

Note: original reviewer comments in black; responses of authors in blue

Reviewer #1 Jan Hennissen

### Overall Comments

In this paper, Shu and co-authors present the compilation of a database, the Global Acritarch Database (GAD), comprising acritarch taxa from the literature and their associated metadata. They build on the database from Palynodata.ru adding data from publications post 2007. In total, the GDA contains 115,947 entries with, where possible, data on palaeolatitude and palaeolongitude. The authors perform basic exploratory data analysis that reveal that the majority of studies originate from the current Northern Hemisphere with a focus on the Palaeozoic Eon.

The manuscript is well-structured and is easy to follow. To improve readability even more, I encourage the authors to avoid the past tense, which is unnecessary in most places where it was used (e.g., P.4 L. 97 and 122).

I find the presentation of the exploratory data analysis in Figure 5–10 effective and I believe it will help researchers identifying knowledge gaps spatially (e.g., Southern Hemisphere) and temporally (e.g., Late Cambrian and Permian).

Author Response: Thank you for your thorough review and valuable feedback on our manuscript, your positive remarks on the exploratory data analysis in Figures 5–10 were greatly encouraging. We have carefully addressed all of your comments and uploaded the revised version to the submission system. In particular, we have revised the tense issues you pointed out (e.g., P.3 L. 85, 88; P.5 L. 125, 151) and adjusted the entire text to a more concise present tense, as suggested.

The GAD is presented as a csv and I imported it into R to interrogate the dataset. Below are a number of comments that I found during this exercise. These should be addressed before the manuscript is formally published.

In addition, I added an annotated pdf of the paper with a non-exhaustive list of mostly editorial comments.

### Specific comments

P. 1, L. 14: “Acritarchs, microfossils with an algal affinity” Rephrase. Acritarchs are an artificial polyphyletic group of microfossils of unknown biological affinity (Downie et al., 1963; Evitt, 1963; Kroeck et al., 2022; Strother, 1996). It is true that many acritarchs **likely** have an algal affinity and only a small number are non-algal which the authors state themselves in lines 27–31. This should be reflected in the opening line of the abstract rather than the definitive statement that is there now.

Author Response: Revised. Regarding the suggestion that the definition of acritarchs should be phrased with greater precision. We have refined the original version by: (1) explicitly defining acritarchs as microfossils of uncertain biological affinities, and (2) specifying that most are considered algal (P. 1, L. 14).

P. 1, L. 24: the link to the Zenodo page for the GAD works but note the spelling mistake in the title of the page “Gobal” instead of “Global”. This spelling mistake is also perpetuated in the filename of the csv file upon downloading the GAD.

Author Response: Revised. We have corrected the “Gobal” spelling error to “Global” on the Zenodo page and in the csv file, and have reuploaded on Zenodo: <https://doi.org/10.5281/zenodo.15208303> (Shu et al., 2025). Many thanks for spotting this, our apologies for the oversight (P. 1, L. 26).

In the GAD CSV:

- check spelling of column title AL which should be “Reference”. Unfortunately, many of the references which contain special symbols have had those replaced with “?”. This is not a critical issue, but it could impede linking to the original publications.

Author Response: Revised. We have corrected the misspelled header “Refence” in column AL, now split into AN “Reference (Author and Year)” and AO “Reference (Title and Journal)”, and restoring the original special characters that were previously replaced by “?” marks to ensure proper linkage to source publications. However, some “?” marks remain unchanged, such as “Ref ID 7499”: “Trace fossils and acritarchs from the Colorado Formation (Upper Ordovician?), Ossa-Morena Zone, SW Iberian Peninsula.”, which has been preserved according to its original publication title.

- To increase ease of use, it would be nice to have doi’s for the references included as a separate column.

Author Response: Added. We have added DOI (AP) in the GAD csv file.

- Genus and Species name are in a separate column which makes it easy to interrogate the database. However, the original name is in a single column (genus and species name together). It would be easier to use if those were separated into two columns as well.

Author Response: Revised. Following the recommendations, the original name have been parsed into separate fields for “Original genus name (H)” and “Original specific epithet (I)”.

- Each entry has an associated Doc number. Is there a separate list with all documents and their numbers?

Author Response: Added. We have added a comprehensive list (“Reference List”) with all

documents and their numbers on Zenodo: <https://doi.org/10.5281/zenodo.15208303> (Shu et al., 2025).

