i) Data Preprocessing

Laura Mary Pugh

09/10/2019

The following shell script was used to trim and screen for poor quality reads.

```
#! /bin/bash
#$ -cwd
#$ -V
#$ -M lmpugh@dundee.ac.uk
#$ -m e
#$ -m a
cp /homes/lmpugh/cluster_home/raw_data/16S_samples/*.fastq.gz $TMPDIR
gunzip $TMPDIR/*.fastq.gz

for k in $TMPDIR/*R1_001.fastq; do trimmomatic PE -basein $k -baseout ${k:0:-14}.fastq
    LEADING:20 TRAILING:20 MINLEN:50 HEADCROP:20 SLIDINGWINDOW:4:15; done

mkdir $TMPDIR/fastqc_check
cat $TMPDIR/*1P.fastq >> $TMPDIR/fastq_check/forward_reads_trimmed.fastq
cat $TMPDIR/*2P.fastq >> $TMPDIR/fastq_check/reverse_reads_trimmed.fastq
mkdir $TMPDIR/fastq_check/forward_reads_trimmed
mkdir $TMPDIR/fastq_check/reverse_reads_trimmed
fastqc $TMPDIR/fastq_check/forward_reads_trimmed.fastq -o $TMPDIR/fastq_check/forward_reads_trimmed
fastqc $TMPDIR/fastq_check/reverse_reads_trimmed.fastq -o $TMPDIR/fastq_check/reverse_reads_trimmed

mkdir $TMPDIR/trimmed_reads
cp $TMPDIR/*P.fastq $TMPDIR/trimmed_reads
for k in $TMPDIR/trimmed_reads; do gzip $k ; done
cp $TMPDIR/trimmed_reads /homes/lmpugh/cluster_home/raw_data/16S_samples/
cp $TMPDIR/fastq_check/forward_reads_trimmed $TMPDIR/fastq_check/reverse_reads_trimmed
    /homes/lmpugh/cluster_home/raw_data/
```