Retrospects and prospects of a section-based stratigraphic and palaeontological database – the Geobiodiversity Database

Hong-He Xu[1,3,*], Zhi-Bin Niu[1,2,*] Yan-Sen Chen[1,3]

[1] State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, 210008 Nanjing, China

2 College of Intelligence and Computing, Tianjin University, 300354 Tianjin, China

3 Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, 210008 Nanjing, China

* The authors contributed equally to this work.

Correspondence to: Hong-He Xu (hhxu@nigps.ac.cn) and Zhi-Bin Niu (zniu@tju.edu.cn)

Abstract

Big data are significant to quantitative analysis and contribute to the data-driven scientific research and discoveries. Here a brief introduction is given on the Geobiodiversity database (GBDB), a comprehensive stratigraphic and palaeontological database and its data. The GBDB includes abundant geological records from China and has supported a serial of scientific studies on the Palaeozoic palaeogeographic, tectonic and biodiversity evolution of China. The data that the GBDB has and newly collected are described in details, the statistical result and structure of the data are given. A comparison between the GBDB, the largest paleobiological database - PBDB, and the geological rock database – Macrostrata, is drawn. The GBDB has been updating since 2019 and the prospects are given. The GBDB and other databases are complementary in palaeontologial and stratigraphical research. The GBDB will continually and assiduously provide users access to the detailed palaeontological and stratigraphical data based on publications. Non-structured data of the palaeontology and stratigraphy will also be included in the GBDB and they will be organically correlated with the existing data of the GBDB, making the GBDB more widely used for both researchers and anyone who are interested in fossils and strata. The GBDB fossil and stratum dataset (Xu, 2020) is freely downloadable from http://doi.org/10.5281/zenodo.3667645.

1. Introduction

Palaeontology and stratigraphy have become a quantitative discipline of geoscience and there has been a subsequent rapid increase in the implementation of numerical methods in palaeontology and stratigraphy that started in the 1960s (Shaw, 1964; Schwarzacher, 1975; Kemple et al., 1989; 1995; Sepkoski, 1992, 2002; Alroy et al., 2001; Hammer and Harper, 2006; Rong et al., 2007). Quantitative analysis based on big data of fossil and stratum records have been more common recently, especially on biodiversity evolution (Alroy et al., 1994; 2001; 2008; Hautmann, 2016; Fan et al., 2020), graphic correlation of strata (Kemple et al., 1989; Fan et al., 2013b), palaeoecology (Muscente et al., 2018), mass extinction (Muscente et al., 2019) and palaeogeography (Ke et al., 2016; Hou et al., 2020). There are professional databases, such as Paleobiology Database (PBDB), Macrostrat (https://macrostrat.org/) and Geobiodiversity Database (GBDB), storing and providing a big volume of fossil record data and making a number of quantitative studies possible. Well-structured stratigraphic and palaeontological databases and user-friendly-, accessible data are significant to the quantitative development of the discipline and furthermore, push forward digital earth science in the era of big data (Guo, 2017). In this paper, we show the update and the improvement of a comprehensive database of stratigraphy and palaeontology biodiversity, Geobiodiversity Database (GBDB), and its data, brief history, development, and improvement. The comparisons between related databases are also given.

2. A brief history of the Geobiodiversity Database

The Geobiodiversity Database (GBDB) was started in 2006 and provided online service since 2007 when there was a strong and urgent demand for the quantitative understanding of fossil and stratum records from China, which was initially supported by the national project of "Organism origination, radiation, extinction and recovery during the key geological ages" (973 Project) (Rong et al., 2006; 2007). At that time the PBDB (Paleobiology Database) had been a large paleontological database that included plenty of fossil occurrence data from the publications of euro-languages, however, fossil and stratum data from China were temporarily ignored for the obstacle of language. The initial purpose of the GBDB was to accommodate fossil and stratum data, geological sections as well as fossil collections from China, and furthermore to recognized biodiversity changed occurring in the geological ages of China (Rong et al., 2006).

Since the start of the GBDB, there used to be at most ten data entry clerks, including master or PhD students, assistant researchers and non-professional employees, digitalizing palaeotological and stratigraphic descriptions "from the page into cyberspace" (Normile, 2019) and aligning these data with standards that are acceptable to international researchers, so that a researcher could quickly link them to carry on quantitative analysis that would likely have omitted Chinese data previously.

The GBDB was designed to facilitate regional and global scientific collaborations focused on palaeobiodiversity, systematics, palaeogeography, palaeoecology, regional correlation, and quantitative stratigraphy.

