**#NarI reference package imported from Hallam lab github**

**#Refpkg file path**

/data/jw/NarI\_seed/NarI\_build.pkl

**#Classify amino acids with TreeSAPP assign**

cd/data/jw/

treesapp assign \

  -n 4 \

  --trim\_align \

  --refpkg\_dir /data/jw/NarI\_seed/final\_outputs/ \

  --fastx\_input UniProt\_NarI.fasta \

  --output /data/jw/UniProt\_NarI\_assign/

**# Download the data package SI072\_sequence\_data.tar.gz from Zenodo and decompress it with tar**

wget https://zenodo.org/record/6323402/files/SI072\_sequence\_data.tar.gz && \

tar -xzvf SI072\_sequence\_data.tar.gz

**# Classify ORFs predicted from genomes (SAGs).**

treesapp assign \

  -n 4 \

  --trim\_align \

  --refpkg\_dir /data/jw/NarI\_seed/final\_outputs/ \

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

  --output SI072\_SAGs\_assign/

**#Skip updating publicly available sequences**

**#Update NarI reference package with SAGs**

treesapp update \

  --fast \

  --headless \

  --overwrite \

  --delete \

  --cluster \

  --trim\_align \

  -n 4 \

  --output NarI\_SAG\_update/ \

  --skip\_assign \

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

  --treesapp\_output SI072\_SAGs\_assign/ \

  --refpkg\_path/data/jw/NarI\_seed/final\_outputs/NarI\_build.pkl

**# Classify ORFs predicted from genomes (MAGs)**

Updating with MAGs

treesapp assign \

-n 4 \

--trim\_align \

--refpkg\_dir /data/jw/NarI\_SAG\_update/final\_outputs \

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

--output /data/jw/SI072\_MAGs\_NarI\_assign\_version2/

treesapp update \

--fast \

--headless \

--overwrite \

--delete \

--cluster \

--trim\_align \

-n 4 \

--output NarI\_MAG\_update\_version2/ \

--skip\_assign \

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

--treesapp\_output SI072\_MAGs\_NarI\_assign\_version2/ \

--refpkg\_path /data/jw/NarI\_build.pkl

Sequence summary:

Number of sequences: 242

Longest sequence length: 757

Shortest sequence length: 451

Mean sequence length: 681.9

Median sequence length: 694.0

Extracting information from headers... done.

Reading cached lineages in '/data/jw/NarI\_MAG\_update\_version2/intermediates/accession\_id\_lineage\_map.tsv'... done.

Clustering sequences with MMSeqs' Linclust... done.

Number of unique lineages:

root 1

domain 2

phylum 15

class 21

order 27

family 45

genus 96

species 165

Unclassified and incomplete lineages account for 31/210 (14.8%) references.

**#Calculating abundance**

**#Assign taxonomic labels to Sannich Inlet metagenomic contigs**

mkdir SI072\_MetaG\_contigs\_NarI\_assign\_3

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

do sample=$( basename $f | sed 's/.fa//g')

treesapp assign \

-n 8 \

--trim\_align \

--refpkg\_dir /data/jw/NarI\_MAG\_update/final\_outputs/ \

-i $f \

--output SI072\_MetaG\_contigs\_NarI\_assign\_3/${sample}\_assign;

done

**#Overwrite previous abundance values with TPM values based on SI072 dataset calculations**

for f in SI072\_MetaG\_contigs\_NarI\_assign\_3/SI072\_\*assign; \

do sample=$( basename $f | sed 's/\_contig\_assign//g')

treesapp abundance \

-n 8 \

--treesapp\_output $f \

--reads /mnt/datasets/2021w/saanich/MetaG\_Trim\_QC\_Reads/${sample}\_pe.1.fq.gz \

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

--report update \

--metric tpm; done

**#Calculate TPM values from metatranscriptomes**

for f in SI072\_MetaG\_contigs\_NarI\_assign\_3/SI072\_\*assign; \

do sample=$( basename $f | sed 's/\_contig\_assign//g')

treesapp abundance \

-n 8 \

--treesapp\_output $f \

--reads /mnt/datasets/2021w/saanich/MetaT\_Raw\_Reads/${sample}\_MetaT\_QC\_Filtered.fastq.gz \

--pairing pe \

--metric tpm \

--report append; done

**# Annotate NarI query sequences with their respective paralog at all 7 depths**

for FILE in /data/jw/SI072\_MetaG\_contigs\_NarI\_assign\_3/\*

do

treesapp layer \

-o $FILE \

--refpkg\_dir /data/jw/NarI\_MAG\_update\_version2/final\_outputs/

done

**# Concatenate 7 classification files**

cat /data/jw/SI072\_MetaG\_contigs\_NarI\_assign\_3/SI072\_\*m\_contig\_assign/final\_outputs/layered\_classifications.tsv | head -n 1 >SI072\_NarI\_layered\_classifications.tsv

tail -q -n +2 /data/jw/SI072\_MetaG\_contigs\_NarI\_assign\_3/SI072\_\*m\_contig\_assign/final\_outputs/layered\_classifications.tsv >>SI072\_NarI\_layered\_classifications.tsv