

Biopython Installation

Brad Chapman, with other contributors

November 26, 2010

Contents

1	Purpose and Assumptions	2
2	C Compiler	2
2.1	Unix	2
2.2	Mac OS X	2
2.3	Windows	2
3	Installing Python	2
3.1	Python installation on UNIX systems	3
3.1.1	RPM and other Package Manager Installation	3
3.2	Python installation on Windows	4
3.3	Python installation on Mac OS X	4
4	Installing Biopython dependencies	4
4.1	Numerical Python (NumPy) (strongly recommended)	4
4.1.1	UNIX and Mac OS X systems	4
4.1.2	Windows systems	5
4.1.3	Making sure it installed correctly	5
4.2	ReportLab (optional)	5
4.2.1	UNIX and Mac OS X systems	5
4.2.2	Windows systems	6
4.2.3	Making sure it installed correctly	6
4.3	Database Access (MySQLdb, ...) (optional)	6
4.4	mxTextTools (no longer needed)	6
4.5	flex (fast lexical analyzer generator) (optional)	6
5	Installing Biopython	6
5.1	Obtaining Biopython	6
5.2	Installing on UNIX and Mac OS X	7
5.2.1	Installation from source on UNIX and Mac OS X	7
5.2.2	Using the Python package index	8
5.2.3	Installation on Mac OS X using the fink package manager	8
5.2.4	Installation on UNIX systems using RPMs	8
5.3	Installing with a Windows Installer	8
5.4	Installing from source on Windows	9
6	Making sure everything worked	9
7	Third Party Tools	10

1 Purpose and Assumptions

For those of you familiar with installing python packages and who don't care for following details instructions can try going to <http://biopython.org/wiki/Download>, installing the relevant prerequisites, and Biopython.

This document describes installing Biopython on your computer. To make things as simple as possible, it basically assumes you have nothing related to Python or Biopython on your computer and want to end up with a working installation of Biopython when you are finished following through this documentation.

Biopython should work on just any operating system where Python works, so these instructions contain directions for installation on UNIX/Linux, Windows and Macintosh machines. The directions assume that you have permission to install programs on the machine (root access on UNIX and Administrator privileges on Windows or Mac machines). While it is certainly possible to install things without these privileges, this is a serious pain and all the tedious workarounds aren't something that I'll go into very much in this documentation.

With all this said, hopefully these directions will make it straightforward to get Biopython installed on your machine so you can begin using it as quick as possible.

2 C Compiler

Although mostly written in Python, Biopython (and some of its dependencies) include C code, which must be compiled to run. If you are going to be installing from source you will therefore need a C compiler. However, in many cases this can be avoided by using pre-compiled packages (which is what we recommend on Windows).

2.1 Unix

We recommend GCC as the C compiler, this is usually available as part of the standard set of packages on any Unix or Linux system.

2.2 Mac OS X

Please install Apple's XCode suite - *including* the optional Mac OS 10.4 SDK which includes important header files.

2.3 Windows

We recommend you install Biopython and its dependencies using the provided pre-compiled Windows Installers. i.e. You don't *need* a C compiler. See Section 5.4 for more details.

3 Installing Python

Python is a interpreting, interactive object-oriented programming language and the home for all things python is <http://www.python.org>. Presumably you have some idea of python and what it can do if you are interested in Biopython, but if not the python website contains tons of documentation and reasons to learn to program in python.

Biopython is designed to work with Python 2.4 or later (but not Python 3 yet). With python, the general rule of thumb is to keep yourself using the latest version, as the development process is very clean and new releases are quite stable. Upgrading bug-fix releases (for example. 2.6.1 to 2.6.2) is incredibly easy and won't

require any re-installation of libraries. Upgrading between versions (e.g. 2.5 to 2.6) is more time consuming since you need to re-install all libraries you have added to python.

Let's get started with installation on various platforms.

3.1 Python installation on UNIX systems

First, you should go the main python web site and head over to the information page for the latest python release. At the time of this writing the latest stable python release is 2.6.2, which is available from <http://www.python.org/download/releases/2.6.2/>. This page contains links to all released files for the given release. For UNIX, we'll want to use the tarred and gzipped file, which is called `Python-2.6.2.tgz` at the time of this writing.

