

SymbiotaR2: An R Package for Accessing Symbiota2 Data

Austin Koontz¹, Benjamin Brandt³, Curtis Dyreson⁴, and William D Pearse^{1,2}

1 Department of Biology & Ecology Center, Utah State University, Logan, Utah, USA **2** Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Rd., Ascot, Berkshire, SL5 7PY UK **3** Northern Arizona University, Arizona, USA **4** Department of Computer Science, Utah State University, Logan, Utah, USA

DOI: [10.21105/joss.02917](https://doi.org/10.21105/joss.02917)

Software

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

Editor: [Kyle Niemeyer](#) ↗

Reviewers:

- [@kyleniemeyer](#)

Submitted: 04 December 2020

Published: 21 December 2020

License

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

Summary

SymbiotaR2 is an R ([R Core Team, 2018](#)) package for easily accessing and handling specimen-based Symbiota2 data within an R environment, allowing anyone to download digitized biological collection data from any established database. Symbiota2 is the updated version of Symbiota ([Gries et al., 2014](#)), a widely used software platform that grants access to data from >750 museums and herbaria worldwide. Through a complete refactoring of the Symbiota code, the structure of Symbiota2 places an emphasis on modularity and accessibility.

Statement of need

The release of SymbiotaR2 is motivated by the ongoing development of Symbiota2. Several R packages already exist for accessing data from standard Symbiota portals; because Symbiota2 is a complete refactoring of the original framework, a new package is required to interface with new Symbiota2 portals.

The goal of this package is to allow users to access digitized biological specimen data hosted via Symbiota2 portals quickly and efficiently. This data could include geographic, taxonomic, or genetic information tied to a recorded specimen, as well as pertinent collection information, images, and publications. By making this data accessible in R, SymbiotaR2 is part of the broader Symbiota effort of creating a collaborative environment to share biodiversity data more widely.

Demonstration

The code below provides an example of a user accessing two different attributes from the same entry within a Symbiota2 database. First, data from a particular entry (here, `id=28`) is pulled into R using the “Occurrences” API endpoint. Then, different attributes associated with this entry (`reproductiveCondition`, `decimalLatitude`, and `decimalLongitude`) are printed. The specification of the `url` argument (at the beginning of the example) allows for users to reference the web address corresponding to any Symbiota2 portal of interest.

```
# Specify web address from which to access Symbiota2 API
> url <- "http://demo.portal.address/api/"
```

```
# Pull occurrence information associated with database entry (id=28)
> test.Occ <- Occurrences(id=28, url=url)

# Find reproductive condition for entry of interest
> print(test.Occ$reproductiveCondition)
[1] "bud"
# Retrieve and concatenate coordinates for entry of interest
> test.coord <- c(test.Occ$decimalLatitude, test.Occ$decimalLongitude)
> print(test.coord)
[1] 50.70 -103.65
```

From here, these R objects be manipulated and used for any downstream analysis. The library provides built in commands for all of the default API endpoints included for any given Symbiota2 portal. These endpoints include access to several different resources: trait data, taxonomic classification, collector information, and much more.

References

- Gries, C., Gilbert, E. E., & Franz, N. M. (2014). Symbiota – a virtual platform for creating voucher-based biodiversity information communities. *Biodiversity Data Journal*, 2(e1114). <https://doi.org/10.3897/BDJ.2.e1114>
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>