

Ejercicio 1: (4 pts)

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**

GNU nano 8.1									
resultadoEjercicio1.txt									
20	14370	rs6054257	G	A	29	PASS	NS=3;DP=14;AF=0.5;DB;H2	GT:GQ:DP:H2	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:H2	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:H2	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 GT:GQ:DP	0/1:35:4 0/2:17:2	1/1:40:3
20	17330	.	T	A	3	q10	NS=3;DP=11;AF=0.017 GT:GQ:DP:H2	0/0:49:3:58,50 0/1:3:5:65,3 0/0:41:3	

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 nano 8.1									
ejercicio2.txt									
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:H2	0/0:48:1:51,51 1/0:48:8:51,51 1/1:43:5:...
20	24530	.	G	A	3	q10	NS=3;DP=11;AF=0.017 GT:GQ:DP:H2	0/0:49:3:58,50 0/1:3:5:65,3 0/0:41:3	
20	1110696	rs6040355	.	A	67	PASS	NS=2;DP=10;AF=0.333,0.667;AA=T;DB	GT:GQ:DP:H2	1/2:21:6:23,27 2/1:2:0:18,2 2/2:35:4
22	1230237	.	T	.	47	PASS	NS=3;DP=13;AA=T GT:GQ:DP:H2	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

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 prokaryotes.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**.

