

LABORATORIO INFORMÁTICA DÍA 1

• EJERCICIO 1

Usa el fichero con formato .vcf file.vcf. Explora el archivo usando el comando head.

El vcf (variant call format) es un formato usado en bioinformática para almacenar aquellas secuencias de genes respecto a un genoma.

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:

Para generar un archivo de texto con los SNPs (los polimorfismos de un solo nucleótido) que han pasado los filtros de calidad (aquellos que tienen el texto PASS), introducimos: `grep "PASS" file.vcf > resultados.txt` y para verlo introducimos el comando: `nano resultados.txt`. Ahora bien, para obtener los que no han pasado el filtro de calidad que son en vez de texto PASS, texto q10, usamos `grep "q10" file.vcf >>resultados.txt`, con esto conseguimos reescribir el archivo PASS debido al doble >> y añadir los que no han pasado el filtro de calidad.

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.

Vamos a coger la cabecera con el comando `grep "^#" file.vcf > head.vcf`, con este cogemos las líneas que empiezan con # que solo es la cabecera y lo guardamos y lo llamamos "head.vcf", para luego con el comando `cat head.vcf file2.vcf > merged.vcf` juntar tanto lo que teníamos en head.vcf como el archivo file2.vcf y lo hemos juntado en un nuevo archivo llamado merged.vcf. Lo podemos ver entero utilizando `nano merged.vcf`.

• EJERCICIO 2

Preparar un directorio de trabajo, llamado ejercicio2.

Lo hacemos con `mkdir ejercicio2`.

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 entramos en ejercicio2 (cd ejercicio 2), luego creamos el archivo akkermansia.txt lo podemos hacer con `nano akkermansia.txt`, así vemos que este vacío. Para obtener la cabecera del archivo original prokaryots.txt usamos: `grep "TaxID" prokaryotes.txt > akkermansia.txt`. Si queremos verlo utilizamos nano. Para introducir los genomas de "Akkermasia" utilizamos: `grep "Akkermansia" prokaryotes.txt >> akkermansia.txt`.

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

- ¿Cómo pueden verse las primeras 10 líneas del archivo? Con `head akkermansia.txt`
- ¿Cuántos genomas de Akkermansia muciniphila hay? Con `grep -c "Akkermansia muciniphila" akkermansia.txt` vemos que hay 1205.
- ¿Cuántos de Akkermasia biwaensis? Con `grep -c "Akkermansia biwaensis" akkermansia.txt` vemos que hay 1.