- The column “Species name” should be replaced with “Species epithet”.

Author Response: Replaced. “Species name” has been changed to “Species epithet” (F).

P. 5, L. 129: “Time Field”. I think “Age” may be more appropriate. Note that it is not a single field that is assigned to the definition of the age of an entry (I count 12 separate columns that are used to define the age).

Author Response: Revised. We have replaced “Time” with “Age” for accuracy. The article has been annotated that 12 separate columns collectively contribute to “Age Field” (P. 6, L. 159).

P. 5, L. 142: Same as above: Location Field. There is more than 1 field that describes the location.

Author Response: Revised. The article has been annotated that 9 separate columns collectively contribute to “Location Field” (P. 6, L. 172).

P. 5, L. 146: Check the correct reference for “Google Satellite Electronic Maps”.

Author Response: Revised. “Google Satellite Electronic Maps” has been changed to “Google Maps (<http://www.gditu.net>)” (P. 6, L. 176).

P. 5, L. 153: Should there not be a paragraph break after addition. The next paragraph should then start with (6) The reference field.

Author Response: Revised. We have implemented the suggested formatting change, now presenting “The reference field” as an independent numbered paragraph (6) (P. 6, L. 186).

P. 5., L. 156: as mentioned above, special characters seem to have caused issues with frequent “?” etc.

Author Response: Revised. We have corrected all potentially erroneous “?” markers by cross-referencing the original literature, while retaining instances such as *Frankea longiuscula?* a doubtful species name in the database (P. 6, L. 189).

P.8, Table 3: “Species name” should be replaced with “Species epithet”. For example in the case of *Pterospermopsis australiensis*. The genus name *Pterospermopsis* the species name is *Pterospermopsis australiensis* while the species epithet is *australiensis*. Note that this should also be changed in the GAD proper as the column now entitled “Species name” contains in fact species epithets.

Author Response: Replaced. “Species name” has been changed to “Species epithet”. Other corresponding modifications are also displayed in the table (P.10, Table 3).

Figure 2: “Too old” is not a reason for not being able to find literature.

Author Response: Revised. The reason has been removed, corresponding to the term structure of “Can find literature”, changed to “Cannot find literature” (Figure 2).

Figure 3: the commas in the “Time” field come out looking like accents.

Author Response: Replaced. Thank you for your keen observation. We have now replaced the Chinese enumeration commas (、) with standard commas (,) in the “Age” field (Figure 3).

P. 10, L. 178: on importing the database and filtering for references I arrive at 7799 unique entries (not 7816) and only 1170 unique sample locations (not 2993).

Author Response: Revised. After rechecking the database, we hereby confirm the following filtered results: “Reference” column (AN+AO): 7791 unique entries, “Location (Detail)” column (AD): 1146 unique sample locations (P.13, L. 212).

Figure 5: “Number of literature” should be replaced with “Number of published studies”

Author Response: Replaced. We have now replaced the “Number of literature” with “Number of published studies” (Figure 5).

P. 17, L. 263: “The large volume and consistent structure of data in GAD allow for a comprehensive analysis of acritarch evolution over geological timescales”. I disagree with this statement. At best, it shows the patterns of research interest for acritarchs in the global research community throughout the geological timescale. As pointed out by the authors in Figures 7 and 8, the number of studies is highly biased, with the majority of studies conducted in areas currently in the Northern Hemisphere. I am not sure if this is in response to acritarch evolution. This is especially hard to prove when dealing with a polyphyletic group like acritarchs where the inclusion of species within this group may be tenuous, especially in the Palaeozoic. I think Figure 10 provides an interesting analysis of the available data in the GAD, but I think there is more to its observed distribution than explained by the authors in Paragraph 3.7.

Author Response: Revised. The issue you raised is of paramount importance. We fully acknowledge that the original phrasing “GAD allow for a comprehensive analysis of acritarch evolution” was insufficiently rigorous. It has now been revised to “GAD provide opportunities to investigate research trends in acritarchs” (P.20, L. 303).

Furthermore, regarding the phenomena shown in Figure 10, we have supplemented the discussion with multi-faceted analysis, including: tectonic movements, sampling biases and inherent taxonomic uncertainties (P.20, L. 310-316).

