::QTrim:BMCBioinformatics 2:FiguresAndTables:Figure1.pdf

Fig 2.1: Systematic work flow of QTrim

QTrim_sequential_steps.pdf

Fig 2.2: Sequential steps describing the trimming process in QTrim. A) Iterative trimming with mean across sequence. B) Iterative trimming with mean in window. C) Iterative trimming with mean in last nucleotide. At step C, the last nucleotide has a score of 25, which satisfies the user mean and QTrim therefore does no further trimming. The final read length is also greater than the user-specified read length. Therefore, the sequence is written to an output file.

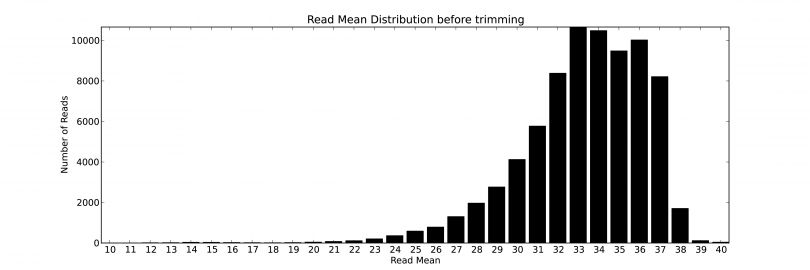


Fig 2.3: QTrim generated plot showing the distribution of sequence reads by specific mean scores.

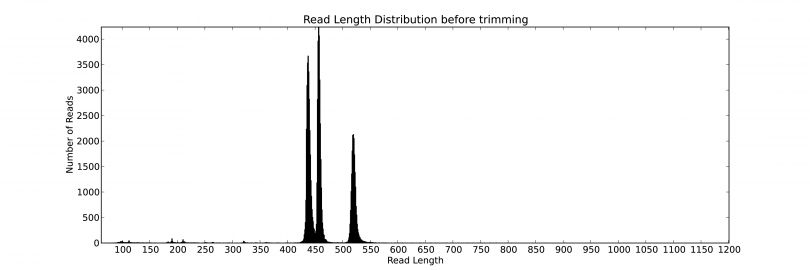


Fig 2.4: QTrim generated plot showing the distribution of sequence reads with read lengths.

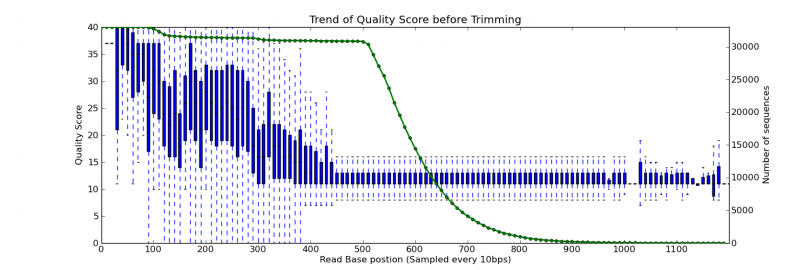


Fig 2.5: QTrim generated box plot showing the average quality score across the sequence reads and the number of sequences contributing to the calculation of the average quality score at every 10th base (Bin size: 10). The green line represents the total number of sequences (represented at the secondary Y axis at the right side) for evaluating the average quality score.

