The Journal of Open Source Software

neotoma2: An R package to access data from the Neotoma Paleoecology Database

Socorro Dominguez Vidaña [1] **and Simon J Goring** [2,3]

1 HT Data **2** Department of Geography, University of Wisconsin – Madison **3** Center for Climatic Research, University of Wisconsin – Madison

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

The `neotoma2` R package is a tool to access and manipulate data from the Neotoma Paleoecology Database (https://www.neotomadb.org) within the R environment. Neotoma is a community curated paleoecological data resource (Williams et al., 2018), containing nearly 9 million unique observations of paleoecological proxies with global coverage from 37 constituent databases. The package uses the Neotoma API v2.0 (Simon J. Goring, 2023) as a tool to import records from the Neotoma database, allowing researchers to examine taxonomic, spatial and temporal patterns across space and time over the last 5.4 million years. The R package allows researchers to download, and create new records using `get_*()` (e.g., `get_sites()`) and `set_*()` functions (e.g., `set_sites()`) respectively. This provides researchers with the opportunity to develop dynamic workflows that include data generated locally, and not yet submitted to the Neotoma database.

The `neotoma2` R package has been under dynamic development for over a year, but has been used for teaching and training (Simon J. Goring & Dominguez Vidaña, 2023). This release of the `neotoma2` R package is a clean release of the package, with all of the core features provided and extensive test coverage implemented.

Statement of Need

The `neotoma` R package (S. Goring et al., 2015) leveraged the Neotoma Paleoecology Database v1.0 API and had been one of the primary tools for researchers working with data from Neotoma (Byun et al., 2021; Kujawa et al., 2016; Wang et al., 2023). Changes to the underlying database and a rebuilding of the API required new data objects within the R package to more closely align to the Neotoma data model (Grimm, 2008). The broad user community for Neotoma (Simon James Goring et al., 2018; Williams et al., 2018) requires a toolset that can access and manage data for each of the more than 40 dataset types within Neotoma and so extensive metadata must be accessed for each record. This package conforms to a `tidyverse` (Wickham et al., 2019) approach for data management, with data objects that more closely resemble the underlying data model within Neotoma (https://open.neotomadb.org/db_schema). Most importantly the `neotoma2` package provides a toolset for paleoecologists, ecologists, conservation ecologists, archaeologists, and others, to access and examine the broad range of fossil data contained within the Neotoma Paleoecology Database.

Acknowledgements

We acknowledge contributions from the Neotoma Paleoecology Community, the participants of our workshops for the European Pollen Database, American Quaternary Association and

the International Association of Limnologists/International Paleolimnology Association, and members of the EarthCube community. This work was funded through a grant to SJG from the National Science Foundation (NSF-1948926).

References

Byun, E., Sato, H., Cowling, S. A., & Finkelstein, S. A. (2021). Extensive wetland development in mid-latitude north america during the bølling–allerød. *Nature Geoscience*, *14*(1), 30–35.

Goring, Simon J. (2023). Neotoma paleoecology database API. In *GitHub repository*. GitHub. https://doi.org/10.5281/zenodo.7893836

Goring, Simon J., & Dominguez Vidaña, S. (2023). Neotoma workshops APD release. In *GitHub repository*. GitHub. https://doi.org/10.5281/zenodo.7849972

Goring, Simon James, Graham, R., Loeffler, S., Myrbo, A., Oliver, J. S., Ormond, C., & Williams, J. W. (2018). *The neotoma paleoecology database: A research outreach nexus*. Cambridge University Press. https://doi.org/10.1017/9781108681582

Goring, S., Dawson, A., Simpson, G., Ram, K., Graham, R., Grimm, E., & Williams, J. W. (2015). Neotoma: A programmatic interface to the neotoma paleoecological database. *Open Quaternary*, *1*(1).

Grimm, E. (2008). Neotoma: An ecosystem database for the pliocene, pleistocene, and holocene. *Illinois State Museum Scientific Papers E Series*, *1*.

Kujawa, E. R., Goring, S., Dawson, A., Calcote, R., Grimm, E. C., Hotchkiss, S. C., Jackson, S. T., Lynch, E. A., McLachlan, J., St-Jacques, J.-M., & others. (2016). The effects of anthropogenic land cover change on pollen-vegetation relationships in the american midwest. *Anthropocene*, *15*, 60–71.

Wang, Y., Pineda-Munoz, S., & McGuire, J. L. (2023). Plants maintain climate fidelity in the face of dynamic climate change. *Proceedings of the National Academy of Sciences*, *120*(7), e2201946119.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Williams, J. W., Grimm, E. C., Blois, J. L., Charles, D. F., Davis, E. B., Goring, S. J., Graham, R. W., Smith, A. J., Anderson, M., Arroyo-Cabrales, J., & others. (2018). The neotoma paleoecology database, a multiproxy, international, community-curated data resource. *Quaternary Research*, *89*(1), 156–177.