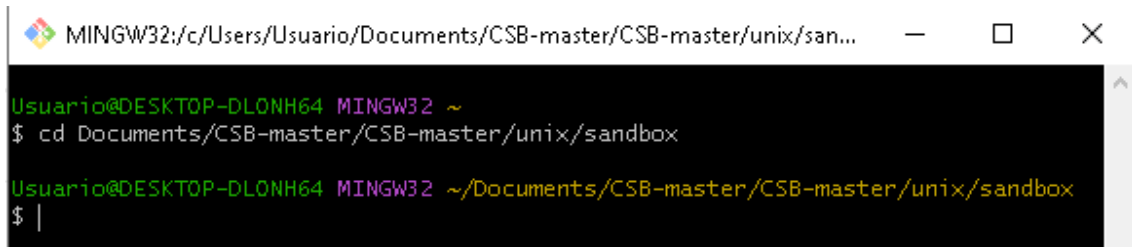# TC3_ForLoop

Jhanina Nicole Torres Montaquiza

1.10 Exercises
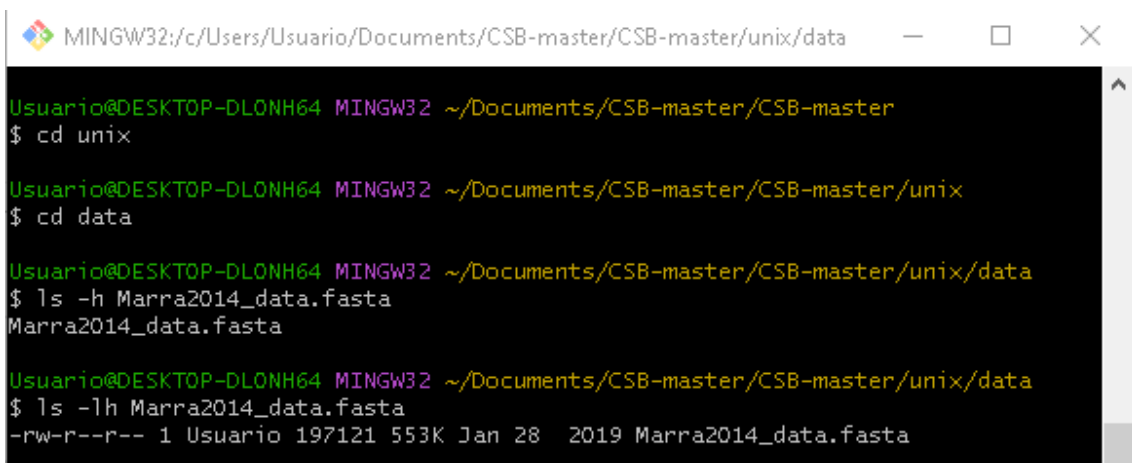
1.10.1 Next Generation Sequencing Data

1. Change directory to CSB/unix/sandbox.
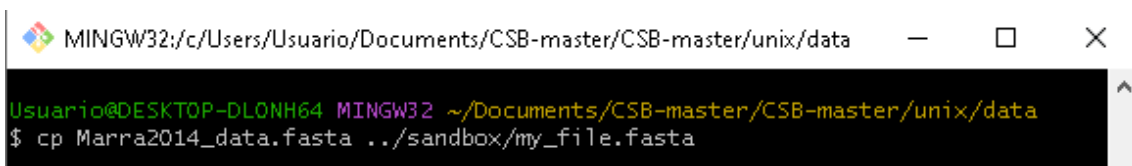
```
Usuario@DESKTOP-DLONH64 MINGW32 ~
$ cd Documents/CSB-master/CSB-master/unix/sandbox

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/sandbox
$ |
```

2. What is the size of the file Marra2014_data.fasta?

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master
$ cd unix

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix
$ cd data

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ ls -h Marra2014_data.fasta
Marra2014_data.fasta

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ ls -lh Marra2014_data.fasta
-rw-r--r-- 1 Usuario 197121 553K Jan 28  2019 Marra2014_data.fasta
```

3. Create a copy of Marra2014_data.fasta in the sandbox and name it my_file.fasta.

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ cp Marra2014_data.fasta ../sandbox/my_file.fasta
```

4. How many contigs are classified as isogroup00036?

```
MINGW32:/c/Users/Usuario/Documents/CSB-master/CSB-master/unix/data        —    □    ×

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ grep isogroup00036 Marra2014_data.fasta
>contig01235   length=561    numreads=21    gene=isogroup00036    status=isotig
>contig01236   length=257    numreads=59    gene=isogroup00036    status=isotig
>contig01238   length=187    numreads=42    gene=isogroup00036    status=isotig
>contig01240   length=276    numreads=22    gene=isogroup00036    status=isotig
>contig01241   length=2154   numreads=171   gene=isogroup00036    status=isotig
>contig01243   length=902    numreads=65    gene=isogroup00036    status=isotig
>contig01244   length=466    numreads=29    gene=isogroup00036    status=isotig
>contig01245   length=2579   numreads=66    gene=isogroup00036    status=isotig
>contig01246   length=540    numreads=24    gene=isogroup00036    status=isotig
>contig01248   length=237    numreads=24    gene=isogroup00036    status=isotig
>contig01250   length=166    numreads=8     gene=isogroup00036    status=isotig
>contig01251   length=176    numreads=8     gene=isogroup00036    status=isotig
>contig01252   length=136    numreads=8     gene=isogroup00036    status=isotig
>contig01253   length=145    numreads=8     gene=isogroup00036    status=isotig
>contig01254   length=145    numreads=42    gene=isogroup00036    status=isotig
>contig01255   length=180    numreads=59    gene=isogroup00036    status=isotig

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ grep -c isogroup00036 Marra2014_data.fasta
16
```

