

Al laboratori m'han demanat que generi un codi que permeti diferenciar d'una llistat d'identificadors de seqüències, quina és la seqüència més curta i quina és la seqüència més llarga. Per fer-ho, he hagut de seguir els passos següents:

In [35]:

```
from Bio import SeqIO #primer obrim el modul Biopython i importem les funcions
de SeqIO que es per analitzar seqüències
from Bio import Entrez #descarreguem les seqüències de GenBank amb el paquet E
ntrez
Entrez.email = "chgodayol@gmail.com"
def calculadora_len(iden): #defineixo una funció que em calculi la longitud de
les diferents seqüències en funció del identificador
    handle = Entrez.efetch(db = "nucleotide", id = iden, rettype="gb", retmode
="text") #creem l'objecte handle per comunicarnos amb Genbank
    GBseq = SeqIO.read(handle, "genbank") #llegim les seqüències de GenBank
    contador = len(GBseq) #guardem la longitud de les cadenes dels diferents i
dentificadors en un objecte anomenat contador
    return contador
lista = ["JX398977", "JX475045", "NM_131329", "JX308815", "JQ712977", "NM_0012
66228", "NM_001135551", "NM_001172751", "JX317624", "JQ011270"] #creem una llis
ta amb els identificadors de les diferents seqüències
longitud = [] #creem una llista buida on posarem les longituds de les diferent
s seqüències
for identificador in lista: #fem un bucle per analitzar cada seqüència de la l
lista d'identificadors (lista)
    longitud.append(calculadora_len(identificador))#afegim a la llista longitu
d les longituds de les diferents cadenes que introduïm amb els seus identifica
dors corresponents

index_seq_curta = longitud.index(min(longitud)) # busquem l'índex on es troba
la seqüència més curta dintre de la llista longitud
sec_final_curta = Entrez.efetch(db = "nucleotide", id = lista[index_seq_curta]
, rettype="fasta", retmode="text") #descarreguem la seqüència més curta que he
m guardat abans com un objecte amb el nom de index_seq_curta

index_seq_llarga = longitud.index(max(longitud)) # busquem l'índex on es troba
la seqüència més llarga dintre de la llista longitud
sec_final_llarga = Entrez.efetch(db = "nucleotide", id = lista[index_seq_llarg
a], rettype="fasta", retmode="text") #descarreguem la seqüència més llarga que
hem guardat abans com un objecte amb el nom de index_seq_llarga

title1 = "Seqüència més curta" #definim la seqüència més curta abans d'imprimi
r-la
print ("\033[1m" + str(title1))

print("\033[0m" + sec_final_curta.read()) #llegim la seqüència més curta amb l
a funció read i la imprimim amb la funció print

title2 = "Seqüència més llarga"#definim la seqüència més llarga abans d'imprim
ir-la
print ("\033[1m" + str(title2))

print ("\033[0m" + sec_final_llarga.read()) #llegim la seqüència més llarga a
```

mb la funció read i la imprimim amb la funció print

```
print("\033[1m" + "Longitud de les seqüències" + "\033[0m" + str(longitud)+"pb") #imrpimim la llista longitud on tenim les diferents longituds de les seqüències que hem buscat a la bbdd amb els seus respectius identificadors. I ho fem en negreta
```

Seqüència més curta

```
>JQ011270.1 Clunio marinus period (per) mRNA, partial cds
CCGCCTTTTTCTGTCTAGTCTTACAACCTTACAAAGCACACTAACAACGTCACAAGCAACAGAATTC
CCTG
CATCGAATTTATTCCCAACAGTTTATTATCTTCCACAAACCATTCACAAATGTATGCAAAGCGTAA
TCTA
TCAACCCTTAATATATTACACCCATCGACTTTCTATCAAAYGAACTTCCAAGCCACTCAATCAAG
TCGT
AATGAATTGAATTACAACACATCCTATCAGTTTGTAAAGAATGCAACCGTTCCGGCACTGCAAAC
CCTG
CTGACTCTGATTCGATTGTGCAAAGTTTATCTTTAAAACGTCCACCGTCACAAGCAACTTCAGTCA
AAGC
TGATATGGGGTCAACATCGGCATCTGTTAACAGGGCGCTTTCAGAACTTCAAGAAAAGGTGTGTT
GGCT
GATTCACCACTCCCGCATGATGGTCCAAGTGGATTAAATATTAATGACAATGAGGCAAGTGGATTG
ATTG
KAACAAGCGAGGATG
```