Basic functions of data input and output were gradually added and enhanced. In 2013, a huge volume of palaeontological and stratigraphic data were included in the GBDB, such as taxonomy, identification features, occurrence, opinion, lithostratigraphy, biostratigraphy, chemostratigraphy, radio isotopic dating, reference and palaeogeographic map (Fan et al., 2013a; Fan et al., 2014). Additionally, there were embedded a few online statistical and visualization tools, such as Time Scale Creator (integrated into GBDB in 2010), a stratigraphic visualization tool designed by Jim Ogg and Adam Lugowski (http://www.tscreator.com), and 2) GeoVisual (integrated in GBDB in 2010 and updated in 2012), a tool used for geographic visualization and preliminary biogeographic analysis.

One of the exclusive features of the GBDB is its abundant geological section data, which are readily exported for the numerical correlation tools, such as Constrained Optimization (CONOP) (Kemple et al., 1995) and SinoCor. SinoCor was designed and updated by Fan et al. (2002) and Fan and Zhang (2000; 2004). Its correlation resembles CONOP but requires a unique file format. SinoCor and CONOP are individual outgrowths of graphic correlation. There are several related professional tools, such as Graphcor, PAST, and CONMAN (see Hammer and Harper, 2006; Fan et al., 2013b).

The GBDB became the formal database of the International Commission on Stratigraphy in August 2012 at the 34^{th} International Geological Congress in Brisbane, Australia, and, as a result, GBDB achieved the goal of integrating stratigraphic standards (e.g. the GSSPs) with comprehensive and authoritative web-based stratigraphic information service for global geoscientists, educators, and the public.

Since 2011, data related to early Paleozoic, especially Ordovician and Silurian periods, stratigraphic and palaeontological records had been quantitatively analyzed and a series of scientific findings were published. The related research themes include the Ordovician and Silurian palaeogeography and tectonic evolution of South China (Chen et al., 2012; 2014b; 2017a), the spatio-temporal pattern of the Ordovician and Silurian marine organisms from China (Chen et al., 2014a; 2017b; Zhang et al., 2014a; 2016), Permian-Triassic transition and extinction (Shen et al., 2011; 2013; Wang et al., 2014; Ke et al., 2016), and the Paleozoic paleogeography evolution of South China (Chen et al., 2018; Zhang et al., 2014b; Hou et al., 2020). Recently, nearly all data of Paleozoic marine organisms of GBDB were used to analyze biodiversity evolution (Fan et al., 2020). Though all data were from China, the Paleozoic geological sections of China cover several palaeocontinents and reflect global biodiversity change.

In 2017, the GBDB became a data partner of the British Geological Survey (BGS) and started to digitalize the fossil and stratum data and establish the datasets for the BGS. This is a time-consuming task and still ongoing by the GBDB data entry team. The BGS has amassed and housed about 3 million fossils gathered over more than 150 years at thousands of sites across the British Islands.

At the end of 2018, the head of the GBDB, Dr. Fan J.X., left the NIGP, CAS and Dr. Xu H.H. took over the GBDB. Besides data collecting, processing and visualization as the GBDB group did during 2007-2018, data of fossil terrestrial organisms, such as insects and plants, were input into the GBDB, the database and the website were re-designed according to the feedbacks collected from the GBDB users, and the GBDB is ushering a new start.

3. The data of the Geobiodiversity Database

The Geobiodiversity database (GBDB) was designed as a stratigraphic and palaeontological database and its input format was designed as geological section-based, which means that data entry clerks or any scientific users must input the metadata for the GBDB according to the geological sections or assumed sections. Every metadata record contains all geological information of a geological section, including its basic unit (or bed or layer), sediment color, lithology, thickness, horizon, locality, palaeo-block, geological age, bio-stratigraphy, geochemistry, palaeo-ecology, radio isotopic age, fossil collection and any available original information of the rock specimens or fossil sample during the fieldwork. An individual geological section normally can be subdivided into dozens of basic units when it is input the GBDB. Such geological section records with much information can be found from stratigraphic and palaeontological literature. Sometimes the geological sections are not easily or directly to obtain and the help from the professional experts is necessary. However, many paleontological descriptions or reports are lacking detailed stratigraphic description, the GBDB includes these records as assumed sections, e.g. they were treated as geological sections, which may have only a very small portion, for example, of a single bed or collection. Borehole core records, many of which are from the oil company and are not open to the public, are also input into the GBDB as assumed sections (Figure 1).

