

Bitacora para

1. Cambiar el nombre de los archivos recibidos de la secuenciación de microalgas.

2. Número de lecturas por archivo

3. Descompresión de los archivos

MUESTRAS de microalgas Chaetoceros de la tesis de Diamanda Tapia Gallardo (<https://biblioteca.cicese.mx/catalogo/tesis/ficha.php?id=25475>)

Elaborado por: Dra. Edith Elizondo Reyna

Como parte de la estancia posdoctoral en el Departamento de Acuicultura bajo la dirección del Dr. Miguel Ángel del Río Portilla, CICESE, Ensenada. Período 2020-2021

Cargado de los paquetes

```
In [ ]: from Bio import SeqIO
import os
from pandas import DataFrame
import pandas as pd
```

```
In [ ]: cd /LUSTRE/bioinformatica_data/lga/edith/data/microalgas/
```

```
In [ ]: ls /LUSTRE/bioinformatica_data/lga/edith/data/microalgas/
```

Los archivos llegaron con el nombre arriba señalado y se les cambió el nombre para algo más parecido a las muestras analizadas

```
In [ ]: lista = os.listdir(path = "./")
        lista.sort()
        lista
```

Cambio de nombre para dejar los datos de las muestras originales y los resultados de la secuenciación..... dudas ???

```
In [ ]: n=0
        print ("renaming ", " \t", " to" )
        for row in lista:
            if row[0]== "." or row[-3:]!=".gz":
                continue
            n+=1
            row1 = row[10:]
            print (n, row, " \t", row1)
            os.rename (row, row1) # quitar el símbolo "#" para
            que se ejecute y cambie el nombre
        print("finished", n, "files")
```

Se corrobora el nombre de los archivos

```
In [ ]: ls -lh
```

```
In [ ]: lista = os.listdir(path = "./")
        lista.sort()
        lista
```

Conteo de las lecturas en cada uno de los archivosno salio resultado??

Se utiliza el comando de linux `zgrep "^@" $file | wc -l >> lecturas.txt`
Observe que se coloca la salida al archivo `lecturas.txt`

```
In [ ]: %%bash
        for file in *.gz
        do
            echo $file >> lecturas.txt
            zgrep "^@" $file | wc -l >> lecturas.txt
        done
```

```
In [ ]: ls
```

```
In [ ]: %%bash
        for file in *.gz
        do
            echo $file >> lecturas.txt
            zgrep "^@" $file | wc -l >> lecturas.txt
        done
```

Visualización del número de lecturas

```
In [ ]: !head -20 lecturas.txt
```

Proceso para generar un archivo 'csv' conteniendo el número de lecturasedith

```
In [ ]: f = open("lecturas.txt", "r")
lineas =[]
n=1
lin =""
print( "\tarchivo\t\t\t\número de secuencias")
for row in f:
    lin = row
    lin = lin.replace("\n", "")
    lin = lin.replace(" ", "")

    if n%2!=0:
        print (n, lin, end = "")
        lin1 = lin
    else:
        #lin = row
        print("\t", lin)
        lin1 = lin1,lin
        lineas.append(lin1)
        lin1=[]
    n+=1
```

Proceso para generar un archivo 'csv' conteniendo el número de lecturas

```
In [ ]: f = open("lecturas.txt", "r")
lineas =[]
n=1
lin =""
print( "\tarchivo\t\t\t\número de secuencias")
for row in f:
    lin = row
    lin = lin.replace("\n", "")
    lin = lin.replace(" ", "")

    if n%2!=0:
        print (n, lin, end = "")
        lin1 = lin
    else:
        #lin = row
        print("\t", lin)
        lin1 = lin1,lin
        lineas.append(lin1)
        lin1=[]
    n+=1
```

```
In [ ]: f1 = DataFrame(lineas, index=None, columns =("archivo",
"numero_de_lecturas"))
f1
```

```
In [ ]: f1.to_csv("numero_lecturas.csv", index= None)
```

```
In [ ]: !head -13 numero_lecturas.csv
```

Descompresión de los archivos.... edith

```
In [ ]: %%bash
for file in *.gz
do
    echo $file
    gunzip $file
done
```

Descompresión de los archivos

```
In [ ]: %%bash
        for file in *.gz
        do
            echo $file
            gunzip $file
        done
```

edith

```
In [ ]: ls -lh
```

```
In [ ]: ls -lh
```