The Journal of Open Source Education

evolved: open software for teaching evolutionary biology through virtual inquiries

Matheus Januario [1,2], **Jennifer P. Auler** [3], **Andressa L. Viol** [1,2], **and Daniel L. Rabosky** [1,2]

1 Museum of Zoology, University of Michigan, Ann Arbor, MI, USA **2** Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA **3** Programa de Pós-Graduação em Ecologia, Universidade de São Paulo, Rua do Matão, 321, Travessa 14, São Paulo, SP, Brazil

DOI: 10.21105/jose.00300

Software

- Review
- Repository
- Archive

Submitted: 23 December 2024
Published: 11 February 2026

Summary

`evolved` (**EVOL**utionary **V**irtual **ED**ucation) is an open-source R package (R Core Team, 2024) designed for graduate or advanced undergraduate courses in evolutionary biology. It provides tools for inquiry-based learning, where students engage in scientific practices to actively build knowledge (Pedaste et al., 2015). The package includes vignettes (tutorials) to facilitate classroom investigations (**Tables 1** & **2**). However, educators are encouraged to develop their own content modules depending on class context and/or learning objectives. Most of evolved's core functions are oriented towards either (i) developing student intuition for classic models in evolutionary biology using simulations, or (ii) analyzing (parameter estimation, hypothesis testing, visualization) datasets associated with key questions in evolutionary biology (**Tables 1** & **2**). Importantly, these datasets correspond structurally (e.g., organization and data types) to real datasets used by scientists, and they are compatible with widely-used software for evolutionary analysis (e.g., `ape` (Paradis & Schliep, 2019)).

`evolved`'s functions can be loosely classified into "high level" and "low level" (**Table 1**), thus providing tools for inquiries that assume different levels of student independence (e.g., as categorized by Banchi & Bell (2008) - see **Table 2**). High-level functions execute complex sequences of operations while hiding many details, thus making minimal assumptions of R proficiency and allowing users to focus on results. They are either optimized for performance, or work as a simplified interface ("wrappers") of simulators from standard scientific packages (e.g., functions from packages `ape` (Paradis & Schliep, 2019), `diversitree` (FitzJohn, 2012), or `phytools` (Revell, 2024)). These functions will likely work better in inquiries with more student freedom and a greater focus on analyzing data (**Table 2**). In contrast, "low level" functions focus on making explicit connections between every sequential operation and are thus more easily applicable to structured inquiries (Banchi & Bell, 2008). They are also internally designed to be readable by students with intermediate skills in R, allowing further engagement with simulation parameters and outcomes.

The package is designed to reduce barriers to data- and simulation-driven inquiries and one vignette (`intro_r`) aims to introduce students/educators to R and its pedagogical capabilities. We do not provide answers to the investigation prompts in the vignettes, but educators can request answer keys from the corresponding author (MJ).

Statement of need

Evolutionary biology is a complex discipline that affords many opportunities for students to practice interrogating, analyzing, and discussing data, but advanced classes still need effective educational resources. `evolved` addresses this need by focusing on core evidence used to study evolution: molecular/fossil data, simulations, and sound mathematical principles in evolutionary biology. The package helps teach programming, analytical, and reasoning skills through interactive, hands-on activities, and its open-source design allows direct engagement with algorithms and methods.

`evolved` began in 2022 as a set of standalone functions used in an advanced evolutionary biology course at the University of Michigan, originally designed for use in hands-on classes of 2 hours per topic module. However, valuable class time was lost on data setup and formatting, reducing student engagement with core material. In 2023, the functions were consolidated into an R package to free up class time by simplifying data loading, now made with a simple command: data(`dataset name`).

Software installation and use

`evolved`'s stable version is available on the Comprehensive R Archive Network (CRAN). The available vignettes draw inspiration from the authors' teaching experiences, but they are not essential for the software to function and should be interpreted as suggestions. Instructors should feel free to adapt evolved's functionalities as they see fit.

To access a vignette, type `vignette("vignette_name")` in R, replacing "`vignette_name`" with the vignette name (see **Table 2**). The development version, more detailed information on the package/vignettes, and instructions for installing and using R/RStudio/RMarkdown are provided at https://github.com/mjanuario/evolved. This information as well as the vignettes can also be acessed in https://mjanuario.github.io/evolved/. Questions should be sent to MJ (correspondence author).

Table 1. Selected functions illustrating the range of topics covered, with notes on their level (low = intended to be opened by students, high = wrapper function). A full list and descriptions of all functions, including helper and plotting functions, are available in the package manual (Januario et al., 2024).

Function name	Associated subfield	Function purpose	Function Level	Vignette name
`calcFossilDivTT()`	Paleobiology	Estimate diversity through time from fossil occurrences	Low	`deeptime_rocks`
`countSeqDiffs()`	Comparative molecular evolution and Phylogenetics	Calculates the number of different sites between two protein sequences	Low	`deeptime_clocks`
`NatSelSim()`	Population genetics	Compute allele frequency dynamics under natural selection	Low	`pongen_selection`
`OneGenHWSim()`	Population genetics	Simulates independent assortment of gametes under Hardy-Weinberg equilibrium in a finite population	Low	`pongen_intro`
`WEDriftSim()`	Population genetics	Simulates allele frequency change through time under genetic drift	Low	`pongen_drift`
`simulateBirthDeath()`	Macroevolution	Simulates the number of species following a birth-death model	High	`birthdeath_deeptime`
`SimulateTree()`	Macroevolution	Simulates a phylogenetic tree following a birth-death model	High	`birthdeath_phylogenies`

Table 2. Different datasets and their suggested applications. The inquiry level follows the terminology and definitions of Banchi & Bell (2008). Legend: occs. = Occurrences; Ver. = Version of the package (stable = hosted on CRAN, Devel. = development version hosted on github).

