

Nombre: Josué Leonardo Chango Gómez

Grupo: 01

Fecha: 18/11/2022

## 1.10 Exercises

### 1.10.1 Next Generation Sequencing Data

- Change directory to CSB/unix/sandbox.

```
MINGW64:/c/Users/TOSHIBA/Downloads/CSB-master/unix/sandbox
```

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~  
$ cd Downloads/CSB-master/unix/sandbox/
```

- What is the size of the file Marra2014\_data.fasta?

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)  
$ ls -lh ../data/Marra2014_data.fasta  
-rw-r--r-- 1 TOSHIBA 197121 553K Jan 28 2019 ../data/Marra2014_data.fasta
```

- Create a copy of Marra2014\_data.fasta in the sandbox and name it my\_file.fasta.

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)  
$ cp ../data/Marra2014_data.fasta my_file.fasta  
  
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)  
$ ls  
'Papers and reviews'/' clase4_until.sh* f1/ f3/ my_file.fasta  
clase4_if.sh* clase4_while.sh* f2/ f4/
```

- How many contigs are classified as isogroup00036?

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)  
$ grep -c isogroup00036 my_file.fasta  
16  
  
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)  
$ grep isogroup00036 my_file.fasta | wc -l  
16
```

- Replace the original "two-spaces" delimiter with a comma.

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)  
$ cat my_file.fasta | tr -s ' ' ',' | head -n 3  
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh  
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAAAGTCAGCTCAAGCAAGAGATTTG  
TTTACAATTAACCCACAAAAGGCTGTTACTGAAGGTGTGGCTTAAGTGTCAGAGCAACAG
```

- How many unique isogroups are in the file?

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)  
$ 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  
  
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)  
$ grep '>' my_file.fasta | cut -d ',' -f 4 | sort | uniq | wc -l  
955
```

- Which contig has the highest number of reads (numreads)? How many reads does it have?

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)
$ 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

TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)
$ 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

TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/sandbox (master)
$ grep '>' my_file.fasta | cut -d ',' -f 1,3 | sort -t '=' -k 2 -n -r | head -n 1
>contig01115 length=6087 numreads=185 gene=isogroup00030 status=isotig
```

### 1.10.2 Hormone Levels in Baboons

- How many times were the levels of individuals 3 and 27 recorded?

MINGW64/c:/Users/TOSHIBA/Downloads/CSB-master/unix/data

```
GNU nano 6.2 Ejercicio_1-10-2.sh
#!/bin/bash

#¿Cuántas veces fueron los niveles de los individuos 3 y 27 grabados?

#Para los individuos 3
cut -f 1 Gesquiere2011_data.csv | grep -w 3 | grep -c 3

#Para los individuos 27
cut -f 1 Gesquiere2011_data.csv | grep -w 27 | grep -c 27
```

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/data (main)
$ nano Ejercicio_1-10-2.sh

TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/data (main)
$ cat Ejercicio_1-10-2.sh
#!/bin/bash

#¿Cuántas veces fueron los niveles de los individuos 3 y 27 grabados?

#Para los individuos 3
cut -f 1 Gesquiere2011_data.csv | grep -w 3 | grep -c 3

#Para los individuos 27
cut -f 1 Gesquiere2011_data.csv | grep -w 27 | grep -c 27

TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/data (main)
$ bash Ejercicio_1-10-2.sh
61
5
```

- Write a script taking as input the file name and the ID of the individual, and returning the number of records for that ID.

MINGW64:/c/Users/TOSHIBA/Downloads/CSB-master/unix/data

```
GNU nano 6.2 Ejercicio_1-10-2.sh
#!/bin/bash

#¿Cuántas veces fueron los niveles de los individuos 3 y 27 grabados?
#Para los individuos 3
cut -f 1 Gesquiere2011_data.csv | grep -w 3 | grep -c 3
#Para los individuos 27
cut -f 1 Gesquiere2011_data.csv | grep -w 27 | grep -c 27
#Luego se crea un vector para los IDs del documento
vdownloads=$(tail -n +2 Gesquiere2011_data.csv | cut -f 1 | uniq)
for x in $vdownloads
do
ids=$(bash Conteo_Ejercicio_1-10-2.sh Gesquiere2011_data.csv $x)
echo "ID:" $x "conteo:" $ids
done
```

MINGW64:/c/Users/TOSHIBA/Downloads/CSB-master/unix/data


```
GNU nano 6.2 Conteo_Ejercicio_1-10-2.sh
#!/bin/bash

#Codigo con sus respectivas variables
cut -f 1 $1 | grep -c -w $2
```

