BED FILES

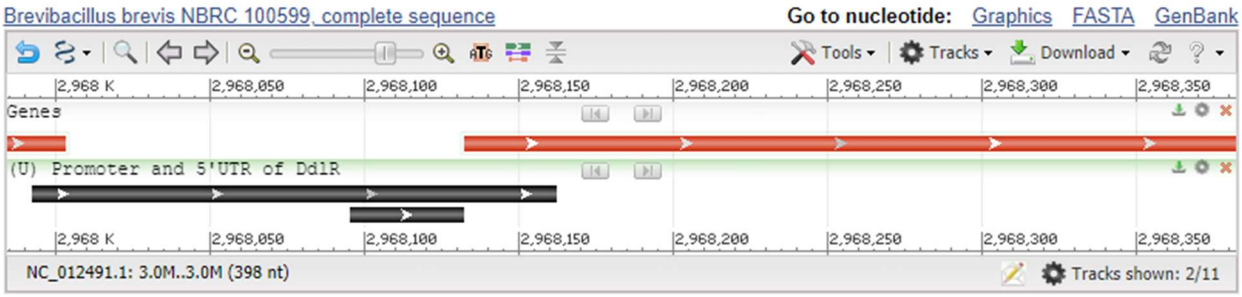
Create a BED6 file with 2 lines based on the attached paper (Takenaka_et_al-2015-FEBS_Journal.pdf). Figure 3 shows the location of transcription factor DdlR binding to the promoter region of the ddlR-ddl operon in Brevibacillus brevis. The chromosomal location of the ddlR CDS is 2968133..2969623. The 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.

Screenshot of BED6 file in NCBI Genome Browser:



SAM FILES

a. What is a SAM file and how is a SAM file generated?
A SAM (Sequence Alignment/Map) file contains information on alignment between two sequences, one of which is normally a reference genome. It is generated by aligning NGS (next-generation sequencing) data with tools such as HISAT, BWA, or Bowtie. It is tab-delimited and must contain the following 11 fields: QNAME, FLAG, RENAME, POS, MAPQ, CIGAR, MRNM, MPOS, ISIZE, SEQ, and QUAL.

b. Upload the SAM file to Galaxy and convert it to a BAM file. List the Galaxy tool(s) you used and the parameter(s) you set to complete the previous question.
Tool: SAM-to-BAM
File to convert: ERR181582.a.sam
Reference Genome: Yeast (Saccharomyces cerevisiae): sacCer3