Download this file and then unpack it with the following commands:

```
$ gunzip Python-2.6.2.tgz
$ tar -xvpf Python-2.6.2.tar
```

Then enter into the created directory:

```
$ cd Python-2.6
```

Now, start the build process by configuring everything to your system:

```
$ ./configure
```

Build all of the files with:

```
$ make
```

Finally, you'll need to have root permissions on the system and then install everything:

```
# make install
```

If there were no errors and everything worked correctly, you should now be able to type `python` at a command prompt and enter into the python interpreter:

```
$ python
Python 2.6.2 (...)
[GCC 3.3.3] on cygwin
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

(The precise version text and details will depend on the version you installed and your operating system.)

3.1.1 RPM and other Package Manager Installation

There are a multitude of package manager systems out there for which python is available. One popular one is the RPM (RedHat Package Manager) system. Each of these package managing systems has its own quirks and tricks and I certainly can't pretend to understand them all so I won't try to describe them all here.

While these package repositories may include Biopython all ready to install, you will typically want to install Biopython from source to get the very latest version.

However, there is one general point which it is important to remember when installing from any of these systems: you need to download and install the development packages for python. A number of distributions contain a "basic" python which contains libraries and enough stuff to run simple python programs. However, they do not contain the python libraries necessary to build third-party python applications (like Biopython and its dependencies). You'll need to install these libraries and header files, which are often found in a separate package called `python-devel` or something similar.

3.2 Python installation on Windows

Installation on Windows is most easily done using handy windows installers. As described above in the UNIX section, you should go to the webpage for the current stable version of Python to download this installer. At the current time, you'd go to <http://www.python.org/download/releases/2.6.2/> and download `Python-2.6.2.msi`.

The installer is an executable program, so you only need to double click it to run it. Then just follow the friendly instructions. On all newer Windows machines you'll need to have Administrator privileges to do this installation.

3.3 Python installation on Mac OS X

Apple includes python on Mac OS X, and while you can use this many people have preferred to install the latest version of python as well in parallel. We refer you to the <http://www.python.org> for more details, although otherwise the UNIX instructions apply.

4 Installing Biopython dependencies

Once python is installed, the next step is getting the dependencies for Biopython installed. Since not all functionality is included in the main python installation, Biopython needs some support libraries to save us a lot of work re-writing code that already exists. We try to keep as few dependencies as possible to make installation as easy as possible.

4.1 Numerical Python (NumPy) (strongly recommended)

The Numerical Python distribution is a fast implementation of arrays and associated array functionality. This is important for a number of Biopython modules that deal with number processing (e.g. `Bio.Cluster` and `Bio.PDB`).

As of release 1.49, Biopython supports the standard NumPy distribution. Previous releases instead used the older Numeric module (which is no longer being maintained).

The main web site for NumPy is: <http://numpy.scipy.org/>.

4.1.1 UNIX and Mac OS X systems

You should download the `tar.gz` file, and follow the standard python build process. Note you will need a C compiler installed (see above):

```
> tar -xzvpf numpy-1.1.1.tar.gz
> cd numpy-1.1.1/
> python setup.py build
```

Once it is built, you should become root, and then install it:

```
> python setup.py install
```

One important note if you use an package system and not installing NumPy from source: you may also need to install the header files which are not included with some packages. As with the main python distribution, this means you'll need to look for something like `python-numpy-devel` and make sure to install this as well as the basic package.

4.1.2 Windows systems

We recommend using the NumPy provided windows installers for your installed version of python. For python 2.5, at the current time this would be `numpy-1.3.0-win32-superpack-python2.5.exe`. You should follow the now-standard procedure of downloading the installer, double clicking it and then following the installation instructions. As before, you will need to have administrator permissions to do this.

4.1.3 Making sure it installed correctly

To make sure everything went okay during the install, fire up the python interpreter and ensure you can import NumPy without any errors:

```
> python2.5
Python 2.5.2 (r252:60911, Apr 27 2008, 11:46:35)
[GCC 4.2.3 (Debian 4.2.3-3)] on linux2
Type 'help', 'copyright', 'credits' or 'license' for more information.
>>> import numpy
>>>
```

Note that for the import statement, NumPy should be in lower case!

4.2 ReportLab (optional)

The ReportLab package is a library for generating PDF documents. It is used in the Biopython Graphics modules, which contains basic functionality for drawing biological objects like chromosomes. If you are not planning on using this, installing ReportLab is not necessary. ReportLab in itself is very useful for a number of tasks besides just Biopython, so you may want to check out <http://www.reportlab.org> before making your decision.

The main download page for ReportLab is <http://www.reportlab.org/downloads.html>. The ReportLab company has some commercial products as well, but just scroll down their page to the Open Source software section for the base ReportLab downloads.

If you want to generate bitmap images, you will also need the ReportLab module renderPM. This in turn requires the [Python Imaging Library \(PIL\)](#).

4.2.1 UNIX and Mac OS X systems

For UNIX installs, you should download the tarred and gzipped version of the ReportLab distribution. At the time of this writing, this is called `ReportLab_2_3.tar.gz`. First, unpack the distribution and change into the created directory:

