A Novel specimen-based Mid-Paleozoic dataset of antiarch placoderms (the most basal jawed vertebrates)

Zhaohui Pan1*, Zhibin Niu²*, Zumin Xian³, Min Zhu¹4.5

¹ Key CAS Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences (CAS), Beijing, China 100044
² College of Intelligence and Computing, Tianjin University, Tianjin, China 300350,
³ Institute of Palaeontology, Yunnan Key Laboratory of Earth System Science, Yunnan University, Kunming, China 650500
⁴ CAS Center for Excellence in Life and Paleoenvironment, Beijing, China 100044
⁵ University of Chinese Academy of Sciences, Beijing, China 100049

*These authors contributed equally to this work.

Correspondence to: Min Zhu (zhumin@ivpp.ac.cn)

Abstract. Antiarcha data are essential to quantitative studies of basal jawed vertebrates. The absence of structured data on key groups of early vertebrates, such as Antiarcha, has lagged in understanding their diversity and distribution patterns. Previous works of early vertebrates usually focused on anatomy and phylogeny, given their significant impacts on the evolution of key characters but lacked comprehensive structured data. Here, we contribute an unprecedented open access Antiarcha dataset covering 60 genera of 6025 specimens from the Ludfordian to the Famennian globally. We have organized an expert team to collect and curate 142 publications spanning from 1939 to 2021. Additionally, we have two-stage quality controls in the process, domain experts examined the literature, and senior experts reviewed the results. In this paper, we give details of the data storage structure, and visualize these antiarch fossil sites on the paleogeographic map. The novel Antiarcha dataset has tremendous research potential, including testing previous qualitative hypotheses in biodiversity changes, spatiotemporal distribution, evolution, and community composition. It is now an essential part of the DeepBone database and will update with the latest publication, also available on https://zenodo.org/record/6536446 (Pan and Zhu, 2021).

1 Introduction

Placodermi is an extinct group of jawed vertebrates that first occurred in the Silurian, then dominated the Devonian and constituted a prevalent biotic component of the marine vertebrate ecosystem from 425.0 to 358.9 million years ago (Carr, 1995; Denison, 1978; Janvier, 1996; Young, 2010; Zhu, 1996).
Recent prevailing phylogenetic hypotheses placed Placodermi as jawed stem-Gnathostomata that is sister to crown-Gnathostomata or modern jawed vertebrates (Brazeau, 2009; Davis et al., 2012; Dupret et al., 2014; Giles et al., 2015; King, 2021; Long et al., 2015; Qiao et al., 2016; Trinajstic et al., 2015; Zhu et al., 2013, 2016). In this scenario, Antiarcha has usually been placed at the most basal position in the Placodermi (Fig. 1), representing the most basal jawed vertebrates. The spatiotemporal distribution of Antiarcha will thus help us understand the origin and early evolution of jawed vertebrates. For example, Sallan et al. (2018) found that vertebrate diversification occurred primarily in nearshore environments by analyzing early vertebrate occurrence and habitat data. Historically, antiarchs resided in various paleo-environments across all paleo-continents, including marine and freshwater environments close to shore. As a successful vertebrate group during the Devonian (Long, 2011; Young, 2010), Antiarcha has contributed significantly to the Devonian stratigraphic correlation. For instance, the biozonation of the East Baltic and southern East Antarctica Devonian succession is partly based upon the antiarchs *Bothriolepis, Asterolepis,* and *Pambulaspis* (Young, 1974, 1988). Lukševičs (1996) identified 14 bothriolepid species (12 *Bothriolepis* and 2 *Grossilepis*) in the Frasnian-Famennian formations of the East European Platform, proposed nine antiarch assemblages, and set up the most detailed zonation of the Main Devonian Field, north-western part of the East European Platform (Latvia and NW Russia).
Collecting and visualizing the data of Antiarcha is a prerequisite to explaining the spatial and temporal distribution of early vertebrates. With the help of data visualization, we could better understand the biogeographic evolution of early vertebrates. Although Zigaite and Blieck (2013) advocated a quantitative analysis to define early vertebrates' biogeographic patterns, efficient quantitative analysis is still lacking to understand the dispersal of early vertebrates. This occurs mainly because no comprehensive data collection of early vertebrates was accomplished. What's more, the disadvantages of unstructured data are clear: the absence of schema and structure makes them difficult to manage, and the lack of predefined attributes makes them difficult to be reused or extended.

