

# BIOINFORMÁTICA

Nombre: Dayan Ochoa Oviedo

Fecha: 18 de noviembre de 2022

## Ejercicio 1:

```
MINGW64:/c/Users
pablo@papacito MINGW64 ~
$ cd desktop

pablo@papacito MINGW64 ~/desktop
$ mkdir CSB

pablo@papacito MINGW64 ~/desktop
$ cd CSB

pablo@papacito MINGW64 ~/desktop/CSB
$ mkdir unix

pablo@papacito MINGW64 ~/desktop/CSB
$ cd unix

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ mkdir sandbox

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio 1.10.1

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio 1.10.1.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio_1.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio_1.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat Ejercicio_1.sh
echo 'Tamaño del archivo Marra2014_data.fasta.'
ls -lh ../../CSB-master/unix/data/Marra2014_data.fasta
echo ''
echo 'Creación de copia de Marra2014_data.fasta en sandbox con nombre my_file.fasta.'
cp ../../CSB-master/unix/data/Marra2014_data.fasta sandbox/my_file.fasta
ls sandbox
echo ''
echo 'Reemplazo los 2 espacios originales por una coma.'
cat sandbox/my_file.fasta | tr -s ' ' ',' | head -n 5
echo ''
echo 'Número de isogrupos unicos en el archivo.'
grep '>' sandbox/my_file.fasta | tr -s ' ' ',' | cut -d ',' -f 4 | uniq | wc -l
echo ''
echo 'Contig con el mayor número y cantidad de lecturas.'
grep '>' sandbox/my_file.fasta | tr -s ' ' ',' | cut -d ',' -f 1,3 | sort -t '=' -k 2 -n -r
| head -n 1
```

```
MINGW64:/c/Users
$ nano Ejercicio 1.10.1.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio_1.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio_1.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat Ejercicio_1.sh
echo 'Tamaño del archivo Marra2014_data.fasta.'
ls -lh ../../CSB-master/unix/data/Marra2014_data.fasta
echo ''
echo 'Creación de copia de Marra2014_data.fasta en sandbox con nombre my_file.fasta.'
cp ../../CSB-master/unix/data/Marra2014_data.fasta sandbox/my_file.fasta
ls sandbox
echo ''
echo 'Reemplazo los 2 espacios originales por una coma.'
cat sandbox/my_file.fasta | tr -s ' ' ',' | head -n 5
echo ''
echo 'Número de isogrupos unicos en el archivo.'
grep '>' sandbox/my_file.fasta | tr -s ' ' ',' | cut -d ',' -f 4 | uniq | wc -l
echo ''
echo 'Contig con el mayor número y cantidad de lecturas.'
grep '>' sandbox/my_file.fasta | tr -s ' ' ',' | cut -d ',' -f 1,3 | sort -t '=' -k 2 -n -r
| head -n 1

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ bash Ejercicio_1.sh
Tamaño del archivo Marra2014_data.fasta.
-rw-r--r-- 1 pablo 197610 553K Jan 28 2019 ../../CSB-master/unix/data/Marra2014_data.fasta

Creación de copia de Marra2014_data.fasta en sandbox con nombre my_file.fasta.
my_file.fasta test.sh

Reemplazo los 2 espacios originales por una coma.
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAGGTCAGCTCAAGCAAGGATTTG
TTTACAATTAACCCACAAAGGCGTTTACTGAAGGTGTGGCTTAAGTGTGACAGCAACAG
CTATGAGTGGAGGAATTTTCTATTACAATATAATTTCATCTCTGGTAAATTGACCAATTA
ACTGGAACATTTTCCAACCTGAAATAAATGGTAAACITTTTATCCACCATTCGCCATCTG

Número de isogrupos unicos en el archivo.
43

Contig con el mayor número y cantidad de lecturas.
>contig00302,numreads=3330

pablo@papacito MINGW64 ~/desktop/CSB/unix
$
```