```
$ gunzip ReportLab_2_3.tar.gz
$ tar -xvpf ReportLab_2_3.tar
$ cd reportlab_2_3/
```

Once again, ReportLab uses the standard python installation system which you are probably feeling really comfortable with by now. So, first build the package:

```
$ python setup.py build
```

Now become root, and install it:

```
$ python setup.py install
```

4.2.2 Windows systems

ReportLab now has graphical windows installers. Nice and easy.

4.2.3 Making sure it installed correctly

If reportlab is installed correctly, you should be able to do the following:

```
$ python
Python 2.4 (#1, Dec 5 2004, 20:47:03)
[GCC 3.3.3] on cygwin
Type "help", "copyright", "credits" or "license" for more information.
>>> from reportlab.graphics import renderPDF
>>>
```

Depending on your version of python and what you have installed, you may get the following warning message: **Warn: Python Imaging Library not available**. This isn't anything to worry about unless you want to produce bitmap images, since the Biopython parts that use ReportLab will work just fine without it.

4.3 Database Access (MySQLdb, ...) (optional)

The MySQLdb package is a library for accessing MySQL databases. It is used in the Biopython GFF module, which allows access to databases created from GFF feature tables. Biopython also includes an accessory module, DocSQL, which provides a convenient interface to MySQLdb. If you are not planning on using Bio.GFF or Bio.DocSQL, installing MySQLdb is not necessary.

Additionally, both MySQLdb and psycopg (a PostgreSQL database adaptor) can be used for accessing BioSQL databases through Biopython (see <http://biopython.org/wiki/BioSQL>). Again if you are not going to use BioSQL, there shouldn't be any need to install these modules.

4.4 mxTextTools (no longer needed)

Historically this was an important Biopython dependency, used extensively in a number of parsers. However, we have gradually phased out its use, and as of Biopython 1.50, mxTextTools is no longer used at all.

mxTextTools is available along with the entire mx-base system (which contains a number of other useful utilities as well) and the latest version is available for download at: <http://www.egenix.com/products/python/mxBase/mxTextTools/>.

4.5 flex (fast lexical analyzer generator) (optional)

The mmCIF parser `Bio.PDB.mmCIF.MMCIFlex` relies on C code which uses flex (fast lexical analyzer generator). At the time of writing, in order to parse mmCIF files you'll have to install flex, then tweak your `setup.py` file to include the `Bio.PDB.mmCIF.MMCIFlex` module, before (re)installing Biopython from source.

5 Installing Biopython

5.1 Obtaining Biopython

Biopython's internet home is at, naturally enough, <http://www.biopython.org>. This is the home of all things Biopython, so it is the best place to start looking around. You have two choices for obtaining Biopython:

1. Release code – We made available releases on the download page (<http://biopython.org/wiki/Download>). The releases are also available both as source and as installers (windows installers right now), so you have some choices to pick from on releases if you prefer not to deal with source code directly.
2. git – The current working copy of the Biopython sources is available via git hosted on github – <http://github.com/biopython/biopython>). Concise instructions for accessing this copy are available at <http://biopython.org/wiki/SourceCode>. Our code in git is normally quite stable but there is always the caveat that the code there is under development.

Based on which way you choose, you'll need to follow one of the following installation options. Read on for the platform you are working on.

5.2 Installing on UNIX and Mac OS X

5.2.1 Installation from source on UNIX and Mac OS X

Biopython uses Distutils, the standard python installation package, for its installation. If you read the install instructions above you are already quite familiar with its workings. Distutils comes standard with Python 1.6 and beyond.

Now that we've got what we need, let's get into the installation:

1. First you need to unpack the distribution. If you got the git version, you are all set to go and can skip on ahead. Otherwise, you'll need to unpack it. On UN*X machines, a tar.gz package is provided, which you can unpack with `tar -xvzf biopython-X.X.tar.gz`. A zip file is also provided for other platforms.
2. Now that everything is unpacked, move into the `biopython*` directory (this will just be `biopython` for git users, and will be `biopython-X.X` for those using a packaged download).
3. Now you are ready for your one step install – `python setup.py install`. This performs the default install, and will put Biopython into the `site-packages` directory of your python library tree (on my machine this is `/usr/local/lib/python2.4/site-packages`). You will have to have permissions to write to this directory, so you'll need to have root access on the machine.
 - (a) This install requires that you have the python source available. You can check this by looking for `Python.h` and `config.h` in some place like `/usr/local/include/python2.5`. If you installed python with RPMs or some other packaging system, this means you'll also have to install the header files. This requires installing the python development libraries as well (normally called something like `python-devel-2.5.rpm`).
 - (b) The distutils setup process allows you to do some customization of your install so you don't have to stick everything in the default location (in case you don't have write permissions there, or just want to test Biopython out). You have quite a few choices, which are covered in detail in the distutils installation manual (<http://www.python.org/sigs/distutils-sig/doc/inst/inst.html>), specifically in the Alternative installation section. For example, to install Biopython into your home directory, you need to type `python setup.py install --home=$HOME`. This will install the package into someplace like `$HOME/lib/python2.5/site-packages`. You'll need to subsequently modify the `PYTHONPATH` environmental variable to include this directory so python will be able to find the installation.
4. That's it! Biopython is installed. Wasn't that easy? Now let's check and make sure it worked properly. Skip on ahead to section [6](#).

5.2.2 Using the Python package index

Another simple option is to use the Python package index (<http://pypi.python.org/pypi>) with the `easy_install` command:

