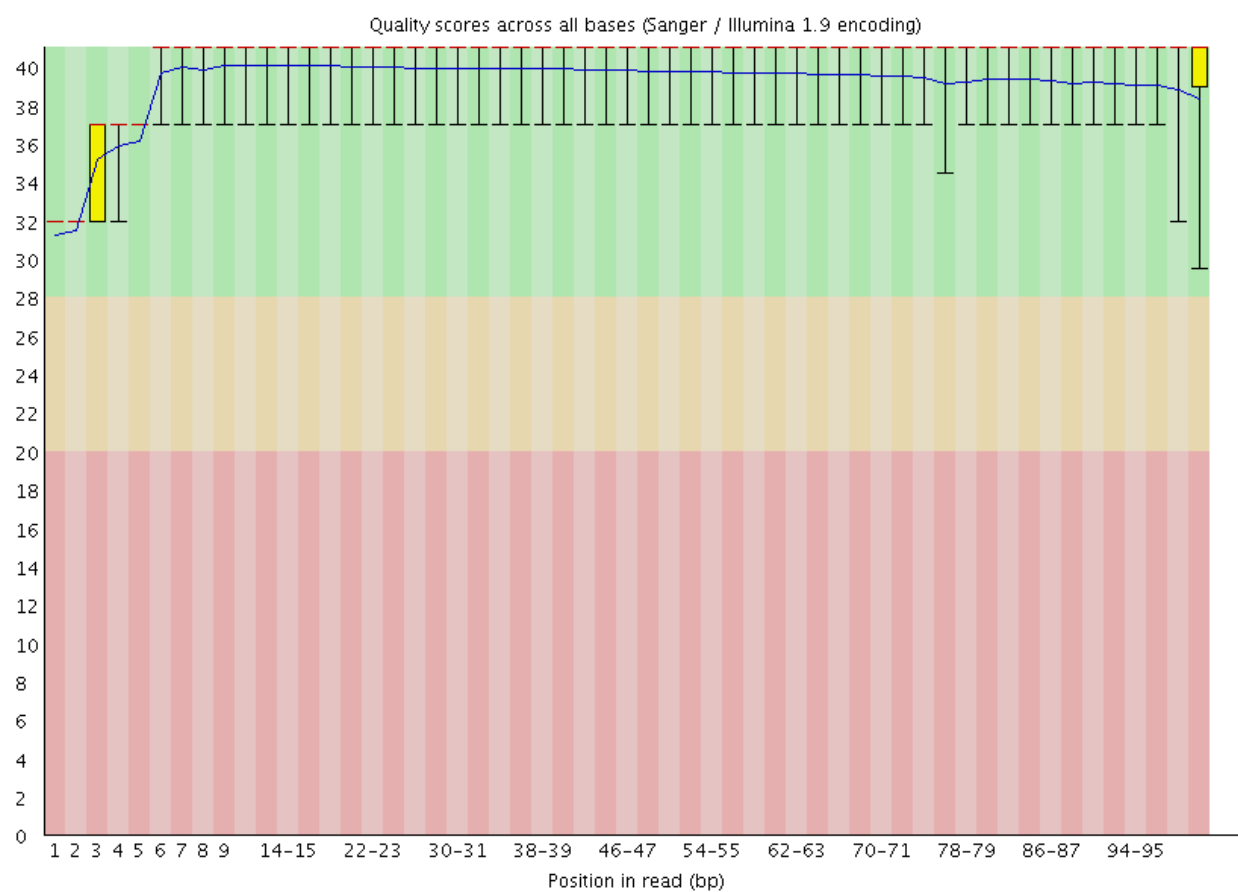


Figure 1: 29_4E_fox_S2 Read1 QScore distribution

