cold\_genes\_HMMs

Kaylah Marcello

1/31/2022

Table of Contents

## Testing Parameters for HMMs

These are the websites and workflows I used:

[Anvio phylogenomics workflow](https://merenlab.org/2017/06/07/phylogenomics/)  
[Making your own HMM database](https://merenlab.org/2016/05/21/archaeal-single-copy-genes/)  
[HMM hits Matrix](https://anvio.org/help/main/programs/anvi-script-gen-hmm-hits-matrix-across-genomes/)  
[EggNOG database](http://eggnog5.embl.de/#/app/results)

### Making HMM file for Anvi’o (x3)

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-5, - E 1e-12, -E 1e-20

Completed - E 1e-30, now testing stringency of noise cutoff  
HMMs downloaded from EggNOG

### Genomes used

#### NCBI cyanobacteria whole genome sequences (WGS), filamentous

[kmrcello@farm](mailto:kmrcello@farm):~/cyanobacteria/outputs/db/filamentous-subset

### Anvi’o

hope (v7.1)

##### Pre analysis - reformat fasta file names:

###### 1. Convert to an “improved” fasta header, saves file-name-key in a tsv

anvi-script-reformat-fasta contigs.fa -o contigs-fixed.fa -l 0 --simplify-names --report-files contigs.tsv

###### 2. Generate an anvio-ready genome database (.db) and run HMMS

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

###### 3. Run HMMs from our own file

for i in `ls ./outputs/db/filamentous-subset/\*db | awk 'BEGIN{FS=".fa"}{print $1}'`  
do  
 anvi-run-hmms -c $i --hmm-profile-dir hmms --just-do-it  
done

###### 4. Use the program anvi-get-sequences-for-hmm-hits to get sequences out of genomes.

anvi-get-sequences-for-hmm-hits --external-genomes external-genomes-filamentous.txt \  
 --hmm-source hmms \  
 -o cold-genes-dna.fasta   
   
anvi-get-sequences-for-hmm-hits --external-genomes external-genomes-filamentous.txt \  
 --hmm-source hmms \  
 --get-aa-sequence \  
 -o cold-genes-aa.fasta

###### 5. Get table of HMM hits

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