Part 1:
- Downloaded tools:
    o bfx server – with Windows 10 will need PuTTY or FileZilla to access.
    o R CRAN download – download R studio here(more user friendly): https://rstudio.com/products/rstudio/download/
- GenBank records
    o Feature table – Standardizes records of DNA/prot sequences, includes features
    o Coding Sequence(CDS) – feature listed in GenBank that denotes which portion of the DNA sequence listed will be translated into a protein
    o N.B. –
        ▪ Complementary strands run 3' to 5', so their listed genetic loci have the opposite order of stop and start codons than that of the coding strand.
        ▪ Exons presented on GenBank records are denoted by a "join" statement that highlights which stretches of nucleotides would make up the coding section of the mRNA.

Part 2:
- Gene Annotation and Prediction
    o Annotations describe the locations of specific features on a genome, such as the CDS regions.
    o Gene prediction automatically annotates records based on their nucleotide sequences.
- Gene Prediction Challenges:
    o 1. 5' & 3' UTRs of mRNA is not easily predictable for these programs, necessitating the inclusion of mRNA locations along with CDS regions for more accurate predictions.
    o 2. Alternative splicing yields more versions of a protein for one gene that a prediction might account for.
    o 3. Often incorrect predictions of start/stop codons and splice sites.
- Annotation Methods:
    o Analysis methods like MAKER use RNA-seq data to annotate records, but mistakes happen and become difficult to fix.
- Annotation Challenges (according to supplementary reading article Yandell and Ence, 2012)
    o Can include 5' and 3' UTRs of the gene's sequence, but relies heavily on the information already gained from Gene prediction, mRNA/EST evidence, and Protein evidence to make all of its assumptions about the possible annotation for the record.

Part 3:

- Features of mRNA vs. genomic DNA:
    o Bacteria:
        ▪ Their mRNA lacks regulatory/promoter regions that its genomic DNA has, and is polycistronic, meaning one mRNA molecule contains the translatable code for multiple genes.
    o Eukaryotes:
        ▪ Eukaryotic genomic DNA contains introns that a mature mRNA molecule does not have. Eukaryotic mRNA is monocistronic unlike bacterial, and but can have multiple differing variations from the same genomic DNA due to alternative splicing.