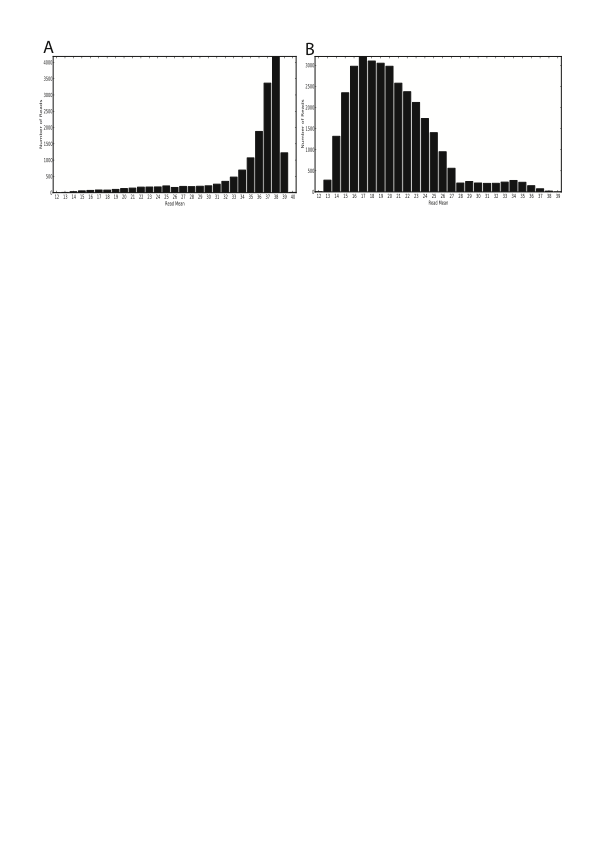


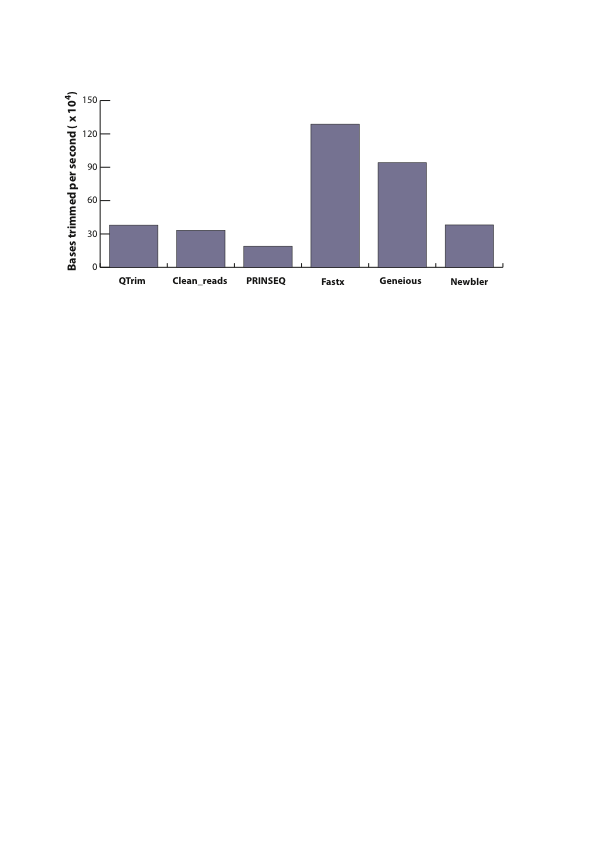
Figure 2.6: Distribution of test data sequence reads by mean quality A) Large number of sequences in test data A have higher mean quality score B) Large number of sequences in test data B have lower mean quality score



Figure 2.7: Distribution of test data sequence reads by read length. A) Large number of sequences in test data A have close range of read length from 450 to 550 and the plot very close distribution of sequences by read length. B) Sequence reads in test data B have wide range of read length and the plot shows wide distribution of sequences by read length

:Figures_for_thesis:qtrim_results.pdf

Fig 2.8: Comparison of QTrim with other methods taking into account the total number of produced reads and mean read length of produced reads using two datasets. The good quality data is presented in A and C, and the poor quality data in B and D. The top panel (A and B) display the results at a mean threshold quality score of 20, whereas the bottom panel shows the results at a mean threshold of 30.

 Figure 2.9: Speed comparison of all the tools with the number of bases trimmed per second. Geneious and Newbler v2.6 are graphical tools and their time of execution is done manually with stopwatch. Other tools are command line and the time is obtained using command line “time” function.

