

# Cistematic 2.5 install page

*A platform for cis-regulatory element analysis within and across multiple genomes.*

---

## What is Cistematic

The core of Cistematic is a Python package with a rich set of API's that simplify the collection and analysis of candidate cis-regulatory elements from a number of different motif-finding programs such as Meme, and cisGreedy (built-in). Cistematic assesses the significance of each motif by comparing it to its prevalence genome-wide.

One of the more useful APIs in Cistematic involved wrapping genes, annotations, and genomic sequences that are stored in a sqlite database. It is on top of this foundation that we have built additional platforms, such as ChIPSeqMini, and ERANGE.

This page has the updated version of the Cistematic code underlying the following papers:

- Comparative genomics modeling of the NR5F/REST repressor network: from single conserved sites to genome-wide repertoire (Mortazavi, 2006)
- Genome-wide mapping of in vivo protein-DNA interactions (Johnson, 2007)
- Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq (Mortazavi, 2008)

## Cistematic Pre-requisites

### Required

Cistematic currently runs on Linux and Macintosh; it also runs under Cygwin in Microsoft Windows. In addition to python, the current version of Cistematic is heavily dependent on sqlite and its python interface, pysqlite (which is now part of python 2.5+). You will therefore need:

- Any Unix-compatible OS such as Linux and MacOS 10.4-10.5
- Python 2.5 or better
- Python Imaging Library: [1.1.6+](#)

Note that earlier versions of python did not include pysqlite, which we rely on heavily. You might be able to get Cistematic to run with older versions of python if you separately install sqlite/pysqlite, but we do not support it any longer.

### Optional (Can be installed later)

In addition to the requirements listed above, three additional packages will allow you to get the most out of Cistematic. psyco, which only runs on Intel 32-bit CPUs and on **ALL** Macintosh Intel platforms, will give you approximately 9-fold speed up running Cistematic code and is highly recommended, if it's available for your platform.

Matplotlib was used to generate the figures in the papers and is hence also recommended.

Weblogo is used to visualize the PSFMs in ERANGE. It is not necessary if you are only interested in RNA-seq, for example.

The actual version numbers are:

- [psyco 1.6+](#) **HIGHLY RECOMMENDED ON INTEL 32-bit AND ALL MAC INTEL PLATFORMS**
- [matplotlib 0.9+](#)
- [weblogo 2.8.2 \(not 3.0\)](#)

Weblogo's seqlogo script should be installed in `$CISTEMATIC_ROOT/programs/weblogo` and that we only save motifs in PNG format. Note that weblogo itself requires ghostscript to be installed for PNG support, which is hence a secondary Cistematic dependency.

## Cistematic core

You will need to download the following packages:

- [cistematic2.5.tgz](#)
- [db2.0.tgz](#)

as well as a set of motif finding binaries (not required for RNA-seq, but definitely for ChIP-seq):

- [programsMin.mac10\\_4.tgz](#)
- [programsMin.linux32.tgz \(32bit\)](#)

For motif-finding with ChIP-seq data, you can now use the bundled cisGreedy program, a python implementation of Consensus (Hertz, 1999), which will work well for short motifs, but isn't optimized for speed. You may also want to download Meme and modify the parameter `memePath` in `$CISTEMATIC_ROOT/cistematic/programs/meme.py` to the appropriate path for the meme top-level script.

To install, create a directory (for example `/proj/genome`), `cd` into it, and unpack each file using `tar xzvf`.

You will need to add the directory in which you installed the Cistematic python code to your `PYTHONPATH` environment!

If you use a root directory different from `/proj/genome`, you will need to tell Cistematic where to find it by setting up the environmental variable `CISTEMATIC_ROOT`.

Some of the code for `cistematic.core.motif` for PSFM and Markov1 motif scanning across mammalian genomes in pure python mode. Therefore some of the key code is *optionally* compilable into a python C-extension. The default distribution contains the source code as well as pre-compiled binaries for Linux (64 bit) as default and MacOS 10.5. Mac users should rename the precompiled binary for the extension to be used:

```
cp $CISTEMATIC_ROOT/cistematic/core/_motif.so $CISTEMATIC_ROOT/cistematic/core/_motif.so.linux
cp $CISTEMATIC_ROOT/cistematic/core/_motif.so.mac $CISTEMATIC_ROOT/cistematic/core/_motif.so
```

You can recompile the extension, if ever necessary and reinstall it:

```
cd $CISTEMATIC_ROOT/cistematic/core
```

```
python setup.py build
```

```
copy the resulting binary in the appropriate build subdirectory back to $CISTEMATIC_ROOT/cistematic/core
```

## Cistematic Genomes

To make full use of Cistematic, you will need to use genomes that are either installed from scratch, or that are installed from the following packages:

- [mouse mm9](#)
- [human hg18](#)
- [dog cf2](#)
- [rat rn4](#)
- [C elegans ce6 \(WS190\)](#)
- [D melanogaster dm3](#)
- [Arabidopsis thaliana \(TAIR8\)](#)

which are unpacked in the same manner as the other Cistematic files.

## **Previous versions of Cistematic**

Earlier versions of Cistematic are now obsolete, but files for the earlier release of Cistematic are available [here](#).

---

Last Modified: 2009/04/20 by Ali Mortazavi