

Ejercicio 1.10.1

HP@David MINGW64 ~

\$ cd Downloads/CSB/unix/sandbox/

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ ls

ExtractionBodyM.sh 'Papers and reviews/' sandbox testl.txt

Marra2014_data.fasta my_file.fasta test.txt

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ rm my_file.fasta

rm: cannot remove 'my_file': No such file or directory

rm: cannot remove 'fasta': No such file or directory

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ rm my_file.fasta

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ ls

ExtractionBodyM.sh 'Papers and reviews/' test.txt

Marra2014_data.fasta sandbox testl.txt

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ mv Marra2014_data.fasta ../data

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ cd ../data

HP@David MINGW64 ~/Downloads/CSB/unix/data

```
$ ls
```

```
BodyM.csv      Gesquiere2011_about.txt Pacifici2013_about.txt data
Buzzard2015_about.txt Gesquiere2011_data.csv Pacifici2013_data.csv miRNA/
Buzzard2015_data.csv Marra2014_about.txt  Saavedra2013/      sandbox
ExtractionBody.sh  Marra2014_data.fasta  Saavedra2013_about.txt
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/data
```

```
$ #ejercicio 1.10.1
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/data
```

```
$ #1.Cambie el directorio a CSB/unix/sandbox
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/data
```

```
$ mv m
```

```
Marra2014_about.txt  Marra2014_data.fasta  miRNA/
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/data
```

```
$ mv Marra2014_data.fasta ../sandbox/
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/data
```

```
$ ls
```

```
BodyM.csv      Gesquiere2011_about.txt Pacifici2013_data.csv miRNA/
Buzzard2015_about.txt Gesquiere2011_data.csv Saavedra2013/      sandbox
Buzzard2015_data.csv Marra2014_about.txt  Saavedra2013_about.txt
ExtractionBody.sh  Pacifici2013_about.txt data
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/data
```

```
$ cd ../sandbox/
```

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ ls

ExtractionBodyM.sh 'Papers and reviews'/ test.txt

Marra2014_data.fasta sandbox testl.txt

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ #2 ¿Cual es el tamaño del archivo Marra2014_dat.fasta?

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ #uso el comando -lh

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ ls -lh Marra2014_data.fasta

-rw-r--r-- 1 HP 197609 553K Jan 28 2019 Marra2014_data.fasta

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ #el tamaño del archivo es de 553k

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ #3. Cree una copia de Marra2014_data.fasta y nombrela como: my_file.fasta

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ #uso el comando cp, que me sirve para copiar

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

\$ cp ../Marra2014_data.fasta my_file.fasta

cp: cannot stat '../Marra2014_data.fasta': No such file or directory

HP@David MINGW64 ~/Downloads/CSB/unix/sandbox

```
$ cp Marra2014_data.fasta my_file.fasta
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
```

```
$ ls
```

```
ExtractionBodyM.sh 'Papers and reviews/' sandbox testl.txt
```

```
Marra2014_data.fasta my_file.fasta test.txt
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
```

```
$ #4 ¿Cuántos contigs están clasificados como "isogroup00036?"
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
```

```
$ #uso el comando grep y lo intercalo con wc -l
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
```

```
$ grep isogroup00036 my_file.fasta | wc -l
```

```
16
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
```

```
$ # tengo 16 contigs
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
```

```
$ #5. Sustituya el delimitador original de " dos espacios por una coma"
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
```

```
$ # uso los comandos cat, tr y head (intercalados)
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
```

```
$ cat my_file.fasta | tr -s ' ', | head -n 3
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox  
$ #5
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ cat my_file.fasta | tr -s ' ','' | head -n 3
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAAAGTCAGCTCAAGCAAGAGATTTG
TTTACAATTAACCCACAAAAGGCTGTTACTGAAGGTGTGGCTTAAGTGTCAAGCAACAG
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ #6. ¿cuantos unique isogroups hay en el archivo?
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ grep '>' my_file.fasta | head -n 3
>contig00001 length=527 numreads=2 gene=isogroup00001 status=it_thresh
>contig00002 length=551 numreads=8 gene=isogroup00001 status=it_thresh
>contig00003 length=541 numreads=2 gene=isogroup00001 status=it_thresh
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ grep '>' my_file.fasta | head -n 2
>contig00001 length=527 numreads=2 gene=isogroup00001 status=it_thresh
>contig00002 length=551 numreads=8 gene=isogroup00001 status=it_thresh
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 4 | head -n 2
>contig00001 length=527 numreads=2 gene=isogroup00001 status=it_thresh
>contig00002 length=551 numreads=8 gene=isogroup00001 status=it_thresh
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 4 | sort | uniq | wc -l
955
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ ¿ que conting tiene el mayor numero de lecturas?
bash: ¿: command not found
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ #7¿ que conting tiene el mayor numero de lecturas?
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ #usamos grep y cut, para aislar el nukm de lecturas
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 1,3 | head -n 3
>contig00001 length=527 numreads=2 gene=isogroup00001 status=it_thresh
>contig00002 length=551 numreads=8 gene=isogroup00001 status=it_thresh
>contig00003 length=541 numreads=2 gene=isogroup00001 status=it_thresh
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ #usamos el comando '=' para dividir segun el signo
```

