

# Ensembl 14.1 Website Installation Instructions

## *Introduction*

This document provides step-by-step instructions for installing your own version of the Ensembl web site. They are intended to be complete; that is, if you install the site simply by following these instructions, it should work.

We will update this document with each new release, so it should always be up-to-date.

If you find any omissions, errors, or just confusing bits in the instructions, please send an email to [webmaster@ensembl.org](mailto:webmaster@ensembl.org), and we'll try to fix things. Thanks.

*Ensembl Web Team*

## *Summary of instructions*

There are four main sections to these instructions:

- ◆ **Site Structure** - this describes the directory structure of the site, and what the various sections do.
- ◆ **Build / Installation** - how to install all the components that make up the site.
- ◆ **Configuration** - instructions for configuring the site for your local set-up.
- ◆ **Running the site** - how to start up, shut down, and generally maintain the site.

## *Before you start*

Running the Ensembl web site can be rather taxing on hardware. At a minimum you will need:

- ◆ A UNIX-like OS (e.g. Tru64, Solaris, Linux, etc.)
- ◆ Approx. 65 Gigabytes of free hard-disk to hold a complete set of MySQL databases, and about the same again for downloading and unpacking the data.

As with most systems, the more memory and processor speed you can get, the better. As an indication for those looking to purchase machines, our web-servers are 8 Compaq ES45s, with 16Gb main memory and 4 Alpha EV67 1GHz processors.

Ensembl has also been successfully installed on Win32, with some modifications.

## ***Instructions***

### **1. Site structure**

The site structure that we are going to build is self-contained within a server-root directory. It doesn't matter where this server-root directory is situated on your system; as long as the files within it are located as described here then everything should work OK.

The site structure is as follows (using "usr/local/ensembl" as an example server-root):

#### **/usr/local/ensembl**

-- <b>bioperl-live</b>	bio-modelling tools that Ensembl is based upon
-- <b>ensembl</b>	Ensembl core modules
-- <b>ensembl-compara</b>	Code for handling cross-species comparisons
-- <b>ensembl-draw</b>	Drawing code for the website
-- <b>ensembl-external</b>	Code for data outside the core database
-- <b>ensembl-lite</b>	Code for accessing denormalised data
-- <b>ensembl-mart</b>	Code to access the mart data-retrieval system
-- <b>perl</b>	Website mod_perl modules
-- <b>modules</b>	Website accessory modules
-- <b>conf</b>	Website configuration files
-- <b>htdocs</b>	Website HTML and image files
-- <b>utils</b>	Utility scripts

## 2. Build / Install instructions

The installation can be split into two sections, non-Ensembl and Ensembl. The first of these involves the installation of applications/modules that are not part of the Ensembl project but are necessary for the web site to work, such as Apache and MySQL. The second involves installing the Ensembl data, modules and web site code.

### Non-Ensembl Applications Build/Install

These applications are not version-specific for Ensembl; that is, if you upgrade your Ensembl installation to a newer version when one becomes available, you probably won't need to install new versions of these applications.

All of this software, like all of Ensembl, is Open Source software and can be downloaded and used free of charge. You should, however, check the documentation for each application to see what license it has been released under, particularly if you are installing Ensembl in a commercial environment.

The following instructions assume you have root access to the installation machine. If you do not, get your systems administrator to install this software for you.

You may have some or all of this software installed already. If you have any problems getting the site running with pre-installed software (in particular Apache with mod\_perl installed from RPMs), we recommend simply installing the latest version using the following instructions.

### CVS

CVS is a software version control system that we use for storing the source code to Ensembl. You will need CVS installed if you want to download the Ensembl source code. It will also help you keep up to date with any bug fixes.

To install CVS:

- 1 Download the latest source from <http://ftp.cvshome.org/LATEST/>. At the time of writing this is version 1.11.5 and the file to download is cvs-1.11.5.tar.gz
- 2 Unpack the source in a working directory with:

```
$ gunzip < cvs-1.11.5.tar.gz | tar xvf -
$ cd cvs-1.11.5
$ ./configure
$ make
$ make install
```

## Perl

You will probably already have Perl installed. You need Perl5, version 5.6.0 or higher to run the website. Type:

```
$ perl -v
```

to see your version number, and to see if you have Perl installed.

If you don't have Perl installed, or need to upgrade, go to <http://www.cpan.org/> and choose the 'source code' install. Follow the installation instructions on the web site.

