**MIP 280A4: Computational Microbiology**

**Sequence Databases, In-class exercise questions**

1. What is the annotated collection date and host of the first SARS-CoV-2 sequence (the RefSeq SARS-CoV-2 genome sequence; 1 pt each)
   1. Collection Date:
   2. Host:

**NCBI Scavenger Hunt!**

1. How many neanderthal protein sequences are there in the NCBI RefSeq database? (neanderthal = *Homo sapiens neanderthalensis*) (1 pt)
2. From what type of genome do these neanderthal protein sequences derive? (1 pt)
3. How many West Nile virus sequences are in the Nucleotide database (the entire database, not just RefSeq)? (1 pt)
4. How many West Nile virus nucleotide sequences are in the RefSeq database? Why are there more than one? (2 pts)
5. There are 2 Zika virus genome sequences in the RefSeq database. What is the isolation source and collection date for each of these? (4 pts)
6. How many mammalian genomes are in the NCBI Genome database? How did you arrive at this answer? (2 pt)
7. The flavivirus genus includes a number of well-known viruses, including West Nile virus, Zika virus, and yellow fever virus. What is the name of another virus in this genus that you hadn’t heard about before? How did you find this answer? (2 pts)