Reply on RC1
Zhaohui Pan

Community comment on "A Novel specimen-based Mid-Paleozoic dataset of antiarch placoderms (the most basal jawed vertebrates)" by Zhaohui Pan et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-394-CC3, 2022

Thanks for the kind comment and revision suggestions of the manuscript. We really benefit. We revised the manuscript carefully.

The title, 'Dataset of antiarch placoderms (the most basal jawed vertebrates) throughout Middle Paleozoic', why Middle palaeozoic? In the text, we know the geological range of placodems is from the late Silurian to the Late Devonian. why not limit the time range before the fossil group? Authors need to give explanation. In the description of the dataset, the geological background is very important but not well given in the text.

Response: "Middle Paleozoic" is a conventional term for Silurian and Devonian, frequently used in the literature. The international geoscience programme also uses this term, like IGCP 491 “Middle Palaeozoic Vertebrate Biogeography, Palaeogeography and Climate”, and IGCP 596 “Mid-Paleozoic climate change and biodiversity patterns”. For a short title, we keep it unchanged. In the Introduction section, we add “throughout Middle Paleozoic” in the sentence “Antiarcha was a diverse and successful group within Placodermi from the late Silurian to the end of Devonian…”

We have added extra text to describe the geological background in the Introduction and 3.1 Data Overview section.

This dataset was extracted from the DeepBone database, or the present dataset is a subset of it. The purpose of this study, as a data description study, is to show the dataset and its potential using, not giving much attention to the analytical result. The explanation of the data elements, its geological background, and data preparation, make up the key contents of the study, which, in this respect, the study should give more information. The using and analyzing data, in this study, are actually only examples.

Response: The present dataset was derived from a subset of the DeepBone database. We are trying to compile all the published data of Vertebrate Paleontology one group by one group based on the DeepBone database. Dataset of Antiarcha is the first finished group. However, this dataset is different from that in DeepBone in the data fields. We deleted the uninformative fields of the original dataset and added the fields of paleo-coordinates for
the practical aim.

We have strengthened the description by explaining the data elements, geological background, and data preparation.

The valuable feature of the present dataset is its unique and abundant records of Silurian to Devonian Antiarcha. A simple comparison is given in this study (section 3.1). But I think that authors can go further.

**Response:** Adopted. We compared the two datasets in more detail in section 3.1.

Fossil occurrence-based dataset is better for analyzing fossil organism diversity and distribution. A lot of paleobiological study just prefer fossil occurrence data. GBDB is geological section based (Xu et al., 2020, ESSD. the publication year is 2020, but in this study it was written as 2021) and better in stratum correlation, but its data can be exported to fossil occurrences.

**Response:** Corrected.

For the present data analyzing examples, I see that authors are still using the fossil occurrence data (figure 5 and related text). What is the unique merit of the fossil specimen-based dataset? Why the present dataset or DeepBone chose the specimen-based data structure?

**Response:** The fossil occurrence data is suitable for data visualization. DeepBone chose the specimen-based data structure for three reasons. Firstly, because classic paleontology is based on the specimen, all the information about the specimen could be digitized with as much as possible. Secondly, a specimen with literature is the hard evidence in Paleontology. Thirdly, specimen-based data structure could cover the fossil occurrence-based data structure and do more than occurrence-based data like fossil abundance analysis, geometric morphometrics, fossil calibrations, and so on.

Additionally, the elements in the table 1 are not all corresponding to those in the first line in the data spreadsheet.

**Response:** Revised.

Line 16 and other, “The dataset consists of 64 genera and 6025 records, covering all antiarch lineages”. Why authors do not mention the number of species? Such thing occurs in all the text. Here “6025 records’, I guess, means 6025 pieces of fossil specimens. I think such causes confusing because that it needs further definition, especially to define the basic unit (element) of the dataset.