## MySQL

MySQL is a very popular Open Source relational database system. You can get pre-compiled binaries from <http://www.mysql.com/downloads/>, or install from source using the instructions given here. Installation of MySQL is slightly more complicated than the other installs, but excellent in-depth instructions can be found on the MySQL site at <http://www.mysql.com/documentation/mysql/bychapter>. We strongly recommend you take a look at the chapter on security.

To install MySQL from source:

- 1 Download the source tarball from <http://www.mysql.com/downloads/mysql-3.23.html>. Get the current stable version - at the time of writing, this is 3.23.55, and the file to download is `mysql-3.23.55.tar.gz`.
- 2 Create a system user and group for the MySQL server to run as. As an example, we will use **mysqldb** in these instructions. Replace `mysqldb` with whatever user you choose.
- 3 Create a directory for MySQL to be installed into. A subdirectory of this will hold the databases, so choose somewhere that has sufficient space free - at least 40 GB. We will use **/data/mysql** as an example. Again, when following these instructions, replace `/data/mysql` with whatever path you choose.
- 4 Copy the tarball to `/data/mysql`
- 5 Unpack the sources with:

```
$ gunzip < mysql-3.23.55.tar.gz | tar xvf -
$ ./configure --prefix=/data/mysql --localstatedir=/data/mysql/var
$ make
$ make install
$ chown -R root /data/mysql
$ cd /data/mysql
$ ./bin/mysql_install_db
$ chown -R mysqldb /data/mysql/var
$ chgrp -R mysqldb /data/mysql
```

You can now start up the MySQL server with:

```
$ /data/mysql/bin/safe_mysqld --user=mysqldb &
```

and bring up a MySQL console (where you can query the database by hand) with:

```
$ /data/mysql/bin/mysql --user=mysqldb
```

## Apache & mod\_perl

Apache is the web server that the Ensembl site runs on. mod\_perl is a module for Apache that enables it to compile perl scripts once rather than each time they are requested, and so makes everything run a lot faster.

Note that while the latest version of Apache is the version 2 series, mod\_perl for Apache 2 is still under development, and so we don't recommend using it. Ensembl may work perfectly well with version 2 of both Apache and mod\_perl, but we haven't tried it. The installation instructions that follow may not work with version 2 of Apache and mod\_perl.

To Install Apache with mod\_perl:

- 1 Download the Apache source tarball from <http://httpd.apache.org/dist/httpd/>. Get the current stable version - at the time of writing, this is 1.3.27, and the file to download is apache\_1.3.27.tar.gz.
- 2 Download the mod\_perl source from <http://www.cpan.org/modules/by-module/Apache/>. Again, get the latest version, currently this is 1.27 and the file to download is mod\_perl-1.27.tar.gz.
- 3 Unpack all the sources in a working directory with:

```
$ gunzip < apache_1.3.27.tar.gz | tar xvf -
$ cd apache_1.3.27
$ ./configure
$ cd ../mod_perl-1.27
```

- 4 Build the perl makefile:

```
$ perl Makefile.PL APACHE_SRC=../apache_1.3.27/src \
  DO_HTTPD=1 USE_APACI=1 EVERYTHING=1
```

- 5 Run the 'make' utility:

```
$ make
```

- 6 ...and install

```
$ make install
```

- 7 Return to the Apache install directory:

```
$ cd ../apache_1.3.27
```

- 8 ...and run the Apache configure script:

```
$ ./configure --enable-module=include \
  --activate-module=src/modules/perl/libperl.a \
  --enable-module=perl
```

- 9 Run 'make', and install:

```
$ make
$ make install
```

## Perl modules

The Ensembl website needs quite a few Perl modules to be installed in order for it to run. These modules can all be downloaded from [www.cpan.org](http://www.cpan.org), and are all installed in much the same way:

Download the module tarball, unpack in a working directory, and install the module:

```
$ gunzip < module.tar.gz | tar xvf -
$ cd module
$ perl Makefile.PL
$ make
$ make test
$ make install
```

The modules that are required are listed below, along with their URLs. The file part of the URL is current at the time of writing - you should install whatever is the latest version of the module.

