Suppdata: Downloading Supplementary Data from Published Manuscripts

William D Pearse¹ and Scott A Chamberlain²

¹ Department of Biology & Ecology Center, Utah State University, Logan, Utah, USA ² rOpenSci

Summary

suppdata is an R (R Core Team 2018) package to provide easy, reproducible access to supplemental materials within R. Thus suppdata facilitates open, reproducible research workflows: scientists re-analyzing published datasets can work with them as easily as if they were stored on their own computer, and others can track their analysis workflow painlessly.

For example, imagine you were conducting an analysis of the evolution of body mass in mammals. Without suppdata, such an analysis would require manually downloading body mass and phylogenetic data from published manuscripts. This is time-consuming, difficult (if not impossible) to make truly reproducible without re-distributing the data, and hard to follow. With suppdata, such an analysis is straightforward, reproducible, and the sources of the data (S. A. Fritz, Bininda-Emonds, and Purvis 2009, Jones et al. (2009)) are clear because their DOIs are embedded within the code:

```r
# Load phylogenetics packages
library(ape)
library(caper)
library(phytools)

# Load suppdata
library(suppdata)

# Load two published datasets
tree <- read.nexus(suppdata("10.1111/j.1461-0248.2009.01307.x", 1))[[1]]
traits <- read.delim(suppdata("E090-184", "PanTHERIA_1-0_WR05_Aug2008.txt", "esa_archives"))

# Merge datasets
traits <- with(traits, data.frame(body.mass = log10(X5.1_AdultBodyMass_g),
species=gsub(" ", ",",MSW05_Binomial)))
c.data <- comparative.data(tree, traits, species)

# Calculate phylogenetic signal
phylosig(c.data$phy, c.data$data$body.mass)
```

The above example makes use of code from the packages ape (Paradis, Claude, and Strimmer 2004), caper (D. Orme et al. 2013), and phytools (Revell 2012).

As suppdata was, originally, part of fulltext (Chamberlain 2018), it is already being used in a number of research projects. One such project is natdb, a package that builds a database of functional traits from published sources. The software is currently available on...
GitHub (https://github.com/willpearse/suppdata), and we plan to distribute it through ROpenSci and CRAN.

References


