# Cold genes HMM

## Carla Greco 27.01.2022

Following anvio phylogenomics tutorial
https://merenlab.org/2017/06/07/phylogenomics/
Making own HMM collection
https://merenlab.org/2016/05/21/archaeal-single-copy-genes/
Get hmm-hits-matrix-txt
https://anvio.org/help/main/programs/anvi-script-gen-hmm-hits-matrix-across-genomes/

## Making HMM file for anvio

HMM models downloaded from EggNOG website. I was unsure regarding noise cutoff so I used E 1e-12.

Custom HMM folder contains:
genes.hmm.gz - concatenated hmm profiles from EggNOG
genes.txt - list of genes + accession + source
kind.txt - gene
target.txt - AA:GENE
noise_cutoff_terms.txt - E 1e-12

## Genomes used

Assembled using metaSPAdes, binned using MaxBin2 and scaffoled using SSPACE. Contamination and completeness estimated with CheckM.

**D1:** Joyce_1_Leptolyngbya
Scaffolds - 152
Largest Scaffold - 355994
N50 - 87779
Contamination - 1.57
Completeness - 97.2%

**D2:** Fryxell_1_Phormidesmis
Scaffolds - 322
Largest Scaffold - 272946
N50 - 31154
Contamination - 0.54
Completeness - 99.1%

**D3:** Fryxell_2_Leptolyngbya
Scaffolds - 232
Largest Scaffold - 237628
N50 - 49850
Contamination - 0.86
Completeness - 99.29%

**D4:** Fryxell_3_Anabaena
Scaffolds - 93
Largest Scaffold - 281882
N50 - 75233

Contamination - 0.88
Completeness - 100%

## On anvio…

Using anvio 7.1

**Pre analysis - Reformatted fasta file names:**

```
anvi-script-reformat-fasta D1.sspace.final.scaffolds.fasta -o D1-contigs-fixed.fa -l 0 --simplify-names
```

**1. Generate contigs database** - each FASTA file should have a file with the same name that ends with ‘.db’.

```
for i in `ls *fa | awk 'BEGIN{FS=".fa"}{print $1}'`
do
    anvi-gen-contigs-database -f $i.fa -o $i.db -T 4
    anvi-run-hmms -c $i.db
done
```

**2. Use the program anvi-get-sequences-for-hmm-hits to get sequences out of genomes.** 'external-genomes.txt' is lsit of genomes and their path.

```
anvi-get-sequences-for-hmm-hits --external-genomes external-genomes.txt \
                                --hmm-source Cold_HMM \
                                -o cold-genes-dna.fasta

anvi-get-sequences-for-hmm-hits --external-genomes external-genomes.txt \
                                --hmm-source Cold_HMM \
                                --get-aa-sequence \
                                -o cold-genes-aa.fasta
```

-
**3. Get table of hits**

```
anvi-script-gen-hmm-hits-matrix-across-genomes --external-genomes external-genomes.txt \
                                               --hmm-source Cold_HMM \
                                               -o output.txt
```

## Final files

output.txt - table of hmm hits for each genome
cold-genes-aa.fasta - amino acid sequences of hmm hits
cold-genes-dna.fasta - dna sequences of hmm hits