Lab04
BIOL638/738
Due 11/7/21 at midnight

Aim
Ancient DNA is the study of old DNA from fossils and environmental samples. Ancient DNA is hard to study because the environment degrades DNA over time. Thus, ancient DNA samples might have low yields (i.e. low density of DNA). In lab04, we will investigate and hypothesize the origin of an ancient DNA sample. This sample, SBS026, was sequenced by Dr. Sabrina Shirazi at UC Santa Cruz. It's a piece of hide that is thought to be either deer (deer), cow (boTau9Y), or sheep (oviscan). The aim of this lab is to figure out which one it most likely is.

The sample was aligned to each of these reference genomes. If the sample is sheep, then the alignment to sheep should have the highest depth of coverage compared to the other two alignments. Some reads, align to multiple references. We are going to disregard these reads. We want to find the number of unique reads aligning to each reference.

Object-oriented programming is not required for this assignment.

Your script should take in three sam files, and output two columns.
The first column is the alignment name, the second column is the number of unique reads aligned to that reference. Note, the numbers listed here are just an example, your numbers will be different.

SBS026-aln-oviscan.sam	1234
SBS026-aln-deer.sam	23
SBS026-aln-bosTau9Y.sam	2345

The first column in the sam file is the read name.

Extra Credit
Week 10, Activity 2
https://docs.google.com/document/d/1FoO13oiFOd8XYqcLPLW_ianOvlcQDWy3vSeoL00IOYM/edit

Submit your script and FASTA output to iLearn as part of Lab04.
To "save" your FASTA output, you can pipe the output to a file using the following command.
python3 script.py fasta gff > output.fa

Here, "> output.fa" will save the output that is normally written out to your screen, to the file output.fa.