

LukProt: A database of eukaryotic predicted proteins designed for investigations of animal origins - 2nd round

The author replied very satisfactorily to all my points and also did extra work to improve the database and manuscript. I consider the article suitable for publication.

I found just two little errors in the text that could be improved.

- line 25 and forward: taxogroup is not a “real” word (although it’s easy to understand what the author means), it may be better to use “taxonomic grouping”
- line 44 “bionformatic tools” -> “bioinformatics tools”

And here are my replies to some of the comments:

“Regarding a repository to reproduce what was done: I have documented the analyses step by step with all the terminal commands recorded. To be completely honest, I have no formal training in computer science and I do not feel comfortable enough with GitHub CLI to put together a repository and release them at this point. If the Reviewer allows, I will release a repository to reproduce all analyses for the next major version of LukProt, as this would be a major effort and cause a long delay”

I think that the author did enough to make it more trustable and reproducible. I hope that for the next release the author will manage to make it fully reproducible but I agree that it is not necessary for the current release.

“I am interested in this, if it is indeed appropriate, I can include these data, but I am not sure if Reviewers are allowed to provide data for the article they are reviewing. It would probably be better if the Reviewer could contact EukProt maintainers to include it. It would be then added automatically to the next versions of LukProt.”

Yes it would be more appropriate if I would contact EukProt’s maintainers for this. Thanks for adding the links to the new proteomes nonetheless.

“I am aware of this and considered including some of the datasets. MateDB seems to be focused on Protostomia, of which there are plenty species in the database already. If the Reviewer can suggest some important taxa that were omitted, they can be added to LukProt”

I agree with the author that it would be better to avoid skewing towards Protostomia and from what I see there is no crucial dataset to add. It is however important to be aware of this new resource in case of future releases.