Dr Lewis A. Jones
Centro de Investigación Mariña,
Grupo de Ecoloxía Animal,
Universidade de Vigo,
36310, Vigo, Spain
Email: LewisAlan.Jones@uvigo.es

Universida_{de}Vigo

1 February 2023

Dear Editor and Reviewers,

We would first like to extend our gratitude for handling and reviewing our manuscript. Below, we respond to each individual point raised. Our modifications to the original manuscript are attached in a tracked format, and a 'clean' version is also submitted. Below, comments from the referees are reproduced here in *italics*, while our detailed responses to each comment are in **bold**.

Yours sincerely,

Dr Lewis A. Jones

Senior Editor

In addition to addressing the comments from the reviewers and the Associate Editor, I also have some comments of my own that should be addressed:
1- As mentioned by the reviewers, the dependencies from other packages should be clarified.

We have now included an additional column in Table 1 documenting (and citing) all the dependencies required by the palaeoverse package and highlighting which ones are available in base R. We have also included additional text where appropriate to denote which packages are required.

2- Also, your unique and novel functions should be clearly described and how they add to the existing toolbox from other packages.

All functions included in palaeoverse are unique and novel whether it be in functionality or implementation. No function has been directly ported from another package into palaeoverse (as suggested by Reviewer 2) and we have avoided duplicating others' efforts. Functions in other packages might have similar goals and aims of those included in palaeoverse. However, to our knowledge, no palaeobiological R package provides the versatility and flexible workflow that palaeoverse does. Nevertheless, we have expanded upon the uniqueness/novelty of the functions within the available word count.

Reviewer: 1

Congratulations on your manuscript! It is extremely well-written, organized, and presented. I have a minor comment on the presentation of the community survey (see below), but overall could not find many aspects of the paper to criticize. I am very sympathetic to the effort of streamlining the research pipeline of colleagues in the field, and I believe palaeoverse will be a fixture of these pipelines going

forward. You've clearly put a lot of work in this package and manuscript, and I am excited to see where it goes.

Firstly, we would like to thank you for taking the time to review our manuscript. We really appreciate the kind words and hope you will find the package useful in your own research.

Community Survey
Might be in the Supplementary material (which, as far as I can tell, I did not receive), but information on the methods of choice for participants of the survey would be important. Also curious about how exactly the Twitter survey was conducted, since if anyone with a Twitter account could answer it, it might bias the results. Either way it's probably fine, just wanted to point it out.

In the supplementary material, we provide explicit details on how the survey was conducted and the questions asked as part of the survey. To summarise, the survey was distributed on Twitter and via a mailing list (we did not include the mailing list for the sake of privacy). We also tried to distribute the survey via several palaeontological organisations (societies/associations). However, they were understandably not willing to do so. We agree with the reviewer that by conducting the survey on Twitter results might be biased in some way, either by only reaching a particular group of people, or by being completed by people who might be unqualified to participate. However, this is why we also distributed the survey via the mailing list, and we did collect names and email addresses to enable us to verify the identity of our respondents. Nevertheless, we would like to highlight that all survey results are typically biased in some way or another, and without a large sample, it can be difficult to control for this. Ultimately, we felt that by having an open survey on Twitter we could extend the reach of our survey. Figures from Twitter Analytics support this notion with over 20,000 impressions (views) and almost 1,000 engagements (interactions) with the tweet introducing the survey.

It also seems like the survey results are rarely mentioned in the section describing package functionalities, so it wouldn't be absurd to cut that part entirely if it'd allow for more details to be included in other functions. A mention that the authors have surveyed members of the community to inform their functionality choices is enough to justify the "community-driven" aspect of the package, in my opinion.

We would have liked to expand upon the survey more in the main text, but we are limited by the word count. We also appreciate the reviewer's perspective but feel there is value in including the survey results to provide context for the development of the package and perhaps useful data for other/future developers. Moreover, it reinforces the community aspect of the package and helps set a precedent for future work/developers.

