**# Raw data:**

/mibi/miseq/miseq\_microbio/210505\_M02529\_0413\_000000000-JLCBY/Alignment\_1/20210507\_235407/Fastq/Run20210505Order35\*

Order351 = Paula

Order352 = Andreas

Order353 = Rodrigo

Ordre354 = Lorrie

**# Load files /check for the sequences number Raw Fastq / simplify samples names / create samples list**

cd /mibi/users/lmaccario/Biorecover/Genomes\_Sequencing/Order354/RawFastq

ln -sf /mibi/miseq/miseq\_microbio/210505\_M02529\_0413\_000000000-JLCBY/Alignment\_1/20210507\_235407/Fastq/Run20210505Order35\* ./

for i in \*R1\_001.fastq.gz ; do mv $i $(echo $i | sed -e 's/\_.\*/\_R1.fastq.gz/' -e 's/Run20210505//') ; done

for i in \*R2\_001.fastq.gz ; do mv $i $(echo $i | sed -e 's/\_.\*/\_R2.fastq.gz/' -e 's/Run20210505//') ; done

for file in \*.gz ; do zcat $file | echo $((`wc -l`/4)) ; done

ls \*.gz > Samples.list

sed -i  -e 's/\_R1.fastq.gz//g' -e 's/\_R2.fastq.gz//g'  Samples.list

 sort Samples.list  | uniq -d  > Samples.txt

cd ..

mv RawFastq/Samples.txt  ./

**# trim-galore for reads trimming : adapters (auto-detection) + quality trim <Q25:**

mkdir Trim\_galore

conda activate /mibi/users/lmaccario/miniconda3/envs/trim-galore/

parallel --xapply " trim\_galore --paired -q 25 RawFastq/{}\_R1.fastq.gz  RawFastq/{}\_R2.fastq.gz  -o Trim\_galore/

--path\_to\_cutadapt /mibi/users/lmaccario/miniconda3/bin/cutadapt " :::: Samples.txt

### Check for Reads removed and remaining and adapters trimmed

for file in Trim\_galore/\*.gz ; do zcat $file | echo $((`wc -l`/4)) ; done

grep -e "Adapter sequence" Trim\_galore/\*.txt

Adapter sequence: 'CTGTCTCTTATA' (Nextera Transposase sequence; auto-detected)

grep -e "Reads with adapters" Trim\_galore/\*.txt

|  |  |  |  |
| --- | --- | --- | --- |
|  | %Trimmed | #Trimmed | #ReadsAfterTrimming |
| Order351Sample001\_R1 | 63.90% | 762,416 | 1158012 |
| Order351Sample001\_R2 | 59.10% | 705,394 | 1158012 |
| Order351Sample002\_R1 | 68.50% | 916,326 | 1311845 |
| Order351Sample002\_R2 | 62.30% | 834,446 | 1311845 |
| Order351Sample003\_R1 | 66.10% | 1,112,348 | 1633052 |
| Order351Sample003\_R2 | 59.80% | 1,007,065 | 1633052 |
| Order351Sample004\_R1 | 65.50% | 1,029,733 | 1535822 |
| Order351Sample004\_R2 | 59.20% | 929,964 | 1535822 |
| Order351Sample005\_R1 | 69.50% | 939,131 | 1324870 |
| Order351Sample005\_R2 | 62.80% | 847,925 | 1324870 |
| Order351Sample006\_R1 | 62.70% | 714,308 | 1113023 |
| Order351Sample006\_R2 | 59.30% | 675,004 | 1113023 |
| Order351Sample007\_R1 | 71.10% | 1,202,633 | 1666931 |
| Order351Sample007\_R2 | 66.90% | 1,130,653 | 1666931 |
| Order351Sample008\_R1 | 66.60% | 929,957 | 1353692 |
| Order351Sample008\_R2 | 60.50% | 844,524 | 1353692 |
| Order351Sample009\_R1 | 54.30% | 589,043 | 1057232 |
| Order351Sample009\_R2 | 49.80% | 540,008 | 1057232 |
| Order351Sample010\_R1 | 61.30% | 747,183 | 1190971 |
| Order351Sample010\_R2 | 56.50% | 688,710 | 1190971 |
| Order352Sample001\_R1 | 54.70% | 604,294 | 1049612 |
| Order352Sample001\_R2 | 52.10% | 575,541 | 1049612 |
| Order352Sample002\_R1 | 63.10% | 690,648 | 1053447 |
| Order352Sample002\_R2 | 58.20% | 637,784 | 1053447 |
| Order352Sample003\_R1 | 57.90% | 18,104 | 29739 |
| Order352Sample003\_R2 | 53.10% | 16,602 | 29739 |
| Order353Sample001\_R1 | 55.20% | 14,191 | 24217 |
| Order353Sample001\_R2 | 50.00% | 12,850 | 24217 |
| Order353Sample002\_R1 | 55.30% | 18,884 | 31941 |
| Order353Sample002\_R2 | 49.10% | 16,760 | 31941 |
| Order353Sample003\_R1 | 63.00% | 20,297 | 30203 |
| Order353Sample003\_R2 | 57.80% | 18,617 | 30203 |
| Order353Sample004\_R1 | 59.20% | 705,938 | 1174937 |
| Order353Sample004\_R2 | 55.00% | 656,423 | 1174937 |
| Order353Sample005\_R1 | 72.00% | 871,927 | 1203011 |
| Order353Sample005\_R2 | 68.50% | 828,716 | 1203011 |
| Order353Sample006\_R1 | 60.40% | 718,760 | 1173873 |
| Order353Sample006\_R2 | 56.60% | 673,841 | 1173873 |
| Order353Sample007\_R1 | 74.50% | 955,792 | 1271674 |
| Order353Sample007\_R2 | 70.60% | 905,816 | 1271674 |
| Order353Sample008\_R1 | 64.80% | 915,369 | 1396103 |
| Order353Sample008\_R2 | 61.90% | 874,645 | 1396103 |
| Order353Sample009\_R1 | 60.60% | 775,703 | 1266450 |
| Order353Sample009\_R2 | 57.80% | 740,109 | 1266450 |
| Order353Sample010\_R1 | 59.50% | 738,789 | 1225833 |
| Order353Sample010\_R2 | 55.40% | 688,640 | 1225833 |
| Order353Sample011\_R1 | 57.20% | 845,199 | 1461070 |
| Order353Sample011\_R2 | 54.00% | 798,002 | 1461070 |
| Order354Sample001\_R1 | 61.10% | 729,931 | 1151547 |
| Order354Sample001\_R2 | 54.90% | 655,642 | 1151547 |
| Order354Sample002\_R1 | 54.30% | 615,105 | 1076467 |
| Order354Sample002\_R2 | 51.60% | 584,427 | 1076467 |
| Order354Sample003\_R1 | 52.30% | 526,106 | 964059 |
| Order354Sample003\_R2 | 47.80% | 481,259 | 964059 |
| Order354Sample004\_R1 | 61.50% | 694,004 | 1097297 |
| Order354Sample004\_R2 | 56.50% | 637,446 | 1097297 |

