Ecoli: 3.2.1

Reference genome: NCBI under the accession number NC\_000913.3

<https://www.ncbi.nlm.nih.gov/nuccore/NC_000913.3>

Illumina

<https://osdr.nasa.gov/bio/repo/data/studies/OSD-84>

GLDS-84\_Illumina sequencing\_mixed\_illumina.fastq.gz

Nanopore

<https://www.ncbi.nlm.nih.gov/sra/?term=SRR6334890>

SRR6334890.fastq

Align ecoli nanopore reads to reference

#!/bin/bash

SBATCH -p batch

#SBATCH -N 1

#SBATCH -n 16

#SBATCH --time=03:45:00

#SBATCH --mem=64GB

# Notification configuration

#SBATCH --mail-type=ALL

#SBATCH --mail-user=joseph.mcconnell@adelaide.edu.au

minimap2 -ax map-ont ecoliRef.sequence.fasta SRR6334890.fastq > rawNanoAln.sam

Get similarity score counts for primary aligned reads.

samtools view -F 4079 rawNanoAln.srt.bam | awk -f ~/awk/sim.awk | awk -f ~/awk/simstat.awk | sort -n -k1,1 | uniq -c > rawNanoHist.txt

sim.awk Calculates similarity between read alignment and reference.

simstat.awk Statistical summary of similarity.

sim.awk and simstat.awk are located in the chapter 3 supplementary folder.

FMLRC:

samtools view -F 4079 fmlrcSRR6334890cor.bam | awk -f ~/awk/sim.awk | awk -f ~/awk/simstat.awk | sort -k1,1V > fmlrcNanoHist

Ratatosk:

samtools view -F 4079 fmlrcSRR6334890cor.bam | awk -f ~/awk/sim.awk | awk -f ~/awk/simstat.awk | sort -k1,1V > fmlrcNanoHist