

```

#!/bin/bash

### Names:
# Stephanie Hauser
# Lissette Hernandez

cat hsp70gene_* > hsp70.refs

# Putting all the hsp reference sequences into one sequence
cat mcrAgene_* > mcrA.refs

# Move hsp and mcrA reference sequences into tools for easier access
mv hsp70.refs ../../Biocomputing/tools
mv mcrA.refs ../../Biocomputing/tools
cd ~/Private/Biocomputing/tools

./muscle -align hsp70.refs -output hsp70.align

# Run muscle for HSP and mcrA reference sequences in order to create an alignment
sequence from the reference
./muscle -align mcrA.refs -output mcrA.align

cd ~/Private/bioinformaticsProject/teomes

# Put proteome sequences using wildcard into a new proteome.fasta file to use for
hmmr
cat proteome_* > proteome.fasta

cd ~/Private/Biocomputing/tools

# Use hmmbuild to make a profile using the reference align sequences and comparing
to each of 50 proteomes
./hmmbuild hsp70.hmm hsp70.align
./hmmbuild mcrA.hmm mcrA.align

# Use hmmsearch to search through each proteome individually and compare to each
query, place output in table with tblout
for file in ~/Private/bioinformaticsProject/teomes/proteome.fasta;
do
    ./hmmsearch --tblout $file.hsp70.hits hsp70.hmm $file;
done

# Above and below, first using a path to the directory where proteome file is
placed, and then redoing hmmr search to search
# proteome for query (hsp70 or mcrA) individually

for file in ~/Private/bioinformaticsProject/teomes/*.fasta;
do
    ./hmmsearch --tblout $file.mcrA.hits mcrA.hmm $file;
done

rm ~/Private/bioinformaticsProject/teomes/proteome.fasta

for file in ~/Private/bioinformaticsProject/teomes/*.fasta; do
    hsp70hits=$(cat "$file.hsp70.hits" | grep -v "#" | tr -s " " | cut -d " " -f 1
| sort -u | wc -l)
    mcrAhits=$(cat "$file.mcrA.hits" | grep -v "#" | tr -s " " | cut -d " " -f 1 |
sort -u | wc -l)
    echo "$file $hsp70hits $mcrAhits"

```

```
(( hsp70hits > 0 && mcrAhits > 0 )) && echo "${basename "$file"}" >> array.txt  
done
```

```
# We ouput a list of proteomes that satisfy significance requirements.  
# We know that presence of mcrA indicates a methanogen and a 'significant' number  
# of hsp70 repeats indicate pH resistance. In this  
# case, we coded the output file to include any proteomes that had an hsp70 count  
# above 0 and mcrA count also above 0. However, this  
# is subject to user input and can easily be adjusted in the for loop.
```