**# PhiX cleaning:**

### Run bbduk in parallel but using one job at the time, because of potential rewritting in fq file (cf or that multiple programs/threads were writing to the same file at the same time (that's the only case I've encountered that gave similar-looking output) http://seqanswers.com/forums/showthread.php?t=42776&page=2.

mkdir PhiX\_removal\_bbduk/

### bbduk scripts

cp /mibi/users/jnesme/Bin/bbmap/bbduk.sh ./

### Run bbduk

parallel --xapply -j 1 " bbduk.sh in=Trim\_galore/{}\_R1\_val\_1.fq.gz in2=Trim\_galore/{}\_R2\_val\_2.fq.gz out=PhiX\_removal\_bbduk/{}\_R1\_PhiXCleaned.fq.gz out2=PhiX\_removal\_bbduk/{}\_R2\_PhiXCleaned.fq.gz outm=PhiX\_removal\_bbduk/{}\_R1\_PhiXReads.fq.gz outm2=PhiX\_removal\_bbduk/{}\_R2\_PhiXReads.fq.gz ref=/mibi/users/jnesme/Bin/bbmap/resources/phix174\_ill.ref.fa.gz k=31 hdist=1 stats=PhiX\_removal\_bbduk/{}\_stats.phixclean.txt " :::: Samples.txt

### Check for Reads removed

grep "#Matched" PhiX\_removal\_bbduk/\*.txt

|  |  |  |
| --- | --- | --- |
| Order351Sample001 | 13 | 0.00% |
| Order351Sample002 | 12 | 0.00% |
| Order351Sample003 | 23 | 0.00% |
| Order351Sample004 | 24 | 0.00% |
| Order351Sample005 | 8 | 0.00% |
| Order351Sample006 | 22 | 0.00% |
| Order351Sample007 | 3 | 0.00% |
| Order351Sample008 | 475 | 0.02% |
| Order351Sample009 | 6 | 0.00% |
| Order351Sample010 | 5 | 0.00% |
| Order352Sample001 | 5 | 0.00% |
| Order352Sample002 | 251 | 0.01% |
| Order352Sample003 | 2 | 0.00% |
| Order353Sample001 | 0 | 0.00% |
| Order353Sample002 | 1 | 0.00% |
| Order353Sample003 | 1 | 0.00% |
| Order353Sample004 | 4 | 0.00% |
| Order353Sample005 | 1 | 0.00% |
| Order353Sample006 | 2 | 0.00% |
| Order353Sample007 | 163 | 0.01% |
| Order353Sample008 | 6 | 0.00% |
| Order353Sample009 | 4 | 0.00% |
| Order353Sample010 | 331 | 0.01% |
| Order353Sample011 | 9 | 0.00% |
| Order354Sample001 | 16 | 0.00% |
| Order354Sample002 | 9 | 0.00% |
| Order354Sample003 | 6 | 0.00% |
| Order354Sample004 | 22 | 0.00% |