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


```
MINGW32:/c/Users/Usuario/Documents/CSB-master/CSB-master/unix/san...        —    □    ×

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/sandbox
$ cat my_file.fasta | tr " " "," | head -n 1
>contig00001,,length=527,,numreads=2,,gene=isogroup00001,,status=it_thresh

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/sandbox
$ |
```

6. How many unique isogroups are in the file?

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/sandbox
$ grep "gene" my_file.fasta | cut -d "=" -f 4 | uniq -c
    147 isogroup00001   status
     47 isogroup00002   status
     30 isogroup00003   status
     28 isogroup00004   status
     26 isogroup00005   status
     27 isogroup00006   status
     27 isogroup00007   status
     26 isogroup00008   status
     15 isogroup00009   status
     21 isogroup00010   status
     21 isogroup00011   status
     17 isogroup00012   status
     22 isogroup00013   status
     21 isogroup00014   status
     15 isogroup00015   status
     23 isogroup00016   status
     18 isogroup00017   status
     20 isogroup00018   status
     16 isogroup00019   status
     12 isogroup00020   status
     21 isogroup00021   status
     15 isogroup00022   status
     22 isogroup00023   status
     18 isogroup00024   status
     19 isogroup00025   status
     18 isogroup00026   status
     20 isogroup00027   status
     16 isogroup00028   status
     16 isogroup00029   status
     17 isogroup00030   status
     13 isogroup00031   status
     14 isogroup00032   status
     19 isogroup00033   status
     16 isogroup00034   status
     15 isogroup00035   status
     16 isogroup00036   status
     17 isogroup00037   status
     16 isogroup00038   status
     19 isogroup00039   status
     17 isogroup00040   status
     16 isogroup00041   status
     13 isogroup00042   status
      3 isogroup00043   status
```

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

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/sandbox
$ grep "contig" my_file.fasta | cut -d "=" -f 1 | sort -n | tail -n 1
>contig01385   length

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/sandbox
$ |
```

1.10.2 Hormone Levels in Baboons

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

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ cat Gesquiere2011_data.csv | cut -f 1 | grep -c 3
408

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ cat Gesquiere2011_data.csv | cut -f 1 | grep -c 27
18

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ |
```

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

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ nano script.sh
```



```
GNU nano 5.8                         script.sh
cut -f 1 $1.csv | grep -c -w $2




^G Help       ^O Write Out ^W Where Is  ^K Cut        ^T Execute   ^C Location
^X Exit       ^R Read File ^\ Replace   ^U Paste      ^J Justify   ^/ Go To Line
```

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 2
2

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 3
61
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ nano script.sh Gesquiere2011_data 39

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 50
21
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 5
28

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 9
4

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 100
13

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 19
3

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 1
10

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 4
46

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 49
19

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 99
2

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data
$ bash script.sh Gesquiere2011_data 98
5
```

3. [Advanced] 17 Write a script that returns the number of times each individual was sampled.

MINGW32:/c/Users/Usuario/Documents/CSB-master/CSB-master/unix/data

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data (mai
n)
$ nano script2.sh

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data (mai
n)
$ cat script2.sh
cut -f 1 $1.csv | sort | uniq -c
```

MINGW32:/c/Users/Usuario/Documents/CSB-master/CSB-master/unix/data   —   □   ✕

```
  GNU nano 5.8                          script2.sh
cut -f 1 $1.csv | sort | uniq -c








^G Help       ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location
^X Exit       ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^/ Go To Line
```

1.10.3 Plant-Pollinator Networks

1. 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.

MINGW32:/c/Users/Usuario/Documents/CSB-master/CSB-master/unix/data/Saavedra2013   —   □   ✕

```
  GNU nano 5.8                          nfyc.sh
wc -l $1.txt
cat $1.txt | head -n 1 | tr -d ' ' | wc -c








^G Help       ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location
^X Exit       ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^/ Go To Line
```

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data/Saavedra2013 (main
)
$ bash nfyc.sh n1
97 n1.txt
81

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data/Saavedra2013 (main
)
$ bash nfyc.sh n2
62 n2.txt
42

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data/Saavedra2013 (main
)
$ bash nfyc.sh n30
8 n30.txt
20

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data/Saavedra2013 (main
)
$ bash nfyc.sh n56
110 n56.txt
208

Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data/Saavedra2013 (main
)
$ bash nfyc.sh n6
9 n6.txt
32
```

```
Usuario@DESKTOP-DLONH64 MINGW32 ~/Documents/CSB-master/CSB-master/unix/data (mai
n)
$ bash script2.sh Gesquiere2011_data
    10 1
    21 10
    13 100
    26 101
    15 102
    26 103
    29 104
     6 105
    46 106
     7 107
    41 108
    28 109
    26 11
     3 110
    24 111
     3 112
     1 113
     1 114
     1 115
    14 116
    23 118
     1 119
    23 12
    42 120
    12 121
     9 122
    39 123
     1 124
    39 125
    15 126
    13 127
    16 13
     1 14
    40 15
    31 16
     3 17
     4 18
     3 19
     2 2
     4 20
    12 21
     5 22
    36 23
    35 24
    35 25
    22 26
```