Pfam curation tools Documentation

Release 1.0

EMBL-EBI

Oct 15, 2020

Contents

Contents:

Requirements

- A UNIX/Linux/Apple computer with a 64-bit architecture (e.g x86_64) with administrative rights

For unsupported architectures the docker image will have to be built from a Dockerfile.

Preparation

2.1 Install Docker

You will need admininstration privileges to do this (or have a sys admin do it). On Debian/Ubuntu flavour machines, install the package docker.io:

```
sudo apt-get install docker.io
```

Ensure user is a member of the docker group:

```
groups
```

and add docker if that does not appear in the output above.

```
sudo adduser USERNAME docker
```

2.2 Download docker image

Download the latest version of the Pfam curation tools container from Dockerhub:

```
docker pull dockerhub.ebi.ac.uk/pfam/pfam-curation
```

2.3 Set up computer

Choose a directory on your machine where you will run the container; this can be anywhere you have write access but it's important that you always use this location.

```
mkdir pfam_curation
cd pfam_curation
```

Within that directory create two additional directories: one will contain the sequence files required by the pfam tools and the other will be a working directory:

```
mkdir seqlib
mkdir pfam_data
```

Also, create a pfam.conf file in the pfam_curation directory. You can use the template.

Obtain the pfamseq (and, optionally) the uniprot fasta file. Either download the pfamseq.gz (and uniprot.gz) files from Pfam ftp or run the download_pfamseq.sh script.

```
bash download_pfamseq.sh
```

If downloading manually you will need to move the files to the seqlib directory and uncompress.

In the directory above pfam_curation directory, get a checkout of the Pfam dictionary:

```
svn co https://xfamsvn.ebi.ac.uk/svn/pfam/trunk/Data/Dictionary/
```

Running the container

```
docker run --rm -it -v $(pwd)/pfam_data:/home/pfam/pfam_data -v $(pwd)/pfam.conf:/home/pfam
```

This will give you a command prompt inside the container where you will have access to all the Pfam curation tools and scripts.

The first time you run you will need to index the sequence file(s) you downloaded previously:

```
esl-sfetch --index /data/seqlib/pfamseq
esl-sfetch --index /data/seqlib/uniprot
```

To exit the container, type control-D.

If you use the directory pfam_data to download and build families, then the files in this directory will be preserved if you exit the container and re-run later. They will also be accessible from the host computer in the directory of that name.

The command above should work on Linux-based machines and will permit applications such as belvu to run correctly. On a Mac the command above will require some modification. If a workaround does not exist, it's possible to install and run these programs from the host computer. If you change to the pfam_data directory first it will be equivalent to running the command within the container.

Troubleshooting

If you see an error similar to this:

```
Temporary failure in name resolution: Unable to connect to a repository at URL 'https:
→//xfamsvn.ebi.ac.uk/svn/pfam/trunk/Data/Families/PF00023' at /opt/Pfam/PfamLib/Bio/
→Pfam/SVN/Client.pm line 232
```

This means your Docker container is unable to access your network. To resolve the issue, you can either disable dnsmasq in NetworkManager, or specify a DNS server for Docker to use.

To disable dnsmasq, comment out the dns line (dns=dnsmasq line -> #dns=dnsmasq) in /etc/NetworkManager/NetworkManager.conf and then restart the network-manager service:

```
sudo vi /etc/NetworkManager/NetworkManager.conf
sudo service network-manager restart
```

To specify a DNS server, add your network's DNS server to /etc/docker/daemon.json, and then restart the docker service:

```
sudo vi /etc/docker/daemon.json
sudo service docker restart
```

An example DNS line is given below:

```
{
  "dns": ["10.0.0.2", "8.8.8.8"]
}
```