

Exonmap installation instructions

MichałJ Okoniewski, Tim Yates, Crispin J Miller

July 16, 2008

Contents

1	Installation of the X:Map database	1
2	Configuration	2
2.1	The configuration directory	2
2.2	Specifying available databases	2
2.3	The db.local subdirectory	3

1 Installation of the X:Map database

The R package requires access to a MySQL database server running a current version of the X:Map database (and an installation of Ensembl). Database tables and installation instructions may be downloaded from <http://xmap.picr.man.ac.uk/download> (the installation instructions can be found inside the database download files). Briefly, what you will need to do is:

1. If it is not already there, download and install MySQL on a database server. This does not necessarily have to be the same machine that R is running on. A common approach is to share a single database installation with a number of machines running R.
2. Download and install Ensembl on this machine
3. Download and install X:Map on the same machine
4. Download and install the *RMySQL* package
5. Install exonmap
6. Create a configuration directory for exonmap to use, and within it, a file 'databases.txt' telling it how to find the X:Map database(s).

The first three steps are described in the X:Map installation instructions; here we describe the next three steps...

Once the server is up and running, you must create the necessary configuration files to tell R how to connect to the database.

2 Configuration

2.1 The configuration directory

Exonmap uses a configuration directory that stores database connection details. This is specified by the environment variable `R_XMAP_CONF_DIR`. If this variable is not set then exonmap tries to find your home directory and looks in a directory within that called `.exonmap` - but will generate a warning complaining that the environment variable is not set up.

So the easiest thing to do is first to create a directory called `.exonmap` in your top level user directory, and then set up the environment variable to point to it (you will probably want to put this in your initial start up scripts). How to do this is platform dependent - refer to the instructions for your particular OS for more details.

2.2 Specifying available databases

Next create a config file telling exonmap where to look for its databases. NB *this differs from previous versions of exonmap*. Create a file `'databases.txt'` in your `.exonmap` directory. The first row contains column headers, and these are followed by a row for each database; e.g.:

Table 1: databases.txt

name	species	db.name	type	host	port	username	password
"human"	"homo_sapiens"	"homo_sapiens_core_47_36i"	"core"	"HOST"	""	"UNAME"	"PWDOR"
"human"	"homo_sapiens"	"homo_sapiens_otherfeatures_47_36i"	"otherfeatures"	"HOST"	""	"UNAME"	"PWDOR"
"human.old"	"homo_sapiens"	"homo_sapiens_core_old.version"	"core"	"HOST"	""	"<ASK>"	"<ASK>"
etc... more rows for additional databases, species, etc.							

The first column (name) is used within R to specify the name of the database to connect to, as used by `xmapConnect` (e.g. `xmapConnect("human")`). Other columns should be self-explanatory - "HOST", "UNAME" and "PWDOR" should obviously be replaced with the name of the database host and the necessary login details. Both "UNAME" and "PWDOR" can be replaced by "<ASK>" in which case, the user will be prompted at the command line to supply the necessary information. 'port' allows the database port to be specified (if it is other than the default).

`exonmap` can potentially make use of two X:Map databases for each genome. The first, 'core', database provides mappings to Ensembl genes, the second 'otherfeatures'

supplies additional mappings for ESTs and gene predictions. It does not need to be installed, but is, of course, required for mappings involving these additional datasets.

In the example above the first two rows both define “human”, but specify different database names (“homo_sapiens_core_47_36i” and “homo_sapiens_otherfeatures_47_36i”) and different types (“core” and “other features”). *exonmap* will look first for “core” databases and then to see if there is an “otherfeatures” database with the same name. If so, it will try to connect to both at the same time.

Before first use of any of the database related functions, the connection function for the Human, Mouse or Rat database must be called using, for example, *xmapConnect*(“human”).

2.3 The db.local subdirectory

If the directory *db.local* exists in *.exonmap*, *exonmap* will try to generate a local datafile within it. This contains the data it needs for the probeset filtering functions (e.g. *select.probewise*). This can result in considerable gains in speed for large queries.

If all goes well, that should be it.