**# Absolute path of nirK reference package (Connor’s and Immanuel’s):**

/root/Refpkgs/Nitrogen\_metabolism/Denitrification/NirK/seed\_refpkg/final\_outputs/

/mnt/datasets/2021w/RefPkgs\_from\_git/Nitrogen\_metabolism/Denitrification/NirK/seed\_refpkg/final\_outputs/

**# Download reference sequences from UniProt using Advanced search parameters ✅**

**# Copy UniProt\_NirK.fasta from local computer to MICB 425 server ✅**

cd /Users/KristiMacBookAir2015/Desktop/saanich/capstone/

scp UniProt\_NirK.gz root@10.32.204.58:/root/data/kristi/NirK\_processed/

**# Unzip the UniProt\_NirK.gz file. Rename to UniProt\_NirK.fasta ✅**

cd /root/data/kristi/NirK\_processed/

gunzip UniProt\_NirK.gz

mv UniProt\_NirK UniProt\_NirK.fasta

**# Classifying amino acid sequences ✅**

cd /root/data/kristi/NirK\_processed/

treesapp assign \

  -n 4 \

  --trim\_align \

  --refpkg\_dir /mnt/datasets/2021w/RefPkgs\_from\_git/Nitrogen\_metabolism/Denitrification/NirK/seed\_refpkg/final\_outputs/ \

  --fastx\_input UniProt\_NirK.fasta \

  --output /root/data/kristi/NirK\_processed/UniProt\_NirK\_assign/

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

cd /root/data/kristi/NirK\_processed/

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

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

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

treesapp assign \

-n 4 \

--trim\_align \

--refpkg\_dir /mnt/datasets/2021w/RefPkgs\_from\_git/Nitrogen\_metabolism/Denitrification/NirK/seed\_refpkg/final\_outputs/ \

--fastx\_input SI072\_sequence\_data/SI072\_MAGs.fa \

--output SI072\_MAGs\_assign/

**# Create a table with sequence names and taxonomic lineage information ✅**

echo -e "SeqID\tOrganism\tDomain\tPhylum\tClass\tOrder\tFamily\tGenus\tSpecies" >UniProt\_NirK.tsv

**# Update the NirK reference package with MAG sequences ✅**

awk -F"\t" '{ OFS="\t"; print $1,$2 }' SI072\_sequence\_data/SI072\_MAGs\_gtdbtk.bac120.summary.tsv | \

sed 's/user\_genome/Organism/g' | \

sed 's/classification/Lineage/g' | \

sed 's/;/; /g' >gtdb\_classifications.tsv

treesapp update \

--fast \

--headless \

--overwrite \

--delete \

--cluster \

--trim\_align \

-n 4 \

--output NirK\_MAG\_update/ \

--skip\_assign \

--seqs2lineage gtdb\_classifications.tsv \

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

--refpkg\_path /mnt/datasets/2021w/RefPkgs\_from\_git/Nitrogen\_metabolism/Denitrification/NirK/seed\_refpkg/final\_outputs/NirK\_build.pkl

**# Annotate features using the NirK reference package updated w/ MAGs ✅**

treesapp assign \

-n 4 \

--trim\_align \

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

--fastx\_input UniProt\_NirK.fasta \

--output UniProt\_NirK\_update\_assign/

treesapp layer \

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

--treesapp\_output UniProt\_NirK\_update\_assign/

**# Colour phylegenies in iTOL ✅ (I deleted the line with CuMMO because it does not exist for NirK in github)**

treesapp colour \

--output\_dir NirK\_MAG\_update/final\_outputs/ \

--refpkg\_path NirK\_MAG\_update/final\_outputs/NirK\_build.pkl \

--unknown\_colour grey

**# Copy all of the iTOL output files to a new folder ✅**

cp NirK\_MAG\_update/final\_outputs/NirK\*.txt UniProt\_NirK\_update\_assign/iTOL\_output/

**# Copy all of the iTOL output files to my local computer ✅**

scp -r root@10.32.204.58:/root/data/kristi/NirK\_processed/UniProt\_NirK\_update\_assign/iTOL\_output/ /Users/KristiMacBookAir2015/Desktop/saanich/capstone/iTOL

**# SAG stuff from Tamara’s script (not working for me)**

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

treesapp assign \

  -n 4 \

  --trim\_align \

  --refpkg\_dir /mnt/datasets/2021w/RefPkgs\_from\_git/Nitrogen\_metabolism/Denitrification/NirK/seed\_refpkg/final\_outputs/ \

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

  --output SI072\_SAGs\_assign/

**# (^ Says I don’t have permissions to manipulate the SAGs fasta file. So I’ll copy the file from /mnt to my own folder)**

cd /mnt/datasets/2021w/saanich/SAGs/Concatenated\_Med\_Plus

cp Saanich\_Med\_Plus\_SAGs.fasta /root/data/kristi/NirK\_processed

cd /root/data/kristi/NirK\_processed/

treesapp assign \

  -n 4 \

  --trim\_align \

  --refpkg\_dir /mnt/datasets/2021w/RefPkgs\_from\_git/Nitrogen\_metabolism/Denitrification/NirK/seed\_refpkg/final\_outputs/ \

  --fastx\_input Saanich\_Med\_Plus\_SAGs.fasta \

  --output SI072\_SAGs\_assign/

# FileExistsError: the input fasta file does not exists

**# This is what I (Jerry) did. I honestly don’t understand what the problem is. We used different paths to different reference packages, but that isn’t what your error is anyway and the permissions on the SAGs.fasta file and the MAGs.faa file are the same so if you’re able to run treesapp assign with the MAGs input idk why you’re not allowed to run it with the SAGs**

**#!/bin/bash**

**# 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/**

**# Update NirK refpkg with SAG sequences**

treesapp update \

  --fast \

  --headless \

  --overwrite \

  --delete \

  --cluster \

  --trim\_align \

  -n 4 \

  --output NirK\_SAG\_update/ \

  --skip\_assign \

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

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

  --refpkg\_path /mnt/datasets/2021w/RefPkgs\_from\_git/Nitrogen\_metabolism/Denitrification/NirK/seed\_refpkg/final\_outputs/NirK\_build.pkl

# FileExistsError: the input fasta file does not exists