

WS4 Files & IGV

Conda env: 03MBIF_v4

Files:

Sanger.ab1
Metagenoma_R1.fastq
Metagenoma_R2.fastq
pacbio_output.fastq
BY4741_Q20_sorted_rmdup.bam
BY4741_Q20_sorted_rmdup.bam.bai
AV16_Q20_sorted_rmdup.bam
AV16_Q20_sorted_rmdup.bam.bai

1_Secuenciación de primera generación:

Archivos Sanger

2_Secuenciación de segunda generación:

Archivo Illumina pair-end

3_Secuenciación de tercera generación:

Archivo PacBio

4_Visualizador Integrative Genome Viewer

1 Archivo Sanger

Vamos a ver un archivo de Sanger, tiene una terminación .ab1 es un archivo propietario

```
less Sanger.ab1
```

less nos indica que puede tratarse de un archivo binario, los archivos binarios no están preparados para ser leídos por humanos, únicamente por máquinas, le decimos que lo muestre igualmente. Los archivos binarios nos los encontraremos más adelante con el BAM y el SAM durante el mapeado o alineamiento de secuencias (.fastq) al genoma de referencia (.fasta). El BAM, de menor tamaño al estar comprimido.

Para ver este archivo, nos vamos al env 03MBIF_v4, en el tenemos un programa que nos permite interactuar con el archivo .ab1

```
conda env list
```

```
conda activate 03MBIF_v4
```

```
abiview -h
```

Podemos modificar el tamaño de la ventana de salida con -window [valor numérico]

vamos a transformar el archivo que ha salido del secuenciador: Sanger.ab1

```
abiview
```

Debemos indicar las opciones

```
$ Display the trace in an ABI sequencer file
```

```
$ ABI sequencing trace file:
```

```
Sanger.ab1
```

```
$ nucleotide output sequence [sanger.fasta]:
```

```
sanger.fasta
```

```
$Graph type [png]:
```

```
png
```

Ha creado los archivos en la carpeta en la que nos encontrábamos.

Abrimos sanger.fasta con pluma, se trata de la secuencia resultado de la secuenciación.

En los archivos .png podemos ver el resultado del cromatograma.

Podemos contar cuántos nucleótidos tiene el archivo:

```
wc -m sanger.fasta
```

```
1276
```

2 Archivo Illumina

Visualizamos:

```
more Metagenoma_R1.fastq
```

visualizar y comparar R1 y R2

```
more Metagenoma_R2.fastq
```

cuantas lecturas tiene mi archivo?

```
grep "@MISEQ" -c Metagenoma_R1.fastq
```

```
R1 : 876589
```

y el archivo R2? ¿cuántas secuencias tienen que contener? que lo hagan ellos

```
grep "@MISEQ" -c Metagenoma_R2.fastq
```

```
R2 : 876589
```

Tengo el mismo número de secuencias en ambos archivos.

Por cierto, ¿que hemos secuenciado?

3 Secuenciación de tercera generación:

Usaremos un ejemplo visto anteriormente:

```
less pacbio_output.fastq
```

4 Visualizador Integrative Genome Viewer

Integrative Genome Viewer es una herramienta de visualización de archivos que podemos correr en web, escritorio y local, ya que corre en java:

<https://software.broadinstitute.org/software/igv/>

Además, IGV tiene en su página excelente tutoriales. Entre ellos un resumen de archivos que además nos lleva a otras páginas e instituciones con buenos tutoriales:

<https://software.broadinstitute.org/software/igv/FileFormats>

vamos a visualizarlo

```
igv
```

El igv por defecto tiene un genoma cargado, en este caso el humano, por tanto, solo acepta archivos que hayan sido indexados usando ese genoma. Nosotros vamos a usar otro genoma de referencia que también encontramos en la base de datos de IGV.

[desplegable/more/search=cer/S. cerevisae \(sacCer3\)](#)

Cargamos el bam con:

[file/load from file/BY4741.sort.bam](#)

¿Posible SNP en YFL040W?