Introduction to HMMER

HMMER is similar to BLAST and is mainly used for sequence alignment.

```
wget ftp://ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam27.0/Pfam-A.hmm.gz

wget ftp://ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam27.0/Pfam-B.hmm.gz

gzip -d Pfam-A.hmm.gz; gzip -d Pfam-B.hmm.gz

Get the HMM file of the PFAM database. The HMM file is a text file, which needs to be converted into a binary format to speed up the operation, while being compressed and built into an index database.

 hmmpress Pfam-A.hmm

 hmmpress Pfam-B.hmm
```

Use hmmscan for Pfam annotation

Each number in the Pfam database represents a protein family. Pfam is divided into two databases, A and B. Database A is a high-quality database that has been manually calibrated. Although database B is of lower quality, it can still be used to find conserved sites in protein families. In the latest v27.0 version of Pfam database, database A contains 14,836 protein family numbers (starting with PF); database B contains 20,000 protein family numbers (starting with PB). Example of Pfam annotation using hmmscan:

```
hmmscan -o out.txt --tblout out.tbl --domtblout out.dom --noali -E 1e-5 /pfam/Pfam-A.hmm query.f
```