**Challenges in Collecting BWA-mem System Metrics**

**Last week’s task:**

1. **Collecting BWA-mem System Metrics**

**Next week’s task:**

1. **Collecting BWA-mem systems metrics**
2. **Analyze correlation between pre-alignment system metrics and elapsed time.**
3. Overview

This report discusses the process of collecting metrics using BWA-mem for paired fastqc files and a genome reference file. I encountered two issues during the execution. Firstly, except for SRR925780, all fastq files scenarios displayed an error indicating different names for paired reads. The suggested solution involved re-running fastqdump or applying a specific command to fix the files. Secondly, in the SRR925780 scenario, the SLURM job did not terminate even after completion, as evidenced by an excessive execution time. However, stopping the process manually resulted in a status change from RUNNING to COMPLETED.

1. How to collect the metrics

To run BWA-mem, we need to input paired fastqc files and a genome reference file that has been indexed by running “bwa index <genome reference file>”. The fastqc files were downloaded by SRA toolkit. We will run the BWA-mem in 5 paired fastqc files scenarios from [SRR15852393](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR15852393), [SRR15852394](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR15852394), [SRR15852395](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR15852395), [SRR15852396](https://trace.ncbi.nlm.nih.gov/Traces/sra?run=SRR15852396), SRR925780. These scenarios were executed by the script below.

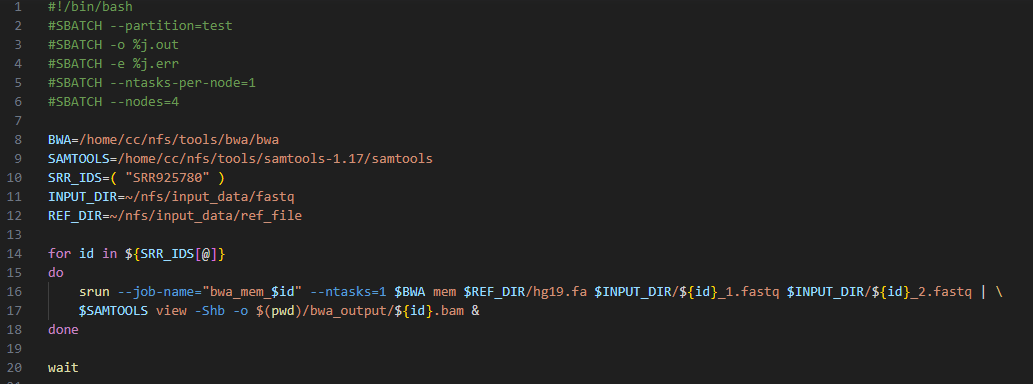


Figure 1. Alignment script

1. Issues

After running the script, there were several problems. Firstly, all fastq files scenarios, unless the SRR925780, showed the same error saying that paired reads have different names. So far I don't understand the cause of this problem, but according to this [page](https://www.biostars.org/p/254155/), the solution to this problem is to re run fastqdump or process the file with the following command:

sed -E "s/^((@|\+)SRR[^.]+\.[^.]+)\.(1|2)/\1/" SRR3239806.fastq > SRR3239806.fixed.fastq

The second problem encountered is when the job in the SRR925780 scenario is successfully executed, SLURM does not kill the job even though the process has actually been completed. This is indicated by an unreasonable execution time. The process, when executed without using SLURM, shows an execution time of 60 minutes. However, if we use SLURM, even up to 3 hours, the job is not finished. When the process is forced to stop then the job status changes from RUNNING to COMPLETED.