Jan Izquierdo

Assignment 1: Identify the origin organism of a fastq sequence

8: Identify the reads origin. (How would you find out from which genome come these reads? To which species they belong? Please describe the method used and the reliability of the results)

```
from Bio import SeqIO
def ex8():
     fastq_file="path/1S-unknown_illumina_2024.fastq" #set path to fastq file
     f=open("path/OUTPUT.fasta", "w")#create new fasta file, with writing permission
     for read in SeqIO.parse(fastq_file, "fastq"):#read the fastq file, and for every read
          print(">"+str(read.id), file=f)#write > and the id
          print(read.seq, file=f)#write the sequence
     f.close()

print(ex8())

#Also can be done using bash command:
#cat 1S-unknown_illumina_2024.fastq | awk '{if(NR%4==1) {printf(">%s\
n",substr($0,2));} else if(NR%4==2) print;}' > OUTPUT.fasta

#Create database
#$ makeblastdb -in ref_seq.fasta -dbtype nucl -out reference_database
#Use it
#$ blastn -query OUTPUT.fasta -db reference_database -out results.txt
```

I transformed the fastq into a fasta, then used BLAST to identify a few sequences of the file chosen at random, I searched for the genome of the most common species, the one that appeared most and had lower E values, first it seemed that it was Limanda limanda, but with some more samples I ended up choosing Podarcis lilfordi because the results had a lower E value.I downloaded the genome of this species, then I transformed the downloaded genome into a BLAST database using the command *makeblastdb* and queried for our sequences in this database using BLAST. In the BLAST search output I could see that most of the searches coincided and that their E values were low, which lets us know that Podarcis lilfordi is very likely to be the genome these reads come from.

The methods used were:

- A python program to transform fastq into fasta
- BLAST to identify and corroborate the origin of the sequences

The species the reads belonged to was Podarcis lilfordi

This are the results of some of the BLAST searches I used to find possible species:

Jan Izquierdo