

A quick guide to the ‘aphid’ R package

Shaun Wilkinson

Analysis with profile hidden Markov models in R

aphid is an R package containing functions for building and using profile hidden Markov models for biological sequence analysis. Functions are included for multiple and pairwise sequence alignment, model construction and parameter optimization, calculation of conditional probabilities (using the forward, backward and Viterbi algorithms), tree-based sequence weighting, sequence simulation, and file import/export compatible with the HMMER software package. **aphid** also includes functions for developing and working with standard hidden Markov models.

This package was written based on the algorithms described in the book Biological Sequence Analysis by Richard Durbin, Sean Eddy, Anders Krogh and Graeme Mitchison. This book offers an in depth explanation of hidden Markov models and profile HMMs for users of all levels of familiarity. Many of the examples and datasets in the package are directly derived from the text, which serves as a useful primer for this package.

Installation

aphid is currently only available as a development version, with a stable release available on CRAN shortly. To download the package from GitHub you will first need to ensure you have a C/C++ compiler and the devtools R package installed. Linux users will generally have a compiler such as **gcc** installed by default; however Windows users will need to download Rtools and Mac OSX users will need Xcode (note that Rtools and Xcode are not R packages). To download and install devtools, run

```
install.packages("devtools")
```

then install and load **aphid** by running

```
devtools::install_github("shaunpwilkinson/aphid")  
library("aphid")
```

Help

An overview of the package and its functions can be found by running

```
?aphid
```

To build the vignette users will need to have LaTeX installed. RStudio recommends MiKTeX Complete for Windows and TexLive 2013 Full for Mac OS X and Linux.

If you experience a problem using this package please feel free to raise it as an issue on GitHub.

Acknowledgements

This software was developed at Victoria University of Wellington with funding from a Rutherford Foundation Postdoctoral Research Fellowship award from the Royal Society of New Zealand.