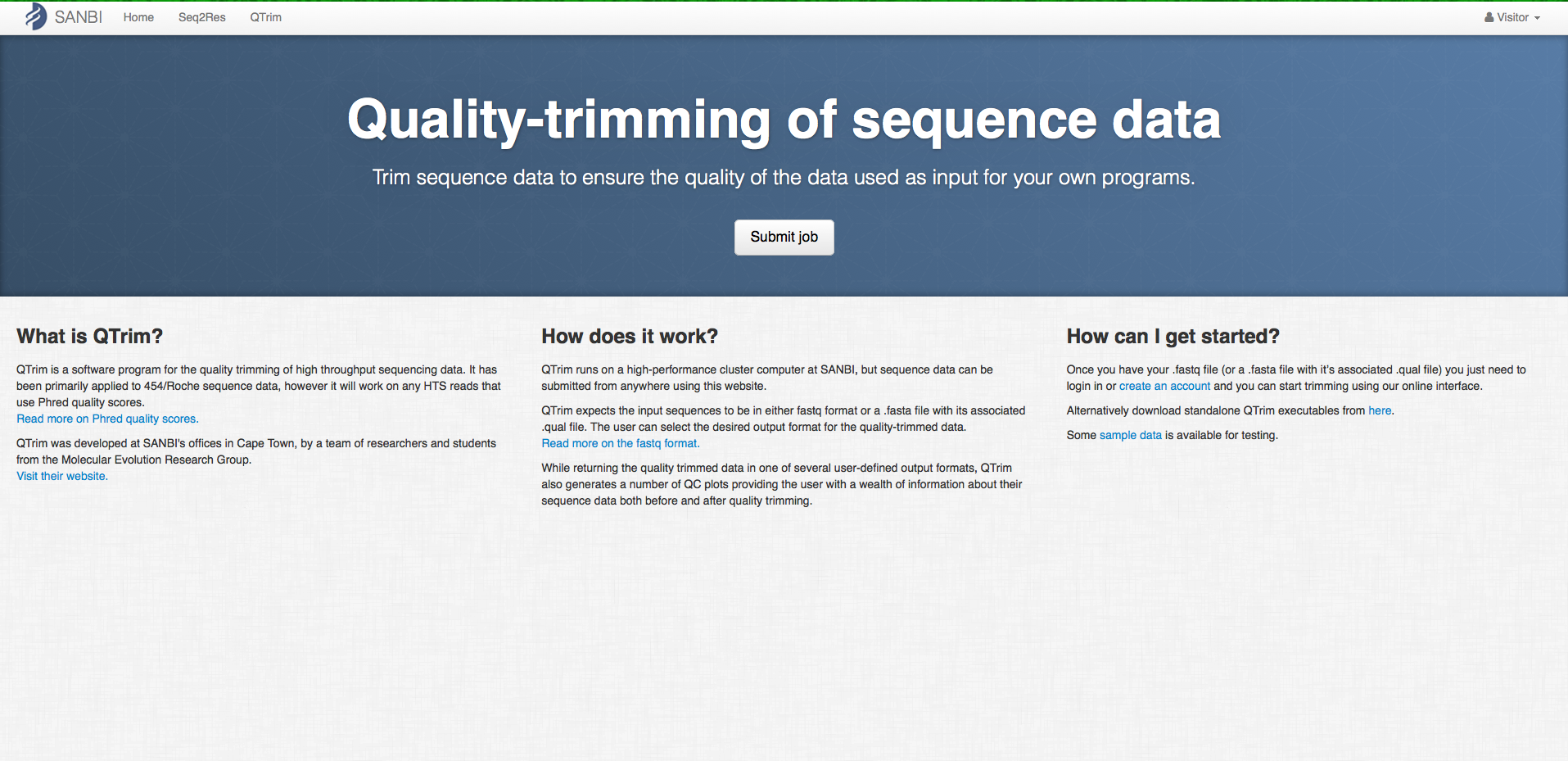


Fig 2.10: Online QTrim home page. Users need to be registered to access and submit sequence file for quality trimming with QTrim.