```
easy_install -f http://biopython.org/DIST/ biopython
```

If Python is installed in the standard location, you will need administrator privileges to do this; the `sudo` command works well:

```
sudo easy_install -f http://biopython.org/DIST/ biopython
```

5.2.3 Installation on Mac OS X using the fink package manager

Instead of installing from source, on Mac OS X you can also use the fink package manager, see <http://fink.sf.net>. Fink should take care of downloading the source code and installing all needed packages for Biopython, including Python itself. Once you have installed fink, you can install biopython using:

```
fink install biopython-pyXX
```

where XX is the python version you would like to use. Currently, python 2.4, 2.5, and 2.6 are available through fink on Mac OS X 10.4, so you would have to replace XX with 24, 25, or 26, respectively. Most likely, you will have to enable the unstable tree of fink in order to install the most recent versions of the package, see also this item in the Fink FAQ: <http://fink.sourceforge.net/faq/usage-fink.php#unstable>. Note that 'unstable' doesn't mean that a package won't work, but only that there has not been feedback to the fink team from users.

5.2.4 Installation on UNIX systems using RPMs

Warning. Right now we're not making RPMs for biopython (because I stopped using an RPM system, basically). If anyone wants to pick this up, or feels especially strongly that they'd like RPMs, please let us know.

To simplify things for people running RPM-based systems, biopython can also be installed via the RPM system. Additionally, this saves the necessity of having a C compiler to install biopython.

Installing Biopython from a RPM package should be much the same process as used for other RPMs. If you need general information about how RPMs work, the best place to go is <http://www.rpm.org>.

To install it, you should just need to do:

```
rpm -i your_biopython.rpm
```

To see what you installed try doing `rpm -qpl your_biopython.rpm` which will list all of the installed files.

RPMs do not install the documentation, tests, or example code, so you might want to also grab a source distribution, so you can use these resources (and also look at the source code if you want to).

5.3 Installing with a Windows Installer

Installing things on Windows with the installer should be really easy (hey, that's why they've got graphical installers, right?). You should just need to download the `Biopython-version.exe` installer from biopython web site. Then you just need to double click and voila, a nice little installer will come up and you can stick the libraries where you need to. No need for a C compiler or anything fancy. You will need to have Administrator privileges on the machine to do the installation.

This does not install the documentation, tests, example code or source code, so it is probably also a good idea to download the zip file containing this so you can test your installation and learn how to use it.

5.4 Installing from source on Windows

This section deals with installing the source (i. e. from git or from a source zip file) on a Windows machine. Much of the information from the UNIX install applies here, so it would be good to read section 5.2 before starting. You will need a suitable C compiler. What you choose may depend on your version of Python.

For Python 2.6 we currently use Microsoft's free VC++ 2008 Express Edition from <http://www.microsoft.com/express/download/>, installation of this is pretty simple. Then go to the Biopython source directory and run:

