Laboratorio I: Shell scripting

Ejercicio 1: (4 ptos)

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.

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.

Primero se usa el comando: grep "PASS" o "q10" según sea el caso para identificarlas.

```
Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)
$ grep "pass" file.vcf

Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)
$ grep "PASS" file.vcf

20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.
5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:...
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.
333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 T .47 PASS NS=3;DP=13;AA=T GT:GQ:DF:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:661:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

Hay que crear el archivo de texto con:

\$ nano resultado.txt

Luego para copiar ambas secuencias dentro usamos:

```
Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)

$ grep "PASS" file.vcf > resultado.txt

Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)

$ grep "q10" file.vcf >> resultado.txt
```

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.

Creamos el archivo de texto con el comando "nano" nuestro caso lo llamamos "file3.vcf".

Extraemos el encabezado del "file.vcf" con el comando "head -n 1" con ">" lo pasamos a un archivo que llamaremos "cabecera.txt" el cuál se creará instantáneamente.

Con el comando "cat" fusionamos el encabezado "cabecera.txt" y "file2.vcf".

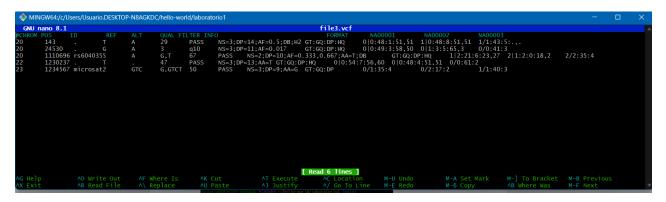
Al usar los signos ">>" los estaremos pasando dentro de file3.vcf

```
Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)
$ nano file3.vcf

Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)
$ head -n 1 file.vcf > cabecera.txt

Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)
$ cat cabecera.txt file2.vcf >> file3.vcf
```

File3.vcf:



Ejercicio 2: (5 ptos)

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

\$ head -n 1 prokaryotes.txt > cabeceraprokaryotes.txt

Todos los genomas de "Akkermansia"

\$ grep "akkermansia" prokaryotes.txt > genesprokaryotes.txt

Para unirlos:

\$ cat cabeceraprokaryotes.txt genesprokaryotes.txt >> akkermansia.txt

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

¿Cómo pueden verse las primeras 10 líneas del archivo?

Head -n 10 akkermansia.txt

¿Cuántos genomas de Akkermansia muciniphila hay?

Hay 1205

```
Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1/ejercicio2 (main) $ grep -c "muciniphila" akkermansia.txt 1205
```

¿Cuántos de Akkermasia biwaensis?

Hay 1

```
Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1/ejercicio2 (main)
$ grep -c "biwaensis" akkermansia.txt
1
```

Ejercicio 3: (1 pto)

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.

```
Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)
$ nano fastq_script.sh

Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)
$ chmod +x fastq_script.sh

Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)
$ ./fastq_script.sh

SRR097977.fastq
996 SRR097977.fastq
Terminado
SRR098026.fastq
996 SRR098026.fastq
Terminado
```

```
#!/bin/bash
for filename in *.fastq
    do echo $filename
    wc -1 $filename
    echo "Terminado"
done
```