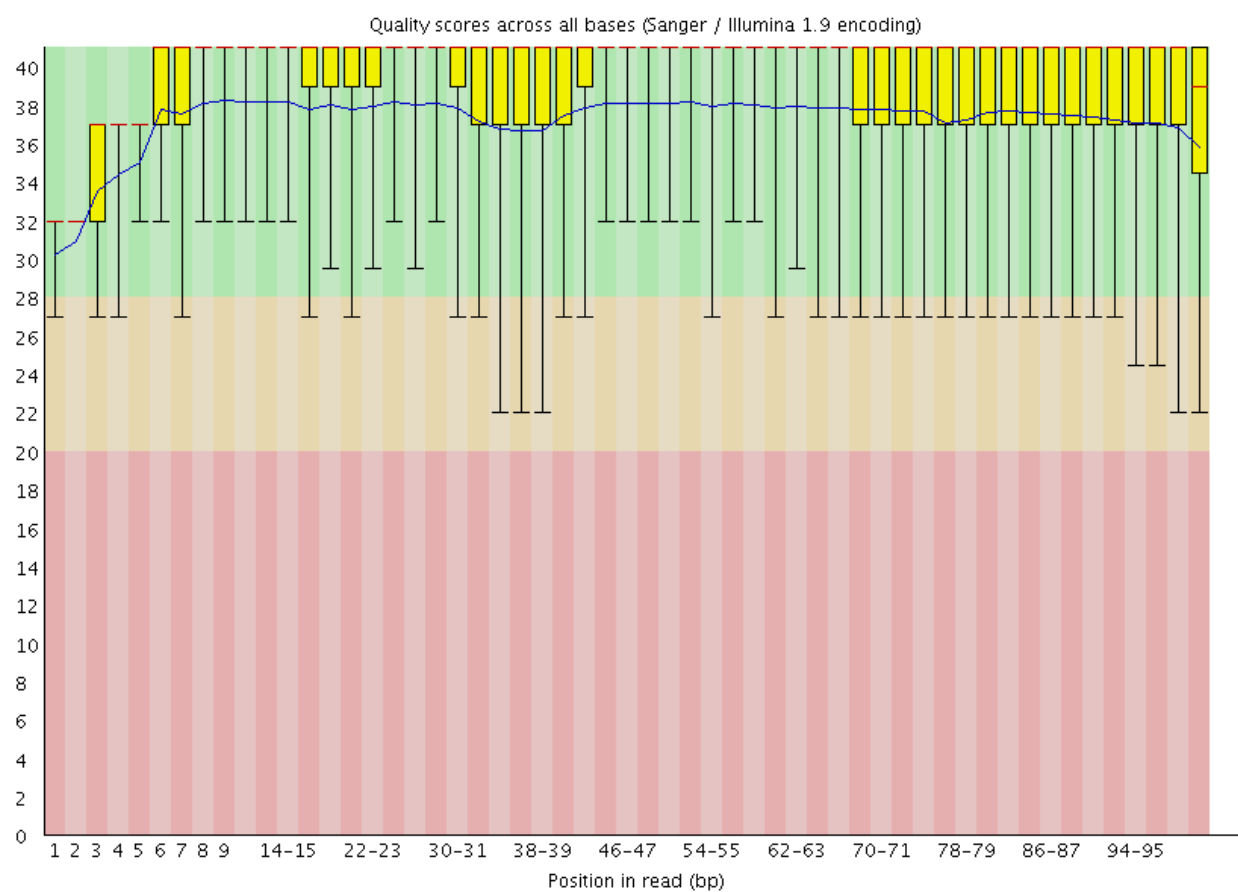


Figure 2: 29_4E_fox_S2 Read2 QScore distribution

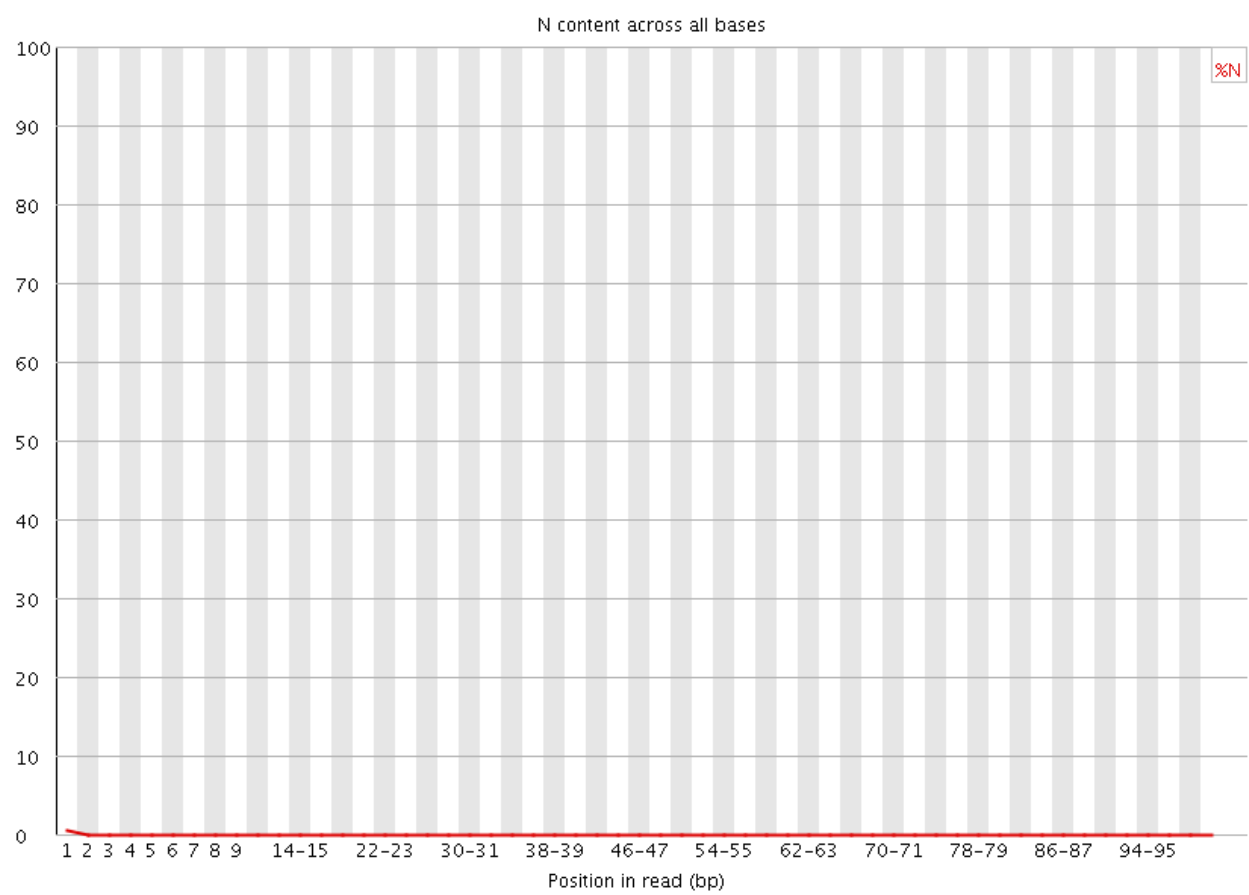


