Installing Bio3D

Xin-Qiu Yao, Lars Skjaerven & Barry J. Grant

September 25, 2016

1 Quick Install

On all platforms (Mac, Linux, and PC) open R (version 3.1.0 or higher) and use the function **install.packages()** at the R command prompt to install the package from your local CRAN site:

```
install.packages("bio3d", dependencies=TRUE)
```

Alternatively, if you prefer to use RStudio, select `Tools → Install Packages..`, Enter **bio3d** in the `Packages` text box (and make sure `Install Dependencies` is selected). Then click `Install`.

1.1 Optional Extras

For full Bio3D functionality you should have *MUSCLE*, *DSSP*, and *NetCDF* (headers and libraries) installed on your system and in the search path for executables. Background details and install instructions for each of these packages on different operating systems is provided further below.

For quick install on a **Mac** we recommend using homebrew:

```
brew install homebrew/science/muscle
brew install homebrew/science/netcdf
brew install homebrew/science/dssp
```

On a **Linux/UNIX** system you should use your appropriate package manager (e.g. *apt-get* for Debian/Ubuntu and *dnf* for Red Hat/Fedora systems, see below). If you experience problems with any of these steps please read on for alternative installation methods.

1.2 Testing your installation

You should now be able to load the Bio3D package into your current R session by typing the usual `library(bio3d)` command at the R Console.

```
library(bio3d)
help(package="bio3d")
vignette(package="bio3d")
```

We now suggest you use the command `demo("pdb")`, `demo("pca")` and `demo("md")` to get a quick feel for some of the tasks that we will be introducing in subsequent vignettes:

```
library(bio3d)
demo("pdb")
demo("pca")
demo("md")
```

2   Detailed Installation Instructions

Before you attempt to install Bio3D you should have a relatively recent version of R installed and working on your system (we recommend at least **R version 3.1.0**). Detailed instructions for obtaining and installing R on various platforms can be found on the R home page http://www.r-project.org.

Bio3D makes use of a number of **additional R packages** including *ncdf*, *bigmemory*, *XML*, *RCurl* and *igraph*. Make sure required headers and libraries are installed on your system prior to installing these R packages. i.e. the R package `ncdf` requires `netcdf` and `libnetcdf-dev`; XML requires `libxml2-dev`; and `RCurl` requires `libcurl3-dev`.

In addition, for full Bio3D functionality you should have *MUSCLE* and *DSSP* installed on your system and in the search path for executables.

Note: If you encounter difficulties in installing any of the suggested packages, note that most functions in Bio3D will work just fine without. e.g. `ncdf` is only required for reading/writing netcdf binary trajectory files; `igraph` for visualization of correlation network analysis; `RCurl` and `XML` for searching and fetching data from various biomolecular databases.

2.1   Obtaining Bio3D

The Bio3D package is available in two forms from CRAN and the Bio3D website

- as platform independent source code (intended primarily for Mac and Unix systems),
- as a compiled binary for Windows.

To install from source requires that your machine has standard compilers and tools such as Perl 5.004 or later. If you run into problems with source installation please refer to section 6.1 of the R Installation and Administration Manual. Below we provide installation instructions for some of the most common OS's.

2.2   MacOS X Installation

R on Mac OS X can be used either on the command-line, like on other UNIX systems, via the R.app GUI (included with your binary R install), or the increasingly popular RStudio IDE.

Regardless of your preferred interface you should be able to find the R command prompt and install the Bio3D package from CRAN using the following command:

```
install.packages("bio3d", dependencies=TRUE)
```

We recommend using the homebrew package manager for installation of MUSCLE, DSSP and NetCDF. For more information on homebrew see http://brew.sh/index.html. Using homebrew these packages can be installed with the following command:

```
brew install homebrew/science/muscle
brew install homebrew/science/netcdf
brew install homebrew/science/dssp
```

Alternatively, you can use the `Packages and Data` menu of the R.app GUI, in particular the sub-item Package Installer: Download the source tar.gz file from above. In the R GUI select `Packages and Data` → `Package Installer` → `Local Source Package`, and press the `Install` button. Select the Bio3D tar file and press `Open`.

