SOP for metaphlan

Installing Metaphlan

1. Install Anaconda or miniconda
2. <https://github.com/biobakery/MetaPhlAn/wiki/MetaPhlAn-4#installation>
3. Supplemental help
   1. <https://omicx.cc/posts/2021-12-17-install-and-setup-metaphlan-3/>
   2. Add channels for conda-forge, bioconda
      1. $ conda config --add channels defaults  
         $ conda config --add channels bioconda  
         $ conda config --add channels conda-forge
   3. Reorder levels channel-proprity flexible
      1. $ conda config –set channel\_priority flexible
   4. Conda create –name mpa –c conda-forge –c bioconda metaphlan
   5. Install bowtie inside metaphlan env

Read end pairs: use <https://github.com/biobakery/MetaPhlAn/wiki/MetaPhlAn-4#installation>

Running Metaphlan

1. Activate metaphlan environment within Conda
   1. conda activate mpa
2. Set directory
   1. Cd /mnt/c/User/pennkey/...
      1. To pull from box … = /Box Sync/file\_with\_fastq\_sequences
3. Run an individual sample
   1. Read pair
      1. Metaphlan file\_1.fastq.gz, file\_2.fastq.gz --bowtie2out sampleID\_metagenome.bowtie2.bz2 --nproc 5 –input\_type fastq –o sampleID\_profiled\_metagenome.txt
4. Run a loop on all sequences in a file
   1. Read pairs
      1. Ex. If samples are labeled “UM1\_R1.fastq.gz” for read 1 and “UM1\_R2.fastq.gz” for read 2
      2. Set the working directory as a file containing all sequences and create 2 new files within this directory manually: “bowtie2” and “profiles”
      3. For I in \*\_R1.fastq.gz; do metaphlan $I,${I/\_R1.fastq/gz/\_R2.fastq.gz} --input\_type fastq –nproc 10 –bowtie2out bowtie2/${I%}.bowtie2.bz2 -o profiles/${I%}\_profiled.txt ; done

Process for SeqCenter fasta file analysis (read pair fastq.gz files)

1. Sync Illumina DNA file (containing all read pair fastq.gz files) to Box Sync on desktop to be used in processing
2. Create new file in Box Sync = “Project name metaphlan output” and within this file “profiles” and “bowtie2”
3. Run metaphlan as described above
   1. conda activate mpa
   2. CD to box folder containing sequences
   3. For paths to profiles and bowtie2 outputs, add in new file paths
   4. E.g. For i in \*\_R1.fastq.gz; do metaphlan $i.${i/\_R1.fastq.gz/\_R2.fastq.gz} -- input\_type fastq –nproc 10 –bowtie2out ../../project name metaphlan output/bowtie2/${i%}.bowtie2.bz2 -o ../../project name metaphlan output/ profiles/${i%}\_profiled.txt; done

Strainphlan

1. Run metaphlan to include sam file outputs
   1. for i in \*\_R1\_001.fastq.gz; do metaphlan $i,${i/\_R1\_001.fastq.gz/\_R2\_001.fastq.gz} --input\_type fastq --nproc 10 -s sams/${i%}.sam.bz2 --bowtie2out bowtie2/${i%}.bowtie2.bz2 -o profiles/${i%}\_profiled.txt ; done
2. Create consensus markers
   1. Mkdir – p consensus\_markers
   2. Sample2markers.py -I sams/\*.sam.bz2 -o consensus\_markers –n 8
3. Make folder with markers
   1. If you don’t know which strains you want to look at
      1. Mkdir –p db\_markers
      2. strainphlan -s consensus\_markers/\*.pkl -o output --print\_clades\_only
   2. Now extract the markers from the Chocophlan (default) database
      1. mkdir -p db\_markers
      2. extract\_markers.py -c t\_\_SGB15332 -o db\_markers/
4. Run strainphlan
   1. mkdir -p output  
      strainphlan -s consensus\_markers/\*.pkl -m db\_markers/t\_\_SGB1877.fna -o output -n 8 -c t\_\_SGB1877 --mutation\_rates