The stratigraphic data in GBDB are based on those published in Chinese literature since the 1920s. By November 2019, all stratigraphic horizons and nearly all published geological sections can be browsed in the GBDB (Figures 2, 3). It is noteworthy that in the GBDB fossil occurrence data are included in the stratigraphic records and can't be queried directly, such was improve in our updating. The palaeontological data are linked to the fossil collections from individual geological sections and borehole cores. The data include taxonomy (species, genus, family, order, class and division), major group, synonym (opinion data with different authors) and description (key features) (Figure 1). Though the GBDB is geological section-based, from which fossil occurrences can be output, it is compatible with fossil occurrence-based databases. Most fossil collections and occurrences of all sections from China are included in the GBDB (Figure 3). Subsequent authors in further study amended a portion of fossil taxa from these sections. In this way, there are also plenty of opinion data in the GBDB.

Since 2017, the GBDB started to record the data of Global Boundary Stratotype Sections and Points (GSSPs) of the International Commission on Stratigraphy, including the detail information of GSSP and some panorama and three-dimensional scanning of individual GSSP, as is example of the Changhsingian GSSP in Changxing, Zhejiang Province, southeastern China (http://www.geobiodiversity.com:8080/Panorama/47/output/).

Since August 2017, the British Geological Survey (BGS) and the GBDB started to collaborate in stratigraphic and palaeontological data processing. The GBDB data working team help to digitalize the geological reports from the BGS archive and to build separated datasets for it.

Since 2019, the GBDB has begun to include the borehole core data of petroleum companies, such as China National Offshore Oil Corporation and China National Petroleum Corporation.

In brief, as much as possible stratigraphic and palaeontological records are collected from the original geological publications. Since establishment, the GBDB data team conscientiously collected and included stratigraphic and palaeontological data from Chinese literature. The detailed statistic outcomes are given here (Table 1) (see Xu, 2020).

For a long time, the biodiversity evolution study was based on marine organism fossil records. For example, the earliest quantitative analysis of the geological time biodiversity that draws the conclusion of the five mass-extinction (Raup and

Sepkoski, 1982) and a serial of related geological biodiversity studies were based on marine organism fossil family or genus records (lot of work was done based on PBDB data, see Jablonski, 1994; Sepkoski, 1992; 2002; Alroy et al., 2001; 2008; Rong et al., 2006; 2007), as well as the quantitative study based on terrestrial organism fossil records (e.g. Alroy, 1998; 2001). There have been quantitative studies on the plant diversity of the Silurian and Devonian periods that was significant for the early plant evolution and diversification (Xiong et al., 2013) and the study on plant diversity change during the Permian-Triassic boundary (Xiong and Wang, 2007). The mass extinction occouring at the end of Permian is the greatest extinction of the geological history and wipes out over 95% marine organisms (Jablonski, 1994). Both plant diversity studies used fossil record data from South China and listed the data as the supplementary material of the published papers. It took the authors of the two studies a few years to complete the data collection, even the data from only the South China palaeo-block.

An inconvenient fact is that the terrestrial organism fossil database is not as good as that of marine ones, and that the non-marine fossil record is necessarily less complete and less widespread. For a lone time, the GBDB focused on the fossil records of marine organisms. Since 2019, GBDB started collect terrestrial fossil and stratum data conscientiously and now has a unique feature for the fossil terrestrial organisms. The fossil plant record dataset has collected 738 Devonian plant species occurrences from global localities and thousands of Mesozoic plant species occurrences from China.

Beside the plant fossil, the terrestrial organism fossil record data of the GBDB are the insect fossil records, which greatly increase after taking over the international fossil insect database of the International Palaeoentomological Society, EDNA (https://fossilinsectdatabase.co.uk/), which holds details of the holotypes of all fossil insects in the world.

The EDNA database was named after Edna Clifford who started the recording of new species on a card index system and was designed as an update of Handlirsch’s 1906-1908 "Die Fossilen insekten und die phylogenie der rezenten formen" which listed all the then known fossil insect species. Handlirsch recorded 5,160 species in 1906. The database is detailed in its contents: it records taxonomic information, synonym details, references for every species (including the page number where it is introduced), and for holotypes site details, stratigraphic information, and geological details are recorded. All the data have been obtained from exhaustive literature searches.

