

Suppdata: Downloading Supplementary Data from Published Manuscripts

William D Pearse¹ and Scott A Chamberlain²

DOI: [10.21105/joss.00721](https://doi.org/10.21105/joss.00721)

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

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Submitted: 01 May 2018

Published: 03 May 2018

Licence

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

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:

```
# 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

Chamberlain, Scott. 2018. *Fulltext: Full Text of 'Scholarly' Articles Across Many Data Sources*. <https://CRAN.R-project.org/package=fulltext>.

Fritz, Susanne A, Olaf RP Bininda-Emonds, and Andy Purvis. 2009. "Geographical Variation in Predictors of Mammalian Extinction Risk: Big Is Bad, but Only in the Tropics." *Ecology Letters* 12 (6):538–49.

Jones, Kate E, Jon Bielby, Marcel Cardillo, Susanne A Fritz, Justin O'Dell, C David L Orme, Kamran Safi, et al. 2009. "PanTHERIA: A Species-Level Database of Life History, Ecology, and Geography of Extant and Recently Extinct Mammals: Ecological Archives E090-184." *Ecology* 90 (9):2648–8.

Orme, David, Rob Freckleton, Gavin Thomas, Thomas Petzoldt, Susanne Fritz, Nick Isaac, and William D Pearse. 2013. *Caper: Comparative Analyses of Phylogenetics and Evolution in R*. <https://CRAN.R-project.org/package=caper>.

Paradis, E., J. Claude, and K. Strimmer. 2004. "APE: Analyses of Phylogenetics and Evolution in R Language." *Bioinformatics* 20:289–90.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Revell, Liam J. 2012. "Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things)." *Methods in Ecology and Evolution* 3 (2):217–23.