In this paper, we present an unprecedented structured dataset of Antiarcha that potentially facilitates understanding the spatiotemporal distribution pattern and quantifying the variety of antiarchs. This dataset is open access and follows the FAIR principles (findability, accessibility, interoperability, and reusability) (Wilkinson et al., 2016). This dataset complements existing fossil records of early vertebrates.
Moreover, it is the first step to accomplishing the global coverage of the vertebrate fossil dataset to analyze the Middle Paleozoic biogeography and paleogeography. Visualizing the distribution of antiarchs in the paleogeographic background with reference to the global paleogeographic reconstructions of Scotese (2021), our preliminary results can also be used to test the hypothesis of paleogeographic reconstructions.

2. Method

2.1 Overview

Comprehensive data are essential for quantitative studies and simulation analysis on early vertebrates. Sallan et al. (2018) pointed out that a lack of early vertebrate fossil data has limited quantitative approaches and hindered the resolution of issues regarding ancestral habitat in vertebrate evolution. To bring the study of Vertebrate Paleontology into the next phase of macroevolutionary, we built the DeepBone database with the implementation of a project entitled "Big Earth Data Science Engineering (CASEarth)" in 2018 (Guo, 2017; Pan and Zhu, 2019).

With continuously refining data, the Antiarcha dataset of the DeepBone database is the first and most comprehensive dataset endorsed by Chinese researchers at the Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences. The Antiarcha dataset of DeepBone differs from that of PBDB (https://paleobiodb.org/) in its basic unit, which is the specimen ID coupled with the occurrence and other detailed data. All the specimens are referenced in taxa and literature to guarantee accuracy. Because the data format was designed as specimen-based, we input the metadata according to the published specimen ID or virtual specimen ID. The literature on classic systematic paleontology always has real specimen IDs. When it handled stratigraphic topics, the authors usually cited fossil records instead of real specimens. We introduce a virtual specimen ID to store the taxon information in this kind of literature containing no real specimens.

Since no satisfactory approach can automatically extract paleontological data from the literature, we recruit several data entry assistants, including researchers, master’s and Ph.D. students, to collect and curate the data. In order to guarantee the quality of the data, we designed a four-step data processing procedure (Fig. 2):

1. Experts who obtained his/her Ph.D. degree in Paleontology collected and sifted the Data source.
2. Data entry assistants read the related references, extracted the antiarch placoderm data, and manually filled them into the online record file under the supervision of vertebrate paleontology experts.

3. According to the references, experts reviewed and cleaned the data line by line as the quality control procedure.

4. Senior experts, who have outstanding achievements in vertebrate paleontology, reviewed the data again to guarantee quality.

5. DeepBone.org published the dataset with visualization. A better user interface helps dissemination.

Next, we provide more details on the data processing and visualization.

---

**Figure 2** Workflow of the data processing. 1, collecting and sifting lectures by experienced experts. 2, data entry assistants digitalize paleontological descriptions from the page into the DeepBone database. 3, According to the reference, experts are accountable for data review and cleaning to implement quality control. 4, senior researchers review the data again. 5, data managers release the data to the public through DeepBone.org.
2.2 Data Source

The data were extracted from the published literature containing information on antiarch specimens. Most of the journals are professional journals on Paleontology. The main journals include Alcheringa, Acta Geologica Polonica, Bulletin of the Geological Society of China, Estonian Journal of Earth Sciences, Journal of Vertebrate Paleontology, Journal of Paleontology, Palaeontologia Electronica, Palaeontology, Palaeoworld, and Vertebrata PalAsiatica. Totally we have collected 142 publications spanning from 1939 to 2021 (see dataset for more details). The satisfactory literature should include an accurate description or revision of the specimen and taxon. We accepted the latest peer-reviewed literature to deal with the inconsistent descriptions of stratigraphy and taxonomy.

2.3 Data Processing and Quality Control

We made a tailored web page that provides a better user interface for data entry assistants to fill in the rows of paleontological data. After that, the other related experts would review the data so that a researcher could quickly access them to perform quantitative analysis (Fig. 2). This workflow was adopted from the Geobiodiversity Database (GBDB) (Xu et al., 2020). Almost all antiarch literature was published in English, Russian, French, German, and Chinese. The data entry assistant could handle the literature in Chinese and English well. Many fossils were documented in French, Russian, and German. We invited paleontology postgraduates, who know French, Russian, or German, to deal with the literature in these languages.

