**SIRE511 Linux for Bioinformatics**

**Assignment: Week 5**

**Maximum score: 10 points, Duration: 1 Oct 2024 – 10 Oct 2024**

### 

1. Write a script to read a tab delimited file **“primer.txt”** containing primer names and sequences. Primer sequences contain 8nt-index at position 1 to 8. Remove the 8nt-index from primer sequences and print out both primer names and edited sequences in FASTA format. Begin your script with **“practice1.sh”** as your guide.
   1. Guideline script: practice1.sh
   2. Input file: primer.txt
   3. Example output in FASTA format:

>DENV1\_1\_LEFT

AGTCTACGTGGACCGACAAGAA

>DENV1\_1\_RIGHT

GGCATCAGCATAAGGAGCATGG

>DENV1\_2\_LEFT

TGGCTAGATGGGGCTCATTCAA

>DENV1\_2\_RIGHT

TCGCCAGTTTGGGAACATGTTC

1. Write a script to read a genome sequence from a “reference.fasta” file. Split the genome sequence into each gene using the gene positions in “gene.txt” file. Redirect all gene names and sequences in a FASTA format to an output file. Begin your script with **“practice2.sh”** as your guide.
   1. Guideline script: practice2.sh
   2. Input 1: reference.fasta
   3. Input 2: gene.txt
   4. Output: gene.fasta