

# cold\_genes\_HMMs

Kaylah Marcello

1/31/2022

## Contents

<b>Testing Parameters for HMMs</b>	<b>1</b>
Making HMM file for Anvi'o (x3) . . . . .	1
CTG various noise cutoff runs . . . . .	1
Output files . . . . .	2
Comparing parameters . . . . .	3

## Testing Parameters for HMMs

These are the websites and workflows I used:

Anvio phylogenomics workflow  
Making your own HMM database  
HMM hits Matrix  
EggNOG database

## 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:~/cyanobacteria/outputs/db/filamentous-subset

## CTG various noise cutoff runs

### 1. Run HMMs from our own file, anvi-run-hmms

```
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
```

Substitute -E 1e-5 -> -12, -20, (-30?) in hmms > noise\_cutoff\_terms.txt

## 2. Use the program anvi-get-sequences-for-hmm-hits

substitute gene name for all genes for GhostKOALA

```
anvi-get-sequences-for-hmm-hits --external-genomes FILAMENTOUS/external-genomes-filamentous-names.tsv \
                                --hmm-source hmms \
                                --gene-names COG1278.faa.final_tree.fa \
                                -o test-params/csp-5-dna.fasta

anvi-get-sequences-for-hmm-hits --external-genomes FILAMENTOUS/external-genomes-filamentous-names.tsv \
                                --hmm-source hmms \
                                --gene-names COG1278.faa.final_tree.fa \
                                --get-aa-sequence \
                                -o test-params/csp-5-aa.fasta
```

Substitute -gene-names for corresponding gene

Substitute -o gene-#-aa.fasta

loop from Anvi'o addition to anvi-get-sequences-for-hmm-hits

```
for gene in $genes
do
    anvi-get-sequences-for-hmm-hits --external-genomes FILAMENTOUS/external-genomes-filamentous-names.t
                                --hmm-source hmms \
                                --gene-name $gene \
                                -o cold-gene-hmm-hits.fa/${gene}
done
```

## 3. Get table of HMM hits

```
anvi-script-gen-hmm-hits-matrix-across-genomes --external-genomes FILAMENTOUS/external-genomes-filament
                                --hmm-source hmms \
                                -o output-5.txt
```

## Output files

output-5.txt

output-12.txt

output-20.txt

Genes: aceE, csp, dnaJ, hupB, recA, otsA

aceE-5-aa.fasta

aceE-5-dna.fasta

aceE-12-aa.fasta

aceE-12-dna.fasta

aceE-20-aa.fasta

aceE-20-dna.fasta

## Comparing parameters

GhostKOALA

upload aa fasta files for each gene to compare target hits.