Figure 3: 29_4E_fox_S2 Per-base N-content

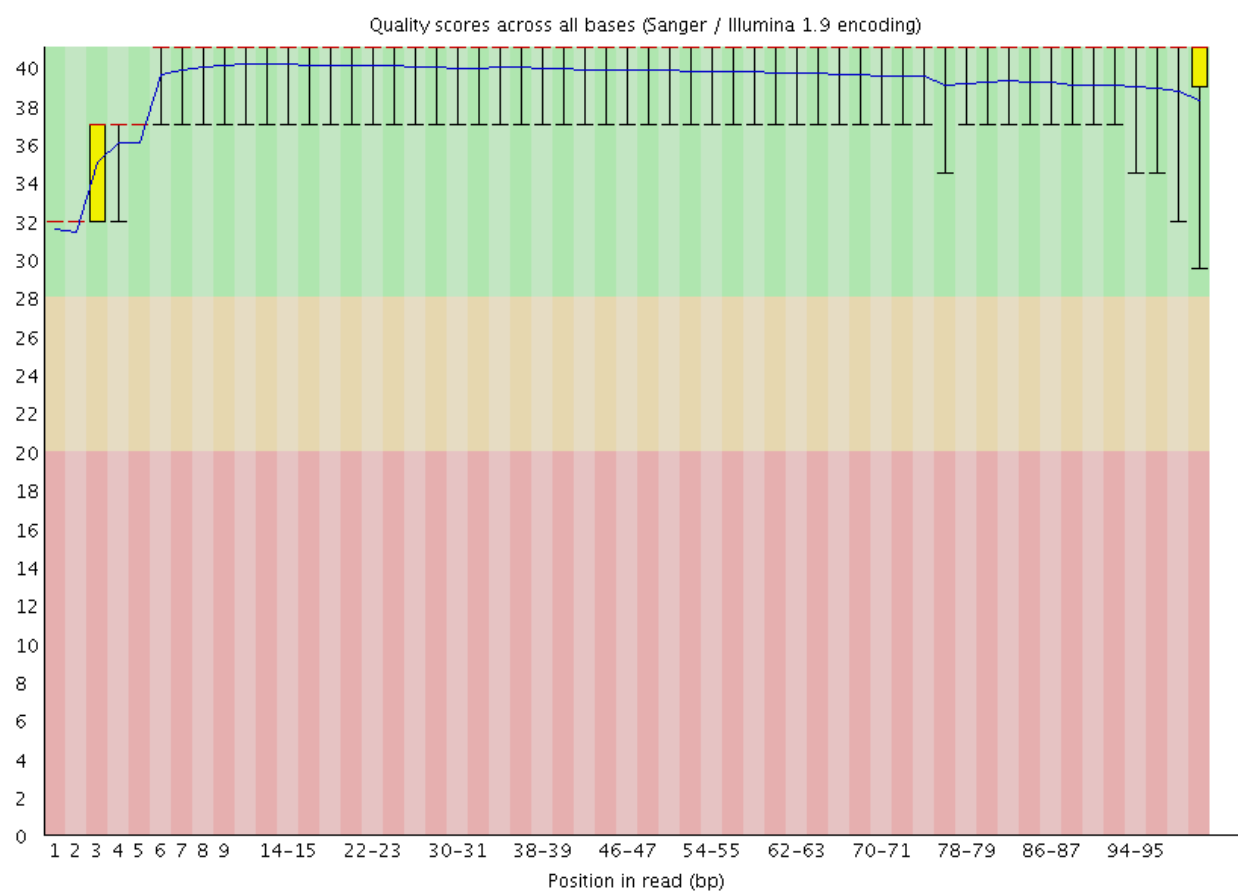


