Christopher Uzokwe, Willow Livengood
ECES 650
Dr. Rosen
5/24/2021

Tutorial 9

Get the input files
1. Download the two files from Picotte
    a. /ifs/groups/eces450650Grp/data/mappings/evol1.sorted.unmapped.R1.fastq
    b. /ifs/groups/eces450650Grp/data/mappings/evol1.sorted.unmapped.R2.fastq
2. Convert the two fataq files into fasta files using the following line of text in the command line
    a. sed -n '1~4s/^@/>/p;2~4p' INFILE.fastq > OUTFILE.fasta
3. These are stored as R1.fasta and R2.fasta in the github

Run DIAMOND on Picotte
1. Make a node using the following line of code
    a. srun --nodes=1 --ntasks=20 --cpus-per-task=1 --mem=120GB --time=48:00:00 --pty /bin/bash
2. Make the database files for both datasets using the following two lines of code
    a. singularity run --bind /ifs/groups/eces450650Grp/ containers/diamond_latest.sif makedb --in ./ECES450650_SP21/Tutorial9/R1.fasta -d ./ECES450650_SP21/Tutorial9/R1db.dmnd
    b. singularity run --bind /ifs/groups/eces450650Grp/ containers/diamond_latest.sif makedb --in ./ECES450650_SP21/Tutorial9/R2.fasta -d ./ECES450650_SP21/Tutorial9/R2db.dmnd
    c. Depending on your permissions in Picotte, this may not be permitted. See Step 2 from section "Run DIAMOND Locally" and download/copy into Picotte the files from the github
3. Run a blastp search on both datasets using the following two lines of code
    a. singularity run --bind /ifs/groups/eces450650Grp/ containers/diamond_latest.sif blastp -q ./ECES450650_SP21/Tutorial9/R1.fasta -d ./ECES450650_SP21/Tutorial9/R1db.dmnd -o ./ECES450650_SP21/Tutorial9/outR1.tsv --very-sensitive
    b. singularity run --bind /ifs/groups/eces450650Grp/ containers/diamond_latest.sif blastp -q ./ECES450650_SP21/Tutorial9/R1.fasta -d ./ECES450650_SP21/Tutorial9/R1db.dmnd -o ./ECES450650_SP21/Tutorial9/outR2.tsv --very-sensitive
4. These output files are stored as outR1.tsv and outR2.tsv

Run DIAMOND Locally
1. Download the tool using the two following commands
    a. wget http://github.com/bbuchfink/diamond/releases/download/v2.0.9/diamond-linux64.tar.gz

b.   tar xzf diamond-linux64.tar.gz
2.   Make a DIAMOND-formatted database file for each fasta file using the following commands OR download R1db.dmnd and R2db.dmnd from the github
    a.   ./diamond makedb --in R1.fasta -d R1db
    b.   ./diamond makedb --in R2.fasta -d R2db
3.   Run a blastp search on each of the datasets using the following code
    a.   ./diamond blastp -q R1.fasta -d R1db.dmnd -o outR1.tsv --very-sensitive
    b.   ./diamond blastp-q R2.fasta -d R2db.dmnd -o outR2.tsv --very-sensitive
4.   These output files are stored in the github as outR1.tsv and outR2.tsv