# Reviewer A:

The Neotoma package is timely: many paleoecologists/paleobiologists use R as a programming language and being able to integrate data downloading and data analysis and hence contribute to repeatability or ?reproducibility? is welcome. My review of the package and this paper describing the package is naturally colored by my experience, so I would like to briefly state some relevant information. I am a regular R user but I have never used the neotoma database for my own research. Am however familiar with the Paleobiology Data and have contributed data to it, used data from it and published with data from it.

I am pleased that neotoma package is now available and hope that my comments/questions will help to improve this ms describing the package.

I mostly have specific comments and since there are no line numbers, I have cited sentences where I have comments and then followed those by my thoughts.

## Abstract

“Paleoecological data are integral to ecological analyses.” – Well, they are also integral to evolutionary analyses! Don't short change paleoecological data!

**We have added the word 'evolutionary' to the abstract.**

“ecological and evolutionary interactions between communities and abiotic environments over long time scales” – This is a bit vague: do you mean interactions among entities in communities, and then also biotic communities with the abiotic environment? Do you assume that information flow is “one way”“, i.e. from physical environment to biota and not the other way around (but totally possible if we are talking about biogeochemical feedbacks).

**We assume that interactions occur both among entities within the community and in a two-way flow between the abiotic elements of the environment and the organisms in the community. I understand the reviewer's comment, but feel like this sentence is acceptable as it stands.**

"study processes that occur infrequently, such as megadroughts, hurricanes, and rapid climate change.” – This is a somewhat weird construction, because ecologists study these as well, since we do have historical data for data associated with these processes.

**We have revised this sentence to read “Second, they allow us to study the long-term outcomes of processes that occur infrequently” to differentiate this from ecological studies of infrequent events.**

“The Neotoma Database provides the cyberinfrastructure”– the database also (very importantly!) provides data, infrastructure is not enough!

**Now reads “…the data and the cyberinfrastructure …”**

“examples of key functions based on the published literature” – this is a bit too abbreviated, sounds like the functions themselves are already published, but it is more that inference has been published with the same data and that you demonstrate how to arrive at the inference in order to show the utility of the neotoma package, right?

**Yes, we have changed the sentence to read: “We show how key functions in the neotoma package can be used, by reproducing analytic examples from the published literature focusing on Pinus migration following deglaciation and shifts in mammal species distributions during the Pleistocene.”**

In general, I strongly recommend that the abstract be rewritten for clarity and correctness.

## Introduction

“e.g. the responses of speciation rates to the five major extinction events in geological history (Peters & Foote 2001; Raup & Sepkoski 1984; Sepkoski 1997)” – please read the references you cited! Peters and Foote is about taxonomic diversity (not speciation rates) and sampling, Raup and Sepkoski is about extinction rates and do not directly look at speciation (or even origination which is what these authors normally look at) rates, Sepkoski 1997 is a review paper. These papers are probably well known among people who work on longer time scales, so it will be a bit embarrassing. I don't know the citations in the second half of the sentence, but please also check, especially if those are well known in the “individualistic response” field.

**Our apologies. I agree with the characterization of these papers, the intention was to provide a range of references representing multiple appraoches to the use of large networks of data, but the phrasing was poor. We have revised this section to read: “allowing us to understand ecological and evolutionary processes and patterns through deep time [@alroy2008phanerozoic; @raup1984periodicity; @sepkoski1997biodiversity; @barnosky2011has]” and added more general references including the more recent Barnosky et al. paper.**

“Constituent databases include, among others” – please just name all the databases or point to the neotoma website where a list of databases can be seen. Are they all 5 My and less? Is that the way it will remain in Neotoma?

**The list is current as of the time of writing, but the Neotoma database is in a period of flux at the moment as more constituent databases are being added. I have added a link to the contituent group webpages to provide more information and indicated in the text that Neotoma is a Pliocene-modern database.**

“The database framework was generalized from the pollen databases (which had identical structures) and the FAUNMAP database to accomodate both macro-and microfossil data as well as other kinds of data such as geochemical, isotopic, and loss-on-ignition. – A citation for the database framework will be good to have. **We've added a reference for that from the original Neotoma manual.**

The URL <http://api.neotomadb.org/v1/apps/geochronologies?datasetid=8> Returned {"success”:1,“data”:[]} for me, can't be right. Please check.

