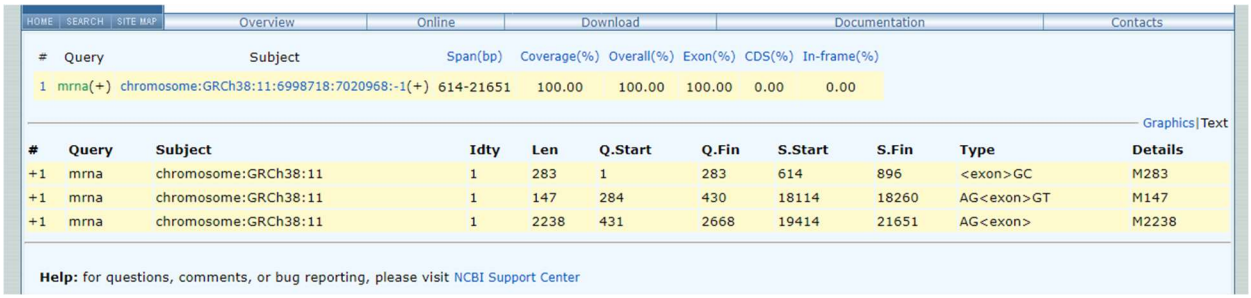
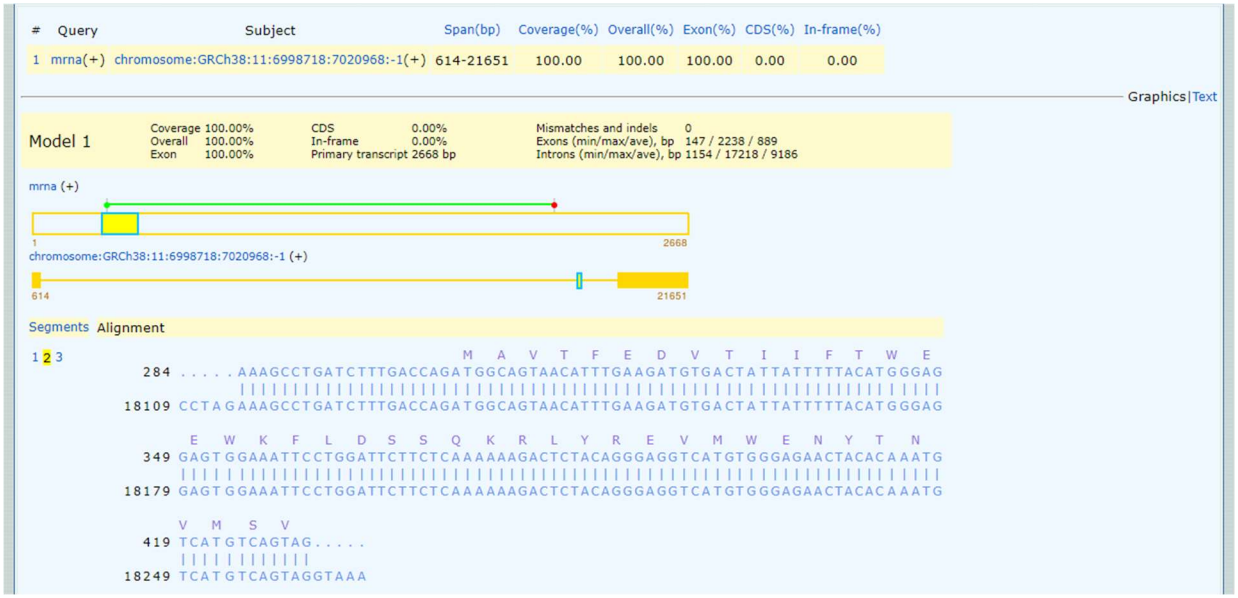
Jaimee Beckett
Units 3-4 Graded Homework

1. There are two attached files: znf214_mrna.txt and znf214_genomic.txt. Use Splign to find the mRNA and CDS coordinates in the genomic DNA.
   a. **mRNA locations:** 614~896, 18114~18260, 19414~21651
   b. **CDS locations:** 18134~18260, 19414~21107

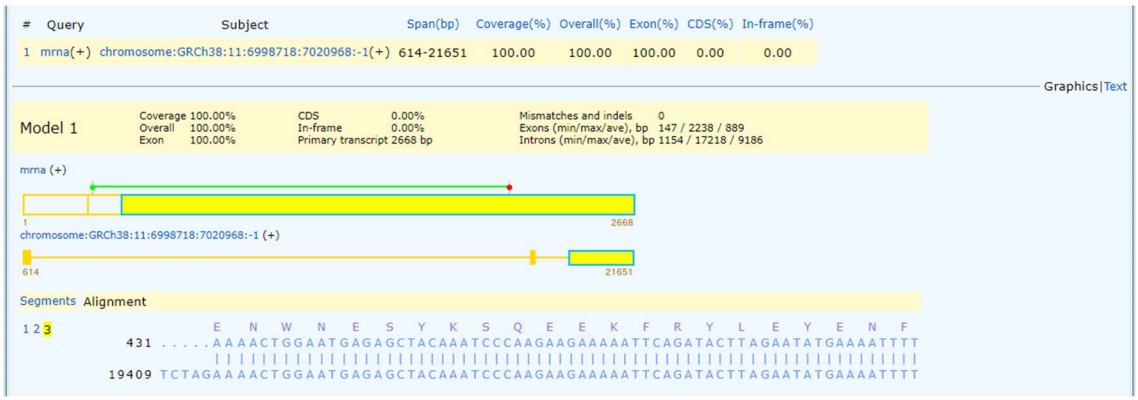
   mRNA locations:



