\*\*\*

We first created the necessary transcriptomics Conda environment and activated it using the code Conda activate transcriptomics.

We then downloaded the data set and placed it in Desktop > Bioinformatics > Transcriptome.

Once we placed it in the folder we ran the command

gunzip fastq/\*.gz

This code will unzip all files in the fastq/ directory ending in .gz.

I will then align my reads using the commands

hisat2 -q -x paeru -1 fastq/plank01\_F.fastq -2 fastq/plank01\_R.fastq -S plank01.sam

hisat2 -q -x paeru -1 fastq/plank02\_F.fastq -2 fastq/plank02\_R.fastq -S plank02.sam

hisat2 -q -x paeru -1 fastq/biofilm01\_F.fastq -2 fastq/biofilm01\_R.fastq -S biofilm01.sam

hisat2 -q -x paeru -1 fastq/biofilm02\_F.fastq -2 fastq/biofilm02\_R.fastq -S biofilm02.sam

Now I will convert my sam files to bam files using these commands

samtools view -bS plank01.sam > plank01.bam

samtools view -bS plank02.sam > plank02.bam

samtools view -bS biofilm01.sam > biofilm01.bam

samtools view -bS biofilm02.sam > biofilm02.bam

Now I will sort my bam files using this list of commands

samtools sort plank01.bam -o plank01.sorted.bam

samtools sort plank02.bam -o plank02.sorted.bam

samtools sort biofilm01.bam -o biofilm01.sorted.bam

samtools sort biofilm02.bam -o biofilm02.sorted.bam

Now I will assemble them into transcripts with the following commands

stringtie plank01.sorted.bam -G paeruginosa.gff -o stringtie/plank01.transcripts.gtf

stringtie plank02.sorted.bam -G paeruginosa.gff -o stringtie/plank02.transcripts.gtf

stringtie biofilm01.sorted.bam -G paeruginosa.gff -o stringtie/biofilm01.transcripts.gtf

stringtie biofilm02.sorted.bam -G paeruginosa.gff -o stringtie/biofilm02.transcripts.gtf

Now I will be creating a txt file for all four conditions and add these into the file

stringtie/plank01.transcripts.gtf

stringtie/plank02.transcripts.gtf

stringtie/biofilm01.transcripts.gtf

stringtie/biofilm02.transcripts.gtf

I then merged all transcript files and counted how many transcripts there are.

There are 5737 transcripts.

I then compared my annotated transcripts to a known transcript

5678 known transcripts compared to 5737

Here is the visualized version of this

A screenshot of a computer

AI-generated content may be incorrect.

Now I will prepare my files for R using these commands

stringtie -e -B -p 8 -G stringtie\_merged.gtf -o ballgown/plank01/plank01.gtf plank01.sorted.bam

stringtie -e -B -p 8 -G stringtie\_merged.gtf -o ballgown/plank02/plank02.gtf plank02.sorted.bam

stringtie -e -B -p 8 -G stringtie\_merged.gtf -o ballgown/biofilm01/biofilm01.gtf biofilm01.sorted.bam

stringtie -e -B -p 8 -G stringtie\_merged.gtf -o ballgown/biofilm02/biofilm02.gtf biofilm02.sorted.bam