**Create the Directory Structure**

busco\_odb10\_gene\_tree\_filtering

round1\_raw

unaln

search

batch

aln

pep

fast\_gene\_trees

cd path/to/your/Cladochytriales\_phylogeny/

mkdir busco\_odb10\_gene\_tree\_filtering

cd busco\_odb10\_gene\_tree\_filtering

mkdir round1\_raw

cd round1\_raw

mkdir unaln

mkdir search

mkdir bach

mkdir aln

mkdir pep

mkdir fast\_gene\_trees

**Getting Going…**

module load hmmer

module load python3.7-anaconda

cat path/to/busco\_odb10/hmms/\*.hmm > fungi\_odb10\_combined.hmm

hmmpress fungi\_odb10\_combined.hmm

**Overall cycle is: Search → Align → Gene tree → start over**

Before you do anything:

export CHYTRID\_PHYLO=/path/to/Chytrid\_Phylogenomics

**1) Search Proteomes for Markers**

For one proteome (For example):

hmmsearch \

--cpu 1 \

-E 1e-5 \

--domtblout Walse1\_GeneCatalog\_proteins\_20100910.aa.named.odb10.domtbl \

/nfs/turbo/lsa-amsesk/database/fungi\_odb10/fungi\_odb10\_combined.hmm \

../pep/Walse1\_GeneCatalog\_proteins\_20100910.aa.named.fasta > Walse1\_GeneCatalog\_proteins\_20100910.aa.named.fasta.hmmsearch.log

For multiple proteomes at once, use 1\_hmmsearch.py:

cd <”batch”\_directory>

python path/to/1\_hmmsearch.py [path/to/pep] [path/to/fungi\_odb10.combined.hmm]

cd <”search”\_directory>

**2) Parse domtbl Files**

# Combine all the peptide files into one

cat pep/\* > combined.pep.fasta

# Index combined.pep.fasta

module load samtools

samtools faidx combined.pep.fasta

~/Chytrid-Phylogenomics/bin/domtbl2unaln \

--cutoffs fungi \

--domtbls path/to/search \

--occupancy\_cutoff 0.75 \

--outdir path/to/search \

--proteins path/to/pep.combined.fasta #from above

**3) Align unaligned FASTA files**

# Move all FASTA files in `unaln dir`

mv search/\*.fasta unaln/

# Go back to batch directory

cd batch

# Generate hmmalign batch scripts

python ~/Chytrid\_Phylogenomics/scripts/slurm/2\_hmmalign.py [path/to/unaln] [path/to/the/fungi\_odb10/hmms]

# Go back to unaln

cd ../unaln

# Load the SCGid environment

scgidenv

-- or --

source ~/home/uniqname/SCGid/scgidenv/bin/activate

# Submit HmAl….sh scripts

sbatch ../batch/HmAl\_0.sh

sbatch ../batch/HmAl\_1.sh

...etc

# Move final alignment files from `unaln` for `aln` and move out

mv \*rmgapped ../aln/.

Cd ../

# Leave SCGid environment

deactivate

**3) Make the gene trees in fasttree**

cd batch

python ~/Chytrid\_Phylogenomics/scripts/slurm/3\_genetrees.py [path/to/aln]

cd ../fast\_gene\_trees/

# Submit FaTr….sh scripts

sbatch ../batch/FaTr\_0.sh

sbatch ../batch/FaTr\_1.sh

...etc