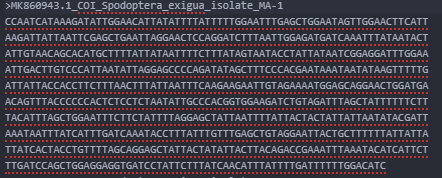
**Punto extra Parcial Bioinformatica**

Valentina Aguilar Cepeda

* Descargue el archivo que contenía la información de las ciudades desde el github con wget https://raw.githubusercontent.com/paula-torres/bioinformatica\_ur/main/files/sequences\_noctuidae\_popset.txt. Lo deje con el mismo nombre.

Lo dividí en cierto numero de pasos

1. En el archivo ya modificado del primer punto del parcial: (sequence.fasta) que tenia esta estructura



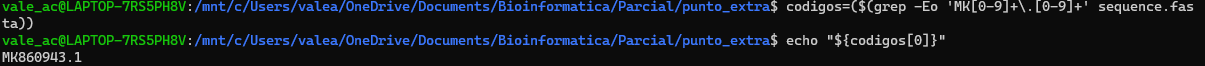
Extraje con:

* + grep -Eo 'MK[0-9]+\.[0-9]+' sequence.fasta, = Permite extraer solo el código, que esta colocado como esa expresión regular, del archivo de secuencias. Con el o, me permite solo tener el código y no el resto de información que se tiene en la fila.

la parte del código del encabezado, para poder buscarlo en el otro archivo lo guarde en un arreglo o lista que llame código:

* + codigos=($(grep -Eo 'MK[0-9]+\.[0-9]+' sequence.fasta))

La cual me permitía llamar cada uno de los códigos de las 10 secuencias por separado = echo "${codigos[0]}"



1. Ahora busque ese código en el archivo que tenia las ciudades
   * declare -A ciudades=()
   * for codigo in "${codigos[@]}"

;do

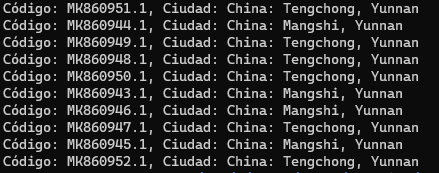
;ciudad=$(grep -A 23 "$codigo" sequences\_noctuidae\_popset.txt | grep -oP '(?<=country=")[^"]+')

;ciudades["$codigo"]=$ciudad

;done

Primero hice que la variable ciudades se volviera un equivalente a los diccionarios de R, en este caso vacío, al inicio. Después de forma manual encontré que la información de la ciudad estaba 23 líneas después de la línea en donde estaba el código. Por lo tanto, con el grep -A23 busque cada uno de los códigos y ubique la línea 23, en donde estaba la ciudad y el país. Con el segundo grep, le pedí que me guardara o imprimiera solo la parte que este entre comillas después del parámetro que dice country. Y eso lo guarde en la variable de ciudad.

Luego agregue una entrada al diccionario que cree antes, con el fin de que quedara el código y su ciudad correspondiente.



1. Posteriormente agregue el nombre de la ciudad y el país en el archivo de sequence.fasta:
   * for codigo in "${codigos[@]}"

;do

;ciudad=${ciudades[$codigo]}

;sed -i "s/^>$codigo\(.\*\)/>$codigo\1\\_$ciudad/" sequence.fasta

;done

Aquí genere otro for que se moviera por cada uno de los códigos. Después llame en el diccionario la ciudad correspondiente a ese código y lo guarde en la variable de ciudad. Con el sed busque el código de la secuencia en el archivo sequence.fasta y en vez de remplazar o quitar alguna parte solo añadí al final el nombre del país y la ciudad.

Texto

Descripción generada automáticamente

Nota: el código que use para imprimir el diccionario que cree fue:

* for codigo in "${!ciudades[@]}"

;do

;echo "Código: $codigo, Ciudad: ${ciudades[$codigo]}"

;done

Texto

Descripción generada automáticamente