Apache::DBI	Allows Apache to pool database connections for speed <a href="http://www.cpan.org/modules/by-module/Apache/ApacheDBI-0.89.tar.gz">http://www.cpan.org/modules/by-module/Apache/ApacheDBI-0.89.tar.gz</a>
CGI	Enables Perl scripts to easily parse cgi <a href="http://www.cpan.org/modules/by-module/CGI/CGI.pm-2.89.tar.gz">http://www.cpan.org/modules/by-module/CGI/CGI.pm-2.89.tar.gz</a>
Compress::Zlib	A compression module for DAS. <a href="http://www.cpan.org/modules/by-module/Compress/Compress-Zlib-1.16.tar.gz">http://www.cpan.org/modules/by-module/Compress/Compress-Zlib-1.16.tar.gz</a>
DBI	A common database interface for Perl <a href="http://www.cpan.org/modules/by-module/DBI/DBI-1.30.tar.gz">http://www.cpan.org/modules/by-module/DBI/DBI-1.30.tar.gz</a>
DBD::Mysql	The MySQL drivers for the DBI interface <a href="http://www.cpan.org/modules/by-module/DBD/Msql-Mysql-modules-1.2219.tar.gz">http://www.cpan.org/modules/by-module/DBD/Msql-Mysql-modules-1.2219.tar.gz</a>
GD	A Graphics library Note: may require additional modules. Please read install docs. <a href="http://www.cpan.org/modules/by-module/GD/GD-2.041.tar.gz">http://www.cpan.org/modules/by-module/GD/GD-2.041.tar.gz</a>
Digest::MD5	calculates a unique checksum for a file <a href="http://www.cpan.org/modules/by-module/Digest/Digest-MD5-2.20.tar.gz">http://www.cpan.org/modules/by-module/Digest/Digest-MD5-2.20.tar.gz</a>
Storable	used to store and restore data structures <a href="http://www.cpan.org/modules/by-module/Storable/Storable-2.05.tar.gz">http://www.cpan.org/modules/by-module/Storable/Storable-2.05.tar.gz</a>
LWP	used by DAS to communicate with DAS sources Note: may require additional modules. Please read install docs. <a href="http://www.cpan.org/modules/by-module/LWP/libwww-perl-5.65.tar.gz">http://www.cpan.org/modules/by-module/LWP/libwww-perl-5.65.tar.gz</a>
XML::Parser	used by DAS to parse DAS sources <a href="http://www.cpan.org/modules/by-module/XML/XML-Parser-2.31.tar.gz">http://www.cpan.org/modules/by-module/XML/XML-Parser-2.31.tar.gz</a>

Note: module is a wrapper around the **expat** library, which should be installed first:

Download the expat source from <http://sourceforge.net/projects/expat/>. Get the latest version (currently [expat-1.95.5.tar.gz](http://sourceforge.net/projects/expat/files/expat-1.95.5.tar.gz)). Run the following commands:

```
$ gunzip < expat-1.95.5.tar.gz | tar xvf -
$ cd expat-1.95.5
$ ./configure
$ make
$ make install
```

Parse-RecDescent	used by Excel exporter <a href="http://www.cpan.org/modules/by-module/Parse/Parse-RecDescent-1.80.tar.gz">http://www.cpan.org/modules/by-module/Parse/Parse-RecDescent-1.80.tar.gz</a>
PDF::API2	used by Image exporter for exporting as PDF <a href="http://www.cpan.org/authors/id/A/AR/AREIBENS/PDF-API2-0.3b49.tar.gz">http://www.cpan.org/authors/id/A/AR/AREIBENS/PDF-API2-0.3b49.tar.gz</a>
Spreadsheet::WriteExcel	used for exporting Excel spreadsheets <a href="http://www.cpan.org/modules/by-module/Spreadsheet/Spreadsheet-WriteExcel-0.39.tar.gz">http://www.cpan.org/modules/by-module/Spreadsheet/Spreadsheet-WriteExcel-0.39.tar.gz</a>
OLE::Storage_Lite	Used for exporting Excel spreadsheets <a href="http://www.cpan.org/modules/by-module/OLE/OLE-Storage_Lite-0.11.tar.gz">http://www.cpan.org/modules/by-module/OLE/OLE-Storage_Lite-0.11.tar.gz</a>
Time::HiRes	Used for code profiling <a href="http://www.cpan.org/modules/by-module/Time/Time-HiRes-1.42.tar.gz">http://www.cpan.org/modules/by-module/Time/Time-HiRes-1.42.tar.gz</a>
HTML::Template	Used by MartView <a href="http://www.cpan.org/modules/by-module/HTML/HTML-Template-2.6.tar.gz">http://www.cpan.org/modules/by-module/HTML/HTML-Template-2.6.tar.gz</a>
File::Temp	Used by MartView <a href="http://www.cpan.org/modules/by-module/File/File-Temp-0.12.tar.gz">http://www.cpan.org/modules/by-module/File/File-Temp-0.12.tar.gz</a>
Mail::Mailer	Used by MartView <a href="http://www.cpan.org/modules/by-module/Mail/MailTools-1.51.tar.gz">http://www.cpan.org/modules/by-module/Mail/MailTools-1.51.tar.gz</a>