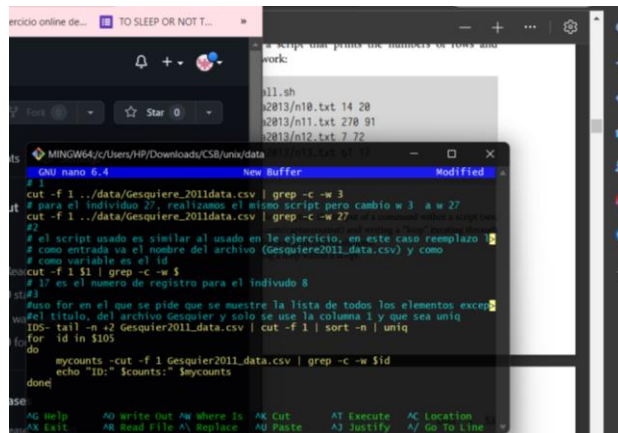
```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 1,3 | sort -t '=' -k 2 -n | head -n 5
>contig00090 length=100 numreads=3 gene=isogroup00001 status=it_thresh
>contig00719 length=100 numreads=49 gene=isogroup00014 status=isotig
>contig01037 length=100 numreads=38 gene=isogroup00027 status=it_thresh
>contig01049 length=100 numreads=20 gene=isogroup00027 status=it_thresh
>contig00309 length=101 numreads=1807 gene=isogroup00002 status=it_thresh
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$ grep '>' my_file.fasta | cut -d ',' -f 1,3 | sort -t '=' -k 2 -n -r | head -n 5
>contig01115 length=6087 numreads=185 gene=isogroup00030 status=isotig
>contig00011 length=5563 numreads=61 gene=isogroup00001 status=it_thresh
>contig00239 length=4674 numreads=48 gene=isogroup00002 status=it_thresh
>contig00238 length=4284 numreads=27 gene=isogroup00002 status=it_thresh
>contig01040 length=4129 numreads=368 gene=isogroup00027 status=it_thresh
```

```
HP@David MINGW64 ~/Downloads/CSB/unix/sandbox
$
```

Especificado en README.md Elegir la misma ruta del repositorio anterior
 Corremos la informacion
 CD Saavedra2013
 # Ejecución el codigo en git bush
 Utilizaremos el for, do, done
 para archivo en \$(ls *.txt) ; hacer wc -l \$archivo ; head -n1 \$archivo | grep
 -o " " | wc -l ; hecho
 # Finalizar obtenemos los resultados de netsize_all.txt con archivos de n59.txt
 Ejercicio 1.10.3

```
GNU nano 6.4 New Buffer Modified
# 1
cut -f 1 ../data/Gesquiere_2011data.csv | grep -c -w 3
# para el individuo 27, realizamos el mismo script pero cambio w 3 a w 27
cut -f 1 ../data/Gesquiere_2011data.csv | grep -c -w 27
#2
# el script usado es similar al usado en le ejercicio, en este caso reemplazo l
# como entrada va el nombre del archivo (Gesquiere2011_data.csv) y como
# como variable es el id
cut -f 1 $1 | grep -c -w $
# 17 es el numero de registro para el indivudo 8
#3
#uso for en el que se pide que se muestre la lista de todos los elementos excep
#el titulo, del archivo Gesquier y solo se use la columna 1 y que sea uniq
IDS- tail -n +2 Gesquier2011_data.csv | cut -f 1 | sort -n | uniq
for id in $(IDS)
do
  mycounts -cut -f 1 Gesquier2011_data.csv | grep -c -w $id
  echo "ID:" $counts:" $mycounts
done
```



nano captura de los codigos corridos