   Segment 2, the first CDS location:



Beginning of segment 3, start of 2nd CDS location:



End of segment 3, end of 2nd CDS location:

2. Create a BED6 file with 2 lines based on the attached paper (Takenaka_et_al-2015-FEBS_Journal.pdf). Figure 3 shows the location of transcription factor DdlR binding to the promoter region of the ddlR-ddl operon in Brevibacillus brevis. The chromosomal location of the ddlR CDS is 2968133..2969623. The zero-based BED6 file should contain the location information of two genomic regions:
   a. The region bound by the DdlR transcription factor, which we will call the promoter. It is 170 bp in length, begins 140 nucleotides upstream from the start codon, and ends 29 nucleotides downstream from the start codon.
      - Begins **<u>140 nucleotides upstream</u>** from the start codon
        - Start site is 2968133
          - 2968133 – 140 = 2967993
        - Compensation of 1bp for BED file format
          - 2967993 – 1 = 2967992
      - Ends **<u>29 nucleotides downstream</u>** from the start codon
        - 2968133 + 29 = 2968162

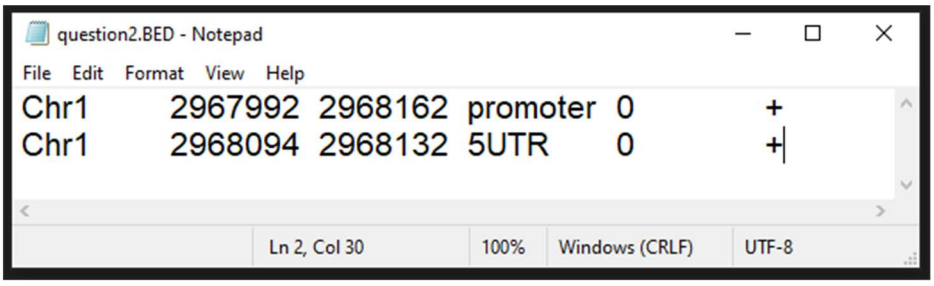
   b. The 5' UTR, noting that the transcription start site, as predicted by BPROM, begins 38 nucleotides upstream from the start codon. The 5' UTR is defined as the region from the transcription start site through the nucleotide that immediately precedes the start codon.
      - Begins **<u>38 nucleotides upstream</u>** from the start codon
        - Start site is 2968133
          - 2968133 – 38 = 2968095
        - Compensation of 1bp for BED file format

        - 2968095 – 1 = 2968094
    - End site **<u>immediately precedes the start codon</u>**
        - 2968133 – 1 = 2968132

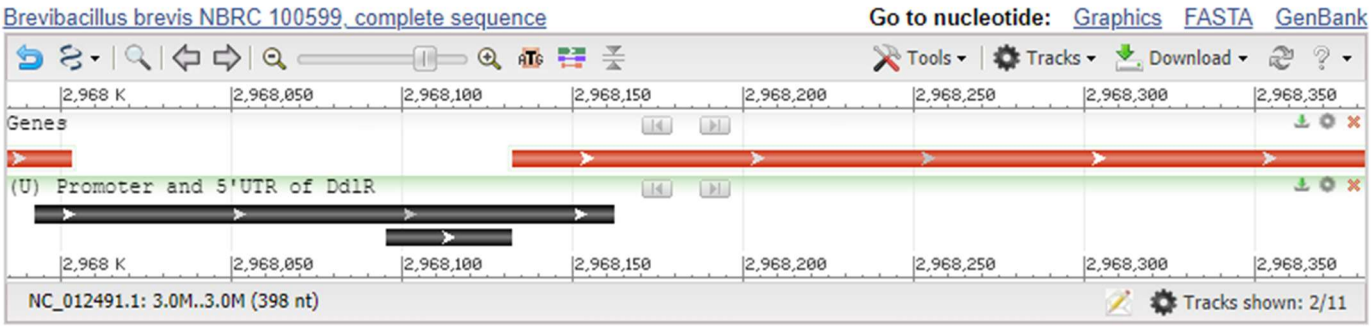
A BED6 file contains the following information separated by a tab: Chromosome, start, end, name, score, and strand. The Chromosome will always be 1 in this case since bacteria only have 1 chromosome. Finally, the direction is + because Figure 3 is pointing to the right.

3. Submit a screenshot of the BED6 from Problem 2. Using the NCBI Genome Browser for Brevibacillus brevis NBRC 100599, load your BED6 file. Take a screenshot showing the entire promoter, 5' UTR region, and CDS of DdlR. Be sure to zoom in so that these regions take up a majority of the shot.

   Screenshot from Problem2:



   Screenshot of BED6 file in NCBI Genome Browser:



4. Use the web-based Biomart in Ensembl to create a dataset and save it as a TSV, CSV, or XLS file. Use the following parameters to make the dataset:

   Dataset:
   Ensembl Genes 100 **(or the latest version)**
   Mouse genes (GRCm38.p6) **(or the latest version)**

   Filters:
   Chromosome 11
   Band E2 only

Transcript count >=7
Limit to genes with RefSeq protein (peptide) IDs only

Attributes:
Default attributes
Add “RefSeq Protein (peptide) ID”

Get all the results, export the results to a file, and submit the file.

- I used the datasets Ensembl Genes 102 (from the archive site) and Mouse Genes (GRCm38.p6). Screenshots of the results are shown below and the results file “mart_export.txt” is submitted.