## Dotter

The Ensembl website contains a page called DotterView which displays a graphical dotplot comparison of two sequences, using the application Dotter. If you wish to use this page on a local installation you will need a local Dotter binary:

Dotter is part of the AceDB package available at:

<http://www.acedb.org/Software/Downloads/supported.shtml>

Pre-built binaries are available for alpha/Tru64 unix and intel/linux. All other flavours will have to be built from sources.

The acedb code makes use of the GTK graphics and GNU readline packages. If you don't have these on your system you can either install them yourself or use the copies distributed with the acedb source.

To install them go to

[www.gtk.org](http://www.gtk.org)

for the GTK package and

<ftp://ftp.cwru.edu/pub/bash/readline-4.3.tar.gz>

for readline.

Follow the GTK and GNU instructions for installing these packages.

If you want to use the versions distributed with the acedb source then follow the instructions for making them in the next section.

Next ungzip/untar the acedb package into an temporary directory:

```
mkdir acedb
mv ACEDB-source.4_91.tar.gz ./acedb/
cd ./acedb
gzip -d < ACEDB-source.4_91.tar.gz | tar xvf -
```

Set an environment variable to tell the build process what platform you are building on - one of:

```
ALPHA_4_DEF
ALPHA_4_GCC_DEF
ALPHA_4_GCC_OPT_DEF
ALPHA_4_LINUX_DEF
ALPHA_4_OLDSTYLE_DEF
ALPHA_4_OPT_DEF
ALPHA_5_DEF
ALPHA_CHRONO_4_DEF
DARWIN_4_DEF
FreeBSD_DEF
FUJITSU_4_DEF
HP_4_DEF
HP_4_GCC_DEF
HP_4_OPT_DEF
IBM_4_3_DEF
IBM_4_DEF
IRIX4_4_DEF
```

```
LINUX_4_DEF
LINUX_MAC_4_DEF
MACOSX_4_DEF
NEC_4_DEF
NEC_4_R10_DEF
NEC_4_R11_DEF
NEXT_4_DEF
POSIX_4_DEF
POSIX_4_GCC_DEF
SGI_4_DEF
SGI_4_GCC_DEF
SGI_4_IRIX5_DEF
SGI_4_PURE_DEF
SGI_5_GCC_DEF
SGI_65_GCC_DEF
SOLARIS_4_DEF
SOLARIS_4_OPT_DEF
SOLARIS_4_RELEASE_DEF
SUN_4_DEF
WIN32_4_DEF
```

without the `_DEF` ending,

For example (in `csh`):

```
setenv ACEDB_MACHINE SOLARIS_4
```

Consult the AceDB documentation if you need more help.

If you wish to use the GTK/readline packages distributed with `acedb` you should build them **before** building any `acedb` code:

```
make gnulibs
```

To build the `dotter` binary issue the command:

```
make dotter
```

(If you just run "make" you will build the whole of AceDB - which is unnecessary).

The build will place the executable at:

```
./bin.${ACEDB_MACHINE}/dotter
```

Check that the binary runs properly and then copy it to your Ensembl shared binaries directory.

DotterView should now work.

## Ensembl Build/Install

This section explains how to install the Ensembl data, perl modules, and web code. It also covers the installation of BioPerl.

### Versioning

In addition to externally maintained software, an Ensembl web site consists of three distinct components:

- The Ensembl API code
- The Ensembl databases
- The Ensembl web code

The Ensembl API is given an integer version number that corresponds to the database schema that it was built for. The current API release is version 14, and has a CVS tag of ' branch-ensembl-14' .

The Ensembl database names consist of the species, the database type, and the database version. The database version consists of the API version (which is equivalent to the database schema), and a further incremental integer to allow for the tracking of data revisions. The current human ' core' database is named `homo_sapiens_core_14_31`. i.e. API version 14, data version 31 (built on the NCBI 31 assembly. Subsequent data releases on this assembly will be suffixed with a lower case letter (a, b, etc.).

The Ensembl web code version consists of the API version, and a further incremental integer to allow for web code revisions between API revisions. The current web code release is version 14-1, and has a corresponding CVS tag of ' branch-ensembl-14-1' .