## Ejercicio 2:

```
MINGW64:/c/Users
pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio_2.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano sandbox/Parte1.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano sandbox/Parte2.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat Ejercicio_2.sh
echo 'Gesquiere et al . (2011)'
head -n 4 ../../CSB-master/unix/data/Gesquiere2011_data.csv
echo ''
echo 'Número de registros de los individuos 3 y 27'
echo '- Número de registros del individuo 3'
cut -f 1 ../../CSB-master/unix/data/Gesquiere2011_data.csv | grep -c -w 3
echo ''
echo '- Número de registros del individuo 27'
cut -f 1 ../../CSB-master/unix/data/Gesquiere2011_data.csv | grep -c -w 27
echo ''
echo 'Introducir nombre e ID:'
bash sandbox/Parte1.sh ../../CSB-master/unix/data/Gesquiere2011_data.csv 3
echo ''
echo 'Introducir ID para conocer el número de muestras tomadas:'
bash sandbox/Parte2.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat sandbox/Parte1.sh
cut -f 1 $1 | grep -c -w $2

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat sandbox/Parte2.sh
IDS=$(tail -n +2 ../../CSB-master/unix/data/Gesquiere2011_data.csv | cut -f 1 | sort -n | uniq)
for id in $IDS
do
    conteo=$(bash sandbox/Parte1.sh ../../CSB-master/unix/data/Gesquiere2011_data.csv $id)
    echo 'ID:' $id ', Número de muestras:' $conteo
done

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ bash Ejercicio_2.sh
Gesquiere et al . (2011)
maleID  GC      T
1      66.9    64.57
1      51.09   35.57
1      65.89   114.28
```

```
MINGW64:/c/Users
pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat sandbox/Parte1.sh
cut -f 1 $1 | grep -c -w $2

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat sandbox/Parte2.sh
IDS=$(tail -n +2 ../../CSB-master/unix/data/Gesquiere2011_data.csv | cut -f 1 | sort -n | uniq)
for id in $IDS
do
    conteo=$(bash sandbox/Parte1.sh ../../CSB-master/unix/data/Gesquiere2011_data.csv $id)
    echo 'ID:' $id ', Número de muestras:' $conteo
done

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ bash Ejercicio_2.sh
Gesquiere et al . (2011)
maleID  GC      T
1      66.9    64.57
1      51.09   35.57
1      65.89   114.28

Número de registros de los individuos 3 y 27
- Número de registros del individuo 3
61

- Número de registros del individuo 27
5

Introducir nombre e ID:
61

Introducir ID para conocer el número de muestras tomadas:
ID: 1 , Número de muestras: 10
ID: 2 , Número de muestras: 2
ID: 3 , Número de muestras: 61
ID: 4 , Número de muestras: 46
ID: 5 , Número de muestras: 28
ID: 6 , Número de muestras: 7
ID: 7 , Número de muestras: 5
ID: 8 , Número de muestras: 17
ID: 9 , Número de muestras: 4
ID: 10 , Número de muestras: 21
ID: 11 , Número de muestras: 26
ID: 12 , Número de muestras: 23
ID: 13 , Número de muestras: 16
ID: 14 , Número de muestras: 1
ID: 15 , Número de muestras: 40
```

