

Setting up an official mirror of PubMLST.org

In order to mirror the PubMLST.org site, you need a Linux system running PostgreSQL (version 7.3 or later), Perl and the Apache web server. The following instructions were written for Bio-Linux 3 (based on RedHat) and Bio-Linux 4 (based on Debian) but the procedure will work on any Linux system, although the paths of various commands may be slightly different. Certain Perl modules are required to be installed – these are already installed on Bio-Linux or they can be installed automatically from CPAN if not, e.g. as root or using sudo, type :

```
perl -MCPAN -e "install XML::Parser::PerlSAX"  
perl -MCPAN -e "install DBI"  
perl -MCPAN -e "install CGI"
```

Required modules: CGI, DBI, XML::Parser::PerlSAX

Optional (adds functionality): BioPerl

1. Register your server IP address with Keith Jolley (keith.jolley@medawar.ox.ac.uk) so that it will be allowed to make an rsync connection to the primary server. You will be assigned a web address for the mirror dependent on your country, e.g. ukmirror2.pubmlst.org for the second UK mirror site. This will be added to the pubmlst.org DNS so that connections to this address are directed to the IP address of your server.
2. Edit the PostgreSQL client authentication file, pg_hba.conf to trust local connections (on Bio-Linux 3 this can be found at /home/db/postgresql/data/pg_hba.conf, on Bio-Linux 4 it is found at /etc/postgresql/pg_hba.conf):

Uncomment the lines:

```
local      all  all                                trust  
host      all  all  127.0.0.1  255.255.255.255  trust
```

Make sure the following line is commented out (with a # at the beginning):

```
#local      all  all  ident          sameuser
```

Please be aware that this will allow any local logged on user access to the databases.

Check that PostgreSQL is accepting TCP/IP connections – check that the line 'tcpip_socket = true' exists and is not commented out in the postgresql.conf file (on Bio-Linux 3 this is found at /home/db/postgresql/data/postgresql.conf, on Bio-Linux 4 it is at /etc/postgresql/postgresql.conf).

Restart (or start) PostgreSQL (either log in as root or use the sudo command):

```
BL3: sudo /etc/rc.d/init.d/postgresql restart
```

```
BL4: sudo /etc/init.d/postgresql restart
```

Create database users 'apache', 'webmaster', and 'remote' using the createuser command. You will need to log in as the postgres user to do this (see the Bio-Linux database FAQ at

<http://envgen.nox.ac.uk/envgen/software/archives/000447.html> for details).

```
createuser apache
createuser remote
createuser webmaster
```

These database users do not need any special privileges (e.g. permission to create new databases or users).

3. Create a user account called 'webmaster' (Either log in as root or use the sudo command).

```
sudo /usr/sbin/useradd -m -g users -s /bin/zsh webmaster
```

The Z-shell (/bin/zsh) is the default on Bio-Linux but any shell will do, e.g. /bin/bash or /bin/csh.

Create a password for the webmaster account:

```
sudo /usr/bin/passwd webmaster
```

4. Ensure that the webmaster has write permissions on the default web cgi-bin directory (usually /var/www/cgi-bin) and create a 'mlstdbnet' subdirectory (on BL4, you may have to create the cgi-bin directory 'sudo mkdir /var/www/cgi-bin'). Either set the owner of this directory to 'webmaster' or add webmaster to an appropriate group with write privileges.

```
sudo chown webmaster /var/www/cgi-bin
sudo mkdir /var/www/cgi-bin/mlstdbnet
sudo chown webmaster /var/www/cgi-bin/mlstdbnet
```

5. Create the /home/httpd directory and make it owned by webmaster

```
sudo mkdir /home/httpd
sudo chown webmaster /home/httpd.
```

6. Log in as webmaster (su webmaster), and copy the site as follows:

```
rsync -av --exclude="/tmp" pubmlst.org::pubmlst
/home/httpd/pubmlst.org (this is one line, with a space after pubmlst.org::pubmlst)
```

Files will be copied over to a directory called pubmlst.org in /home/httpd.

7. Create a directory for temporary files used by the web scripts

```
mkdir /home/httpd/pubmlst.org/tmp
chmod 777 /home/httpd/pubmlst.org/tmp
```

8. Copy the script directory and create the databases:

```
cd (ensure you're in the webmaster home directory)
/home/httpd/pubmlst.org/mirror/scripts/updatemirror
```

The databases are all prefixed 'pubmlst_' to avoid clashing with other databases you may have on your system.