The EDNA database aims to be a complete, fully interactive, list of all the species of insects named from the fossil record, with the site, geological age, and reference for each holotype. Updating and checking will be ongoing, and the data available will be greatly improved if details of omissions and errors are sent to the administrator for incorporation. The data comes from an exhaustive literature search and in the 2019 edition contains 28,439 species names (including synonyms) extracted from 5,218 references (Figure 3d). The data is held in 38 fields, all of which are searchable, independently or in combination, and the output can contain any one or more as required.

Fields include: generic and specific names, citation, subfamily, family, superfamily, division, suborder and order: Author, title, journal, and date of publication, and page on which the species is first described: Age data including stage, epoch, subperiod, period and era and age (range) in millions of years: Bed, member, formation, and group: Site name, nearest feature (town, river etc.) county, state, country and continent (Figure 4). For all taxonomic ranks, citations can be included and both junior and senior synonyms displayed. Natural History Museum London Library call numbers are also included.

4. Database comparisons and discussions

The comparison is made between the GBDB and fossil occurrence-based Paleobiology Database (PBDB), which was founded in 1998 and became the largest paleobiological database. Data of the PBDB include fossil taxa, collection, opinions (paleobiological views from different authors) and related publications. The data volume of the PBDB is larger than the GBDB (Table 1). The noticeable difference lies in that the PBDB has little information about geological sections. The GBDB is known for its large number of geological sections.

By November 2019, 26,450 geological sections were recorded in the GBDB, the geological age of which range from Ediacaran to Cenozoic (Table 1). These include nearly all sections and some of borehole cores from China, worldwide sections and borehole cores from open publications and reports of the British Geology Survey. Every record is based on published literature or internal reports.

As we mentioned, the GBDB is geological section-based; every record was subdivided into detailed parts when being input in the database. The fossil occurrence and collection data can also be exported from the GBDB, just as those in the PBDB. Nevertheless, the fossil taxon count recorded in the GBDB is about 30% of that in the PBDB, whilst the fossil occurrence records in the GBDB is about 40% of that in the PBDB (Table 1). This is because the two databases have different histories, the PBDB was founded in 1998, the GBDB, in 2007 (Figure 1). The PBDB has a history of comprehensive backup, mirror sites and multiple portals (e.g. Fossilworks: Fossilworks.org), and user-training guide, all of which are the thing that the GBDB need to improve.

The stratigraphic record in the GBDB are reminiscent of Macrostrat (https://macrostrat.org/), which is a platform for the aggregation and distribution of geological data relevant to the spatial and temporal distribution of sedimentary, igneous, and metamorphic rocks as well as data extracted from them. Macrostrat aims to become a community resource for the addition, editing, and distribution of new stratigraphic, lithologic, environmental, and economic data. By November 2019, Macrostrat records 1,534 regional rock columns, 35,163 rock units, and 2,484,619 geologic map polygons. Macrostrat has a lot of exclusive data of composite geological sections, e.g., the section that are compiled from several places in one basin and may have completeness and thickness that never accumulated in one place. It is also worth noting that Macrostrat records mostly geological data from North America, whilst the GBDB includes nearly all stratigraphic data of sediments from China, igneous and metamorphic rocks were also recorded if they were reported in sediment units.

5. Updates and prospects

Since the GBDB website started online in 2007, there have been few updates. During the management change of the GBDB at the end of 2018. A survey was carried within some users of the GBDB and a lot of feedbacks were received. According to these suggestions and feedbacks we sorted the existing problems of the GBDB and its website, and we comprehensively updated the server and the website of the GBDB, making the database a safe data bank and the website a new and friendly portal (GBDB 2.0, relative to the previous version). The new website has optimized input and output of data, the search engine, and the data examination system.

During the process of data inputting, the raw data will be checked by registered authorizers, such action aims to make sure that the data valid but not to the authorizer's own point of view. Today knowledge is updating quickly, it is normal to have a mixture of valid and obsolete information to a certain point, such as the taxonomical synonymies, and the implementation of a better decay constant to recalculate old radio-isotopic dates. The GBDB shows only the data bank but not supports any academic points. The authorizers make the data valid but the user choose the data to use and analyze. In the GBDB a huge volume of opinion data is remained.

Data visualization is developed. All data are plotted on the world map of the homepage that also displays the volume of the all data in the right up corner. The view center is the map of China and the map can be zoomed in or out using mouse scroll. Geological sections are showed as individual spots and their rough or detailed information can be checked easily. The different colors of spots on the map correspond to various geological stages of the International Chronostratigraphic Chart that is show in disk-shaped in the right lower corner and can be hidden manually.