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/data (main)
$ bash Ejercicio_1-10-2.sh
61
5
ID: 1 conteo: 10
ID: 2 conteo: 2
ID: 3 conteo: 61
ID: 4 conteo: 46
ID: 5 conteo: 28
ID: 6 conteo: 7
ID: 7 conteo: 5
ID: 8 conteo: 17
ID: 9 conteo: 4
ID: 10 conteo: 21
ID: 11 conteo: 26
ID: 12 conteo: 23
ID: 13 conteo: 16
ID: 14 conteo: 1
ID: 15 conteo: 40
ID: 16 conteo: 31
ID: 17 conteo: 3
ID: 18 conteo: 4
ID: 19 conteo: 3
ID: 20 conteo: 4
ID: 21 conteo: 12
ID: 22 conteo: 5
ID: 23 conteo: 36
ID: 24 conteo: 35
ID: 25 conteo: 35
ID: 26 conteo: 22
ID: 27 conteo: 5
ID: 29 conteo: 33
ID: 30 conteo: 63
ID: 31 conteo: 1
ID: 32 conteo: 3
ID: 33 conteo: 1
ID: 34 conteo: 16
ID: 35 conteo: 5
ID: 36 conteo: 39
ID: 37 conteo: 38
ID: 38 conteo: 1
ID: 39 conteo: 3
ID: 40 conteo: 32
ID: 41 conteo: 53
ID: 42 conteo: 5
ID: 43 conteo: 2
ID: 44 conteo: 56
ID: 45 conteo: 1
ID: 46 conteo: 24
```


### 1.10.3 Plant–Pollinator Networks

- Write a script that takes one of these files and determines the number of rows (pollinators) and columns (plants). Note that columns are separated by spaces and that there is a space at the end of each line. Your script should return.

 MINGW64:/c/Users/TOSHIBA/Downloads/CSB-master/unix/data

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/data (main)
$ nano netsize.sh

TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/data (main)
$ bash netsize.sh
Archivo: n1.txt
Numero de filas:
97
Numero de columnas:
80
Archivo: n59.txt
Numero de filas:
663
Numero de columnas:
130
```

 MINGW64:/c/Users/TOSHIBA/Downloads/CSB-master/unix/data


```
GNU nano 6.2
#!/bin/bash

cd 'C:\Users\TOSHIBA\Downloads\CSB-master\unix\data\Saavedra2013'

echo "Archivo:" n1.txt
echo "Numero de filas:"
wc -l n1.txt| cut -d " " -f1 $1
echo "Numero de columnas:"
head -n 1 n1.txt|grep -o " "|wc -l $2

echo "Archivo:" n59.txt
echo "Numero de filas:"
wc -l n59.txt| cut -d " " -f1 $1
echo "Numero de columnas:"
head -n 1 n59.txt|grep -o " "|wc -l $2
```

- Write a script that prints the numbers of rows and columns for each network:

 MINGW64: c:/Users/TOSHIBA/Downloads/CSB-master/unix/data

```
TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/data (main)
$ ls
Buzzard2015_about.txt      Marra2014_data.fasta      filtrado.txt
Buzzard2015_data.csv      Pacifici2013_about.txt    mirNA/
Conteo_Ejercicio_1-10-2.sh* Pacifici2013_data.csv      netsize.sh*
Ejercicio_1-10-2.sh*      Saavedra2013/             netsize_all.sh
Gesquiere2011_about.txt   Saavedra2013_about.txt    netsize_all.txt
Gesquiere2011_data.csv    clase4_for.sh*
Marra2014_about.txt       clase4_for.sh.save*

TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/data (main)
$ bash netsize_all.sh

TOSHIBA@DESKTOP-UDFJCR5 MINGW64 ~/Downloads/CSB-master/unix/data (main)
$ cat netsize_all.txt
El numero de filas y columnas de Saavedra2013/n1.txt son: 97, 80.
El numero de filas y columnas de Saavedra2013/n10.txt son: 14, 20.
El numero de filas y columnas de Saavedra2013/n11.txt son: 270, 91.
El numero de filas y columnas de Saavedra2013/n12.txt son: 7, 72.
El numero de filas y columnas de Saavedra2013/n13.txt son: 61, 17.
El numero de filas y columnas de Saavedra2013/n14.txt son: 35, 15.
El numero de filas y columnas de Saavedra2013/n15.txt son: 38, 11.
El numero de filas y columnas de Saavedra2013/n16.txt son: 118, 24.
El numero de filas y columnas de Saavedra2013/n17.txt son: 76, 31.
El numero de filas y columnas de Saavedra2013/n18.txt son: 13, 14.
El numero de filas y columnas de Saavedra2013/n19.txt son: 10, 16.
El numero de filas y columnas de Saavedra2013/n2.txt son: 62, 41.
El numero de filas y columnas de Saavedra2013/n20.txt son: 18, 7.
El numero de filas y columnas de Saavedra2013/n21.txt son: 19, 45.
El numero de filas y columnas de Saavedra2013/n22.txt son: 19, 36.
El numero de filas y columnas de Saavedra2013/n23.txt son: 179, 26.
El numero de filas y columnas de Saavedra2013/n24.txt son: 80, 28.
El numero de filas y columnas de Saavedra2013/n25.txt son: 17, 16.
El numero de filas y columnas de Saavedra2013/n26.txt son: 82, 40.
El numero de filas y columnas de Saavedra2013/n27.txt son: 27, 5.
El numero de filas y columnas de Saavedra2013/n28.txt son: 90, 19.
El numero de filas y columnas de Saavedra2013/n29.txt son: 61, 25.
El numero de filas y columnas de Saavedra2013/n3.txt son: 25, 36.
El numero de filas y columnas de Saavedra2013/n30.txt son: 8, 19.
El numero de filas y columnas de Saavedra2013/n31.txt son: 28, 25.
El numero de filas y columnas de Saavedra2013/n32.txt son: 45, 21.
El numero de filas y columnas de Saavedra2013/n33.txt son: 70, 20.
El numero de filas y columnas de Saavedra2013/n34.txt son: 79, 25.
El numero de filas y columnas de Saavedra2013/n35.txt son: 14, 8.
El numero de filas y columnas de Saavedra2013/n36.txt son: 40, 169.
El numero de filas y columnas de Saavedra2013/n37.txt son: 44, 13.
El numero de filas y columnas de Saavedra2013/n38.txt son: 51, 99.
```