Quick guide to installing and running some essential programs

Phylogenetic inference has traditionally depended on computational methods and algorithms, since it usually involves the inference of an optimal tree from hundreds or thousands of character states, many times using complex statistical models. In fact, Joe Felsenstein, a famous alumnus of the UChicago E&E department, and one of the founding fathers of modern phylogenetics, claims to have stored the code for his PhD thesis, titled *Statistical inference and the estimation of phylogenies*, in punched cards that ran on computers with slightly higher processing power than a singing birthday card.

We will be using a few programs for the class, so here is a quick guide on how to install and run them. This guide assumes that you are able to perform basic functions on a Unix terminal, such as navigating between folders and decompressing files. If you need any help at all with this, please don't hesitate to contact the TA (you can ignore this if you use Windows).

Clustal X

Clustal X is a widely used alignment program, which implements the Clustal W algorithm in a nice Graphical User Interphase (GUI). To install it, go to http://www.clustal.org/download/current/, and download the appropriate file for your operating system.

For Mac download: `clustalx-2.1-macosx.dmg`
For Windows download: `clustalx-2.1-win.msi`
For Linux download: `clustalx-2.1-linux-i686-libcppstatic.tar.gz`

The Mac and Windows versions should be installed and run just like most other programs are. The linux binary should work on most distributions. However, if you are having problems, you can download the source code (`clustalx-2.1.tar.gz`), and compile it yourself as follows:

After decompressing the tarball, navigate to the resulting folder. Once there type

```
$ qmake
```

This should generate a makefile, which you can use to compile ClustalX by typing

```
$ make
```

This generates a ClustalX binary file that runs the program.

MUSCLE

MUSCLE is another alignment algorithm that is reputedly very accurate and fast. It is available as a standalone binary that can be run from the command line. It is also implemented in commercial software packages, such as Geneious. It can be downloaded from http://

www.drive5.com/muscle/downloads.htm. Again, download the appropriate file for your OS. You should find an executable binary called 'muscle,' which you can store in a convenient directory.

Muscle is run from the command line. To do so, simply navigate to the directory containing the program and execute the binary file. For example, on Mac OSX type:

```
$ ./muscle3.8.31_i86darwin64
```

This command should show you the usage instructions for the program. The problem set for week 1 has further instructions for running MUSCLE.

PAUP*

This is was perhaps the most widely used phylogenetics program 10 years ago, and continues to be very popular. It is especially useful for parsimony and distance-based analyses. It is the only program for this course that is not free, but, thankfully, it is being re-developed into a more commercial package, so a test version is available for free at the moment.

To install go to http://people.sc.fsu.edu/~dswofford/paup_test/, and download the appropriate file for your operating system. The problem sets in this class are written assuming that you will be using the version of PAUP* with a GUI, so we advise that you download this version. If you are using a Mackintosh OS, then unzipping the downloaded file should give you a ready-to-go app, which you can move wherever you like and execute by double-clicking on its icon. If you are running Windows, download the installer, and let the Wizard guide you into installing PAUP*. If you use linux, then you will not be able to use the GUI version. Download one of the pre-compiled binaries available further down the page, and move them to the desired directory. Then decompress the binary, and, from the terminal, change its privileges to make it an executable:

```
$ chmod a+x paup4a147_ubuntu64
```

You may need to install the LAPACK library, as PAUP* depends on it. You can get them here http://www.netlib.org/lapack/, or install them using a package manager such as `apt-get.` To run the program just execute it from the terminal:

```
$ paup4a147_ubuntu64
```


You should get a comand prompt that loos like this:

```
P A U P *
Version 4.0a147 for Unix/Linux (built on Jan  6 2016 at 04:39:29)
Mon Mar 28 23:48:51 2016

Running on Intel(R) 64 architecture
SSE vectorization enabled
Multithreading enabled for likelihood using Pthreads
Compiled using GNU C compiler (gcc) 4.4.7

            ─────────────────────────NOTICE───────────────────────────
            This is an alpha-test version prepared for the exclusive
            use of course and workshop participants, as well as other
            authorized testers.  It will expire on 1 Jul 2016.

            Please report bugs to david.swofford@duke.edu
            ────────────────────────────────────────────────────────────

paup>
```

If you are using linux, please do let us know, so that we can adapt the problem sets to include command-line specific information for PAUP*.