Ensembl web site releases are versioned according to the combination of API, databases and web code versions. The current human web site release is ' 14.31.1' , where ' 14' is the API, ' 31' is the database, and ' 1' is the web code revision.

The idea is that components with the same API version should work together - i.e. any web site built with ' version 14' API, ' version 14' web code, and ' version 14' databases should work correctly.

### Installing the Ensembl Data

The Ensembl data is provided on the Ensembl FTP site in the form of tab-delimited text files for importing into MySQL. `ftp://ftp.ensembl.org/pub` contains a directory for each release for each species. The latest versions are named `current_species`, e.g. `current_human`, `current_mouse`, etc. The directory structure below this is as follows (using `current_human` as an example):

#### `ftp.ensembl.org/pub/current_human`

```
|-- data
    |-- fasta          cDNA, DNA and peptide dumps
    |-- flatfiles      EMBL and GenBank format dumps
    |-- golden_path    masked and unmasked chromosome seq dumps
    |-- mysql          Database dumps
```

For installing the Ensembl data, we are interested in the contents of the **mysql** directory. This contains a directory for each database we want to install, e.g:

**ftp.ensembl.org/pub/current\_human/data/mysql**

```
|-- homo_sapiens_core_14_31    Core ensembl database
|-- homo_sapiens_disease_14_31 Disease database
...etc...
```

Each database directory contains a checksum file (using a UNIX "sum" utility), so you can verify that the data has downloaded correctly, a data file for each table in that database, and an SQL file that contains the SQL commands necessary to build that database' s table structure.

### Multi-species data

In addition to the individual species data, there is a directory containing multi-species databases - i.e. databases which either affect the site as a whole (e.g. ensembl\_web\_user\_db), or which describe meta-information about multiple species (e.g. ensembl\_compara).

**ftp.ensembl.org/pub/current\_multispecies/data/mysql**

```
|-- ensembl_compara_14_1
|-- ensembl_family_14_1
|-- ensembl_help_14_1
|-- ensembl_web_user_db_14_1
...etc...
```

**Regardless of which species you choose to install, for a full installation you should also install the multi-species databases as well. i.e. compara, family, help and web\_user\_db all need to be installed.**

**NB** The FTP site will ideally be laid out as described. If, however, for reasons of space or maintainability, files are not located as described then check the ftp site for README files which should explain where the data can be found.

### Ensembl data

A final multi-species database can be downloaded from:

**ftp.ensembl.org/pub/current\_mart/data/mysql**

```
|-- ensembl_mart_14_1
```

ensembl-mart is a query-optimised database used for Ensembl' s martview interface

### To install the Ensembl Data:

- 1 Download the directories in ftp.ensembl.org/pub/current\_organism/data/mysql for whatever organism you want to install. Note that the ensembl directory contains several files for the DNA and feature tables - these are very large tables, so the dump file is split into smaller chunks for easier downloading.
- 2 Each table file is gzipped so unpack the data into working directories, keeping separate directories for each database.

For each database you have downloaded, cd into the database directory and perform steps 3-5. **For illustration, we will use homo\_sapiens\_core\_14\_1 as the database - you need to change this appropriately for each database you install.** Remember, you also need to download and install the multi-species databases.

- 3 Start a MySQL console session (see the Installing MySQL section above if necessary) and issue the command:

```
> create database homo_sapiens_core_14_1
```

- 4 Exit the console session, and issue the following command to run the ensembl SQL file, which should be in the directory where you unpacked the downloaded data. This creates the schema for the empty database you created in step 3.

Note that we are using the example MySQL settings of /data/mysql as the install directory, and mysqldb as the database user. Note that here mysqldb is a MySQL account, which is not the same as a system user. See the MySQL documentation for instructions on creating/administering users.

```
$ /data/mysql/bin/mysql -u mysqldb homo_sapiens_core_14_1 <
homo_sapiens_core_14_1.sql
```

- 5 Load the data into the database structure you have just built with the following command.

```
$ /data/mysql/bin/mysqlimport -u mysqldb homo_sapiens_core_14_1
*.txt.table
```

You have now created and loaded the core Ensembl database.

Note that all the databases except the ensembl\_web\_user\_db database only require read access for the website to work. The ensembl\_web\_user\_db requires a MySQL user with delete/insert/update permissions. Also note that because its the only database that the website writes data into, the ensembl\_web\_user\_db has no .table (data) files to download.