In RStudio, select `Tools` → `Install Packages..`, Enter **bio3d** in the `Packages` text box, make sure `Dependencies` is selected and click `Install`.

2.3 Linux Systems (Ubuntu & Fedora)

On a Debian system (e.g. Ubuntu) most required packages and programs can be installed directly through the official package manager system with the `apt-get install` command:

```
apt-get install r-base-core netcdf-bin libnetcdf-dev libxml2-dev \
                libcurl3-dev seaview muscle pymol
```

For Red Hat based systems (e.g. Fedora) the equivalent command is:

```
dnf install R-base R-devel netcdf-devel netcdf libxml2-devel \
            libcurl-devel seaview pymol
```

The Bio3D package can be obtained and installed via CRAN. Start R by issuing the command `R` and then from the R prompt install the Bio3D package:

```
install.packages("bio3d", dependencies=TRUE)
```

Alternatively, Bio3D can be downloaded as source code e.g. from https://bitbucket.org/Grantlab/bio3d/downloads and installed with the command:

```
install.packages("bio3d_2.3-0.tar.gz")
```

Note that **MUSCLE** is not available from the Fedora package manager, but can be installed by:

```
wget http://www.drive5.com/muscle/downloads3.8.31/muscle3.8.31_i86linux64.tar.gz
tar xzvf muscle3.8.31_i86linux64.tar.gz
mv muscle3.8.31_i86linux64 /usr/local/bin/muscle
chmod a+x /usr/local/bin/muscle
```

DSSP is also not available from a number of the package managers, but can be installed by:

```
wget ftp://ftp.cmbi.ru.nl/pub/software/dssp/dssp-2.0.4-linux-amd64 -O /usr/local/bin/dssp
chmod a+x /usr/local/bin/dssp
```

Installing a minimal version of Bio3D

A minimal version of Bio3D with reduced functionality (i.e. for reading/writing binary trajectory files, and fetching data from various databases) will require only the R base installed (i.e. no additional packages needed). Thus, only the R base will be required. In Ubuntu this can be obtained with the following two commands:

```
apt-get install r-base-core
```

```r
install.packages("bio3d", dependencies=FALSE)
```

Installing required R packages individually

The Bio3D dependencies can be installed from within R with the command `install.packages`:

```r
# install only the XML package
install.packages("XML")

# install all required
install.packages(c("XML", "RCurl", "ncdf", "igraph", "bigmemory"),
                 dependencies=TRUE)
```

2.4   Windows Installation

To install the Bio3D package on Windows download the compiled binary .zip file from above.

Start R and from GUI click `Packages → Install Package(s) from local zip file` then simply select your downloaded Bio3D zip file and click `Open` to finish the installation.

2.5   Installing the development version of Bio3D

For the majority of users we recommend the use of the last stable release available from the main Bio3D website. The development version is available from our bitbucket repository and typically contains new functions and bug fixes that have not yet been incorporated into the latest stable release.

There are several ways to download and install the development version of Bio3D. The simplest method is to install directly from our bitbucket repository using the R function `install_bitbucket()` from the `devtools` package.

```
install.packages("devtools")
library(devtools)
install_bitbucket("Grantlab/bio3d", subdir = "ver_devel/bio3d/")
```

Alternative installation methods and additional instructions are posted to the wiki section of our bitbucket repository.

3 Additional utilities

There are a number of additional packages and programs that will either interface directly with Bio3D (MUSCLE, DSSP and STRIDE), or that we consider generally invaluable for working with biomolecular structure and sequence data (e.g. VMD, PyMOL, and SEAVIEW). A brief description of how to obtain these additional packages is given below.

3.1 Required for full Bio3D functionality

MUSCLE:

Muscle is a fast multiple sequence alignment program available from the muscle home page http://www.drive5.com/muscle. The Bio3D functions `seqaln()` and `pdbaln()` currently calls the MUSCLE program, hence MUSCLE must be installed on your system and in the search path for executables if you wish to use this function.

