

My script is as follows:

```
#
# Usage: myscript.sh SRAnumber LIMIT windowSize requiredQuality
#the LIMIT, windowSize, and requiredQuality are optional inputs.
# Example: bash myscript.sh SRR1553607 1000 4 30
#
# This program gets data from SRA and trims it with trimmomatic

#To run this script, one needs to install fastq-dump, trimmomatic and fastqc

#Stop on undefined variables and on errors and also print commands as these are executed
set -uex

#Get the SRA run number from the first input while execution and save it in SRR.
SRR=$1

#Set the output name.

READ2=trimmed.fq

#set the LIMIT variable to the second parameter $2 or 1000 if that was not specified:

LIMIT=${2:-1000}

#Get data from SRA.

fastq-dump -X $LIMIT --split-files $SRR

windowSize=${3:-4}
requiredQuality=${4:-30}

# Run trimmomatic in single end mode

#SLIDINGWINDOW:<windowSize>:<requiredQuality>
#windowSize: specifies the number of bases to average across
#requiredQuality: specifies the average quality required.

trimmomatic SE ${SRR}_2.fastq ${READ2} SLIDINGWINDOW:${windowSize}:${requiredQuality}

# Generate fastqc reports on both datasets.
fastqc ${SRR}_1.fastq ${READ2}
```

The output for SRA number: SRR1553607

```
sug82@submit-001 /storage/work/sug82/applied_bioinformatics/HW6
$ ls
myscript.sh
(bioinfo)
sug82@submit-001 /storage/work/sug82/applied_bioinformatics/HW6
$ bash myscript.sh SRR1553607
+ SRR=SRR1553607
+ READ2=trimmed.fq
+ LIMIT=1000
+ fastq-dump -X 1000 --split-files SRR1553607
Read 1000 spots for SRR1553607
Written 1000 spots for SRR1553607
+ windowSize=4
+ requiredQuality=30
+ trimmomatic SE SRR1553607_2.fastq trimmed.fq SLIDINGWINDOW:4:30
TrimmomaticSE: Started with arguments:
SRR1553607_2.fastq trimmed.fq SLIDINGWINDOW:4:30
Automatically using 4 threads
Quality encoding detected as phred33
Input Reads: 1000 Surviving: 934 (93.40%) Dropped: 66 (6.60%)
TrimmomaticSE: Completed successfully
+ fastqc SRR1553607_1.fastq trimmed.fq
Started analysis of SRR1553607_1.fastq
Approx 100% complete for SRR1553607_1.fastq
Analysis complete for SRR1553607_1.fastq
Started analysis of trimmed.fq
Analysis complete for trimmed.fq
(bioinfo)
sug82@submit-001 /storage/work/sug82/applied_bioinformatics/HW6
$ ls
SRR1553607_1.fastq SRR1553607_1_fastqc.html SRR1553607_1_fastqc.zip SRR1553607_2.fastq myscript.sh trimmed.fq trimmed_fastqc.html trimmed_fastqc.zip
(bioinfo)
sug82@submit-001 /storage/work/sug82/applied_bioinformatics/HW6
$
```

Comparison:

The resulting plot shows the original file on the left and the quality-trimmed version on the right.