**NB** MySQL needs quite a lot of temporary space to load the databases. It is quite possible that your /tmp directory (which MySQL uses by default) is too small, in which case you might see an Error 28 (use the MySQL tool `pererror` to see what these error numbers mean). Fortunately, you can force MySQL to write temporary files to another location. See the MySQL docs for details:

[http://www.mysql.com/doc/T/e/Temporary\\_files.html](http://www.mysql.com/doc/T/e/Temporary_files.html). The simplest solution is to start mysqld with the argument `--tmpdir my_spacious_tmp_location`.

## GO data

The Ensembl ftp site now includes a copy of the GO database as ensembl\_go\_14\_1. Install this if you want local GO information.

## Installing the Ensembl, Web, and BioPerl modules

If you review the Site Structure part of this document, you will recall that the site is based around a single server-root directory. The Ensembl, BioPerl and web modules are all installed into this directory. Choose a suitable location, and create your server-root directory. For the purposes of illustration, we will use `/usr/local/ensembl`. When following these instructions, replace `/usr/local/ensembl` with your chosen server-root.

## To install the Ensembl modules

1. Go to the server-root directory:  
`$ cd /usr/local/ensembl`
2. Log into the Sanger CVS server (using a password of: CVSUSER):  
`$ cvs -d :pserver:cvsuser@cvsro.sanger.ac.uk:/cvsroot/CVSmaster login`
3. Check the ensembl API and web code out of CVS (downloads the code from the Sanger CVS server onto your local machine):

```
$ cvs -d :pserver:cvsuser@cvsro.sanger.ac.uk:/cvsroot/CVSmaster co -r \  
branch-ensembl-14 ensembl-api  
$ cvs -d :pserver:cvsuser@cvsro.sanger.ac.uk:/cvsroot/CVSmaster co -r \  
branch-ensembl-14-1 ensembl-web
```

A listing of your server-root should now look something like:

```
conf/                ensembl-draw/       ensembl-mart/       perl/  
ensembl/            ensembl-external/  htdocs/             utils/  
ensembl-compara/   ensembl-lite/       modules/
```

## To install the BioPerl modules

1. Go to the server-root directory:  
`$ cd /usr/local/ensembl`
2. Log into the BioPerl CVS server (using a password of: cvs):
3. `$ cvs -d :pserver:cvs@cvs.open-bio.org:/home/repository/bioperl login`
4. Check-out the BioPerl code:  
`$ cvs -d :pserver:cvs@cvs.open-bio.org:/home/repository/bioperl co -r branch-1-2 bioperl-live`

You should now have all the Ensembl website code and data installed and ready to configure.

### 3. Configuration

This section explains the configuration changes you have to make so that Ensembl will run on your local set-up.

#### File Configuration

The configuration for the website is split up into site-wide configuration (Apache config, global settings) and species-specific configuration (database names, and so on). Site-wide configuration is stored in SiteDefs.pm, while the species-specific config is kept in *species\_name*.ini files (e.g. Homo\_sapiens.ini). There needs to be a species-specific .ini file for each species you want to display on your Ensembl website.

In addition to these configuration files, there is a DEFAULTS.ini file and a MULTI.ini file. DEFAULTS.ini holds default values for the species-specific .ini files; i.e. unless you override these settings in a particular species ini file, they will apply to all species. This is simply to save having to make the same setting in multiple species ini files.

MULTI.ini contains multi-species settings. This currently contains database connection information for the ensembl\_compara, ensembl\_family and ensembl\_mart databases.

These configuration files can all be found in the "conf" subdirectory of your server-root.

#### SiteDefs.pm

If you open this file in a text editor, such as vi, you will see it is actually a Perl script. The first part of the script exports a number of variables that control how the site works. The part you want to edit is the section after the label:

```
#####  
##### LOCAL CONFIGURATION VARIABLES #####  
#####
```

This section consists of a list of variables (of the form \$ENSEMBL\_whatever), and a value for the variable, and a comment (after a # symbol) about the variable. You will need to change the values of several of these variables to match your installation.

#### General configuration

Change the value of \$ENSEMBL\_SERVERROOT to your server-root. For example, if you installed the Ensembl site in /usr/local/ensembl, then you should change this line of SiteDefs to read:

```
$ENSEMBL_SERVERROOT = '/usr/local/ensembl';
```

You should only change the values, that is the parts between single quotes.

#### Configuration of the Apache web server

Change \$ENSEMBL\_SERVERNAME to the web name of the server - e.g. "www.yoursite.org"

Change \$ENSEMBL\_USER and \$ENSEMBL\_GROUP to the system user and group you want the Apache web server to run as. Usually, for security, this is a special user (such as "nobody") who has

very few permissions.

