

Reviewer 1

This paper describes a Julia package for identifying and standardizing species names in text, with the purpose of ensuring that species are not over-counted, mis-identified, or misunderstood. While I have not been able to check the software, I note that the code and its repository looks nicely engineered and that it is in use in more than one system setup by the authors. While not a new concept, the novelty with the software is its combination of features.

We appreciate the feedback by the reviewer, and have addressed all of their comments in the revision.

My general impression of the paper is that it describes useful software but that the text needs work. It seems the abstract and background has been given far less attention than the rest of the paper. I am confident that more readers will be found if those parts are reworked.

We have updated the abstract, and clarified the background section of the manuscript, notably in the last paragraph. We hope that this will help readers understand the purpose of the package.

Specific comments

Abstract

I don't think "the NCBI taxonomic backbone" is an established term and, as such, should not be used in the abstract. When googling, at least the top three hits are to the authors' own papers and a preprint of the present manuscript.

Corrected as part of the abstract changes

I have been programming my whole life and I struggle with the following sentence: "The basic search functions are coupled with quality-of-life functions including case-insensitive search and custom fuzzy string matching to facilitate the amount of information that can be extracted automatically while allowing efficient manual curation and inspection of results." In particular, "quality-of-life functions" and "custom fuzzy string matching" is not helpful for anyone curious about your work.

Corrected as part of the abstract changes. We now provide longer list of functionalities.

Is relying on the Apache Arrow format a dependency or a feature? If it is an implementation detail, I would say it does not belong in the abstract.

Both - we have kept it in the abstract as it allows high-performance access to the data.

The abstract does not speak to a broader public. What are the applications of the software? The phrase "to facilitate the reconciliation and cleaning of taxonomic names" probably only makes sense to a quite narrow audience.

We have clarified the list of common issues in taxonomic names that the software is intended to correct.

Background

The first paragraph reads as a copy of the abstract.

The first paragraph has been reworked.

Please define "the NCBI taxonomic backbone" before use.

We define taxonomic backbones in the first paragraph.

“Unambiguously identifying species” should be “Unambiguously identifying species names in text”.

Thank you for the suggestion, fixed.

Avoid “presented below”. Write “presented in Table 1” instead. You cannot assume the table in print ends up where you expect it.

Fixed.

I note that Table 1 has a column “Reference”, which is good, but it is empty.

Our apology for the omission, it has been filled-in.

Language

Opening your submitted file in Word, I get spell and grammar warnings on quite trivial mistakes, for example “occurring”, “litterature”, “to the point were”, and more.

These have been fixed

I also note simple mistakes that are hard for Word to notice: “a string of character”

These have been fixed

Code and code access

I am not a Julia user, but from a general (programming language agnostic) standpoint it looks like well-structured code.

Thank you.

The Zenodo page is either not existing or it is not accessible to the public.

There was an issue with the link, it has been fixed in the revision.

The GitHub repository is accessible. It is also setup for and invites for collaboration. There are no instructions for how to install and get started, from what I can find. How much of a Julia user does one need to try this package out? It would be nice with some basic install and get-started instructions. That is extra work I do not want to demand, but it would certainly help with “pickup” of users. For example, I have text would be curious to test your package on, so what would I do?

Installation instructions have been added to the README, and more detailed “getting started” instructions are in the documentation.

It does not strike me as important to have details about error handling in the article. It is good programming and it should be boasted as a feature, but such programming details belongs in the package documentation (or README if you want to make it more public), in my humble opinion.

We feel strongly that keeping this code snippet in the text is important, as it will help users adopt it as basis to build pipelines that use the error catching system.

Reviewer 2

Poisot and colleagues present their software package aimed at making taxonomic classifications/searches more efficient on a local copy of the NCBI database. This appears to be a useful tool that the bioinformatics community will appreciate.

Minor editorial comments:

P3: improve what? Maybe ‘improve classifications such as conservation outcomes’.

Clarified as part of changes to the background section.

P4: should be 'string of characters'.

Fixed.

P4: literature is misspelled.

Fixed

P4: there are '(ii)'s, one of them should be '(iii)'.

Fixed

P5: what is a 'raxonomi'?

Fixed

P6: 'the the' is incorrect.

Fixed

P7: 'possible' should be 'possibly'.

Fixed

P7: 'table has' should be 'table currently has'.

Not fixed, "at the time of writing" is specified immediately after in the sentence

P7: omit 'at the time of writing'.

Not fixed

P7: replace 'search faster' with 'searches faster'.

Fixed

P7: replace 'search are' with 'searches are'

Fixed