User system is optimized; a personal profile and user-favorite feature can be customized. Users can search and choose data to download and analyze, the user account is needed to store search results, to run private data through CONOP or SinoCore, and to give comments. The old version of the GBDB remains available and has an entrance on the homepage, for the users who prefer the old format and hope to use the GBDB in the way they have learned. The GBDB also developed the applications for mobile devices, users can examine the data of the GBDB and give comments through the mobile devices.

In the next step, more data visualization, and analytic tools (Figure 5) will be embedded in the new GBDB website publicly, for stratigraphic and palaeontological research.

GBDB and PBDB are complementary in their great volumes of geological section and fossil occurrence data. Through the geological sections, the GBDB data records the thickness of individual fossil samples and have the important evidence of fossil organism co-existence. Fossil taxa of the two databases contain not only the widely-distributed and endemic fossils, but also those published in both English (and others) and Chinese languages. GBDB and Macrostrat are complementary in the stratigraphic study to some extent. The data of the two databases contain records from both North America and China. Data from these databases, therefore, provide the possibility to conduct various stratigraphic and paleontological analyses.

The GBDB, just as the PBDB and the Macrostrat, will continually and assiduously provide users access to the detailed palaeontological and stratigraphical data based on publications. Multiple and compatible formats for common software, such as CONOP and SinoCor, will be downloadable in the GBDB. Statistical and analytical tools will be easily used in the GBDB. Additionally, the GBDB is collecting non-structured data of the palaeontology and stratigraphy, including fossil specimens' images and three-dimensional models, geological section panorama images, tomographic image stacks, and references. We will build the organic correlations between these non-structured data and the palaeontological and stratigraphic data that the GBDB has collected for years. All-around information will be shown after searching an individual item thar is related to fossils or strata, making the GBDB more widely used for both researchers and anyone who are interested in fossils and strata.

Author contribution: HX and ZN designed the project, developed the model, and performed the simulations. HX prepared and revised the manuscript with contributions from ZN. Y-SC gave technician supports.

Competing interests: The authors declare that they have no conflict of interest.

Data availability: All data are downloadable from the website portal https://www.geobiodiversity.com/ or http://doi.org/10.5281/zenodo.3667645 (Xu, 2020).

Acknowledgments

We thank Prof. W. Kiessling, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany; Drs. Na Lin, Li Qijian and Wang Bo, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences (CAS); Dr. Pan Zhaohui, Institute of Vertebrate Paleontology and Paleoanthropology, CAS, and Mr. Wu Junqi, College of Intelligence and Computing, Tianjin University, for constructive suggestions and helps. This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grants XDA19050101 and XDB26000000) and National Natural Science Foundation of China (Grants 41772012 and 61802278).

References

Alroy, J., Aberhan, M., Bottjer, D. J., Foote, M., Fürsich, F. T., Harries, P. J., Hendy, A. J., Holland, S. M., Ivany, L. C., and Kiessling, W.: Phanerozoic trends in the global diversity of marine invertebrates. Science, 321, 97-100, 2008.

Alroy, J., Marshall, C. R., Bambach, R. K., Bezusko, K., Foote, M., Fusich, F. T., Hansen, T. A., Holland, S. M., Ivancy, L. C., Jablonski, D., Jacobs, D. K., Jones, D. C., Kosnik, M. A., Lidgard, S., Low, S., Miller, A. I., Gottshall, P. M., Olszewski, T. D., Patzkowsky, M. E., Raup, D. M., Roy, K., Sepkoski, J. J. Jr., Sommers, M. G., Wagner, P. J., and Webber, A.: Effects of sampling standardization on estimates of Phanerozoic marine diversification. Proceedings of Natural Academic

Sciences USA. 98, 6261-6266, 2001.

Alroy, J.: A multispecies overkill simulation of the end-Pleistocene megafaunal mass extinction. Science, 292, 1893–1896, 2001.

Alroy, J.: Cope's rule and the dynamics of body mass evolution in North American fossil mammals. Science, 280, 731-734. 1998.

Chen, Q., Fan, J., Melchin, M. J., and Zhang, L.: Temporal and spatial distribution of the Wufeng Formation black shales (Upper Ordovician) in South China. GFF, 136, 55-59, 2014a.

Chen, Q., Fan, J., Zhang, L., and Chen, X.: Paleogeographic evolution of the Lower Yangtze region and the break of the "platform-slope-basin" pattern during the Late Ordovician. Science China Earth Sciences, 61, 625-636, 2018.