9. As root (or using the sudo command) create a directory called /usr/local/mlstdbnet and copy the newly downloaded mlstdbnet.conf file to it:

```
sudo mkdir /usr/local/mlstdbnet

sudo cp /home/httpd/pubmlst.org/mirror/conf/mlstdbnet.conf
/usr/local/mlstdbnet (this is one line)
```

The configuration file is suitable for use with Bio-Linux. Check through the configuration and change any settings which are not appropriate, for example if you don't have BioPerl, BLAST or EMBOSS installed you can switch these options off. You may need to change the paths of programs such as BLAST to whatever is used on your system. Create a virtual host for the web address provided (e.g. ukmirror2.pubmlst.org):

Bio-Linux 3 (RedHat):

Add the following to the virtual host section of the apache configuration file httpd.conf (/etc/httpd/conf/httpd.conf):

```
NameVirtualHost 163.1.73.21 (use your IP address)
<VirtualHost 163.1.73.21> (and again)
ServerName ukmirror2.pubmlst.org (use address provided)
DocumentRoot /home/httpd/pubmlst.org
CustomLog logs/pubmlst.org_access_log combined
</VirtualHost>

<Directory /home/httpd/pubmlst.org>
  DirectoryIndex index.shtml
  Options +Includes
</Directory>
```

If you create one virtual host you also need to add one for the default web site (served from /var/www/html) so also add the following to the httpd.conf file, just above the one for the new mirror (the order determines which is the default site if called by the IP address of the machine)

```
NameVirtualHost 163.1.73.21 (use your IP address)
<VirtualHost 163.1.73.21> (and again)
ServerName atlas.medawar.ox.ac.uk (use your host name)
```

```
DocumentRoot /var/www/html
CustomLog logs/access_log combined
</VirtualHost>
```

Restart the apache server process:

```
sudo /etc/rc.d/init.d/httpd restart
```

Bio-Linux 4 (Debian):

Add the following to the virtual host section of the apache configuration file httpd.conf (/etc/apache/httpd.conf):

```
NameVirtualHost 163.1.73.21          (use your IP address)
<VirtualHost 163.1.73.21>           (and again)
ServerName ukmirror3.pubmlst.org    (use your host name)
DocumentRoot /home/httpd/pubmlst.org
CustomLog /var/log/apache/pubmlst.org_access.log combined
ScriptAlias /cgi-bin/ /var/www/cgi-bin/
Alias /images/ /home/httpd/pubmlst.org/images/
  <Directory /home/httpd/pubmlst.org>
    AllowOverride All
    Options +Includes
    DirectoryIndex index.shtml
  </Directory>
  <Directory /var/www/cgi-bin>
    AllowOverride All
    Options ExecCGI -MultiViews +SymLinksIfOwnerMatch
    Order allow,deny
    Allow from all
  </Directory>
</VirtualHost>
```

If you create one virtual host you also need to add one for the default web site (served from /var/www) so also add the following to the httpd.conf file, just above the one for the new mirror (the order determines which is the default site if called by the IP address of the machine)

```
<VirtualHost 163.1.73.21>
ServerName atlas.medawar.ox.ac.uk
DocumentRoot /var/www
CustomLog /var/log/apache/access.log combined
</VirtualHost>
```

Server side includes are not enabled by default on BL4, you will need to enable them as follows:

```
sudo /usr/sbin/apache-modconf apache enable mod_include
```

Restart the apache server process:

```
sudo /etc/init.d/apache restart
```

10. Make the EMBOSS .libs directory writable to the web server (if EMBOSS installed):

```
sudo chmod a+w /usr/software/EMBOSS/EMBOSS/emboss/.libs
```

11. If you have a firewall installed on the server, it will need to be set to allow web traffic through, i.e. open port 80 (see Bio-Linux 3 FAQ

<http://envgen.nox.ac.uk/envgen/software/archives/000412.html#apache> or BL4 FAQ

<http://envgen.nox.ac.uk/envgen/software/archives/000512.html#firewall>). If your institute has a perimeter firewall it may also be necessary to request web traffic be allowed through to your server.

12. It will take a while (~24 h) for the DNS updates to propagate but in the meantime you can test that the system is working by adding the web and IP addresses to the C:\WINDOWS\system32\drivers\etc\hosts file of a Windows computer and navigating to the new address using a web browser.

13. To keep the databases and sites up-to-date run the following scripts from the webmaster account using a nightly cron job:

```
/home/httpd/pubmlst.org/mirror/scripts/updatemirror  
/home/httpd/pubmlst.org/mirror/scripts/clean.pl > /dev/null
```

The 'clean' script removes temporary files that are older than 24 hours from your pubmlst.org temp directory.