**# Unicycler assembly**

mkdir UnicyclerAssembly/

conda activate /mibi/users/russel/Software/miniconda2/envs/unicycler/

parallel --xapply " unicycler -1 PhiX\_removal\_bbduk/{}\_R1\_PhiXCleaned.fq.gz  -2 PhiX\_removal\_bbduk/{}\_R2\_PhiXCleaned.fq.gz -o UnicyclerAssembly/{}  -t 30  " :::: Samples.txt

### Check for contigs number

grep -c ">" UnicyclerAssembly/\*\*/assembly.fasta

UnicyclerAssembly/Order351Sample001/assembly.fasta:86

UnicyclerAssembly/Order351Sample002/assembly.fasta:36

UnicyclerAssembly/Order351Sample003/assembly.fasta:64

UnicyclerAssembly/Order351Sample004/assembly.fasta:23

UnicyclerAssembly/Order351Sample005/assembly.fasta:100

UnicyclerAssembly/Order351Sample006/assembly.fasta:57

UnicyclerAssembly/Order351Sample007/assembly.fasta:105

UnicyclerAssembly/Order351Sample008/assembly.fasta:116

UnicyclerAssembly/Order351Sample009/assembly.fasta:96

UnicyclerAssembly/Order351Sample010/assembly.fasta:133

UnicyclerAssembly/Order352Sample001/assembly.fasta:100

UnicyclerAssembly/Order352Sample002/assembly.fasta:98

UnicyclerAssembly/Order352Sample003/assembly.fasta:3

UnicyclerAssembly/Order353Sample001/assembly.fasta:686

UnicyclerAssembly/Order353Sample002/assembly.fasta:871

UnicyclerAssembly/Order353Sample003/assembly.fasta:694

UnicyclerAssembly/Order353Sample004/assembly.fasta:39

UnicyclerAssembly/Order353Sample005/assembly.fasta:55

UnicyclerAssembly/Order353Sample006/assembly.fasta:55

UnicyclerAssembly/Order353Sample007/assembly.fasta:47

UnicyclerAssembly/Order353Sample008/assembly.fasta:43

UnicyclerAssembly/Order353Sample009/assembly.fasta:53

UnicyclerAssembly/Order353Sample010/assembly.fasta:44

UnicyclerAssembly/Order353Sample011/assembly.fasta:47

UnicyclerAssembly/Order354Sample001/assembly.fasta:84

UnicyclerAssembly/Order354Sample002/assembly.fasta:32

UnicyclerAssembly/Order354Sample003/assembly.fasta:108

UnicyclerAssembly/Order354Sample004/assembly.fasta:205

**# Prokka Annotation**

mkdir ProkkaAnnotation

### Run Prokka

conda activate /mibi/users/lmaccario/miniconda3/envs/prokka/

parallel --xapply   " prokka UnicyclerAssembly/{}/assembly.fasta  --outdir ProkkaAnnotation/{} --prefix {} " :::: Samples.txt

**# Calculate Coverage**

mkdir Coverage

### Run BBmap

cp /mibi/users/jnesme/Bin/bbmap/bbmap.sh  ./

parallel --xapply   " bbmap.sh in=PhiX\_removal\_bbduk/{}\_R1\_PhiXCleaned.fq.gz in2=PhiX\_removal\_bbduk/{}\_R2\_PhiXCleaned.fq.gz ref=UnicyclerAssembly/{}/assembly.fasta nodisk covstats=Coverage/{}\_stats.txt covhist=Coverage/{}\_histogram.txt " :::: Samples.txt

**# Rename assembly output**

cd/mibi/users/lmaccario/Biorecover/Genomes\_Sequencing/Order354/UnicyclerAssembly

mkdir Test

cp -r Order35\* Test/

cd Test

rename  's:/:-:g' \*/assembly.fasta

cd ..

mkdir All\_assemblies

mv Test/\*.fasta All\_assemblies/

rm -r Test/

**# Extract 16S sequences from annotation and rename fasta**

cd/mibi/users/lmaccario/Biorecover/Genomes\_Sequencing/Order354/ProkkaAnnotation

for file in \*\*/\*.ffn ; do sed -n -e "/16S ribosomal RNA/,/>/p" $file | sed -e '$d' ; done > all\_16S.fa

grep "16S ribosomal RNA" \*\*/\*.ffn

while read n k; do sed -i  "s/$n/$k/g" all\_16S.fa ; done < IDs.list

**# Programm version**

Trim\_galore version 0.6.4\_dev

Cutadapt v3.4

BBMap version 38.76

Unicycler v0.4.8

Prokka v1.14.6