GMI nano 8.1										akkersmansia.txt									
Organism	Accession	Project	Accession	Project	Accession	Project	Accession	Project	Accession	Genes	Proteins	Release Date	Modify Date						
Akkermansia muciniphila	239935	PRJ088988	590660	PVC group	Verrucomicrobiota	2.87846	55.5886	chromosome: NZ_CP021898.1	1	plasmid p30893	NZ_CP021899.1	2013/05/31	2013/12/27						
Akkermansia muciniphila	239935	PRJ088988	590660	PVC group	Verrucomicrobiota	2.87846	55.5886	chromosome: NZ_CP021898.1	1	plasmid p30893	NZ_CP021899.1	2013/05/31	2013/12/27						
Akkermansia muciniphila	CAG:154	1263034	PRJEB726	206650	PVC group	Verrucomicrobiota	2.72183	55.4	CBW01	81	2422	23083	2013/05/31	2015/01/30					
Akkermansia glycaniphila	1679444	PRJEB15121	341763	PVC group	Verrucomicrobiota	3.07408	57.6	chromosome: 1/NZ_L1629973.1	1	1629973.1	1629973.1	2016/10/27	2016/10/27						
Akkermansia bacterium	UBA985	1951376	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.38158	51.2	DBB01	19	-	2017/09/12	2017/09/15						
Akkermansia bacterium	UBA956	1951375	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.38158	51.9	DBCO01	40	-	2017/09/12	2017/09/15						
Akkermansia bacterium	UBA1009	1951347	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.57113	51.1	DBA01	147	-	2017/09/12	2017/09/15						
Akkermansia bacterium	UBA1011	1951348	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.20522	52.9	DBA01	380	-	2017/09/12	2017/09/15						
Akkermansia bacterium	UBA1315	1951350	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.09047	55	DBT01	50	-	2017/09/15	2017/09/18						
Akkermansia bacterium	UBA1090	1951349	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.74617	53.7	DCD01	213	-	2017/09/20	2017/09/20						
Akkermansia bacterium	UBA1507	1951352	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.86348	53.6	DCQ01	230	-	2017/09/21	2017/09/21						
Akkermansia bacterium	UBA1518	1951353	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.6117	51.3	DCQ01	137	-	2017/09/21	2017/09/21						
Akkermansia bacterium	UBA1518	1951353	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.75936	48.5	DCQ01	229	-	2017/09/21	2017/09/21						
Akkermansia bacterium	UBA1506	1951351	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.54126	51.2	DCQ01	142	-	2017/09/21	2017/09/21						
Akkermansia bacterium	UBA1955	1951355	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.23057	55	DBL01	271	-	2017/09/22	2017/09/22						
Akkermansia bacterium	UBA1982	1951356	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.38727	53.2	DDK01	387	-	2017/09/22	2017/09/22						
Akkermansia bacterium	UBA2381	1951358	PRJNA348753	348753	PVC group	Verrucomicrobiota	5.91998	58.1	DSV01	60	-	2017/09/22	2017/09/22						
Akkermansia bacterium	UBA2367	1951357	PRJNA348753	348753	PVC group	Verrucomicrobiota	4.37724	65.4	DDT01	110	-	2017/09/22	2017/09/22						
Akkermansia bacterium	UBA3385	1951360	PRJNA348753	348753	PVC group	Verrucomicrobiota	2.56253	53.9	DEPT01	75	-	2017/09/26	2017/09/26						
Akkermansia bacterium	UBA3157	1951359	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.64389	48.3	DEYN01	237	-	2017/09/26	2017/09/26						
Akkermansia bacterium	UBA4145	1951361	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.47953	53.9	DFW01	376	-	2017/09/27	2017/09/27						
Akkermansia bacterium	UBA4581	1951362	PRJNA348753	348753	PVC group	Verrucomicrobiota	1.64722	38.4	DGK01	57	-	2017/09/28	2017/09/28						
Akkermansia bacterium	UBA589	1951363	PRJNA348753	348753	PVC group	Verrucomicrobiota	1.48955	40.1	DEJ01	63	-	2017/09/28	2017/09/28						
Akkermansia bacterium	UBA5020	1951364	PRJNA348753	348753	PVC group	Verrucomicrobiota	5.37353	61.3	DSX01	144	-	2017/09/28	2017/09/28						
Akkermansia bacterium	UBA5689	1951366	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.80918	54.3	DIHT01	184	-	2017/10/03	2017/10/03						
Akkermansia bacterium	UBA5688	1951365	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.74961	48	DIHU01	269	-	2017/10/03	2017/10/03						
Akkermansia bacterium	UBA6138	1951367	PRJNA348753	348753	PVC group	Verrucomicrobiota	4.2486	60.2	DIV01	92	-	2017/10/04	2017/10/04						
Akkermansia bacterium	UBA6541	1951368	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.35428	53.4	DJKN01	368	-	2017/10/04	2017/10/04						