Chen, X., Fan, J., Chen, Q., Tang, L., and Hou, X.: Toward a stepwise Kwangsian Orogeny. Science China Earth Sciences, 57, 379-387, 2014b.

Chen, X., Fan, J., Wang, W., Wang, H., Nie, H., Shi, X., Wen, Z., Chen, D., and Li, W.: Stage-progressive distribution pattern of the Lungmachi black graptolitic shales from Guizhou to Chongqing, Central China. Science China Earth Sciences, 60, 1133-1146, 2017a.

Chen, X., Zhang, Y., Fan, J., Tang, L., and Sun, H.: Onset of the Kwangsian Orogeny as evidenced by biofacies and lithofacies. Science China Earth Sciences, 55, 1592-1600, 2012.

Chen, Z., Männik, P., and Fan, J.: Llandovery (Silurian) conodont provincialism: An update based on quantitative analysis. Palaeogeography, Palaeoclimatology, Palaeoecology, 485, 661-672, 2017b.

Fan, J. and Zhang, Y.: SinoCor 1.0, a biostratigraphic program for Graphic Correlation. Acta Palaeontologica Sinica, 39, 573-583, 2000. (in Chinese with English abstract).

Fan, J. and Zhang, Y.: SinoCor 3.0, a biostratigraphic program for graphic correlation. Erlanger geologische Abhandlungen Sonderband, 5, 35–36. 2004.

Fan, J., Chen, Q., Hou, X., Miller, A. I., Melchin, M. J., Shen, S., Wu, S., Goldman, D., Mitchell, C. E., Yang, Q., Zhang, Y., Zhan, R., Wang, J., Leng, Q., Zhang, H., and Zhang, L.: Geobiodiversity Database: a comprehensive section-based integration of stratigraphic and paleontological data. Newsletters on Stratigraphy, 46, 111-136, 2013a.

Fan, J., Chen, Q., Melchin, M. J., Sheets, H. D., Chen, Z., Zhang, L., and Hou, X.: Quantitative stratigraphy of the Wufeng and Lungmachi black shales and graptolite evolution during and after the Late Ordovician mass extinction. Palaeogeography, Palaeoclimatology, Palaeoecology, 389, 96-114, 2013b.

Fan, J., Chen, X., and Zhang, Y.: Quantitative biostratigraphy of Upper Ordovician to lowermost Silurian on the Yangtze platform—with the designing of SinoCor 2.0, a software for graphic correlation. Memoirs of the Association of Australasian Palaeontologists, 27, 53-58, 2002.

Fan, J., Hou, X., Chen, Q., Melchin, M. J., Goldman, D., Zhang, L., and Chen, Z.: Geobiodiversity Database (GBDB) in stratigraphic, palaeontological and palaeogeographic research: graptolites as an example. GFF, 136, 70-74, 2014.

Fan, J., Shen, S., Erwin, D. H., Sadler, P. M., McLeod, N., Cheng, Q., Hou, X., Yang, J., Wang, X., Wang, Y., Zhang, H. Chen, X., Li, G., Zhang, Y., Shi, Y., Yuan, D., Chen, Q., Zhang, L., Li, C., and Zhao, Y: A high-resolution summary of Cambrian to Early Triassic marine invertebrate biodiversity. Science, 367, 272-277, 2020.

Guo, H.: Big Earth data: A new frontier in Earth and information sciences. Big Earth Data, 1, 4-20, 2017.

Hammer, O. and Harper, D. A. T.: Paleontological Data Analysis. Blackwell Publishing, 2006.

Hautmann, M.: Diversification and diversity partitioning. Paleobiology, 40, 162-176, 2016.

Hou Z S, Fan J X, Henderson C M, Yuan D X, Shen B H, Wu J, Wang Y, Zheng Q F, Zhang Y C, Wu Q, Shen S Z. Dynamic Palaeogeographic Reconstructions of the Wuchiapingian Stage (Lopingian, Late Permian) for the South China Block. Palaeogeography, Palaeoclimatology, Palaeoecology, 546, 109667, 2020.

Huang, B., Rong, J., and Cocks, L. R. M.: Global palaeobiogeographical patterns in brachiopods from survival to recovery after the end-Ordovician mass extinction. Palaeogeography, Palaeoclimatology, Palaeoecology, 317-318, 196-205, 2012.

