

## Annotating Proteins

Predicting function of proteins.

- [Pfam site docs](#)
- [Pfam manual](#)

## Finding homologs

For Protein to Protein searches BLASTP, phmmer ([HMMER](#)), FASTA

```
module load fasta
fasta36 query database > results.FASTA
fasta36 -m 8c -E 1e-3 query database > results.FASTA.tab
```

## To Find Domains

See Overview lecture [Domains lecture](#)

Searching with [HMMer](#) against [Pfam](#)

See the HMMER [tutorial](#)

Searching with [Interpro](#)

## Searching Interpro on HPCC

Note this can be slow.

```
#!/bin/bash -l
#SBATCH -p epyc -N 1 -n 1 -c 8 --mem 16G --time 24:00:00
#SBATCH -J iprscan
#SBATCH -o logs/%x.%j.log
module load iprscan

set -euo pipefail
CPU=${SLURM_CPUS_PER_TASK:-1}
interproscan.sh --goterms --pathways -f tsv -i PROTEINFILE.fa --cpu $CPU > SEARCH.log
```

Run `mkdir -p logs` before submitting. See [UNIX IV](#) for how to choose `-c`, `--mem` and `--time`.

The results will contain information like

Gene Ontology <http://geneontology.org/>

## Running Analyses on Biocluster

```
# in a job script or an srun session with -c 4 (or more)
module load hmmer
module load db-pfam
CPU=${SLURM_CPUS_PER_TASK:-1}
hmmsearch --cpu $CPU --domtblout domtbl_results.out $PFAM_DB/Pfam-A.hmm proteins.fa > proteins.hmmsearch
hmmsearch --cpu $CPU --domtblout hmmsearch_domtbl_results.out $HMM protein-db.fa > protein.hmmsearch

Pfam2GO - http://current.geneontology.org/ontology/external2go/pfam2go
```

## Workshop in class

Let's compare the genomes of cyanobacteria and identify if there are differences in gene content based on Protein domains.

Many papers investigating the evolution and genomes of cyanobacteria.

- <https://journals.asm.org/doi/10.1128/mbio.00561-19>
  - <https://bmcecolevol.biomedcentral.com/articles/10.1186/1471-2148-10-24>
1. Searching for Pfam domains in a set of proteins from several species (start with 3)
    - Let's download some genomes/proteomes from [NCBI](#)
  2. Parsing report files and count the number of domains per species (in python)
  3. Summarize the content comparing with a table sorted by counts