

Universidad Regional Amazónica Ikiam

Bioinformática – GBI6

Evidencias de la actividad TC3\_ForLoop

Nombre: Josué Leonardo Chango Gómez

Grupo: 02


Fecha: 09/05/2023

*Computing Skills for Biologist*

1.10 Exercises

*1.10.1 Next Generation Sequencing Data*


- Change directory to CSB/unix/sandbox.

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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~
$ echo "Josué Leonardo Chango Gómez"
Josué Leonardo Chango Gómez
```

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~
$ cd Downloads/CSB-master/CSB-master/unix/sandbox/
```


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

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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ echo "Josué Leonardo Chango Gómez"
Josué Leonardo Chango Gómez
```

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ 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.


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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ echo "Josué Leonardo Chango Gómez"
Josué Leonardo Chango Gómez
```

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ cp ../data/Marra2014_data.fasta my_file.fasta
```

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ ls
Actividad.csv      Hola.csv          contador.sh       familia.txt       josue.txt        my_file.fasta    orden.txt        pikachu.csv
Buzzard2015.sh    'Papers and reviews'  contador2.sh     g02.sh           miRNA/          orden.csv        otro.txt        tarea1/
```

- How many contigs are classified as isogroup00036?


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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ echo "Josué Leonardo Chango Gómez"
Josué Leonardo Chango Gómez

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ grep -c isogroup00036 my_file.fasta | wc -l
1

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ grep -c isogroup00036 my_file.fasta
16
```


- Replace the original “two-spaces” delimiter with a comma.

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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ echo "Josué Leonardo Chango Gómez"
Josué Leonardo Chango Gómez

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ cat my_file.fasta | tr -s " " "," | head -n 3
>contig00001,length=527,numreads=2,gene=isogroup00001,status=it_thresh
ATCCTAGCTACTCTGGAGACTGAGGATTGAAGTTCAAAGTCAGCTCAAGCAAGAGATTTG
TTTACAATTAACCCACAAAAGGCTGTTACTGAAGGTGTGGCTTAAGTGTGAGAGCAACAG
```

- How many unique isogroups are in the file?

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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ echo "Josué Leonardo Chango Gómez"
Josué Leonardo Chango Gómez

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ 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?

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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ echo "Josué Leonardo Chango Gómez"
Josué Leonardo Chango Gómez

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/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

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d "," -f 1,3 | sort -t "-" -k 2 -n
.gitignore          contador.sh          miRNA/              pikachu.csv
Actividad.csv       contador2.sh        my_file.fasta       tarea1/
Buzzard2015.sh      familia.txt         orden.csv           orden.txt
Hola.csv            g02.sh             orden.txt
Papers and reviews/ josue.txt           otro.txt

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d "," -f 1,3 | sort -t "-" -k 2 -n | head -n 5
>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
>contig00004 length=291 numreads=3 gene=isogroup00001 status=it_thresh
>contig00005 length=580 numreads=12 gene=isogroup00001 status=it_thresh

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/sandbox
$ grep '>' my_file.fasta | cut -d "," -f 1,3 | sort -t "-" -k 2 -n -r | head -n 1
>contig01385 length=1965 numreads=7 gene=isogroup00043 status=isitig
```

### 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/CSB-master/unix/data

```
GNU nano 7.2 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
```

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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/data
$ nano Ejercicio_1_10_2.sh

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/data
$ 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-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/data
$ 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/CSB-master/unix/data

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/data
$ 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
```

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

```
GNU nano 7.2 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/CSB-master/unix/data

```
GNU nano 7.2 Conteo_Ejercicio_1-10-2.sh
#!/bin/bash

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

### 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/CSB-master/unix/data

```
GNU nano 7.2
#!/bin/bash

echo "Filename:"
echo $1
echo "Number of rows:"
cat $1 | wc -l
echo "Number of columns:"
head -n 1 $1 | tr -d ' ' | tr -d '\n' | wc -c
```

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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/data
$ bash netsize.sh Saavedra2013/n1.txt
Filename:
Saavedra2013/n1.txt
Number of rows:
97
Number of columns:
80
```

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

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

```
GNU nano 7.2 netsize_all.sh
#!/bin/bash

FILES=./data/Saavedra2013/*.txt
for f in $FILES
do
    myrow='cat $f | wc -l'
    mycol='head -n 1 $f | tr -d ' ' | tr -d '\n' | wc -c'
    echo $f $myrow $mycol
done
```

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

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/data
$ bash netsize_all.sh
../data/Saavedra2013/n1.txt 97 80
../data/Saavedra2013/n10.txt 14 20
../data/Saavedra2013/n11.txt 270 91
../data/Saavedra2013/n12.txt 7 72
../data/Saavedra2013/n13.txt 61 17
../data/Saavedra2013/n14.txt 35 15
../data/Saavedra2013/n15.txt 38 11
../data/Saavedra2013/n16.txt 118 24
../data/Saavedra2013/n17.txt 76 31
../data/Saavedra2013/n18.txt 13 14
../data/Saavedra2013/n19.txt 10 16
../data/Saavedra2013/n2.txt 62 41
../data/Saavedra2013/n20.txt 18 7
../data/Saavedra2013/n21.txt 19 45
../data/Saavedra2013/n22.txt 19 36
../data/Saavedra2013/n23.txt 179 26
../data/Saavedra2013/n24.txt 80 28
../data/Saavedra2013/n25.txt 17 16
../data/Saavedra2013/n26.txt 82 40
../data/Saavedra2013/n27.txt 27 5
../data/Saavedra2013/n28.txt 90 19
../data/Saavedra2013/n29.txt 61 25
../data/Saavedra2013/n3.txt 25 36
../data/Saavedra2013/n30.txt 8 19
../data/Saavedra2013/n31.txt 28 25
../data/Saavedra2013/n32.txt 45 21
../data/Saavedra2013/n33.txt 70 20
../data/Saavedra2013/n34.txt 79 25
../data/Saavedra2013/n35.txt 14 8
../data/Saavedra2013/n36.txt 40 169
../data/Saavedra2013/n37.txt 44 13
../data/Saavedra2013/n38.txt 51 99
../data/Saavedra2013/n39.txt 33 25
../data/Saavedra2013/n4.txt 101 11
```

#### 1.10.4 Data explore

Write a script that, for a given CSV file and column number, prints

- the corresponding column name;
- the number of distinct values in the column;
- the minimum value;
- the maximum value.

```
MINGW64:/c/Users/TOSHIBA/Downloads/CSB-master/CSB-master/unix/data
GNU nano 7.2 explore.sh
#!/bin/bash

# $1 is the file name
# $2 is the column of interest

echo "Column name"
cut -d ',' -f $2 $1 | head -n 1
echo "Number of distinct values:"
cut -d ',' -f $2 $1 | tail -n +2 | sort | uniq | wc -l
echo "Minimum value:"
cut -d ',' -f $2 $1 | tail -n +2 | sort -n | head -n 1
echo "Maximum value:"
cut -d ',' -f $2 $1 | tail -n +2 | sort -n | tail -n 1
```

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

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/data
$ nano explore.sh

TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/data
$ bash explore.sh Buzzard2015_data.csv 7
Column name
biomass
Number of distinct values:
285
Minimum value:
1.048466198
Maximum value:
14897.29471
```

```
TOSHIBA@DESKTOP-MSHF1G3 MINGW64 ~/Downloads/CSB-master/CSB-master/unix/data
$ bash explore.sh Buzzard2015_data.csv 9
Column name
SLA
Number of distinct values:
144
Minimum value:
Maximum value:
369.6
```