Figure 4: 2_2B_control_S2 Read1 QScore distribution

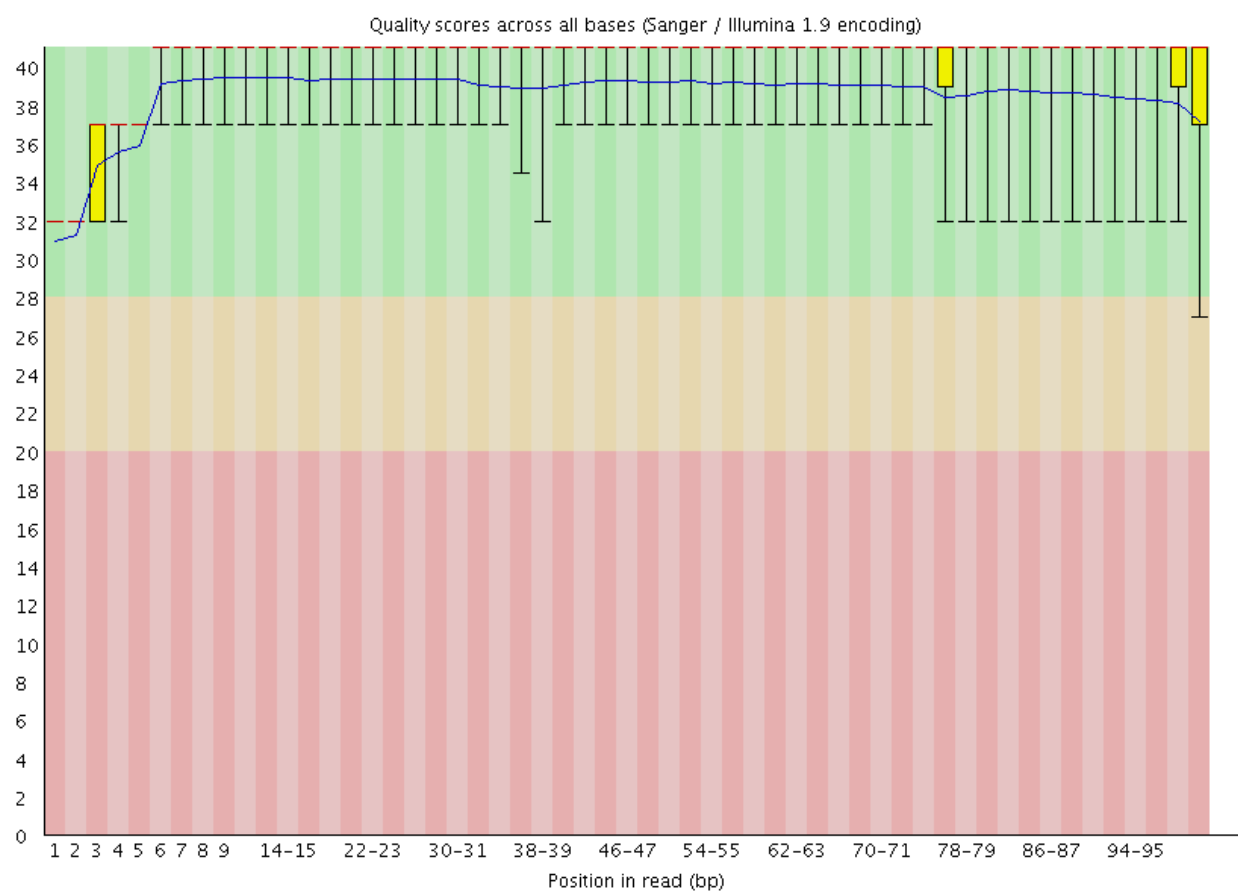


Figure 5: 2_2B_control_S2 Read1 