### Ejercicio 3:

```
pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio_3.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano sandbox/Parte3.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat Ejercicio_3.sh
echo 'Número de filas y columnas en Saavedra2013 n1'
echo '- Número de filas:'
wc -l ../CSB-master/unix/data/Saavedra2013/n1.txt
echo ''
echo '- Número de columnas'
head -n 1 ../CSB-master/unix/data/Saavedra2013/n1.txt | wc -w
echo ''
echo 'Número de filas y columnas para cada archivo'
bash sandbox/Parte3.sh
echo ''
echo 'Archivo con más filas'
bash sandbox/Parte3.sh | sort -n -r -k 2 | head -n 1
echo 'Archivo con más columnas'
bash sandbox/Parte3.sh | sort -n -r -k 3 | head -n 1

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat sandbox/Parte3.sh
datos=../CSB-master/unix/data/Saavedra2013/*.txt
for item in $datos
do
    filas=`cat $item | wc -l`
    columnas=`head -n 1 $item | tr -d ' ' | tr -d '\n' | wc -c`
    echo $item $filas $columnas
done

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio_3.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ bash Ejercicio_3.sh
Número de filas y columnas en Saavedra2013 n1
- Número de filas:
97 ../CSB-master/unix/data/Saavedra2013/n1.txt

- Número de columnas
80

Número de filas y columnas para cada archivo
../CSB-master/unix/data/Saavedra2013/n1.txt 97 80
../CSB-master/unix/data/Saavedra2013/n10.txt 14 20
../CSB-master/unix/data/Saavedra2013/n11.txt 270 91
```

```
../CSB-master/unix/data/Saavedra2013/n25.txt 17 16
../CSB-master/unix/data/Saavedra2013/n26.txt 82 40
../CSB-master/unix/data/Saavedra2013/n27.txt 27 5
../CSB-master/unix/data/Saavedra2013/n28.txt 90 19
../CSB-master/unix/data/Saavedra2013/n29.txt 61 25
../CSB-master/unix/data/Saavedra2013/n3.txt 25 36
../CSB-master/unix/data/Saavedra2013/n30.txt 8 19
../CSB-master/unix/data/Saavedra2013/n31.txt 28 25
../CSB-master/unix/data/Saavedra2013/n32.txt 45 21
../CSB-master/unix/data/Saavedra2013/n33.txt 70 20
../CSB-master/unix/data/Saavedra2013/n34.txt 79 25
../CSB-master/unix/data/Saavedra2013/n35.txt 14 8
../CSB-master/unix/data/Saavedra2013/n36.txt 40 169
../CSB-master/unix/data/Saavedra2013/n37.txt 44 13
../CSB-master/unix/data/Saavedra2013/n38.txt 51 99
../CSB-master/unix/data/Saavedra2013/n39.txt 33 25
../CSB-master/unix/data/Saavedra2013/n4.txt 101 11
../CSB-master/unix/data/Saavedra2013/n40.txt 28 18
../CSB-master/unix/data/Saavedra2013/n41.txt 12 10
../CSB-master/unix/data/Saavedra2013/n42.txt 42 8
../CSB-master/unix/data/Saavedra2013/n43.txt 55 29
../CSB-master/unix/data/Saavedra2013/n44.txt 56 9
../CSB-master/unix/data/Saavedra2013/n45.txt 36 61
../CSB-master/unix/data/Saavedra2013/n46.txt 58 17
../CSB-master/unix/data/Saavedra2013/n47.txt 139 41
../CSB-master/unix/data/Saavedra2013/n48.txt 118 49
../CSB-master/unix/data/Saavedra2013/n49.txt 47 23
../CSB-master/unix/data/Saavedra2013/n5.txt 21 7
../CSB-master/unix/data/Saavedra2013/n50.txt 45 46
../CSB-master/unix/data/Saavedra2013/n51.txt 8 15
../CSB-master/unix/data/Saavedra2013/n52.txt 33 7
../CSB-master/unix/data/Saavedra2013/n53.txt 34 13
../CSB-master/unix/data/Saavedra2013/n54.txt 126 25
../CSB-master/unix/data/Saavedra2013/n55.txt 14 50
../CSB-master/unix/data/Saavedra2013/n56.txt 110 207
../CSB-master/unix/data/Saavedra2013/n57.txt 14 11
../CSB-master/unix/data/Saavedra2013/n58.txt 678 90
../CSB-master/unix/data/Saavedra2013/n59.txt 663 130
../CSB-master/unix/data/Saavedra2013/n6.txt 9 31
../CSB-master/unix/data/Saavedra2013/n7.txt 16 25
../CSB-master/unix/data/Saavedra2013/n8.txt 19 33
../CSB-master/unix/data/Saavedra2013/n9.txt 0 0

Archivo con más filas
../CSB-master/unix/data/Saavedra2013/n58.txt 678 90
Archivo con más columnas
../CSB-master/unix/data/Saavedra2013/n56.txt 110 207

pablo@papacito MINGW64 ~/desktop/CSB/unix
$
```

## Ejercicio 4

```
pablo@papacito MINGW64 ~/desktop/CSB/unix
$ nano Ejercicio_4.sh

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ cat Ejercicio_4.sh
echo 'Nombre de la columna correspondiente'
cut -d ',' -f $2 $1 | head -n 1
echo ''
echo 'Número de los distintos valores en la columna'
cut -d ',' -f $2 $1 | tail -n +2 | sort | uniq | wc -l
echo ''
echo 'Valor mínimo'
cut -d ',' -f $2 $1 | tail -n +2 | sort -n | head -n 1
echo ''
echo 'valor máximo'
cut -d ',' -f $2 $1 | tail -n +2 | sort -n -r | head -n 1

pablo@papacito MINGW64 ~/desktop/CSB/unix
$ bash Ejercicio_4.sh ../../CSB-master/unix/data/Buzzard2015_data.csv 7
Nombre de la columna correspondiente
biomass

Número de los distintos valores en la columna
285

Valor mínimo
1.048466198

valor máximo
14897.29471

pablo@papacito MINGW64 ~/desktop/CSB/unix
$
```