1. Analyzing Sequences Across Databases
   1. Look up accession number KPZ68957 in GenBank. What is the name of the result?
   2. What is the sequence of the protein in Fasta format?
   3. Search the PDB database for the protein structure most similar to this sequence. What is the PDB number and description for the closest match? Search the Pfam database using this sequence. What domains do you see?
2. Querying Entrez
   1. What query string would you use to search Entrez for human alcohol dehydrogenase entries published between January 2000 and December 2010? (You will need to read the Entrez Help documentation (http://www.ncbi.nlm.nih.gov/books/NBK3837/#EntrezHelp.Entrez\_Searching\_Options) to complete this section)
   2. How many nucleotide and protein sequences are returned?
   3. What is the PMID of the top hit in the PubMed database to this query?
3. Querying RCSB-PDB
   1. Search for "zinc finger" in RCSB-PDB. How many structures are available for Homo sapiens? Escherichia coli? Bacteria?
   2. How many structures were determined by X-ray crystallography? NMR?
   3. What is the PDB number of the best hit from Mus musculus only?
   4. What date was this structure deposited?
4. Querying Entrez Using rentrez
   1. Write an R script that does the following:
      1. Retrieves the top 20 hits from the GEO profiles database for "Clostridium difficile"
      2. Plot the number of instances of "Clostridium difficile" in PubMed from 2010 to 2018
      3. Determine the number of links to the top hit from a) in Pubmed. List the journal name, title, authors and publication date for the manuscript.