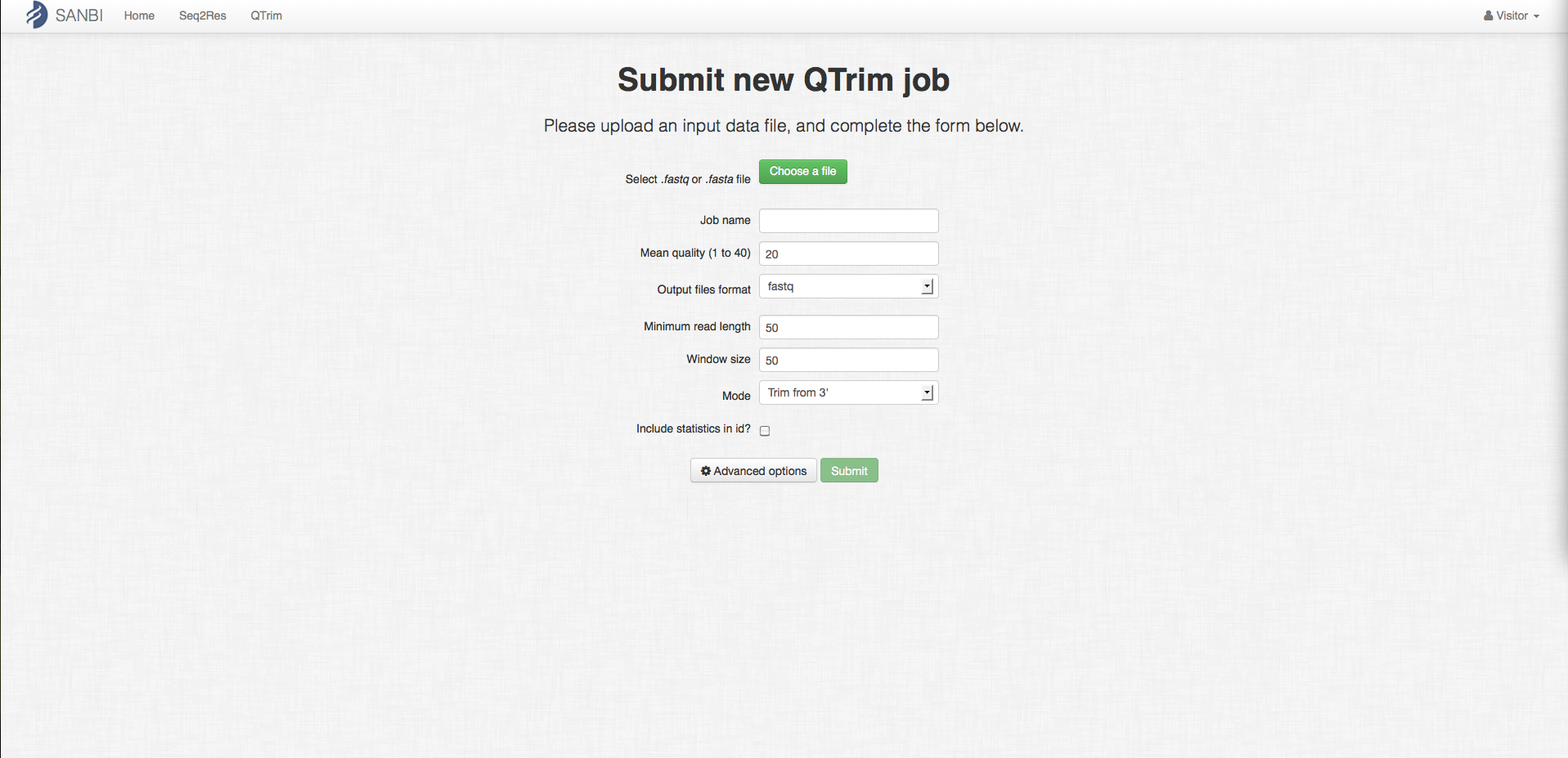


Fig 2.11: Online QTrim job submission page. Sequence file can be uploaded and required parameters can be set before running QTrim online.