Akkermansia bacterium	UBA6946	1951369	PRJNA348753	348753	PVC group	Verrucomicrobiota	2.23303	42.6	DJZ01	265	-	2017/10/05	2017/10/05						
Akkermansia bacterium	UBA7331	1951371	PRJNA348753	348753	PVC group	Verrucomicrobiota	2.12465	40.3	DRP01	13	-	2017/10/06	2017/10/06						
Akkermansia bacterium	UBA7329	1951370	PRJNA348753	348753	PVC group	Verrucomicrobiota	2.73417	47.5	DKP01	315	-	2017/10/06	2017/10/06						
Akkermansia bacterium	UBA7811	1951373	PRJNA348753	348753	PVC group	Verrucomicrobiota	3.72067	53.1	DLB01	304	-	2017/10/06	2017/10/06						
Akkermansia bacterium	UBA7421	1951372	PRJNA348753	348753	PVC group	Verrucomicrobiota	1.98106	49.8	DKMF01	241	-	2017/10/06	2017/10/06						
Akkermansia bacterium	UBA7890	1951374	PRJNA348753	348753	PVC group	Verrucomicrobiota	5.04358	59.7	DKM01	215	-	2017/10/10	2017/10/10						
Akkermansia bacterium	2562705	PRJNA642781	642781	PVC group	Verrucomicrobiota	4.38984	61	chromosome: CP059069.1	1	3755	3676	2021/04/06	2021/04/06						
Candidatus Akkermansia intestinalis	2838431	PRJEB53581	850761	PVC group	Verrucomicrobiota	2.33206	63	CALXNG01	59	-	2022/06/19	2022/06/19							
Candidatus Akkermansia intestinalis	2838430	PRJEB53581	850761	PVC group	Verrucomicrobiota	2.50661	64.4	CALXST01	49	-	2022/06/19	2022/06/19							
Akkermansia biwaensis	2946555	PRJDB13654	874222	PVC group	Verrucomicrobiota	3.17552	56.8	chromosome: NZ_CP025943.1	1	2615	2536	2022/08/25	2022/11/1						
Akkermansia massiliensis	2927224	PRJNA715455	715455	PVC group	Verrucomicrobiota	3.11959	58	chromosome: NZ_CP072022.1	1	2624	2558	2021/06/11	2021/06/11						
Candidatus Akkermansia timonensis	308329	PRJNA715455	715455	PVC group	Verrucomicrobiota	3.20874	56.7	chromosome: CP072047.1	1	2666	2583	2021/06/11	2021/06/11						
Candidatus Akkermansia timonensis	308329	PRJNA715455	715455	PVC group	Verrucomicrobiota	3.20872	56.7	chromosome: CP072049.1	1	2676	2583	2021/06/11	2021/06/11						
Akkermansia massiliensis	2927224	PRJNA715455	715455	PVC group	Verrucomicrobiota	3.30971	57.8	chromosome: NZ_CP072027.1	1	2800	2733	2021/06/11	2021/06/11						
Candidatus Akkermansia intestinalis	2838430	PRJNA543206	543206	PVC group	Verrucomicrobiota	2.31341	64.9	DXEH01	92	1836	1794	2021/07/05	2021/07/05						
Candidatus Akkermansia intestinalis	2838431	PRJNA543206	543206	PVC group	Verrucomicrobiota	2.12813	63.4	DBTQ01	174	1813	1770	2021/07/05	2021/07/05						
Akkermansia bacterium	UBA1315	1951350	PRJEB41592	691855	PVC group	Verrucomicrobiota	2.1502	58.1	CAJ1Z01	73	-	2021/01/25	2021/01/25						
Akkermansia glycaniphila	1679444	PRJNA288964	288964	PVC group	Verrucomicrobiota	3.15785	57.7	LIGX01	41	2649	2583	2016/07/12	2022/04/21						
Akkermansia sp.	EB-400x43	3073964	PRJNA1012816	1012816	PVC group	Verrucomicrobiota	3.2085	57.8	chromosome: CP133761.1	1	2848	2756	2023/09/12	2023/09/12					
Akkermansia muciniphila	239935	PRJNA559704	559704	PVC group	Verrucomicrobiota	2.66404	55.8	chromosome: NZ_CP042830.1	1	2213	2135	2019/08/18	2022/01/0						
Akkermansia muciniphila	ATCC BAA-835	349741	PRJNA20089	20089	PVC group	Verrucomicrobiota	2.6641	55.8	chromosome: NC_010655.1	1	2210	2134	2008/05/0	2008/05/0					
Akkermansia sp.	RCC-129D	3115152	PRJNA1066260	1066260	PVC group	Verrucomicrobiota	3.16876	55.9	chromosome: NZ_CP143889.1	1	2681	2603	2024/02/06	2024/02/06					
Akkermansia glycaniphila	1679444	PRJNA62466	62466	PVC group	Verrucomicrobiota	3.01093	57.7	JACVNO1	15	2526	2452	2021/06/08	2022/04/21						
Akkermansia bacterium	2562705	PRJEB41592	691855	PVC group	Verrucomicrobiota	4.36185	53.4	CAJ1BQ01	2	-	2021/01/25	2021/01/25							
Akkermansia massiliensis	2927224	PRJNA715455	715455	PVC group	Verrucomicrobiota	3.16305	57.6	chromosome: NZ_CP072017.1	1	2639	2568	2021/06/11	2021/06/11						

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

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

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

5. ¿Cuántos genomas de Akkermansia muciniphila hay?

Usamos el comando **grep -c "Akkermansia muciniphila" akkersmansia.txt**. Hay **1205** genomas de Akkermansia muciniphila.

6. ¿Cuántos de Akkermasia biwaensis?

Usamos el comando **grep -c "Akkermasia biwaensis" akkersmansia.txt**. 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`**