

## 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:DP
:HQ      0|0:54:7:56,60 0|0:48:4:51,51 0|0:61:2
20      1234567 microsat1      GTC      G,GTCT      50      PASS      NS=3;DP=9;AA=G G
T:GQ:DP 0|1:35:4      0|2:17:2      1|1:40:3
```

```
Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1 (main)
$ grep "q10" file.vcf
20      17330      .      T      A      3      q10      NS=3;DP=11;AF=0.017      G
T:GQ:DP:HQ      0|0:49:3:58,50 0|1:3:5:65,3 0|0:41:3
```

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:

```
MINGW64/c:/Users/Usuario.DESKTOP-N8AGKDC/hello-world/laboratorio1
GNU nano 8.1 file3.vcf
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA000001 NA000002 NA000003
20 143 . T 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|1:43:5:1...
20 24530 . G A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0|0:41:3
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
22 1230237 . T A 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0|0:61:2
23 1234567 microsat2 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

[ Read 6 Lines ]
^G Help ^O Write Out ^F Where Is ^K Cut ^T Execute ^C Location M-U Undo M-A Set Mark M-] To Bracket M-B Previous
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify ^/_ Go To Line M-E Redo M-6 Copy M-^ Where Was M-F Next
```

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 prokaryotes.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

```
Usuario@DESKTOP-N8AGKDC MINGW64 ~/hello-world/laboratorio1/ejercicio2 (main)
$ head -n 10 akkermansia.txt
#organism Name TaxID BioProject Accession BioProject ID Group SubGroup Size
(Mb) GC% Replicons WGS Scaffolds Genes Proteins Release DateM
Entry Date Status Center BioSample Accession Assembly Accession Reference F
FTP Path Pubmed ID Strain
Akkermansia muciniphila 239935 PRJDB8988 590660 PVC group Verrucomicrobiota 2
.87846 55.5886 chromosome:NZ_AP021898.1/AP021898.1; plasmid pj30893;NZ_AP021899.1/AP021899.1
- 2 2426 2344 2019/11/28 2021/12/16 Complete Genome Laboratory fo
r Microbiome Sciences, Center for Integrative Medical Sciences, RIKEN SAMD00192834 GCA_0
09731575.1 REPR ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/009/731/575/GCF_009731575.
1_ASM973157v1 32054713 JCM 30893
Akkermansia sp. N21116 3040764 PRJNA956280 956280 PVC group Verrucomicrobiota 3
.49718 52.1 chromosome:NZ_CP139035.1/CP139035.1 - 1 2639 2562 2023/
11/27 2023/12/19 Complete Genome Guangxi Academy of Sciences, China SAMN34146709G
CA_029854705.2 - ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/029/854/705/GCF_029854705.
2_ASM2985470v2 - N21116
Akkermansia muciniphila CAG:154 1263034 PRJEB726 206650 PVC group Verrucomicrob
iota 2.72183 55.4 - CBDW01 81 2422 2376 2013/05/31 2015/01/30 S
caffold Technical University of Denmark SAMEA3138704 GCA_000436395.1 - ftp://ftp.ncb
i.nlm.nih.gov/genomes/all/GCA/000/436/395/GCA_000436395.1_MGS154 -
Akkermansia glycaniphila 1679444 PRJEB15121 341763 PVC group Verrucomicrob
iota 3.07408 57.6 chromosome I:NZ_LT629973.1/LT629973.1 - 1 2571 24942
016/10/27 2022/04/19 Complete Genome WAGENINGEN UNIVERSITY, LABORATORY OF SYSTEMS
AND SYNTHETIC BIOLOGY, SAMEA4378782 GCA_900097105.1 REPR ftp://ftp.ncbi.nlm.nih.gov/ge
nomes/all/GCA/900/097/105/GCA_900097105.1_WK001 ApyET
Akkermansia bacterium UBA985 1951376 PRJNA348753 348753 PVC group Verru
comicrobiota 1.94859 50.2 - DBBL01 199 - 2017/09/12 2017/
09/15 Scaffold University of Queensland SAMN06457232 GCA_002292185.1 -
28894102 UBA985
Akkermansia bacterium UBA956 1951375 PRJNA348753 348753 PVC group Verru
comicrobiota 3.38158 51.9 - DBCO01 40 - 2017/09/12 2017/
09/15 Scaffold University of Queensland SAMN06451665 GCA_002291645.1 -
28894102 UBA956
Akkermansia bacterium UBA1009 1951347 PRJNA348753 348753 PVC group Verru
comicrobiota 3.57113 51.1 - DBAK01 147 - 2017/09/12 2017/
09/15 Scaffold University of Queensland SAMN06455474 GCA_002291675.1 -
28894102 UBA1009
Akkermansia bacterium UBA1011 1951348 PRJNA348753 348753 PVC group Verru
comicrobiota 3.20522 52.9 - DBAL01 380 - 2017/09/12 2017/
09/15 Scaffold University of Queensland SAMN06450466 GCA_002291645.1 -
28894102 UBA1011
Akkermansia bacterium UBA1315 1951350 PRJNA348753 348753 PVC group Verru
comicrobiota 3.09047 55 - DBTN01 50 - 2017/09/15 2017/
09/18 Scaffold University of Queensland SAMN06454786 GCA_002307065.1 -
28894102 UBA1315
```

¿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 -l $filename
echo "Terminado"
done
```