```
c:\python26\python setup.py build
c:\python26\python setup.py test
c:\python26\python setup.py install
```

For older versions of Python, we use mingw32 installed from cygwin (<http://www.cygwin.com>). Once everything is setup (which is a bit complicated), you would again get the source, and from that directory run:

```
c:\python25\python setup.py build --compiler=mingw32
c:\python25\python setup.py test
c:\python25\python setup.py install
```

Previously (back on Python 2.0), Brad has also managed to use Borland's free C++ compiler (available from <http://www.inprise.com/bcppbuilder/freecompiler/>), but this required extra work.

Now that you've got everything installed, carry on ahead to section 6 to make sure everything worked.

6 Making sure everything worked

First, we'll just do a quick test to make sure Biopython is installed correctly. The most important thing is that python can find the biopython installation. Biopython installs into top level Bio and BioSQL directories, so you'll want to make sure these directories are located in a directory specified in your \$PYTHONPATH environmental variable. If you used the default install, this shouldn't be a problem, but if not, you'll need to set the PYTHONPATH with something like `export PYTHONPATH = $PYTHONPATH:/directory/where/you/put/Biopython'` (on UNIX). Now that we think we are ready, fire up your python interpreter and follow along with the following code:

```
$ python
Python 2.5 (r25:51908, Nov 23 2006, 18:40:28)
[GCC 4.1.1 20061011 (Red Hat 4.1.1-30)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from Bio.Seq import Seq
>>> from Bio.Alphabet.IUPAC import unambiguous_dna
>>> new_seq = Seq('GATCAGAAG', unambiguous_dna)
>>> new_seq[0:2]
Seq('GA', IUPACUnambiguousDNA())
>>> new_seq.translate()
Seq('DQK', HasStopCodon(IUPACProtein(), '*'))
>>>
```

If this worked properly, then it looks like Biopython is in a happy place where python can find it, so now you might want to do some more rigorous tests. The `Tests` directory inside the distribution contains a number of tests you can run to make sure all of the different parts of biopython are working. These should all work just by running `python test_WhateverTheTestIs.py`.

If you didn't do this earlier, you should also all of our tests. To do this, you just need to be in the source code installation directory and type:

```
python setup.py test
```

You can also run them by typing `python run_tests.py` in the Tests sub directory. See the main Tutorial for further details (there is a whole chapter on the test framework).

If you've made it this far, you've gotten Biopython installed and running. Congratulations!

7 Third Party Tools

Note that Biopython includes support for interfacing with or parsing the output from a number of third party command line tools. These are not required to install Biopython, but may be of interest. This includes:

- NCBI Standalone BLAST, which can be used with the `Bio.Blast` module.
- EMBOSS tools, which can be invoked using the `Bio.Emboss` module. The `Bio.AlignIO` module can also parse some alignment formats output by the EMBOSS suite.
- ClustalW, which can be invoked using the `Bio.Clustalw` module.
- SIMCOAL2 and FDist tools for population genetics can be used via the `Bio.PopGen` module.
- Bill Pearson's FASTA tools output can be parsed using the `Bio.AlignIO` module.
- Wise2 includes the useful tool `dnal`.

See also the listing on <http://biopython.org/wiki/Download> which should include URLs for these tools, and may also be more up to date.

8 Notes for installing with non-administrator permissions

Although I mentioned above that I wouldn't go much into installing in non-root directories, if you are stuck installing Biopython and its dependencies into your home directory here are a few notes and tricks to keep you going:

- Building some C modules, such as `Bio.Cluster` require that the NumPy include files (normally installed in `your_dir/include/python/Numeric`) be available. If the compiler can't find these directories you'll normally get an error like:

```
Bio/Cluster/clustermodule.c:2: NumPy/arrayobject.h: No such file or directory
```

Followed by a long messy list of syntax errors. To fix this, you'll have to edit the `setup.py` file to let it know where the include directories are located. Look for the line in `setup.py` that looks like:

```
include_dirs=["Bio/Cluster"]
```

and adjust it so that it includes the include directory where the NumPy libraries were installed:

```
include_dirs=["Bio/Cluster", "your_dir/include/python"]
```

Then you should be able to install everything happily.

Yes, it's a bit of a mess installing lots of packages in non-standard locations. The best solution is to talk with your friendly system administrator and get them to assist with the installation of at least the required packages (they are generally quite useful for any python install) before going ahead with Biopython installation.