Mesquite

Mesquite is an "all-in-one" type of program that allows users to performe a very wide array of phylogenetic analyses from a single GUI. We will only use it marginally in this class, but it is very useful to perform many standardized phylogenetic inference and phylogenetic comparative analyses. Its website (http://mesquiteproject.wikispaces.com/Installation) has very detailed installation/troubleshooting instructions for all popular operating systems, so it doesn't make sense to repeat them here.

PhyML

This is a very good, fast, and simple program for inferring maximum-likelihood trees. It is run from the command line, and can be used in an interactive way.

To download it go to http://www.atgc-montpellier.fr/phyml/versions.php, and download the zip archive that contains version 3.1 (and not 3.0). Inside the archive you will find the binaries for Windows, Mac OS and Linux systems. For windows just extract the binary and double-click it to run the program. In OSX and Linux, move the file into the desired directory, and run it from the terminal:

```
$ PhyML-3.1_macOS-MountainLion
```

You should get the following command prompt:

```
ooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo

                        ---  PhyML 20120412  ---

        A simple, fast, and accurate algorithm to estimate large phylogenies by
                            maximum likelihood

                        Stephane Guindon & Olivier Gascuel

                        http://www.atgc-montpellier.fr/phyml

                        Copyright CNRS - Universite Montpellier II


ooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo

. Enter the sequence file name >
```

You may need to change the binary's privileges to make in an executable, using the same command as for PAUP*.

MrBayes

This is the most widely used program for Bayesian inference of phylogenies. Go to http://mrbayes.sourceforge.net/download.php, and download the appropriate binaries for your operating system. You will get (again) a zip file with the binary executables inside it, which can be run in the same way as PhyML. If you are using a Mac, use the executable called mb, and not mb.mpi, as the second one is meant for use with multiple processors, which is not needed for any of the analyses that we will be doing in this class. When you run it you should get the following command prompt:

```
MrBayes v3.2.6 x64

                        (Bayesian Analysis of Phylogeny)

                    Distributed under the GNU General Public License


                    Type "help" or "help <command>" for information
                            on the commands that are available.

                        Type "about" for authorship and general
                            information about the program.


MrBayes >
```

The MrBayes binaries will sometimes not run on older versions of OSX. If this is your case, then follow the Linux compilation instructions below.

A pre-compiled binary of MrBayes is not provided for Linux distributions. To install it on a linux machine download the source code from the same website provided above. After unpacking the archive with the code, open a terminal window and navigate to the `src` folder and type:

```
$ autoconf
```

This will generate a configuration script that you can run as:

```
$ ./configure —with—beagle=no
```

After the configuration script is done running type:

```
$ make
```

This should compile Mr.Bayes, generating the `mb` executable, which you can run in the same way as PhyML or PAUP.

Figtree

Figtree is a very good tree visualization program, which we suggest you use to view the trees generated in this class. It can be downloaded here http://tree.bio.ed.ac.uk/software/figtree/. It is installed in the way most programs are (as a Windows executable OSX .dmg file, or Linux-compatible java file), so we will not go into details about how to install it.

Links for other software that may be useful:

Garli - https://code.google.com/archive/p/garli/ an (10 year) old GUI can be found here http://www.bio.utexas.edu/faculty/antisense/garli/garli.html

HyPhy - http://hyphy.org/w/index.php/Download

Tracer - http://tree.bio.ed.ac.uk/software/tracer/

RAxML - http://sco.h-its.org/exelixis/web/software/raxml/index.html, https://github.com/stamatak/standard-RAxML

jModelTest - https://github.com/ddarriba/jmodeltest2

PAML - http://abacus.gene.ucl.ac.uk/software/paml.html