

Análisis MTB

Preparación de muestras

```
mkdir 01_raw_reads  
ls 01_raw_reads/ *fq.gz
```

```
01_raw_reads/130621253_ESFP210029014-1a_HV7YNDX2_L4_1.fq.gz  
01_raw_reads/182028377_ESFP210029015-1a_HV7M3DSX2_L2_2.fq.gz  
01_raw_reads/130621253_ESFP210029014-1a_HV7YNDX2_L4_2.fq.gz  
01_raw_reads/182291946_ESFP210029006-1a_HV7M3DSX2_L2_1.fq.gz  
01_raw_reads/140630054_ESFP210029013-1a_HV7M3DSX2_L2_1.fq.gz  
01_raw_reads/182291946_ESFP210029006-1a_HV7M3DSX2_L2_2.fq.gz  
01_raw_reads/140630054_ESFP210029013-1a_HV7M3DSX2_L2_2.fq.gz  
01_raw_reads/192069454_ESFP210029007-1a_HV7M3DSX2_L2_1.fq.gz  
01_raw_reads/170699855_ESFP210029008-1a_HV7M3DSX2_L2_1.fq.gz  
01_raw_reads/192069454_ESFP210029007-1a_HV7M3DSX2_L2_2.fq.gz  
01_raw_reads/170699855_ESFP210029008-1a_HV7M3DSX2_L2_2.fq.gz  
01_raw_reads/192087440_ESFP210029020-1a_HV7YNDX2_L1_1.fq.gz  
01_raw_reads/172148124_ESFP210029010-1a_HV7M3DSX2_L2_1.fq.gz  
01_raw_reads/192087440_ESFP210029020-1a_HV7YNDX2_L1_2.fq.gz  
01_raw_reads/172148124_ESFP210029010-1a_HV7M3DSX2_L2_2.fq.gz  
01_raw_reads/192127496_ESFP210029011-1a_HV7YNDX2_L4_1.fq.gz  
01_raw_reads/172211651_ESFP210029019-1a_HV7YNDX2_L4_1.fq.gz  
01_raw_reads/192127496_ESFP210029011-1a_HV7YNDX2_L4_2.fq.gz  
01_raw_reads/172211651_ESFP210029019-1a_HV7YNDX2_L4_2.fq.gz  
01_raw_reads/192155610_ESFP210029017-1a_HV7YNDX2_L1_1.fq.gz  
01_raw_reads/180653933_ESFP210029009-1a_HV7M3DSX2_L4_1.fq.gz  
01_raw_reads/192155610_ESFP210029017-1a_HV7YNDX2_L1_2.fq.gz  
01_raw_reads/180653933_ESFP210029009-1a_HV7M3DSX2_L4_2.fq.gz  
01_raw_reads/193507993_ESFP210029012-1a_HV7M3DSX2_L4_1.fq.gz  
01_raw_reads/182003358_ESFP210029016-1a_HV7YNDX2_L4_1.fq.gz  
01_raw_reads/193507993_ESFP210029012-1a_HV7M3DSX2_L4_2.fq.gz  
01_raw_reads/182003358_ESFP210029016-1a_HV7YNDX2_L4_2.fq.gz  
01_raw_reads/202217271_ESFP210029018-1a_HV7M3DSX2_L4_1.fq.gz  
01_raw_reads/182028377_ESFP210029015-1a_HV7M3DSX2_L2_1.fq.gz  
01_raw_reads/202217271_ESFP210029018-1a_HV7M3DSX2_L4_2.fq.gz
```

Cambiamos el nombre para quedarnos solo con el número de petición (para testear rename : rename -v -n)

```
rename 's/HC_//' 01_raw_reads/*fq.gz
rename 's/_ESFP[0-9]{9}-1a_HV7YNDXSX2_L[0-9]//' 01_raw_reads/*fq.gz
rename 's/_ESFP[0-9]{9}-1a_HV7M3DSX2_L[0-9]//' 01_raw_reads/*fq.gz
```