The faunistic elements in the communities are used herein at the genus level for their distributions because many Bothriolepis and Asterolepis species were described based on isolated plates lacking diagnostic characters (Blieck and Janvier, 1993; Downs, 2011). Identifying a specimen depends on the ability to recognize species in a way that is coherent within a particular genus and through broader groups. This is very difficult for fossil material by two especially intractable problems: practically, of the fragmentary nature of the fossil, and philosophically by questions with the criteria by which on demarcates fossil species (Nelson, 1984; Thomson and Thomas, 2001). For example, Thomson and Thomas (2001) reviewed the previous study on Bothriolepis and proposed that B. nitida, B. minor, B. virginiensis, B. darbiensis, and B. coloradensis could not be consistently distinguished. Weems (2004) questioned the validity of B. virginiensis. Since there is no consensus on the species
level of Bothriolepis and Asterolepis, the former researchers only used the evidence of Antiarcha on the genus level to discuss the biostratigraphic significance (Lelievre and Goujet, 1986; Pan, 1981; Young et al., 2010; Young and Lu, 2020). To keep accuracy and consistency, here we choose the genus of Antiarcha to perform data visualization.

2.4 Data Visualization

Due to the easy access of the paleo-geographic coordinates calculator (PointTrack version 7.0) (Scotese, 2021) and its widely used in Paleontology (Ke et al., 2016; Kiel, 2017), we decided to use Scotese's paleocontinent reconstruction to perform the plot map, although many paleogeographic reconstructions were proposed (Heckel and Witzke, 1979; Li and Powell, 2001). Using the TrackPoint software, we converted the excavation locations from the current GPS to paleo-GPS and visualized the locations using the Web Mercator algorithm (Battersby et al., 2014).

The timescale follows the International Commission on Stratigraphy International Chronostratigraphic Chart version 2021/07 (Cohen et al., 2021).

3. Results

3.1 Data Overview

This dataset consists of 60 genera of 6025 specimens, covering all known antiarch lineages. The observed quantity of genus and species in our dataset over time were summarized in histograms (Fig. 3). The 6025 specimens include 5867 fossil specimens that have been systematically described and documented and 158 virtual specimens introduced to describe the taxon information when no specimen was assigned for the referred fossil records. Each specimen has at least one reference within our dataset, and the specimens lacking precise age are excluded. We followed the lithostratigraphic information of the original authors, except we found a revision. We accepted the latest revision in the literature to modify our dataset. The amendments were linked to the latest reference as an endorsement. We took the geological background data in our dataset unless it was missing from the original literature. We transferred the unstructured data from the literature to structured data for further research in as much detail as possible. Table 1 shows the
data structure of our present dataset. Among all the referred specimens, 6.51% belong to Yunnanolepidoidei, 2.86% belong to Sinolepidoidei, 78.92% belong to 'Bothriolepidoidei', 11.71% belong to Asterolepidoidei. We plotted all the fossil sites of the constituent groups in Figure 4.

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specimen ID</td>
<td>Genus</td>
</tr>
<tr>
<td>Lithostratigraphy, formation</td>
<td>Species</td>
</tr>
<tr>
<td>Lithostratigraphy, member</td>
<td>Custodian institute</td>
</tr>
<tr>
<td>Fossil site</td>
<td>Custodian country</td>
</tr>
<tr>
<td>Discovery country</td>
<td>System / Period</td>
</tr>
<tr>
<td>Latitude</td>
<td>Series / Epoch</td>
</tr>
<tr>
<td>Longitude</td>
<td>Stage / Age</td>
</tr>
<tr>
<td>Paleo-latitude</td>
<td>Reference age to calculate paleo-coordinates</td>
</tr>
<tr>
<td>Paleo-longitude</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 The structure of Antiarcha dataset.

Figure 3 Histogram of specimen number. (a) The genus number at different time intervals, and (b) the species number at different time intervals.
Figure 4 Antiarch placoderm spatial distribution. The tree topology is simplified from the phylogenetic result in Pan et al. (2018). Numbers under the pie chart represent each group’s relative amounts of specimens. Terminal groups are linked with their geographic distributions. Each node represents a single specimen. Specimens from one locality overlap each other.