Dataset name	Short description	Suggested inquiry level	Possible topic in inquiry	Name of vignette with example activity	Data source	Ver.
`data_whales`	Cetacean body size and speciation rates	Structured inquiry	Statistical non-independence among closely related species	`birthdeath_phylogenies`	(Jones et al., 2009; Rabosky et al., 2014)	Stable
`whale_nbvlo`	Cetacean molecular phylogeny	Structured inquiry	Handling of phylogenies in analysis software	`birthdeath_phylogenies`	(Steeman et al., 2009)	Stable
`cvtOxidase`	Metazoan (homologous) protein sequence	Open inquiry	Comparative analysis of molecular data	`deeptime_clocks`	(Benson et al., 2012)	Stable
`birds_spp`	Extant species list of birds	Structured inquiry	Richness differences among clades	`deeptime_rocks`	(Jetz et al., 2012)	Stable
`mammals_spp`	Extant species list of mammals	Structured inquiry	Richness differences among clades	`deeptime_rocks`	(Upham et al., 2019)	Devel.
`ammonoidea_fossil`	Ammonoid fossil occs.	Open inquiry	Large-scale biodiversity patterns in the fossil record	`deeptime_rocks`	https://paleobiodb.org	Devel.
`mammals_fossil`	Mammal fossil occs.	Open inquiry	Large-scale biodiversity patterns in the fossil record	`deeptime_rocks`	https://paleobiodb.org	Devel.
`dinos_fossil`	Dinosaur fossil occs.	Open inquiry	Large-scale biodiversity patterns in the fossil record	`deeptime_rocks`	https://paleobiodb.org	Stable
`trilob_fossil`	Trilobite fossil occs.	Open inquiry	Large-scale biodiversity patterns in the fossil record	`deeptime_rocks`	https://paleobiodb.org	Devel.

Dataset name	Short description	Suggested inquiry level	Possible topic in inquiry	Name of vignette with example activity	Data source	Ver.
`timeseries_fossil`	Timeseries of fossil species numbers for many clades	Guided inquiry	Large-scale biodiversity patterns in the fossil record	`birthdeath_deeptime`	(Rabosky & Benson, 2021)	Stable

Author contributions:

MJ and DLR designed most of the code, internal functions, and educational components. JPA joined later, contributing to visualization functions, documentation, implementation improvements, and preparing the package for CRAN standards. MJ, DLR, and ALV developed the vignettes. MJ wrote the first drafts of the paper, help pages, and vignettes, with all authors reviewing the material afterward.

Acknowledgements

We thank Theodore Matel, Tristan Schramer, and Yu Kai Tan for their suggestions. Matheus Januario and his work related to this publication are supported by the National Science Foundation under Grant No. 1754398 and the Rackham Predoctoral Fellowship from the University of Michigan.

References

Banchi, H., & Bell, R. (2008). The many levels of inquiry. *Science & Children*, *46*(2). http://www.jstor.org/stable/43174976.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, *41*(D1), D36–D42. https://doi.org/10.1093/nar/gkw1070

FitzJohn, R. G. (2012). Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, *3*(6), 1084–1092. https://doi.org/10.1111/j.2041-210X.2012.00234.x

Januario, M., Auler, J., Viol, A., & Rabosky, D. (2024). *Package 'evolved'*. https://doi.org/10.32614/CRAN.package.evolved

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448. https://doi.org/10.1038/nature11631

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., & others. (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), 2648–2648. https://doi.org/10.1890/08-1494.1

Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

R Core Team. (2024). R: A language and environment for statistical computing. *Foundation for Statistical Computing, Vienna, Austria*. https://doi.org/10.32614/r.manuals

Rabosky, D. L., & Benson, R. B. (2021). Ecological and biogeographic drivers of biodiversity cannot be resolved using clade age-richness data. *Nature Communications*, *12*(1), 2945. https://doi.org/10.1038/s41467-021-23307-5

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., Huang, H., & Larson, J. G. (2014). BAMM tools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, *5*(7), 701–707. https://doi.org/10.1111/2041-210X.12199

Revell, L. J. (2024). Phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*, *12*, e16505. https://doi.org/10.7717/peerj.16505

Steeman, M. E., Hebsgaard, M. B., Fordyce, R. E., Ho, S. Y., Rabosky, D. L., Nielsen, R., Rahbek, C., Glenner, H., Sørensen, M. V., & Willerslev, E. (2009). Radiation of extant cetaceans driven by restructuring of the oceans. *Systematic Biology*, *58*(6), 573–585. https://doi.org/10.1093/sysbio/syp060

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biology*, *17*(12), e3000494. https://doi.org/10.1371/journal.pbio.3000494