Review of: "metaPR2: a database of eukaryotic 18S rRNA metabarcodes with an emphasis on protists"

Enrique Lara[1]

1 Spanish National Research Council

Potential competing interests: The author(s) declared that no potential competing interests exist.

The authors have incorporated metabarcodes collected from a high number of studies, and propose a new tool to obtain OTU/ASV sequences more easily to incorporate them directly in posterior studies. This is an excellent initiative, and I think it will be useful for the community. The weakpoints I see is that has basically the same defaults as the PR2 database which serves as a comparison point for sequence assignation. The taxonomy of PR2 is 10 years old (outdated) and includes invalidated clades such as “Hacrobia” or “Excavata” for instance, or groups such as Eukaryota_X. Alveolata_X, Fungi _X etc which have been named since. Then, the taxa ranks follow the “PR2 systematics” which might be more logical but is not valid. These make the output hard to interpret from the taxonomical point of view.

Then, the functions associated to ASV:

It sounds weird to classify “animals” as a function. Moreover, I am not convinced that the reference paper for functional assignation is the best adapted reference, as it is centred on marine planktonic diversity. Basidiomycetes appear as “phagotrophs”, which is technically wrong as they do not practice phagocytosis. I would advise using the references given in Adl et al., 2019 which are more general. Then, some major clades have been assigned to a certain functional type but should be subdivided more; for instance, all Discoba are not parasites. In this group, parasites do not even constitute the majority of the group (i.e. all bodonids, most euglenida and heterolobosea, all jakobids are free-living). This table should be verified.

Then, in general, there is obviously a strong bias due to the composition of samples, which are marine to a large extent. This is why dinoflagellates are so over-represented.

Specific comments:

I refer to sentences in the text, authors can look for them

“suggesting the existence of a lot of unknown taxa”… this might also be due to faster evolving genes; this possibility should be evoked at least

“It is possible through reactive menus to navigate the taxonomy tree down to the cASV level (below the species) that potentially corresponding to cryptic or subspecies” for most eukaryotes 18S does not have the resolution power to differentiate close related species. This marker is valid only for broad taxonomy, and does not work at the subspecies level.