Laboratorio 1 03/10/24

Ejercicio 1: (4 ptos)

Laura González López y Álvaro Morro Pérez

Usa el fichero con formato .vcf file.vcf. Explora el archivo usando el comando head. Aquí hay una explicación del tipo de archivo, aunque se comentará en clase.

1. Selecciona aquellos SNPs que han pasado todos los filtros (Tienen el texto PASS). Selecciona también aquellos SNPs que no han pasado el filtro de calidad (Tienen el texto q10). Indica que comandos debes usar si quieres generar un archivo de texto con esa información.

Los comandos usados para obtener únicamente el texto que tiene PASS y luego enviarlo a un archivo de texto es **grep "PASS" file.vcf > resultadoEjercicio1.txt.** Luego para añadir el texto que tenga únicamente q10 y pasarlo al mismo archivo de texto sin sobrescribirlo es **grep "q10" file.vcf >> resultadoEjercicio1.txt** 

```
| Pass |
```

2. Tenemos un nuevo fichero file2.vcf que ha perdido la información de la cabecera. Sabemos que la cabecera del fichero file.vcf nos vale, añadir la cabecera de file.vcf al fichero file2.vcf creando un nuevo fichero. Indica los comandos usados.

```
GNU namo 8.1

#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NAMOMOL NAMOMO2 NAMOMO3

20 143 . T A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 010:48:1:51,51 110:48:85:51,51 11/1:43:5:...

20 24530 . G A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 010:48:1:51,51 110:48:85:1,51 11/1:43:5:...

20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 112:21:62:23,27 2[12:2:0:18,2 2/2:35:4]

22 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 010:54:7:56,6 010:48:4:51,51 07:61:2

23 1234567 microsat2 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4

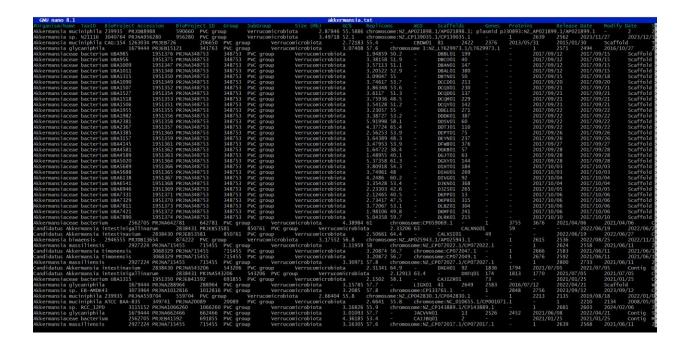
PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4

O/2:17:2 1/1:40:3
```

Primeramente, usamos el comando grep "REF" file.vcf > cabecera.txt de tal manera que extraemos la cabecera del file.vcf en un documento aparte. Luego utilizamos el comando cat cabecera.txt file2.vcf >> ejercicio1.txt para fusionar la cabecera con el file2.vcf y ponerlo en un documento de texto.

3. En el mismo directorio crea un nuevo archivo de texto llamado akkermansia.txt donde se seleccionen los siguientes aspectos (Indica los comandos usados) La cabecera del archivo original prokaryots.txt. Todos los genomas de "Akkermansia".

Primero extrajimos la cabecera del archivo original con el comando grep "TaxID" prokaryotes.txt > akkermansia.txt. Posteriormente, añadimos todos los genomas de akkermansia al archivo sin sobrescribir lo anterior con el comando grep "Akkermansia" prokaryotes.txt >> akkermansia.txt.



Sobre el archivo akkermansia.txt conteste indicando el comando usado:

4. ¿Cómo pueden verse las primeras 10 líneas del archivo?

Se puede usar el comando head -n 10 akkermansia.txt o con head akkermansia.txt .

5. ¿Cuántos genomas de Akkermansia muciniphila hay?

<u>Usamos el comando **grep -c "Akkermansia muciniphila" akkermansia.txt**. Hay **1205** genomas de Akkermansia muciniphila.</u>

6. ¿Cuántos de Akkermasia biwaensis?

<u>Usamos el comando grep -c "Akkermasia biwaensis" akkermansia.txt</u>. Hay 1 genoma de Akkermansia biwaensis.

7. Por último, elimina el archivo prokaryots.txt

Usamos el comando rm prokaryotes.txt

Ejercicio 3: (1 pto)

8. Escribe un script llamado fastq\_script.sh que muestre todos los nombres de archivos .fastq luego cuente el número de líneas en cada archivo y al finalizar diga un mensaje de "Terminado". Recuerda hacerlo ejecutable.

Para generar el script: nano fastq\_script.sh

Para que muestre todos los nombres de archivos .fastq, luego cuente el número de líneas en cada archivo y al finalizar diga un mensaje de "Terminado":

#/bin/bash
for filename in \*.fastq
do echo \$filename
wc -l \$filename
echo "terminado"
done

Para hacerlo ejecutable: chmod +x fastq\_script.sh