Week 7 assignment

Get the NCBI accession numbers for between 8 and 20 NCBI records of homologous sequences that are each under 50,000 bp in length

- Homologous sequences by definition share a common ancestral sequence. It should be possible to align them and build an evolutionary tree
- One useful method is to search google scholar for evolutionary papers that include "accession" in the text. These may then give you a list of the accessions of the sequences they used.

Write a python program that:

- downloads your NCBI records using Entrez
- prints the lengths of the sequences to the screen or a file (avoid datasets where the sequences are of very different lengths)
- Writes a file in Fasta format of the records.
- Aligns the sequences using ClustalW
- Build a phylogeny using PhyML
- Writes an ASCII version of the tree to a file.

Turn in :

- A zip (or gzip) archive of your work, including 5 files:
    o Your program
    o Your file of fasta records
    o Your alignment file generated by ClustalW
    o Your newick format file generated by PHyml
    o Your file with the ASCII version of the tree