Jablonski, D.: Extinctions in the fossil record. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 344, 11-17, 1994.

Ke, Y., Shen, S.-Z., Shi, G.R., Fan, J.-X., Zhang, H., Qiao, L. and Zeng, Y.: 2016. Global brachiopod palaeobiogeographical evolution from Changhsingian (late Permian) to Rhaetian (late Triassic). Palaeogeography, Palaeoclimatology, Palaeoecology, 448, 4–25. 2016.

Kemple, W. G., Sadler, P. M., and Strauss, D. J.: Extending graphic correlation to many dimensions: stratigraphic correlation as constrained optimization, 1995.

Kemple, W.G., P.M. Sadler, and D.J. Strauss. A prototype constrained optimization solution to the time correlation problem. In Agterberg, F.P. & G.F. Bonham-Carter (eds.), Statistical Applications in the Earth Sciences, pp. 417–25. Geological Survey of Canada, Paper 89–9. 1989.

Muscente, A. D., Bykova, N., Boag, T. H., Buatois, L. A., Mangano, M. G., Eleish, A., Prabhu, A., Pan, F., Meyer, M. B., Schiffbauer, J. D., Fox, P., Hazen, R. M., and Knoll, A. H.: Ediacaran biozones identified with network analysis provide evidence for pulsed extinctions of early complex life. Nature Communication, 10, 911, 2019.

Muscente, A. D., Prabhu, A., Zhong, H., Eleish, A., Meyer, M. B., Fox, P., Hazen, R. M., and Knoll, A. H.: Quantifying ecological impacts of mass extinctions with network analysis of fossil communities, Proc Natl Acad Sci U S A, 115, 5217-5222, 2018.

Normile, D.: Earth scientists plan a 'geological Google', Science, 363, 917, 2019.

Raup, D. M. and Sepkoski, J. J.: Mass extinctions in the marine fossil record, Science, 215, 1501-1503, 1982.

Rong, J. Y., Fan, J., Miller, A. I., and Li, G. X.: Dynamic patterns of latest Proterozoic-Palaeozoic-early Mesozoic marine biodiversity in South China, Geological Journal, 42, 431-454, 2007.

Rong, J., Fan, J., and Li, G.: Patterns of latest Proterozoic to early Mesozoic marine biodiversity changes in South China. Science Press Beijing, 2006.

Schwarzacher, W.: Sedimentation models and quantitative stratigraphy. Elsevier. 1975.

Sepkoski, J. J. Jr.: A compendium of fossil marine animal families. 2nd edition, Milwaukee Public Museum Contributions to Biology and Geology. 83, 1-156, 1992.

Sepkoski, J. J. Jr.: A compendium of fossil marine animal genera. Bulletins of A merican Paleontology, 363, 1-563, 2002.

Shaw, A. B.: Time in Stratigraphy. McGraw-Hill, New York. 1964.

Shen, S. Z., Crowley, J. L, Wang, Y., Bowring, S. A., Erwin, D. H., and Jin, Y.-G.: Calibrating the End-Permian Mass Extinction. Science, 334, 1367-1372, 2011.

Shen, S. Z., Zhang, H., Shi, G. R., Li, W., Xie, J., and Fan, J.: Early Permian (Cisuralian) global brachiopod palaeobiogeography. Gondwana Research, 24, 104-124, 2013.

Wang, Y., Sadler, P. M., Shen, S.-Z., Erwin, D. H., Zhang, Y., and Henderson, C. M.: Quantifying the process and abruptness of the end-Permian mass extinction. Paleobiology, 40, 113-129, 2014.

Xiong, C. and Wang, Q.: Was There a Mass Extinction of Land Plants at the Permian —Triassic Boundary(PTB)?, Geological Review, 53, 9, 2007.

Xiong, C., Wang, D., Wang, Q., Benton, M. J., Xue, J., Meng, M., Zhao, Q., and Zhang, J.: Diversity dynamics of silurian-early carboniferous land plants in South china, PLoS One, 8, e75706, 2013.

Xu, H.-H.: Retrospect and prospect of a section-based stratigraphic and palaeontological database -- Geobiodiversity Database (Version 31Dec2019). Zenodo. http://doi.org/10.5281/zenodo.3667645, 2020.

Zhang, L. N., Fan, J., and Chen, Q.: Geographic distribution and palaeogeographic reconstruction of the Upper Ordovician Kuanyinchiao Bed in South China, Chinese Science Bulletin, 61, 11, 2016.