P. 18, L. 285: GAD and PBDB: write out abbreviations in full again in the conclusions.

Author Response: Revised. The full names have now been appended: Global Acritarch

Database (GAD)、Paleobiology Database (PBDB) (P. 21, L. 331).

Reviewer #2 Anonymous

This represents a great effort and I hope that the publication of the database will help to promote and stimulate future research on the acritarchs. The text does not present the topic as I, personally, would have approached it – the introduction does not really address the history of the creation and use of acritarch databases or summaries, which really began with Helen Tappan, back in the 1970s. And the mention of several important compilations is missing altogether, for example the Fensome et al. 1990 index, or, for that matter, the Acritax on-line database that is based on the John Williams Library (housed at the Natural History Museum in London). Nevertheless, having an update to Palynodata is important and I am glad that the authors have been willing and able to do so. But these comments are more in the way of an opinion – the introductory text as it stands, is not necessarily incorrect.

**Author Response: Added.** We sincerely appreciate your thorough review of our manuscript and your highly valuable suggestions! The issues you raised are indeed crucial, the introduction should include a discussion on the development and application history of acritarch databases and compilations.

Following your advice, we have revised and expanded the introduction to incorporate these discussions, (1) Fensome et al. (1990) compiled a comprehensive taxonomic index of acritarchs, improving the standardization of classification criteria in this field (P. 2, L. 42), (2) the pioneering work of Helen Tappan in acritarch databases, while exhibited relatively coarse temporal resolution, (3) Between 1971 and 2010, John Williams compiled the Acritax on-line database, documented 1577 genera, (4) In the 1990s, with support from the Geological Survey of Canada (GSC), the Palynodata was developed and stopped updating in 2006 (P. 2, L. 64-74).

Additionally, we will fully consider the acritarch data in the Acritax online database in our next database update and integrate it organically with our research. We believe sufficient time and a well-developed classification system will further enrich our data resources and provide more comprehensive support for related studies.

There are perhaps two reasons to use the GAD: one is to aid in systematic work, the other is to explore any number of research projects in geobiology. But in either case, one cannot just use the raw data as it is extracted from the literature without referring back to the original papers. This is because the taxonomic assignments vary considerably and many generic-level taxa, especially in the Precambrian, are so broadly interpreted as to be biologically meaningless – *Leiosphaeridia* is a good example of this problem. The value and quality of research based on the GAD will be highly dependent upon the specifics and nature of filtering the data, prior to

constructing any result. And it will be dependent upon the individual researchers to define very succinctly how data has been filtered. As to systematics, the database should be helpful in constructing synonymy lists, although it would be helpful if the date (year) and authors of the references had been presented as separate fields in the GAD.

Author Response: Added. We have expanded the discussion on the potential characteristics of acritarchs in the paper, incorporating additional evidence and references. This includes studies on animal embryos/diapause cysts, giant sulfur bacteria, and a holozoan affinity, which collectively provide clearer insights into the diverse biological features of acritarchs (P. 2, L. 35-41).

We place paramount importance on data quality and have established a rigorous screening protocol for the GAD, (1) Strict cross-verification with original publications. Each taxonomic entry in the database is traced back to its original publication to validate taxonomic reliability. (2) Synonyms and spelling errors are systematically checked to ensure nomenclatural accuracy. (3) For taxonomically contentious groups (e.g., broadly defined taxa such as *Leiosphaeridia*), we have filtered out, as far as possible based on the original literature, those questionable or illegitimate taxa, invalidly named taxa, taxa retained in open nomenclature, etc. (4) All filtered data will be accompanied by publicly accessible source references and revision logs for peer verification (P. 5, L. 147-157).

And we have accordingly addressed the data standardization issues you raised by implementing distinct dedicated fields for the date (year) and authors of the references. In the GAD csv file: 1. Column AN: Reference (Author and Year), 2. Column AO: Reference (Title and Journal).

There are some edits need, but I have not addressed these. Well, the abstract should be re-written to better reflect what acritarchs are (they should not be defined as algae, even though many surely are, but such an assertion is misleading for the Precambrian in particular).

Author Response: Revised. We have enhanced the original version by providing a precise definition of acritarchs as “organic-walled microfossils of uncertain biological affinities”, with the clarification that the majority are currently interpreted as algal (P. 1, L. 14).

Thank you again for your constructive feedback. We have carefully addressed your comments to improve the manuscript and database.