GBDB is a stratigraphic and palaeontological database, but no antiarch record exists. Compared to the 138 records of Antiarcha in the Paleobiology Database (PBDB, 2021-08-12), this is the most comprehensive dataset of Antiarcha up to now (Table 2). Only taxon rank, reference, and occurrence location are available in PBDB. DeepBone dataset has more fields on the structured information of the specimen than that is in the PBDB, such as lithostratigraphic fields (Table 1). Some records in PBDB are not stored at the genus or species level. There are some typing errors in PBDB, for instance, 'Jiangxilepus', 'Bothriolepidoidei' and 'Pterichthys'. Jiangxilepis, 'Bothriolepidoidei' and Pterichthyodes are correct spellings. Macrodontophion is not a genus of Antiarcha, but PBDB adopts it in antiarchs. PBDB also adopts Silurolepis as an antiarch ignoring the latest research of Zhu et al. (2019). To ensure accuracy, every specimen of DeepBone is endorsed by the latest publication and reviewed by the experts.
who have focused on Antiarcha. It is open to access through the website of the DeepBone database or https://doi.org/10.5281/zenodo.6478602.

<table>
<thead>
<tr>
<th></th>
<th>DeepBone Database</th>
<th>Paleobiology Database(PBDB)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type</strong></td>
<td>specimen-based</td>
<td>fossil-occurrence-based</td>
</tr>
<tr>
<td>No. of references</td>
<td>142</td>
<td>19</td>
</tr>
<tr>
<td>No. of genera</td>
<td>60</td>
<td>26</td>
</tr>
<tr>
<td>No. of species</td>
<td>187</td>
<td>98</td>
</tr>
<tr>
<td>No. of Specimens/occurrences</td>
<td>6025</td>
<td>138</td>
</tr>
<tr>
<td>Found in</td>
<td>2018</td>
<td>1998</td>
</tr>
<tr>
<td>Website</td>
<td><a href="http://www.deepbone.org">www.deepbone.org</a></td>
<td><a href="https://paleobiodb.org">https://paleobiodb.org</a></td>
</tr>
</tbody>
</table>

(last access: 22 October 2021)

Table 2 Comparison of Antiarcha data in two paleontological databases.

### 3.2 The Geospatial Distribution of the Antiarcha Dataset

The Geospatial Distribution of the Antiarcha is shown in Figure 4. Yunnanolepidoidei is endemic in the South China block (comprising southern China and northern Vietnam) regarding the fossil site distribution. Sinolepidoidei is limited in South China and Australia (East Gondwana). In contrast, 'Bothriolepidoidei' and Asterolepidoidei are cosmopolitan, especially *Bothriolepis*. The heat map of fossil sites (Fig. 5) shows that Europe, Australia, and China account for the most fossil sites globally, partly due to their long research history.
3.3 The Paleogeographic Distribution of the Antiarcha Dataset

As Young (1990) mentioned that biogeographic data must be interpreted in the context of paleogeographic hypotheses, we plot our data on a paleogeographic atlas (Fig. 6) only to generate an outline of their past. The further interpreting data studies will be published separately. The continental reconstructions of Scotese place Baltica, China, and Australia in the tropic and subtropic near the paleo-equator from Llandovery to Famennian. We excluded the Silurian Shimenolepis, because it is the earliest record of Yunnanolepidoidei and the only documented antiarch specimen before the Devonian (Wang, 1991; Zhao et al., 2016). Most of the fossil sites were positioned around the paleo-equator. In the present scenario, the suborder Yunnanolepidoidei apparently originated as early as the Silurian in the South China block, forming a highly endemic fauna. All fossil sites of Yunnanolepidoidei lay in southern China and northern Vietnam (Wang et al., 2010). From Ludlow (Silurian) to the Early Devonian, Yunnanolepidoidei formed dominant antiarchs. Sinolepidoidei and 'Bothriolepidoidei' first appeared in Pragian in South China, and Asterelepidoidei first evolved in Emsian in Australia or East Gondwana. During the middle Devonian, along with lessened isolation of South China, Yunnanolepidoidei became...
extinct. Euantiarcha ('Bothriolepidoidei' + Astrolepidoidei) dominated Middle and Late Devonian antiarchs, and only a few members of Sinolepidoidei coexisted with them in China and Australia. In Eifelian, Astrolepidoidei suddenly bloomed in Baltica without any clue from the older horizons based on existing research. The distribution of Antiarcha reached a peak in Givetian. 'Bothriolepidoidei' and Astrolepidoidei represent the main groups of Antiarcha in Givetian, comprising five bothriolepidoid genera with 42 fossil locations and nine astrolepidoid genera with 49 fossil locations.