QScore distribution

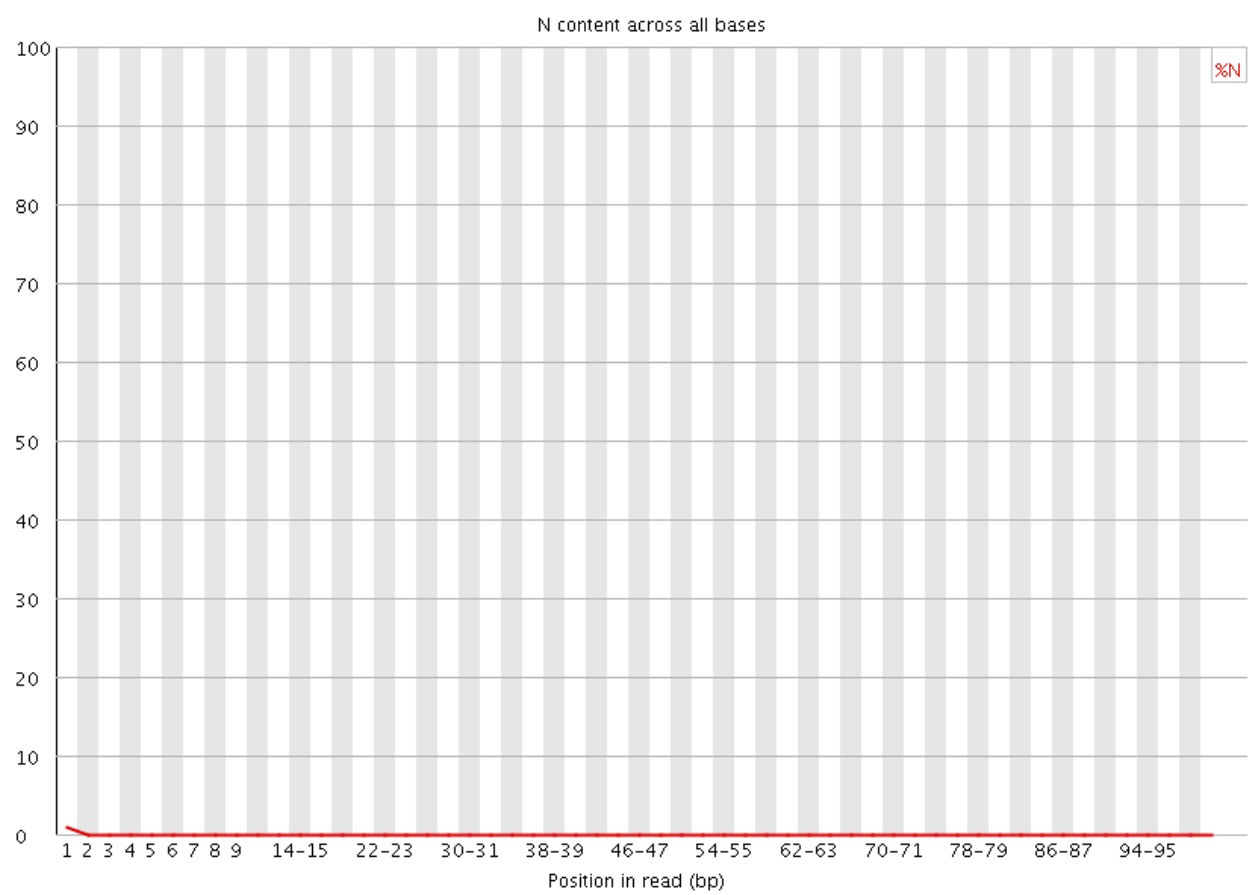


Figure 6: 2_2B_control_S2 Per-base N-content