**Response:** This is a historical problem on the *Bothriolepis* and *Asterolepis*, the largest two groups of Antiarcha. Identifying a specimen depends on the ability to recognize species in a way that is coherent within a particular genus and through the broader groups. This is
very difficult to fossil material by two especially intractable problems: practically, by the fragmentary nature of the fossil, and philosophically by questions with the criteria by which on demarcates fossil species (Nelson, 1999; Thomson and Thomas, 2001). For example, Thomson and Thomas (2001) reviewed the previous study on *Bothriolepis* proposed that *B. nitida*, *B. minor*, *B. virginiensis*, *B. Darbiensis*, and *B. colocadensis* could not be consistently distinguished. Weems (2004) questioned the validity of *B. virginiensis*. No consensus on the species level of *Bothriolepis* and *Asterolepis*. Thus, the former researchers only used the evidence of Antiaracha on genus level to discuss the biostratigraphic significance (Lelievre and Goujet, 1986; Pan, 1981; Young et al., 2010; Young and Lu, 2020).

Here ‘6025 records’ are 5867 fossil specimens and 158 virtual specimens. Virtual specimens are introduced to store the taxon information when no precise specimen was referenced in the literature.

In the sections 2.4 and 2.5, figures 3-5, what are the Antiaracha records? Are they individual species, localities? Or specimens? It is only obvious that basic element of the diversity analyses is the fossil taxa (figure 6).

**Response:** Revised. In sections 2.4 and figures 3-5, they are individual specimens. In section 2.5, biodiversity is calculated on the number of genera and species. We revised the description to ensure clarity.

Line 19 and other, “data of Antiaracha”, “structured data of...: what does this mean? What data? here also need definition.

**Response:** Revised. Data of antiarch is the information on antiarchs, which is usually the unstructured data in the literature text. Structured data is extracted from the text into a predefined format. We rewrote the description in the Abstract.

Line 21, “including testing hypotheses”, actually, using data is not ‘testing’ something but showing something.

**Response:** We were trying to introduce the potential implementation of structured data of Antiaracha. For example, we could calculate the similarity among different areas to test the ‘stepping-stone’ hypotheses on the dispersal of antiarchs proposed by Li et al. (1993). Because this is a data description paper, we delete this part to meet the scope of ESSD.

Section 1, authors should emphasize the significant of the present dataset, not only the fossil group. Such two points are closely related but different.

**Response:** Revised.

Lines 49-50, “Explaining the spatial and temporal distribution of early vertebrates is the prerequisite to understand their biogeographic exchange”. The normal sequence is, collecting data – analyzing and showing the distribution – recognizing pattern, the last
step is probably the explaining you called here.

**Response:** Revised.

Line 116 and others, the TrackPoint V 7.0, I only searched this software in the method part of Xu et al., 2020. Palaeogeography, Palaeoclimatology, Palaeoecology. 560. 110029.

**Response:** The TrackPoint V 7.0 was first introduced by Ke et al. (2016) in the caption of Fig.4 on page 11. It was developed by Christopher R. Scotese based on modern geographical coordinates of brachiopod localities.

Figure 5 needs to be improved, currently it is not clear and hard to get information.

**Response:** Revised. We adjusted the contrast and saturation of the images. And we added more description in the figure caption.

Line 221, “Based on our dataset, the oldest record of”, are you sure that using dataset can conclude the time range result of a fossil? The section 4.1 seems not quite related to the present study. please reconsider it.

**Response:** The time range of a kind of fossil is based on the stratigraphic horizons of their specimens. Paleontologists, stratigraphers, and systematic biologists are always interested in the earliest fossil record because it can be applied to stratigraphic correlation and molecular clock dating.

Line 256, Eem event, needs explanation.

**Response:** Revised.

Section 6, specific and definite conclusion is needed.

**Response:** Revised.

**Reference:**


Lekuevre, H. and Goujet, D., Biostratigraphic significance of some uppermost Devonian palcoderm, Annales de la Société géologique de Belgique, edited by Ministry of Economic


Please also note the supplement to this comment: https://essd.copernicus.org/preprints/essd-2021-394/essd-2021-394-CC3-supplement.pdf