Trimming\_loop

Matilda

2025-03-10

This doc explains how to run Trimmomatic on multiple sets of paired end read files without having to manually write and execute a command for each set.

First you need to create a file with a command for each set of files. This can be done with a for loop.

The following loop creates an executable file with a command for each set of paired read files:

for R1 in \*R1\*;  
do  
 R2=${R1//R1\_001.fastq.gz/R2\_001.fastq.gz}  
 R1paired=${R1//.fastq.gz/\_paired.fastq.gz}  
 R1unpaired=${R1//.fastq.gz/\_unpaired.fastq.gz}   
 R2paired=${R2//.fastq.gz/\_paired.fastq.gz}  
 R2unpaired=${R2//.fastq.gz/\_unpaired.fastq.gz}   
 echo "apptainer exec ../containers/trimmomatic\_0.39.sif trimmomatic PE -phred33 $R1 $R2 ../trimmomatic/trimmomatic\_out/$R1paired ../trimmomatic/trimmomatic\_out/$R1unpaired ../trimmomatic/trimmomatic\_out/$R2paired ../trimmomatic/trimmomatic\_out/$R2unpaired ILLUMINACLIP:../trimmomatic/overrep\_seq.fa:2:30:10:2:True LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36" >> ../scripts/trimmomatic.sh;  
done;  
echo “echo “complete”” >> ../scripts/trimmomatic.sh

Make sure the pattern “R1\_001.fastq.gz” matches all the files you want to run - the code wont change the names to R2 etc if it doesn’t match.

The settings used here for Trimmomatic are very general, but can easily be changed - just edit the loop.

Keep in mind that the paths are relative and may need to be changed to fit your file system. As they are written you need to be in the directory containing the seq files.

The file with sequences to trim (overrep\_seq.fa) is specific to my pusposes, you will need to replace that, either with a general option available with Trimmomatic (for example TruSeq3-PE.fa) or with your own file.

This uses a container with Trimmomatic (apptainer system) - installed versions have other commands.

You just copy the loop into a new .sh file and execute it with:

bash <file\_name>.sh

Then you just need to execute the file with the commands:

bash ../scripts/trimmomatic.sh

It may take some time to run. The output will be four files per paired set. Recommend that you run QC after to make sure the trimming worked.

echo "apptainer exec containers/fastqc\_0.12.1.sif fastqc -o fastqc/trimmomatic\_out trimmomatic/trimmomatic\_out/\*.fastq.gz  
apptainer exec containers/multiqc\_1\_27\_1.sif multiqc fastqc/trimmomatic\_out/ -o multiqc/trimmomatic\_out/" >> scripts/qc.sh

bash ../scripts/qc.sh