Seqüència més llarga

```
>NM_001135551.1 Gallus gallus glucagon like peptide 1 receptor (GLP1R), mRNA
CCGCCGGGGCTCCGCTGGGAACGGGCTGCTCGTTCCGCCGGGGCCCCGCGGACAGGGCCTGCCCCGG
CTAT
GCCCCGTCGCTCCGGCCGTGCCCCCTCGTCCTGCTGCTGGCAGCCGCGGGCAGAGCCGCCCCCGCCAG
GCGT
AGTACAGATGGATCCCTCTCTGGGGTTGTGCAGAAGTGGAAGGAGTACCAGCTCCAATGCCTGAAA
TACT
TGTATGAGGCACCTCCTATTGCAGCAGAAGGCAAGTTCTGCAACAGAACATTTGACAACTATGCCT
GCTG
GCCTGATGGACTGCCTGGCACCTATGTGAACGTCAGCTGCCCCCTGGTATTTACCTTGGGCTAACAC
AGTG
CTCCATGGGCAGGTGTACCGGTTCTGCACCTCCGAGGGTACCTGGCTGCTAAAGGAGAACTCGACT
CTGC
CCTGGAGGAACCTGACAGAGTGTGAGGCCTCTGACCAGGATGCACCAGAAGAACAGCTCCTGAATT
TGTC
CATCATATACACCATTTGGATATGCTCTTTCCCTTCTCTGCCCTTGTAATTGCAACTGCCATTCTTCT
TGGA
TTCAGACATCTTCATTGCACAAGGAATTATATTCACCTGAACCTCTTCACCTCTTTTATCCTCCGT
GCTA
TATCCGTCTTCATTAAAGACTCAGTGGTGAAGTGGATGTACAGCACAGCCACACAAGAGCACCAGT
GGGA
AGGACTTATCTCCTTCCAGGAATCCCTTAGCTGCCGTTTGGTGTGTTGTGATGATGCAGTACTGTGT
GGCT
GCAAACACTACTACTGGTTGCTGGTGGAAGGCATGTACCTGTACACGCTGCTGGTACTGTCTGTCTTT
TCTG
AGCAGAGAATTTTTTCGGCTGTATCTCTGCATTGGCTGGGGTGTGCCAATGCTGTTTGTATCCTCT
GGGG
```

AACTGTCAAGTACCTTTATGAAGATGAGGGCTGCTGGAGCAGGAACTATAACATGAATTACTGGCT
GATC
ATCAGACTGCCCATTTCTAATCGCCATTGGGGTAAATTTTCCTGATCTTTATCAGAGTTATATGTATC
ATCA
TCTCCAAGCTACAAGCTAATCTGATGTGCAAACTGACATAAAATGCAGGCTGGCCAAGTCTACCT
TGAC
ATTGATCCCACTCCTGGGAACCCATGAGGTCATCTTTGCCTTTATTACTGATGAACACGCCAGAGG
GATG
CTGCGCTTTGTGAAGCTTTTTACTGAGCTCTCCTTTGCTTCTTTCCAGGGGCTGATGGTTGCAATT
CTCT
ACTGTTTCATCAATAATGAGG TTCAGATGGAGTTTCGGAAAAGTTGGGAGCGCTGGAGATTGGAAC
ATTT
GTATGTCCAGAGAGACAGCAGCATGAAACCTCTGAAGTGTCCAGCCAACAGCATCAGTAGTGGAGG
CACG
GTAGGCAGTAGTGTTTACGCTGCCACTTGCCAGGCTACCTTCAGCTAAGATTTCTGCAGGTGAACC
AGAC
T

Longitud de les seqüències[1271, 637, 1043, 651, 762, 512, 1471, 1
098, 1217, 505]pb

In []: