

Title: A multi-dimensional dataset of Ordovician to Silurian graptolite specimens for virtual examination, global correlation and shale gas exploration

Author(s): Hong-He Xu et al.

Authors' response to reviewing, underlined text, by Honghe Xu ([hhxu@nigpas.ac.cn](mailto:hhxu@nigpas.ac.cn))

Reviewer's Background and Summary Review:

I was trained as a taxonomist, but for trilobites and conodonts, not graptolites. I read graptolite systematics extensively as I build and maintain a high-resolution, age-calibrated, global time-line of the species-level macroevolution of the entire graptolite clade. My composite timeline has supported age calibration of the Ordovician and Silurian time scales in addition to several macroevolutionary studies. For the time-line project, I wrote my own data-management software that supports the stratigraphic sequencing program that I co-wrote. Both programs are in Fortran.

Although I rely almost entirely upon co-authors to validate our selection of the graptolite publications that I compile and to vet the quality of the taxonomy, it is likely that my experience is above-average among the potential users that the authors describe for their database. The database should surely be enormously valuable for my time-line projects. I trust this statement clarifies which components of this manuscript that I might be best qualified to review.

Reply: the reviewer, of course is qualified, and kindly give many constructive suggestions and helped greatly improve the manuscript.

This is the fourth version of the image database that I have tested. The authors have made a succession of substantive and successful changes. My previous struggles with the download speed and file format have been essentially alleviated in this latest version. **The database is surely ready for a wide user population.**

Reply: thank the editor and the reviewer.

The bulk of this third review focuses on the accompanying manuscript. I will deal with its content, to some extent, but focus primarily on the quality of the English text. The troublesome errors are not substantively improved by the latest edits. Although the quality of English written by non-native English speakers should not reflect upon the quality of the database, I fear that readers might lose confidence in the authors' attention to detail in the database too. That would be most unfortunate; the database surely deserves better.

In most respects, my prior criticism of the text is still valid. Previously I mentioned the common types of grammatical errors and poor word choices. The one example of subject-verb disagreement that I cited from the introduction was not fixed. The authors' reply mentions grammar-checking software; it seems to be quite inadequate. This time I have added a marked-up a copy of the PDF file with more edits, corrections, suggestions and questions to guide the authors' revisions.

Reply: in this version we accept reviewer's suggestion and annotated PDF file. The writing is obviously improved. We also have the manuscript checked by colleagues.

In summary, the database files are ready for use by experts, amateurs and students, just as the authors explicitly intend. The manuscript is not yet ready for publication. Obvious inattention to detail in the text file could undermine users' confidence in the authors' quality control in the database itself.

Reply: thank the reviewer, in the revision, we note this problem. Opinion or comment only represent a few authors' view point. Actually, we give only comments of 1-2 graptolite experts, of which might not be accepted by others. So, in the revision we claimed this point. We respect the tagged name (label) of the individual specimen. Our dataset only provides the platform or access to researchers or ones who are interested in our fossils. However, the revised comments of some specimens are recorded in our uploaded excel file. We can also make related explanation in the manuscript.

#### Importance of the Image Database

Paleontologic research has been advanced considerably by on-line availability of larger collections of publications than are housed as paper copies in most institutional libraries. This image database is a major step toward an exciting parallel development for paleontological museum collections. The on-line images will not entirely replace expensive travel to examine unique physical specimens or the risky loaning of unique specimens via mail services. They will surely, however, allow more effective preliminary evaluations than the limited photographic plates in printed journals.

Reply: thanks. We put this as one of the key contributions of this study. We will go on to enlarge and improve our database.

#### Revisions of the Image Database

The large image collection is not set up for on-line browsing. Instead, users download the files to their own computers. Many users will surely welcome this. Three substantive changes to the database structure have made the downloading and searching of the database faster and easier.

1. The image collection has been divided into 49 zipped folders. Individual folders can be downloaded in a fraction of the time that was needed for the single folder; such long download times tended to crash personal computers, even with robust institutional connections to the internet.
2. An intermediate problem with the naming of compressed folders has been corrected. Folder contents can now be extracted by basic components of common operating systems.

Reply: thank the editor's work

3. The folder names and file names now include genus names to indicate the folder and file contents. Users no longer need to browse through numerous folder contents to locate species of interest. This elegant convention obviates my suggestion that the xlxs spreadsheet could add a column that mapped specimens to the appropriate folder. I thank the editor, Kirsten Elger, for guiding me and the authors through these

successive improvements.

Reply: the spreadsheet file and image folders were both improved for the sake of user's convenience.

Most of the image files are paired. An image of the entire rock sample now includes a ruler for scale. For the close-up images of the graptolite specimens on the rock surface, the dimensions of morphologic features can readily be estimated by comparison. Earlier versions did not provide scales and this was not possible.

Reply: this was done and greatly improve the uploaded dataset. I believe this is one of the key contributions or innovations of our work.

#### The Accompanying Manuscript Document

I have annotated the PDF file with edits, questions and suggestions. Many simply involve subject-verb agreement or the use of definite and indefinite articles (the, an, a). Here are some more significant issues.

Reply: the annotated file is very useful to improve the text. Most suggestions were followed in the revised manuscript.

1. Some species names have been emended from the original publication. Either an erroneous published identification has been corrected, or the published name has been synonymized with a previously named species, or the specimen has been re-assigned to a newer species that was established after publication of the paper to which the museum specimen is connected. The spreadsheet provides both names, emended and "tagged," but does not distinguish between the three possible reasons to emend the name. Some of this uncertainty could be clarified if both the names included author and year. Those two terms are, after all, requirements of a valid taxon name. Readers may be unsure what the authors mean by "tagged" in this context.

Reply: this is quite important. We noted this point and make explanations in the text. Emendation or academic correction is relative and quite personal. We do not want to emphasize this in the dataset description study. Our focus should be the dataset itself.

2. The manuscript states that paleontologists (plural) have provided these taxonomic corrections or updates. Only one author (XM) is credited with this kind of contribution; nobody else is credited for this in the acknowledgements. It would seem that any user wishing to reference an emended name in a synonymy list would, therefore, need to credit the name to the current authors; i.e. Xu et al. 2022. No detailed justifications for the amendments are provided in the xlxs file.

Reply: same answer to the previous one.

3. Users of the database might wish to construct or correct range charts for the localities. They would need to know whether an emended name applies only to one specimen from that locality, to all specimens at one horizon, or to all uses of the name in the cited reference. This kind of uncertainty can arise from traditional publications too. It is difficult to update range charts, but frequently desirable.

Reply: the dataset with detailed information of every specimen is the first and quite crucial step. We then develop the visualizer model (software) to fulfil the functions of data query and retrieval.

4. Some users might not have easy access to the cited Chinese literature. Is it possible for the database to include at least the systematic taxonomy sections from those papers? I imagine there would be copyright issues.

Reply: this will also be fulfilled in the visualizer model.

5. The manuscript has the problem of addressing experts, amateurs and students – three stated potential users of the database with potentially very different levels of sophistication. Some parts, like the first paragraph, are not needed for experts and tend to be oversimplified for students. Other parts omit advanced considerations, like the formal treatment of synonymy and revision.

Reply: now we are more clear the purpose and significance of our study. It is a data description study. It provide the data access and also multiple usage of the dataset.

This part was revised according to reviewer's suggestions.