

Assignment 1: introduction about RabbitQC

- To high-speed scalable quality control for sequencing data , We present RabbitQC.
- the problem is ,the generated datasets are error-prone, downstream applications usually require quality control methods to pre-process FASTQ files. However, existing tools for this task are currently not able to fully exploit the capabilities of computing platforms leading to slow runtimes.
- to solve this problem We present RabbitQC, an extremely fast integrated quality control tool for FASTQ files, which can take full advantage of modern hardware. It includes a variety of operations and supports different sequencing technologies (Illumina, Oxford Nanopore and PacBio).
- The importance of this solution RabbitQC achieves speedups between one and two orders-of-magnitude compared to other state-of-the-art tools.

☐ the paper will study:

- Thread Scalability on Multi-Core CPUs ,like Thread scalability of RabbitQC in comparison to other multithreaded tools for processing the Illumina dataset SRR2496716 1.
- Design of RabbitQC
employ a producer-consumer pattern to process FASTQ input files, where a single thread is used to read data from file(s) and others to process the data. Reading and formatting operations may take a long time in order to provide formatted reads to consumers. This would result in consumer threads waiting for work which would severely limit thread scalability. We have addressed this issue with our fast I/O framework as follows

- Integrated Function Modules for Quality Control and Pre-Processing , In order to improve the quality of reads, RabbitQC supports sliding window trimming when the average quality falls below a given threshold.