

COG annotation guide

Yu-Wei Wu
Graduate Institute of Biomedical Informatics
Taipei Medical University
yuwei.wu@tmu.edu.tw

Introduction

This tutorial provides the detailed steps for annotating COGs for prokaryotic genomes.

Prerequisite:

- HMMER (<http://hmmer.org/>)
 - o Download the latest HMMER for your platform
- Prodigal (<http://prodigal.ornl.gov/>)
-

Steps

1. Download the HMM files from eggNOG website (<http://eggnogdb.embl.de/#/app/downloads>) and unzip it (may take a while as the file is big—we are only use a small fraction of it afterward)

EggNOG 4.5.1

Search protein or OG

Navigation

- Home
- Sequence search
- eggNOG-mapper (genome-wide functional annotation)
- Downloads
- API
- Methods
- Viral OGs

Downloading EggNOG raw data

Browse EggNOG raw data available: /download/

Or download a specific dataset

	members	annotations	raw_algs	trimmed_algs	trees	hmm
All organisms (NOG)	(55.3MB)	(1.8MB)	(2.5GB)	(1.9GB)	(130.3MB)	(5.0GB)
Acidobacteria (aciNOG)	(94.0KB)	(81.4KB)	(4.8MB)	(4.4MB)	(219.4KB)	(125.3MB)
Acidobacteriia (acidNOG)	(50.5KB)	(60.8KB)	(2.4MB)	(2.3MB)	(99.7KB)	(75.4MB)
Aconoidasida (acoNOG)	(80.0KB)	(50.9KB)	(6.4MB)	(5.6MB)	(143.1KB)	(217.1MB)
Actinobacteria (actNOG)	(2.6MB)	(405.7KB)	(92.8MB)	(80.7MB)	(7.4MB)	(765.3MB)
Agaricales (agaNOG)	(112.7KB)	(92.9KB)	(6.5MB)	(5.7MB)	(157.5KB)	(211.1MB)
	members	annotations			trees	hmm

You can also use command “wget” to get the hmm file (remember to use the real link address since eggNOG database will not always stay at version 4.5)

```

root@fe38f3ae0b9c:/#
root@fe38f3ae0b9c:/# wget http://eggnogdb.embl.de/download/eggnog_4.5/data/NOG/NOG.hmm.tar.gz
--2017-07-10 04:09:02-- http://eggnogdb.embl.de/download/eggnog_4.5/data/NOG/NOG.hmm.tar.gz
Resolving eggnogdb.embl.de (eggnogdb.embl.de)... 194.94.44.96
Connecting to eggnogdb.embl.de (eggnogdb.embl.de)|194.94.44.96|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 5397320832 (5.0G) [application/octet-stream]
Saving to: 'NOG.hmm.tar.gz'

NOG.hmm.tar.gz          100%[=====] 5.03G
2017-07-10 04:16:17 (11.9 MB/s) - 'NOG.hmm.tar.gz' saved [5397320832/5397320832]

root@fe38f3ae0b9c:/# tar zxvf NOG.hmm.tar.gz

```

2. There are A LOT OF hmm files in the unzipped folder—possibly over a million. We only want those associated with COGs for now—I still do not know how to make use of other hmm files.

```

root@fe38f3ae0b9c:/#
root@fe38f3ae0b9c:/# cd eggNOG/
root@fe38f3ae0b9c:/eggNOG# ls NOG_hmm/NOG.COG* | wc -l
4866
root@fe38f3ae0b9c:/eggNOG# cat NOG_hmm/NOG.COG* > COG_all.hmm
root@fe38f3ae0b9c:/eggNOG#


```

Count the number of COG files: 4866

Concatenate all COG hmm files into a big HMM file

One good attribute about HMMER files is that you can simply concatenate the files into a big file. Later we need to compress this file for faster search.

3. Clone the COGmapper repository from the github website
 - git clone <https://github.com/yuwwu/COGmapper.git>
4. Download and install HMMER and Prodigal
 - HMMER
 - Download your HMMER distribution from the HMMER download page (If you are using Linux, most likely the viable choice is Linux/Intel x86_64)
 - <http://hmmer.org/download.html>



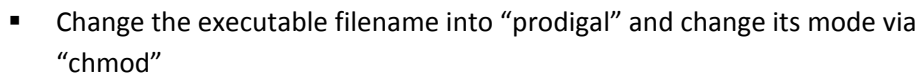
[DOWNLOAD](#)
[DOCUMENTATION](#)
[SEARCH](#)
[PUBLICATIONS](#)
[BLOG](#)

The **current version** is HMMER **3.1b2** (05 March 2015).

Source:		
with Linux/Intel ia32 binaries:	[hmmer-3.1b2.tar.gz]	5.8 MB
with Linux/Intel x86_64 binaries:	[hmmer-3.1b2-linux-intel-ia32.tar.gz]	18.1 MB
with Linux/Intel x86_64 binaries:	[hmmer-3.1b2-linux-intel-x86_64.tar.gz]	20.2 MB
with MacOSX/Intel binaries:	[hmmer-3.1b2-macosx-intel.tar.gz]	13.5 MB

- You should be able to find all executables in your HMMER binary folders unless you downloaded source files.

- Prodigal
 - Download the most suitable executable from Prodigal github website (<https://github.com/hyattprodigal/releases/>)



- 5. Setup HMMER and Prodigal executable locations
There are two ways to setup the executables.
 - (1) Simply place their paths into the system paths so that you can run the programs anywhere.
 - (2) Add their paths into the setting file
 - Note: you need to choose either (1) or (2) but not necessarily both. COGmapper will automatically find the programs and will only report errors if both (1) and (2) fails.

6.