**Trimming reads using Trimmomatic**

*Script*

java -jar /usr/local/bin/Trimmomatic-0.36/trimmomatic-0.36.jar PE -threads 2 -trimlog trim\_CELMISIA.log -phred33 \  
CBR7PANXX-1919AC-20-41-1\_S16\_L001\_R1\_001.fastq CBR7PANXX-1919AC-20-41-1\_S16\_L001\_R2\_001.fastq \  
R108\_R1.fq.gz.P.qtrim R108\_R1.fq.gz.U.qtrim \  
R108\_R2.fq.gz.P.qtrim R108\_R2.fq.gz.U.qtrim \  
ILLUMINACLIP:/usr/local/bin/Trimmomatic-0.36/adapters/TruSeq3-PE-2.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:34 HEADCROP:9

java -jar /usr/local/bin/Trimmomatic-0.36/trimmomatic-0.36.jar PE \ -phred33 -threads 64 -baseout oyster\_trim \ UC2L\_R1.fastq \ UC2L\_R2.fastq \ UC2L\_R1\_trim.fastq\ UC2L\_R2\_trim.fastq\ ILLUMINACLIP:/usr/local/bin/Trimmomatic-0.36/adapters/TruSeq3-PE-2.fa:2:30:10 \ LEADING:5 \ TRAILING:5 \ SLIDINGWINDOW:4:5 \ MINLEN:26\HEADCORP:9

*Parameter info*

Website with information about Trimmomatic parameters: <http://www.usadellab.org/cms/?page=trimmomatic>

You need to change file names in blue need to your respective file names for each library that you trim

trim\_R108.log = log of the trimming process (not 100% needed, but good to have in case something goes wrong)

CBR7PANXX-1919AC-20-41-1\_S16\_L001\_R1\_001.fastq = your read 1 input file; can be zipped, but you need to rename the file if it is

CBR7PANXX-1919AC-20-41-1\_S16\_L001\_R2\_001.fastq = your read 2 input file; can be zipped, but you need to rename the file if it is

R108\_R1.fq.gz.P.qtrim = read 1 output file; contains reads that are paired with reads in R108\_R2.fq.gz.P.qtrim

R108\_R1.fq.gz.U.qtrim = read 1 output file; contains reads that do not have a mate in the R2 file

R108\_R2.fq.gz.P.qtrim = read 2 output file; contains reads that are paired with reads in R108\_R1.fq.gz.P.qtrim

R108\_R2.fq.gz.U.qtrim = read 2 output file; contains reads that do not have a mate in the R1 file

ILLUMINACLIP:2:30:10 = remove adapter sequences that are in the TruSeq3-PE-2.fa file; 2 = maximum mismatch count which will still allow a full match to be performed; 30 = how accurate the match between the two 'adapter ligated' reads must be for PE palindrome read alignment; 10 = how accurate the match between any adapter etc. sequence must be against a read.

SLIDINGWINDOW:4:15 = scan the read with a 4-base wide sliding window, cutting when the average quality per base drops below 15.

LEADING:5 = remove leading low quality or N bases (below quality 5)

TRAILING:5 = remove trailing low quality or N bases (below quality 5)

MINLEN:34 = drop reads below 34bp in length

HEADCROP:9 = remove 9 bases from the start of the read (=Bioo Scientific randomized 8-mer molecular indexes and T from the 5' end of each read)

**Trimming reads using BBDuk**

The location of BBDuk (in bbmap folder) needs to be pathed before it can run

export PATH=/usr/local/bin/bbmap:$PATH

*Script*

bbduk.sh -Xmx1g in1=CBR7PANXX-1919AC-20-41-1\_S16\_L001\_R1\_001.fastq in2=CBR7PANXX-1919AC-20-41-1\_S16\_L001\_R2\_001.fastq out1=R108\_R1.fastq out2=R108\_R2.fastq minlen=25 qtrim=r trimq=10 ktrim=r k=25 mink=11 ftl=9 ref=adapters.fa hdist=1 outm=discarded\_R108.fastq

*Parameter info*

Website with information about BBDuk parameters: <https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/>

You need to change file names in blue need to your respective file names for each library that you trim

outm=discarded\_R108.fastq = fastq file of reads that are discarded

in1=CBR7PANXX-1919AC-20-41-1\_S16\_L001\_R1\_001.fastq = your read 1 input file; can be zipped, but you need to rename the file if it is

in2=CBR7PANXX-1919AC-20-41-1\_S16\_L001\_R2\_001.fastq = your read 2 input file; can be zipped, but you need to rename the file if it is

out1=R108\_R1.fastq = read 1 output file; contains reads that are paired with reads in R108\_R2.fastq

out2=R108\_R2.fastq = read 2 output file; contains reads that are paired with reads in R108\_R1.fastq

-Xmx1g = 1GB of memory is used by BBDuk for trimming

minlen=25 : drop leads below 25bp in length

qtrim=r trimq=10 : this will quality-trim to Q10 using the Phred algorithm, which is more accurate than naive trimming. “qtrim=r” means it will trim the right side only

ktrim=r : once a reference kmer is matched in a read, that kmer and all the bases to the right will be trimmed, leaving only the bases to the left; this is the normal mode for adapter trimming

k=25 : BBDuk will store all 25-mers in the reference, and try to match them against 25-mers in the queries (reads). However, for adapter-trimming, this may not be desirable – for example, if the last 12bp of a read are adapter sequence, it will not match a reference 25-mer, because it is too short. Therefore need to include mink parameter

mink=11 ; will additionally look for shorter kmers with lengths 24 to 11 (in this case)

ftl=9 : This will trim the leftmost 9 bases (Bioo Scientific randomized 8-mer molecular indexes and T from the 5' end of each read)

ref=adapters.fa : the adapters.fa file must be in the same folder as your fastq files (copy it from: /usr/local/bin/bbmap/resources into your scratch file containing your fastq files)