**S1 Text. Shell script 18Sclean.sh**

#!/bin/bash

for j in $(ls \*R1.fastq)

do i=${j%%\_R1.fastq}

**#assemble reads with PEAR v.0.9.6**

pear -f ${i}\_R1.fastq -r ${i}\_R2.fastq -o pear.${i}

**#filter for quality**

usearch9 -fastq\_filter pear.${i}.unassembled.forward.fastq -fastq\_maxee 1.0 -fastaout pear.unmerged.r1.${i}.fas -fasta\_cols 0

usearch9 -fastq\_filter pear.${i}.unassembled.reverse.fastq -fastq\_maxee 1.0 -fastaout pear.unmerged.r2.${i}.fas -fasta\_cols 0

**#get names in forward read**

cat pear.unmerged.r1.${i}.fas | grep ">" > names\_forward.${i}

**#get names in both**

gawk -v STUB=${i} 'BEGIN {RS=">"}FILENAME=="names\_forward."STUB{a[$1]=$1}FILENAME=="pear.unmerged.r2."STUB".fas" {if(a[$1]){printf(">%s\n",$1)}}' names\_forward.${i} pear.unmerged.r2.${i}.fas> names\_both.${i}

**#get forward reads**

gawk -v STUB=${i} 'BEGIN {RS=">"}FILENAME=="names\_both."STUB{a[$1]=$1}FILENAME=="pear.unmerged.r1."STUB".fas" {if(a[$1]){printf(">%s\n",$1);printf("%s\n",$3)}}' names\_both.${i} pear.unmerged.r1.${i}.fas > pear.unmerged.qfiltered.r1.${i}.fas

**#get reverse reads**

gawk -v STUB=${i} 'BEGIN {RS=">"}FILENAME=="names\_both."STUB{a[$1]=$1}FILENAME=="pear.unmerged.r2."STUB".fas" {if(a[$1]){printf(">%s\n",$1);printf("%s\n",$3)}}' names\_both.${i} pear.unmerged.r2.${i}.fas > pear.unmerged.qfiltered.r2.${i}.fas

**#join both reads with an N**

paste pear.unmerged.qfiltered.r1.${i}.fas pear.unmerged.qfiltered.r2.${i}.fas | awk '{if($1~/>/){printf("%s\n",$1)}else{printf("%sN",$1);printf("%s\n",$2)}}' > combined.seq.${i}

**#change sequence names**

cat combined.seq.${i} | sed '1s/\>//' | gawk -v STUB=${i} 'BEGIN{RS=">"}{printf(">%s\n",STUB"."NR);printf("%s\n",$2)}' > ${i}\_nb.fas

rm pear\*

rm names\*

done

**#merge sequences in a single file**

cat \*\_nb.fas > merged0.fas

**#renumber all sequences**

cat merged0.fas | sed '1s/\>//'| gawk 'BEGIN{RS=">"}{{sub(/\..\*/,"."NR,$1)};printf(">%s\n",$1);printf("%s\n",$2)}'>merged.fasta

rm pear\*

rm names\*