### ***Mail configuration - error messages***

If you want errors to be automatically emailed to you, change \$ENSEMBL\_MAIL\_ERRORS to the value 1, and change \$ENSEMBL\_ERRORS\_TO to your email address. If you don't want errors mailed, set \$ENSEMBL\_MAIL\_ERRORS to 0.

### ***User database - Database and cookie configuration***

Change the values of \$ENSEMBL\_USERDB\_NAME, \$ENSEMBL\_USERDB\_HOST, \$ENSEMBL\_USERDB\_USER, \$ENSEMBL\_USERDB\_PASS, to the details for your web\_user database, remembering that this database needs a user with update/insert/delete privileges. Additionally you may wish to change the encryption keys used to "protect" the cookies - \$ENSEMBL\_ENCRYPT\_0 should be a six digit hex-number, \$ENSEMBL\_ENCRYPT\_1, \_2 and \_3 should each contain two alphanumeric characters. Unless you are particularly concerned about people changing cookies, the default values will probably do.

### ***Enabling Species***

\$ENSEMBL\_SPECIES\_ALIASES contains a hash of species aliases against species names. For example:

```
$ENSEMBL_SPECIES_ALIASES = {  
    ' human'   => ' Homo_sapiens' ,  
    ' mouse'   => ' Mus_musculus' ,  
    ' mosquito' => ' Anopheles_gambiae' ,  
};
```

The values of the hash (the species names on the right hand side) defines the list of valid species for this website. The keys of the hash (the aliases on the left) can be used in URLs instead of the longer species names. i.e. <http://my.ensembl.site/mouse/contigview> is identical to [http://my.ensembl.site/Mus\\_musculus/contigview](http://my.ensembl.site/Mus_musculus/contigview).

To configure a species it has to be present as a value in the \$ENSEMBL\_SPECIES\_ALIASES hash, and it has to have an appropriately-named .ini file in the conf directory.

You might also want to change \$ENSEMBL\_PERL\_SPECIES - this is a special alias for URLs of the form <http://my.ensembl.site/perl/contigview>, for purely historical reasons. It will probably be used in the future to set the default species for multi-species pages/functions.

### ***Temporary Files***

There are three temporary file locations that can be configured:

\$ENSEMBL_TMP_DIR	General storage for temporary files
\$ENSEMBL_TMP_DIR_IMG	Storage for image files
\$ENSEMBL_TMP_DIR_BLAST	Storage for blast files

The values for these should be set to an appropriate filesystem path.

Some temporary files need to be referenced by URL from web pages. The first two tmp directories

above therefore have URL aliases, also configured within SiteDefs.pm. You should not need to edit these.

\$ENSEMBL_TMP_URL	URL alias for \$ENSEMBL_TMP_DIR
\$ENSEMBL_TMP_URL_IMG	URL alias for \$ENSEMBL_TMP_DIR_IMG

There are two further temporary file configuration options available:

\$ENSEMBL_TMP_CREATE	If set, then at apache startup the server will attempt to create any temporary directories that have been configured, but which don't already exist. It also changes their ownership to \$ENSEMBL_USER.\$ENSEMBL_GROUP
----------------------	--

\$ENSEMBL_TMP_DELETE	If set, then at apache startup the server will attempt to delete the contents of \$ENSEMBL_TMP_DIR and \$ENSEMBL_TMP_DIR_IMG.
----------------------	---

### **[Species\_name].ini**

The format of the species-specific .ini files are similar to a standard windows ini file, i.e. split into sections identified by a header in square brackets, which contain key = value entries. When the Apache server is started, these files are parsed and stored in a file called config.packed in the conf directory. This config.packed file is regenerated whenever the server is restarted.

### **Database configuration**

In the [general] section, change the values of ENSEMBL\_DBUSER and ENSEMBL\_DBPASS to the username and password you want MySQL to be accessed by. Set ENSEMBL\_HOST and ENSEMBL\_HOST\_PORT to be the database server and port that MySQL is running on.

These are default settings - you can override them by adding a section for a particular database. For example, adding a section like the following would override the defaults for the Help database:

```
[ENSEMBL_HELP]
USER=mysqluser2
PASS=helppass
HOST=mysqlserver2
PORT=4444
```

### **Names of databases**

In the [databases] section, change the values of ENSEMBL\_DB, ENSEMBL\_EST, ENSEMBL\_LITE, etc. to match the names of the core, EST, lite, etc. databases you created.