Reviewer: 2

The manuscript presents a new R package, palaeoverse, which combines some existing functions from other packages (e.g., trimming taxa from a phylogeny) with new and very useful functions (e.g., latitudinal binning of occurrences) relevant to analyzing fossil occurrence data. The new functions provided by palaeoverse, particularly pertaining to spatiotemporal binning of occurrences, will prove to be very useful to researchers because they do not, to my knowledge, exist in other commonly-used packages. I, as many of the folks they cite in the paper, have also written (terribly inefficient) code to do some of the same things and, in the past, not shared it on Github etc. So, this package will be very useful in the field.

Overall, the manuscript is well-written and informative. I am really excited to start using it in my own work and that of my students.

Firstly, we would like to thank you for taking the time to review our manuscript. We appreciate your enthusiasm for the project and hope you will find the package useful in your own work.

As a general note, there are obviously a few functions included in the palaeoverse package that already exist in other packages (e.g., ape, geiger, claddis). I appreciate that bringing the various functions together is really useful, particularly for folks just starting out with these types of analyses, but I would also like to see explicit acknowledgement of those dependencies in the text. I say this only because those authors should receive due credit in the form of citations here. It will also give users a jumping off point (e.g., I have used this function in palaeoverse but would like to do some more in depth work and now that I know which package to go to). Perhaps it would be good to present a table with package names upon which palaeoverse depends and their relevant citations.

We have not directly included any functions from other packages, and we have taken care to not duplicate other efforts. Some functions in other packages such as 'assign_taxa_to_bins' in claddis are somewhat similar to our 'bin_time' function but do not have the broad functionality implemented in palaeoverse (i.e. different and standardised binning approaches). Several dependencies do exist in palaeoverse for working with certain data types/analyses (e.g. ape, sf), but packages such as geiger and claddis are not used in any form. However, we agree with the reviewer wholeheartedly that authors should receive due credit for their development efforts and have now included citations for all dependencies in Table 1.

I appreciate that users will have to edit PBDB data in only very minimal ways to use these functions (necessary taxonomy checks notwithstanding). As a test of the functions in the text, I downloaded data from PBDB pertaining to occurrences of Equidae in North America. The downloaded Equidae data seem to work well with the functions in the manuscript and I made zero modifications to the download (aside from removing the metadata at the top and removing some punctuation in the genus name column).

Thank you for taking the time to test the functionality. We are glad to hear about your positive experience when using the package.

The help documentation is useful. I read through them for most of the functions and found them informative.
I appreciate that time_bins takes user defined time bins, so one can apply non-international systems such as the North American Land Mammal ages.
I appreciated the warning when using bin_lat that some occurrences fell on the latitudinal boundaries. It was useful in that it informed me of how those occurrences were dealt with.

Thank you. We are glad to hear that.

It was useful that tax_unique identified punctuation in the genus column, which will force folks to look at the more closely at the taxonomy rather than just blindly plugging in downloaded data (like I did, in preparing this review).

Thank you, we agree. We would like to make researchers' lives easier, but feel it is important to encourage taking a closer look at the data.

Specific Comments

Line 65 – e.g., Quental and Marshall, 2013…
e.g. should be used in these cases because the authors obviously cannot provide every existing citation.
Line 195 is another example but there are several throughout

Thank you, we have implemented these suggestions.

A note that I received the following error message with tax_range_time. I am certain the authors are aware.
Warning message:
`getScaleData()` was deprecated in deeptime 1.0.0.
Please use `get_scale_data()` instead.

The deeptime package (maintained by one of the co-authors) was recently updated (v. 1.0.0). This note has already been addressed in the development version of the palaeoverse package which can be installed via:

```
# Load devtools
library(devtools)
# Install development version
install_github("palaeoverse-community/palaeoverse")
# Load palaeoverse
library(palaeoverse)
# Grab internal data
occdf <- tetrapods
# Remove NAs
occdf <- subset(occdf, !is.na(order) & order != "NO_ORDER_SPECIFIED")
# Temporal range
tax_range_time(occdf = occdf, name = "order", plot = TRUE)
```

This version (1.1.1) of the package will be submitted to CRAN in the coming weeks.