# **Key Bioinformatics Procedures**

# #Demultiplex

# This is done in the Ubuntu Lab computer, using Claident::clsplitseq function

$ Clsplitseq --runname=nectar\_survey --truncateN=enable --index1file=survey\_index1 --index2file=survey\_index2 --primerfile=forward.primer.ITS1F.txt --reverseprimerfile=reverse.primer.ITS2.txt --minqualtag=30 --numthreads=8 lane1\_NoIndex\_L001\_R1\_001.fastq.gz lane1\_NoIndex\_L001\_R2\_001.fastq.gz lane1\_NoIndex\_L001\_R3\_001.fastq.gz lane1\_NoIndex\_L001\_R4\_001.fastq.gz Demultiplexed2.ITS

# # Transfer raw data files (including demultiplexed sequences) to Stanford Sherlock (mobaxterm)

$ cd d:/Research/Dhami-Miseq-data/raw\_bcl/survey\_fastq/Analysis.Dhami

$ rsync -rltvPh Demultiplexed2.16S ytwu@login.sherlock.stanford.edu:/home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/01\_Data/Dhami-Miseq-data/Analysis.Dhami/

$ rsync -rltvPh Demultiplexed2.ITS ytwu@login.sherlock.stanford.edu:/home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/01\_Data/Dhami-Miseq-data/Analysis.Dhami/

# # Check if there are any bugs in fastq files

$ cd /home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/01\_Data/Dhami-Miseq-data/Analysis.Dhami/Demultiplexed2.ITS/

$ for f in \*.fastq.gz; do gunzip -t -v $f; done 2>"debug.ITS.txt"

$ cd /home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/01\_Data/Dhami-Miseq-data/Analysis.Dhami/Demultiplexed2.16S/

$ for f in \*.fastq.gz; do gunzip -t -v $f; done 2>"debug.16S.txt"

# # Check if any files have number of sequences = zero

$ cd /home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/01\_Data/Dhami-Miseq-data/Analysis.Dhami/Demultiplexed2.ITS/

$ for sample in \*.fastq.gz; do

printf '%s\t' "$sample"

zcat "$sample" | awk '!(NR % 4){k++}END{print k}'

done >"ITS1.seq.number.sum.txt"

$ cd /home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/01\_Data/Dhami-Miseq-data/Analysis.Dhami/Demultiplexed2.16S/

$ for sample in \*.fastq.gz; do

printf '%s\t' "$sample"

zcat "$sample" | awk '!(NR % 4){k++}END{print k}'

done >"B16S.seq.number.sum.txt"

# # Do DADA2 pipeline for ITS and 16S separately in R

# Please refer to R script “20230202\_ITS1\_DADA2\_CONSTAXtaxa.Rmd” for ITS1 and “20230207\_16S\_DADA2\_SILVAtaxa.Rmd” for 16S sequences.

# # Use CONSTAX below to do taxonomy assignment for ITS1

(Tutorial of CONSTAX: <https://constax.readthedocs.io/_/downloads/en/latest/pdf/>; <https://constax.readthedocs.io/en/latest/tutorial1.html>)

# look at the fasta file for the ASVs (this is from the 20230202\_ITS1\_DADA2\_CONSTAXtaxa.Rmd script)

$ cat /home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/02\_Analysis/dada2/dada2\_ITS1/unpooled/unpooled.ASVs.fa

# run CONSTAX

$ srun -p bigmem --mem 512G -n 1 -N 1 -c 12 -t 0-08:00 --pty /bin/bash

$ cd /home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/02\_Analysis/CONSTAX/CONSTAX\_ITS1/

$ constax \

--num\_threads 12 \

--mem 512000 \

--db /home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/02\_Analysis/dada2/dada2\_ITS1/sh\_general\_release\_dynamic\_s\_all\_29.11.2022.fasta \

--train \

--input /home/groups/fukamit/ytwu/Wu\_mimulus\_analysis2022/02\_Analysis/dada2/dada2\_ITS1/unpooled/unpooled.all\_ASVs.fa \

--trainfile training\_files/ \

--tax taxonomy\_assignements/ \

--output taxonomy\_assignements/ \

--conf 0.5 \

--blast \

--make\_plot \

--pathfile pathfile.txt

# Then download output files to local computer:

$ cd c:/Users/Amanda/Desktop/CA\_oak\_pilot\_study/02\_Analysis/Sherlock\_cluster/itsxpress\_dada2\_FungalTraits/CONSTAX

$ rsync -rltvPh [ytwu@login.sherlock.stanford.edu:/home/groups/fukamit/ytwu/Wu\_CA\_oak\_pilot/02\_Analysis/itsxpress\_dada2\_FungalTraits/CONSTAX/](mailto:ytwu@login.sherlock.stanford.edu:/home/groups/fukamit/ytwu/Wu_CA_oak_pilot/02_Analysis/itsxpress_dada2_FungalTraits/CONSTAX/) :

# Go back to the R script “20230202\_ITS1\_DADA2\_CONSTAXtaxa.Rmd”, section “Assign taxonomy”, to continue running the rest of the bioinformatics steps.