cd /data/jhe/napA

# All commands were run in a script after defining the shell #!/bin/bash

# command was run as follows:

nohup bash <script> &

# Classify SAGs with TreeSAPP assign

treesapp assign \

-n 4 \

--trim\_align \

--refpkg\_dir /root/Refpkgs/Nitrogen\_metabolism/Denitrification/NapA/seed\_refpkg/final\_outputs \

--fastx\_input /mnt/datasets/2021w/saanich/SAGs/Concatenated\_Med\_Plus/Saanich\_Med\_Plus\_SAGs.fasta \

--output SI072\_SAGs\_assign/

# Classify MAGs with TreeSAPP assign

treesapp assign \

-n 4 \

--trim\_align \

--refpkg\_dir /root/Refpkgs/Nitrogen\_metabolism/Denitrification/NapA/seed\_refpkg/final\_outputs \

--fastx\_input /mnt/datasets/2021w/saanich/MAGs/Concatenated/All\_SI072\_Metawrap\_MAGs.fa \

--output SI072\_MAGs\_assign/

# Updating reference packages with SAGs

treesapp update \

--fast \

--headless \

--overwrite \

--delete \

--cluster \

--trim\_align \

-n 4 \

--output NapA\_SAG\_update/ \

--skip\_assign \

--seqs2lineage /mnt/datasets/2021w/saanich/seq2lineage\_Tables/Med\_Plus\_SAGs\_GTDB\_Taxonomies.tsv \

--treesapp\_output SI072\_SAGs\_assign \

--refpkg\_path /root/Refpkgs/Nitrogen\_metabolism/Denitrification/NapA/seed\_refpkg/final\_outputs/NapA\_build.pkl \

# Updating reference packages with MAGs (after SAG update)

treesapp update \

--fast \

--headless \

--overwrite \

--delete \

--cluster \

--trim\_align \

-n 4 \

--output NapA\_MAG\_update/ \

--skip\_assign \

--seqs2lineage /mnt/datasets/2021w/saanich/seq2lineage\_Tables/SI072\_MAGs\_All\_GTDB\_taxonomies.tsv \

--treesapp\_output SI072\_MAGs\_assign \

--refpkg\_path /data/jhe/napA/NapA\_SAG\_update/final\_outputs/NapA\_build.pkl \

# Check purity of updated reference packages (most recent update was with MAGs)

treesapp purity \

-n 4 \

-r NapA\_MAG\_update/final\_outputs/NapA\_build.pkl \

--extra\_info /data/jhe/tutorial/TIGRFAM\_info.tsv \

-i /data/jhe/tutorial/TIGRFAM\_seed\_named.faa \

--output NapA\_purity

for FILE in /mnt/datasets/2021w/saanich/MetaG\_Assemblies/SI072\_\*m\_contig.fa

do

sample=$( basename $FILE | sed 's/.fa//g');

treesapp assign \

--fastx\_input $FILE \

--refpkg\_dir /data/jhe/napA/NapA\_MAG\_update/final\_outputs/ \

--output /data/jhe/napA/SI072\_MetaG\_contigs\_NapA\_assign/${sample}\_assign \

--trim\_align \

-n 8

done

# Calculate TPM metrics for metgenome reads

for FILE in /mnt/datasets/2021w/saanich/MetaG\_Trim\_QC\_Reads/SI072\_\*m\_pe.1.fq.gz

do

sample=$( basename $FILE | sed 's/.1.fq.gz//g');

depth=$( basename $FILE | sed 's/\_pe.1.fq.gz//g');

treesapp abundance \

-n 4 \

--treesapp\_output SI072\_MetaG\_contigs\_NapA\_assign/${depth}\_contig\_assign/ \

--reads $FILE \

--reverse /mnt/datasets/2021w/saanich/MetaG\_Trim\_QC\_Reads/${sample}.2.fq.gz \

--report update \

--metric tpm

done

# Calculate TPM metrics for metatranscriptome reads

for FILE in /mnt/datasets/2021w/saanich/MetaT\_Raw\_Reads/\*

do

depth=$( basename $FILE | sed 's/\_MetaT\_QC\_Filtered.fastq.gz//g');

treesapp abundance \

-n 4 \

--treesapp\_output SI072\_MetaG\_contigs\_NapA\_assign/${depth}\_contig\_assign/ \

--reads $FILE \

--pairing pe \

--report append \

--metric tpm

done

# Annotate NapA query sequences with their paralogs

for FILE in /data/jhe/napA/SI072\_MetaG\_contigs\_NapA\_assign/\*

do

treesapp layer \

-o $FILE \

--refpkg\_dir NapA\_MAG\_update/final\_outputs/

done