

HMM analyses using eggNOG data

*LJM*

*2019-11-20*

# Contents

<b>1</b>	<b>Searching with HMMs</b>	<b>2</b>
<b>2</b>	<b>Utility script for changing delimiters</b>	<b>3</b>
<b>3</b>	<b>Session info</b>	<b>5</b>
<b>4</b>	<b>References</b>	<b>6</b>

# Chapter 1

## Searching with HMMs

```
#ml bioinfo-tools hmmer/3.2.1
DATA="../analyses/HMM_scan_using_eggNOG_HMMs/eggNOG_data/HMMs/"
HMMer_OUTPUT_DIR="../analyses/HMM_scan_using_eggNOG_HMMs/hmmsearch_out_viruses/"
fOTUproteoms_DIR="../analyses/fOTU_proteomes/"

read -r -a fOTUproteoms <<< $( find $fOTUproteoms_DIR -name "*.faa" -and -type f -print0 | xargs -0 echo )

for TAXID in $DATA*/ ; do
    TAXID_NO=$(echo "$TAXID" | awk -F "/" '{print $(NF-1)}')
    read -r -a HMMs <<< $( find $TAXID -name "*.hmm" -and -type f -print0 | xargs -0 echo )
    for HMM in "${HMMs[@]}"; do
        HMM_FILE=$(echo $(basename "$HMM"))
        HMM_VARIANT=$(echo $(basename "$HMM") | awk -F "." '{print $1}')
        for fOTUproteom in "${fOTUproteoms[@]}"; do
            fOTU_FILE=$(echo $(basename "$fOTUproteom"))
            fOTU=$(echo $(basename "$fOTUproteom") | awk -F "." '{print $1}')
            #echo $fOTUproteom
            hmmsearch --cpu 4 --noali -E 0.01 --domE 0.01 --pfamtblout "$HMMer_OUTPUT_DIR"$TAXID_NO
            #hmmsearch --cpu 4 --noali -E 0.01 --domE 0.01 --incE 0.001 --incdomE 0.001 --notextw
            #hmmsearch --cpu 3 --noali -E 0.01 --domE 0.01 --incE 0.001 --incdomE 0.001 --notextw
            #hmmsearch --cpu 3 --noali -E 0.01 --domE 0.01 --incE 0.001 --incdomE 0.001 --notextw
        done
    done
done
```

## Chapter 2

# Utility script for changing delimiters

```
SCRIPT="../scripts/eggNOGdelimiter.awk"  
TSV_FILE="Fuselloviridae_10474_annotations.tsv"  
cat $TSV_FILE | awk -v delimiter=";" -v last_col="4" -f $SCRIPT
```

```
## 10474;4QHRA;S;integral to membrane  
## 10474;4QHRB;S  
## 10474;4QHRC;S  
## 10474;4QHRD;S;integral to membrane  
## 10474;4QHRE;S;integral to membrane  
## 10474;4QHRE;S;nuclease activity  
## 10474;4QHRE;S;integral to membrane  
## 10474;4QHRE;S  
## 10474;4QHRI;S  
## 10474;4QHRJ;S  
## 10474;4QHRK;S  
## 10474;4QHRM;S  
## 10474;4QHRN;S;ATP-dependent DNA helicase activity  
## 10474;4QHRP;S;DNA integration  
## 10474;4QHRQ;S  
## 10474;4QHRR;S  
## 10474;4QHRS;S  
## 10474;4QHRT;S;integral to membrane  
## 10474;4QHRT;S;regulation of transcription, DNA-dependent  
## 10474;4QHRT;S;Viral/Archaeal nuclease  
## 10474;4QHRT;S;ATP binding  
## 10474;4QHRT;S;Pfam:B277  
## 10474;4QHRY;S
```

```
## 10474;4QHRZ;S;Pfam:C166
## 10474;4QHS0;S;integral to membrane
## 10474;4QHS1;S;structural molecule activity
## 10474;4QHS2;S;nucleic acid binding
## 10474;4QHS3;S
## 10474;4QHS4;S
## 10474;4QHS5;S
## 10474;4QHS6;S;nucleic acid binding
## 10474;4QHS7;S;integral to membrane
## 10474;4QHS8;S;nucleic acid binding
## 10474;4QHS9;S
## 10474;4QHSA;S
## 10474;4QHSB;S;Pfam:C166
## 10474;4QHSC;S
## 10474;4QHSD;S;integral to membrane
## 10474;4QHSE;S;Pfam:DUF5517
## 10474;4QHSF;S;Pfam:DUF5493
## 10474;4QHSG;S;sequence-specific DNA binding transcription factor activity
## 10474;4QHSH;S;Ribbon-helix-helix protein, copG family
## 10474;4QHSI;S;Pfam:DUF5489
## 10474;4QHSJ;S;regulation of transcription, DNA-dependent
```

## Chapter 3

### Session info

```
sessionInfo()
```

```
## R version 3.6.1 (2019-07-05)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Debian GNU/Linux 9 (stretch)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib/libopenblas-r0.2.19.so
##
## locale:
##  [1] LC_CTYPE=C.UTF-8      LC_NUMERIC=C           LC_TIME=C.UTF-8
##  [4] LC_COLLATE=C.UTF-8    LC_MONETARY=C.UTF-8    LC_MESSAGES=C
##  [7] LC_PAPER=C.UTF-8      LC_NAME=C              LC_ADDRESS=C
## [10] LC_TELEPHONE=C        LC_MEASUREMENT=C.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] compiler_3.6.1  magrittr_1.5    bookdown_0.13   tools_3.6.1
##  [5] htmltools_0.3.6 yaml_2.2.0      Rcpp_1.0.2      stringi_1.4.3
##  [9] rmarkdown_1.15  knitr_1.24      stringr_1.4.0   xfun_0.9
## [13] digest_0.6.20   evaluate_0.14
```

## Chapter 4

## References