Creamos el archivo list_samples.txt

```
ls 01_raw_reads/*_1.fq.gz | sed "s#01_raw_reads/##" | sed "s/_1.fq.gz//"
> list_samples.txt
```

Análisis de Calidad: fastqc + multiqc

```
conda activate QUALITY
fastqc 01_raw_reads/*fq.gz
multiqc 01_raw_reads/
```

Ensamblaje (Unicycler)

```
conda activate UNICYCLER
mkdir 02_assembly
```

Prueba con 1 muestra:

```
unicycler -1 01_raw_reads/130621253_1.fq.gz \
-2 01_raw_reads/130621253_2.fq.gz \
-o 02_assembly/130621253_unicycler
```

Loop para todas las muestras:

Los kmeros deben ser impares. Valor máximo: 127

```
f_path=/home/elrubio/Documentos/WGS/WGS_NTM/
kmers="85,115,127"

for i in $(cat list_samples.txt); do
    unicycler -1 ${f_path}/01_raw_reads/${i}_1.fq.gz \
    -2 ${f_path}/01_raw_reads/${i}_2.fq.gz \
    -o ${f_path}/02_assembly/${i}_unicycler \
    --kmers $kmers \
    -t 6
    mv ${f_path}/02_assembly/${i}_unicycler/assembly.fasta
    ${f_path}/02_assembly/${i}_assembly.fasta
done
```

Analizamos los assemblies con BUSCO

https://busco.ezlab.org/busco_userguide.html#interpreting-the-results

```
busco --auto-lineage-prok -i ./assemblies/flye_barcode01.fasta -o  
busco_flye_barcode01 -m genome -c 18
```

"--auto-lineage-prok" és per a què ell et busqui quina database
utilitzar dins les de prokaryotes
"-i" genoma input
"-o" directori output
"-m" tipus de material, en aquest cas genome
"-c" seria igual que "-t", els threads

Análisis con 1 muestra:

```
f_path=/home/elrubio/Documentos/WGS/WGS_NTM/  
  
mkdir 03_busco  
  
busco -l corynebacteriales_odb10 \  
-i ${f_path}/02_assembly/130621253_assembly.fasta \  
--out_path ${f_path}03_busco/ \  
-o 130621253 -m genome -c 6
```

Loop con todas las muestras:

```
f_path=/home/elrubio/Documentos/WGS/WGS_NTM/  
  
for i in $(<list_samples.txt); do  
    busco -l corynebacteriales_odb10 \  
    -i ${f_path}/02_assembly/${i}_assembly.fasta \  
    --out_path ${f_path}03_busco/ \  
    -o ${i} -m genome -c 6  
    mv ${f_path}03_busco/${i}/*.txt ${f_path}/03_busco/  
done
```

Obtenemos una figura resumen de los resultados de busco:

```
py_path=/home/elrubio/miniconda3/envs/UNICYCLER/bin/  
python3 ${py_path}generate_plot.py -wd ${f_path}03_busco/
```

Quast

Descargamos el genoma de referencia de Genebank:

https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_002219285.1/

ASM221928v1

```
quast 02_assembly/*_assembly.fasta -o quat_fast_icarus
```

```
quast 02_assembly/*_assembly.fasta -o quat_fast_icarus_reference -r  
Reference_genomes/GCF_002219285.1/ncbi_dataset/data/GCF_002219285.1/GCF_  
002219285.1_ASM221928v1_genomic.fna --circos --glimmer --threads 6
```

Otra referencia: https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_020412405.1/

```
r_path=Reference_genomes/GCF_020412405.1/ncbi_dataset/data/GCF_020412405  
.1/
```

```
quast 02_assembly/*_assembly.fasta -o quat_fast_icarus_reference2 \  
-r ${r_path}GCF_020412405.1_ASM2041240v1_genomic.fna \  
-g ${r_path}genomic.gff \  
--circos --glimmer --threads 6
```