A note for Mac and Unix users:

After downloading MUSCLE, it should be unzipped and renamed to just "muscle" and placed in a directory such as "/usr/local/bin/" (i.e. in your PATH).

DSSP:

DSSP a popular secondary structure analysis program which should be installed on your system as an executable called "dssp" or "mkdssp" and be in the search path for executables. DSSP is available from a number of sources including:

- http://www.cmbi.ru.nl/dssp.html

- http://swift.cmbi.ru.nl/gv/dssp/DSSP_5.html.

3.2   Optional

STRIDE:

STRIDE is another secondary structure analysis program available from the EMBL-Heidelberg. Stride is similar in functionality to the more prevalent DSSP (see above). However, stride is often much easier to setup on different computer systems as you may be able to simply copy or link to the stride executable distributed within every version of VMD (see below).

SEAVIEW:

SEAVIEW is a graphical multiple sequence alignment editor. Download information and documentation are available from PBIL http://pbil.univ-lyon1.fr/software/seaview.html. I use Seaview to manually check and edit protein sequence alignment files prior to detailed analysis. I believe this should be done with every alignment regardless of how accurate the various automatic tools are supposed to be.

Clustal Omega:

Clustal Omega is multiple sequence alignment program that can be used as an alternative to MUSCLE (needed e.g. for functions `seqaln()` and `pdbaln()`). Clustal Omega is available from http://www.clustal.org/omega/.

VMD:

VMD is a molecular visualization program for displaying, animating, and analyzing large biomolecular systems using 3-D graphics. Visit the VMD website for download information and documentation http://www.ks.uiuc.edu/Research/vmd/.

PyMOL:

PyMOL is another visualization program. Bio3D functions `pymol.dccm()` and `pymol.modes()` require PyMOL to be in your search path. PyMOL is available from http://www.pymol.org.

4   A Note on Calling External Programs from R/Bio3D

Ideally, as mentioned previouly, MUSCLE and DSSP should installed on your system and be in the search path for executables. To test this you should be able to call these programs from the command line with just their name from any directory.

For **Mac and Linux** you can find out whats in your **PATH** by launching your favorate Terminal program (on Mac one called Terminal can be found in Applications/Utilities folder) and entering:

```
echo $PATH
```

And the result should be like this. . .

```
/usr/bin:/bin:/usr/sbin:/sbin:/usr/local/bin
```

So this is stating that you can run Unix style applications located in 5 default locations of path in the file system:

```
/usr/bin
/bin
/usr/sbin
/sbin
/usr/local/bin
```

You can add extra locations to your path by creating or editing an existing **.bash_profile** file in your home directory. This file should contain a line like the following:

```
export PATH="/my/new/path:$PATH"
```

You can now put muscle and dssp in any of the locations listed by echo $PATH, including /my/new/path/, which of course you should change to something sensible for you.

For **Windows**, right click "My computer" -> click "Change settings" -> Advanced -> Environment Variables -> From "System variables" list find "Path" and click "Edit" -> Add the path to your programs at the end of the line.

5 Where to next

If you have read this far, congratulations! We are ready to have some fun and move on to other package vignettes that describe various analysis including Molecular Dynamics Trajectory Analysis, Correlation Network Analysis (where we will build and dissect dynamic networks form different correlated motion data), enhanced methods for Normal Mode Analysis (where we will explore the dynamics of large protein families and superfamilies), and advanced Comparative Structure Analysis (where we will mine available experimental data and supplement it with simulation results to map the conformational dynamics and coupled motions of proteins). Happy Bio3Ding!

6 Session Information

The version number of R and packages loaded for generating the vignette were:

```
sessionInfo()
```

```
## R version 3.3.1 (2016-06-21)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora 24 (Twenty Four)
```

```
## 
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
## 
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
## 
## other attached packages:
## [1] rmarkdown_1.0
## 
## loaded via a namespace (and not attached):
##  [1] magrittr_1.5   formatR_1.4    tools_3.3.1    htmltools_0.3.5
##  [5] yaml_2.1.13    Rcpp_0.12.7    stringi_1.1.1  knitr_1.14
##  [9] stringr_1.0.0  digest_0.6.10  evaluate_0.9
```