Figure 6 The distributions of Antiarcha during Devonian. Each spot represents the location of one specimen on the paleogeographic map. Specimens from the same locality and the same age overlap each other. The paleo-coordinates are calculated by TrackPoint. Colors denoting respective groups follow Figure 4.

3.4 First Appearance Record

A taxon's first appearance record or lineage is important in Paleontology and Evolutionary Biology. It renders a hard minimum constraint on molecular clock calibration for a taxon (Benton and Donoghue, 2007; Benton et al., 2009; Donoghue and Benton, 2007). Based on our dataset, the oldest record of
Shimenolepis granifer from the Xiaoxi Formation at Shanmen Reservoir, Lixian County, Hunan, China. Shimenolepis was first described as the oldest known placoderm, dated as Telychian of Llandovery (Janvier, 1996; Wang, 1991). However, after a detailed stratigraphic work, Zhao et al. (2016) suggested that the age of Shimenolepis is late Ludlow rather than late Llandovery. Janvier and Tông-Dzuy (1998) also documented an indeterminate yunnanolepidoid (Antiarcha gen. sp. indet.) from the Do Son Formation of northern Vietnam, which could be another earliest antiarch potentially.

The oldest sinolepid is Liujiangolepis suni, from the Nakaoling Formation (Pragian), Xiangzhou, Guangxi, China (Wang, 1987). The oldest bothriolepidid is Houershanaspis zhangi, documented from the Danlin Formation (Pragian) in Houershan, Dushan county, Guizhou, southwestern China, based on a bothriolepid-like anterior median dorsal plate (Lu et al., 2017). The earliest asterolepidoid records are represented by Wurungulepis and some disarticulated specimens, which have been documented from the Broken River Formation, Broken River, Australia. The age of the Broken River Formation was first referred to Eifelian and then reassigned to Emsian (serotinus Zone) (Burrow, 1996; De Pomeroy, 1996; Young, 1984, 1990b).

4 Data Availability

The current dataset achieved via Zenodo represents a static version of the dataset in June 2022: https://zenodo.org/record/6536446 (Pan and Zhu, 2021). The latest version of the dataset is always freely available via https://deepbone.org/ (last access: June 2022).

5 Conclusions

We presented here an open-access dataset of Antiarcha, the most basal jawed vertebrate, from the late Silurian to the Late Devonian. This dataset significantly expands the previously available data of antiarch fossils. Paleontologists, stratigraphers, and evolutionary biologists could import the tab-delimited file for future research studies, especially for biodiversity analysis, stratigraphic correlation, and molecular clock calibration. With the information of 6025 specimens, our Antiarcha dataset is far more comprehensive than the other sources in lithostratigraphy and specimen details. Data are significant for quantitative analysis and potentially contribute to data-driven paleontology research. We performed a visualization of the data to show the spatiotemporal distribution of Antiarcha. In brief, Antiarcha first appeared in the Pan-Cathaysia province during the late Ludlow and then boomed worldwide. At the end of Devonian,
Antiarcha was extinct along with the traditional placoderms. The available Antiarcha data may be just the tip of the historical reality due to the incomplete fossil record. We will continue to update the dataset with the latest publication.

260 **Author contributions.** MZ supervised the project. ZHP and ZBN performed the visualization. ZHP prepared and revised the manuscript with contributions from MZ and ZBN. ZHP and ZMX prepared the data.

265 **Competing interests.** The authors declare that they have no conflicts of interest.

267 **Acknowledgment.** We thank Wenjin Zhao, Zhikun Gai, and Tuo Qiao, Institute of Vertebrate Paleontology and Paleoanthropology (CAS); Junxuan Fan, School of Earth Sciences and Engineering, Nanjing University; Honghe Xu, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences (CAS); Junqi Wu, Yaqi Xue, College of Intelligence and Computing, Tianjin University for discussion and help.

272 **Financial support.** This project was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA19050102, XDB26000000), National Science Foundation of China (No. 42002015), Youth Innovation Promotion Association CAS (No. 2021070), State Key Laboratory of Palaeobiology and Stratigraphy (Nanjing Institute of Geology and Palaeontology, CAS) (No. 193121).

275 **Reference**


Kiel, S.: Using network analysis to trace the evolution of biogeography through geologic time: A case study, 45(8), 711-714, 2017.


Pan, Z. H. and Zhu, M, Dataset of antiarch placoderms (the most basal jawed vertebrates) throughout Middle Paleozoic (Version 31-Aug-2021), Zenodo, 10.5281/zenodo.5347006, 2021.