**The API was revised, we have replaced this example with one for datasets. This example is also more useful.**

## Examples

I find the data download extremely slow (I have a new computer!). Is there someway to make the download more efficient? For example, let the default downloads have less information be downloaded (verbal descriptions etc) and allow the user to ?turn on? more information if they want it at a later stage. Especially in the data exploration phase, it is a bit annoying to have to see long descriptions. I gave up running the mammal example because it was taking forever, so you might want to shorten the example or have a shorter version for people who only want a quick demo of mammal data.

**The limitation with the speed is in the API itself, not the R package. The limitation is the time to send the message to the neotoma and then recieve the reply. We are working on modifying the API to allow multiple sites to be downloaded at the same time. We will include the .RData file for the mammal data as a Supplement to enable faster analysis.**

I think it is important to be “correct” in examples. You say (for instance) ?Marion Lake (red, Figure 3) maintains much higher proportions of Alnus throughout its history, but this is an observation, not the “truth”. If preservation for Louise is worse for Alnus, then it will seem like Marion has higher proportions. Better to say “Marion Lake (red, Figure 3) maintains much observed higher proportions of Alnus throughout its history”. As programmers, I assume you are more quantitatively inclined than most users and should assume the responsibility for not being misleading. There might be other places in the ms like that.

**In some cases I would agree, however the specific ecological setting of these two sites is well known, as is the species make up of the parent vegetation. Higher proportions of Alnus pollen at Marion Lake do explicitly represent higher proportions of Alnus at the site. Regardless, I have changed the caption to generalize the example adding “… much higher proportions of Alnus pollen …”**

You used other pacakges (e.g. reshape2). Please state versions. Might be important later on.

**I've added the version numbers into the references. They are included in the bib file we use, but they were not rendered in the final version. I have modified the bib style file to allow these to be rendered.**

Lee Hsiang Liow

# Reviewer B:

1) General comments and summary of recommendation Describe your overall impressions of the submission, how it fits within the scope of the journal, and your recommendation.

It's a very nice implementation. It's well built, rational, and user friendly. It solves some of the problems associated with paleobiological research, especially the never ending quest for multiple-source, well interpolated data presented in a modern fashion. Needless to say, it also have some of the limitations intrinsic to this approach. The wealth of data present is very much oriented towards particular regions, hence taxa and records, of the world. For instance, as of mammals, it would benefit from including datasets, such as the Now database and PANGEA, that also deal with the old world. I know that there are problems and limitations depending on intellectual property rights and the like, and won't blame the authors for this drawback. At the very least, neotoma will help encouraging more integration in our field, as it is currently happening in the field of ecology.

**The Neotoma group is working toward building these relationships. This publication intends to provide a tool to access the Neotoma database, not build the database, as the reviewer recognizes. We certainly hope that by presenting the neotoma database as an option for accessing data we can increase end user buy-in.**

1) It's very well done. I consider this a significant, modern, and wasy to use tool for paleoecological research. **No comment neccessary.**

2) The abs is ok **No comment neccessary.**

3) Figures are clear and easy to understand **No comment neccessary.**

4) the language is ok **No comment neccessary.**

5) N/A **No comment neccessary.**

# Reviewer C:

1) General comments and summary of recommendation

The paper gives an introduction to the neotoma package, an R package for interfacing with the neotoma database, a rich source of data for scientific research. The topic is within the scope of the journal and the paper is in general well written and clear and the package will clearly be of use to many researchers, however I suggest some minor changes.

**No comment neccessary.**

2) Content: I cannot speak to the paleoecological relevance and rigor of the paper as this is not my area of expertise. The R package however is well designed and implemented. Since R is widely used in the research community this seems an appropriate platform on which to have implemented this tool.

**No comment neccessary.**

3) Structure and argument: The introduction and abstract are appropriate though the relationship to ROpenSci could be made more explicit. The structure of the main text, progressing through examples, is logical and gradually introduces the user to more features of the package **We have added the sentence: “The neotoma package was developed in conjunction with ROpenSci and represents one of two ROpenSci packages that interact with paleoecological data along with the paleobioDB package [@varela2014pbdbR].” to make the relationship with ROpenSci more explicit.**

4) Figures/tables: The figures are key to helping readers interpret the package and the objects it creates. Figure 1 is the most important and should be explained more clearly in the caption. **We have added more clarification to figure 1: “The class download contains multiple lists, including dataset, which is a list and defined class. dataset itself contains lists, one of which is also a class, site. neotoma can interact directly with classes through the use of special methods for the various functions described here. These objects and classes are described in more detail below.”**

5) Language: The language is good. **No comment neccessary**

6) Ethical approval: NA **No comment neccessary**

8) Open Reviewer name