### **DAS Proxy**

If you wish to display external DAS sources on your ensembl installation, and are behind a firewall, then you will need to set a value for ENSEMBL\_DAS\_PROXY. This should probably be set in DEFAULTS.ini, as the proxy is likely to be the same for all species. The value should be your usual web proxy setting, e.g.

ENSEMBL_DAS_PROXY	= http://webproxy.mycompany.com:80
-------------------	------------------------------------

## **MULTI.ini**

Configure the [general] section for connection to the MySQL server, as for the species-specific ini files. In the [databases] section, change the values of ENSEMBL\_COMPARA and ENSEMBL\_MART to match the name of the ensembl\_compara and ensembl\_mart databases that you created.

In addition, if you chose to install a local GO database you can configure it here by adding an entry: ENSEMBL\_GO = your\_go\_database\_name to the [databases] section along with ENSEMBL\_COMPARA and ENSEMBL\_MART.

As in the other species.ini files, you can override the database connection settings for a particular database by adding a section similar to the following:

```
[ENSEMBL_MART]
USER=mysqluser2
PASS=dbpass
HOST=mysqlserver2
PORT=4444
```

## Note on Ensembl Web Site Search

The Ensembl web site uses the AltaVista search engine (<http://www.altavista.com/>). This software requires a user license and cannot be distributed as part of the Ensembl web code. If you would like to install and use AltaVista locally, please contact AltaVista Support: [ensemblav@av.com](mailto:ensemblav@av.com). AltaVista are currently considering introducing preferential licensing terms for academic users of Ensembl/AV search.

As an alternative to using AltaVista, two external contributors, Chen Peng and Dyfed Lloyd-Evans, have built a system using the new full-text indexing capabilities of MySQL4. You can find the code and instruction for use in the `contrib/mysql_indexer` directory. We haven't tested it here, but it is working for the Ensembl Fugu site in Singapore.

By default, local installations of Ensembl use a simple search page called Unisearch that just does SQL searches against the database. This method is rather slow and crude, however.

## Site Preparation

cd into your server root and run the following commands:

```
$ mkdir logs
$ chown -R $ENSEMBL_USER:$ENSEMBL_GROUP
```

where `$ENSEMBL_USER` and `$ENSEMBL_GROUP` are the web server user & group you configured in `SiteDefs`.

Your Ensembl site should now be ready to start up.

## DISCLAIMER

Please note that the suggested set-up for Apache, `mod_perl`, MySQL and all Ensembl modules is not intended to provide high-level security. We strongly recommend that you get your systems administrator to audit any system that you provide to others. In particular, note that the `perl.startup` file in the `conf` directory may be run as root, so be careful about permissions on this file.

We are not responsible for any damage or loss of data resulting or arising from running the Ensembl website on your own machines.

## 4. Maintenance

This section outlines how to go about the day-to-day running of an Ensembl website.

### Starting the Ensembl website

1. Check that MySQL is running. If not, you can start it:

```
$/data/mysql/bin/safe_mysqld --user=mysqldba &
```

You should change the `/data/mysql` and `mysqldba` in the command above to the install path and user you chose when you installed MySQL.

2. Start the web server:

```
$/usr/local/apache/bin/httpd -d /usr/local/ensembl
```

where `/usr/local/ensembl` is your server-root.

Note that it is important that you start MySQL before you start the website. The startup scripts for the site rely on the databases being available, as they compile lists of available data at this point.

You should now be able to point a web-browser at your server and see the Ensembl homepage.

### Stopping the Ensembl website

1. To stop the web server:

```
$/ kill `cat /usr/local/ensembl/logs/HOSTNAME.httpd.pid`
```

where `/usr/local/ensembl` is your server-root, and `HOSTNAME` is whatever the hostname is of the machine you are running the server on. Note the punctuation marks are backticks and not single-quotes.

2. To stop MySQL:

```
$/data/mysql/bin/mysqladmin shutdown
```

### Troubleshooting

The best way of troubleshooting is to watch the Apache `error_log` and look for exactly what the problem is. You can do this with the command:

```
$/ tail -f logs/error_log
```

from the server-root. It is often obvious what the error is, especially if it is due to file permissions or a missing Perl module.

If you have email errors turned on, then you will be receiving some additional information via that route. Please email us at [helpdesk@ensembl.org](mailto:helpdesk@ensembl.org) if you need assistance. Any details of the error, relevant sections from the `error_log`, etc, will be of help in tracking down your problem.