Zhang, L., Fan, J., Chen, Q., and Melchin, M. J.: Geographic dynamics of some major graptolite taxa of the Diplograptina during the Late Ordovician mass extinction in South China, GFF, 136, 327-332, 2014a.

Zhang, L., Fan, J., Chen, Q., and Wu, S. Y.: Reconstruction of the mid-Hirnantian palaeotopography in the Upper Yangtze region, South China, Estonian Journal of Earth Sciences, 63, 4, 2014b.

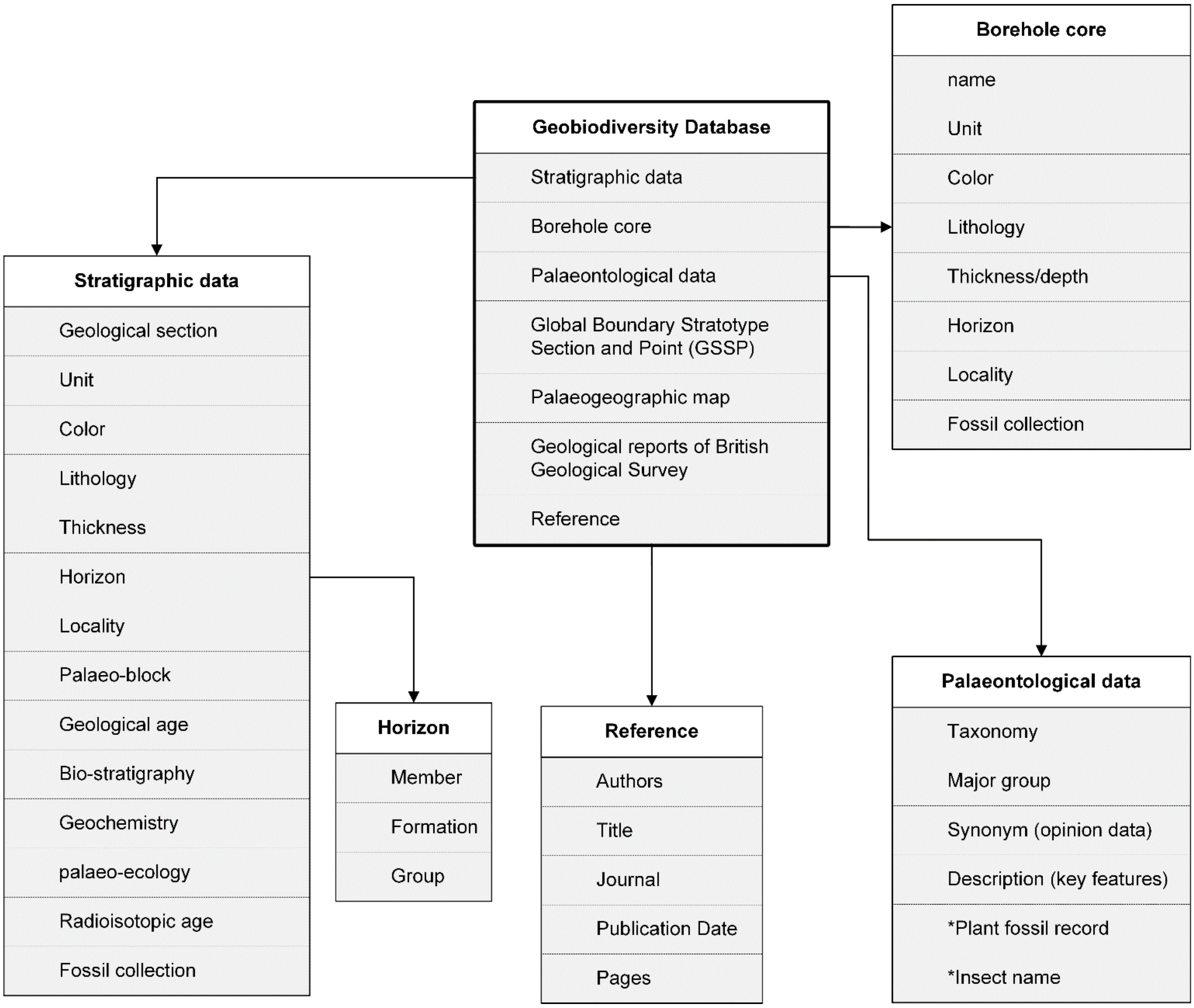

Figure 1. The data structure of the Geobiodiversity Database (GBDB). * refers the newly-added datasets.

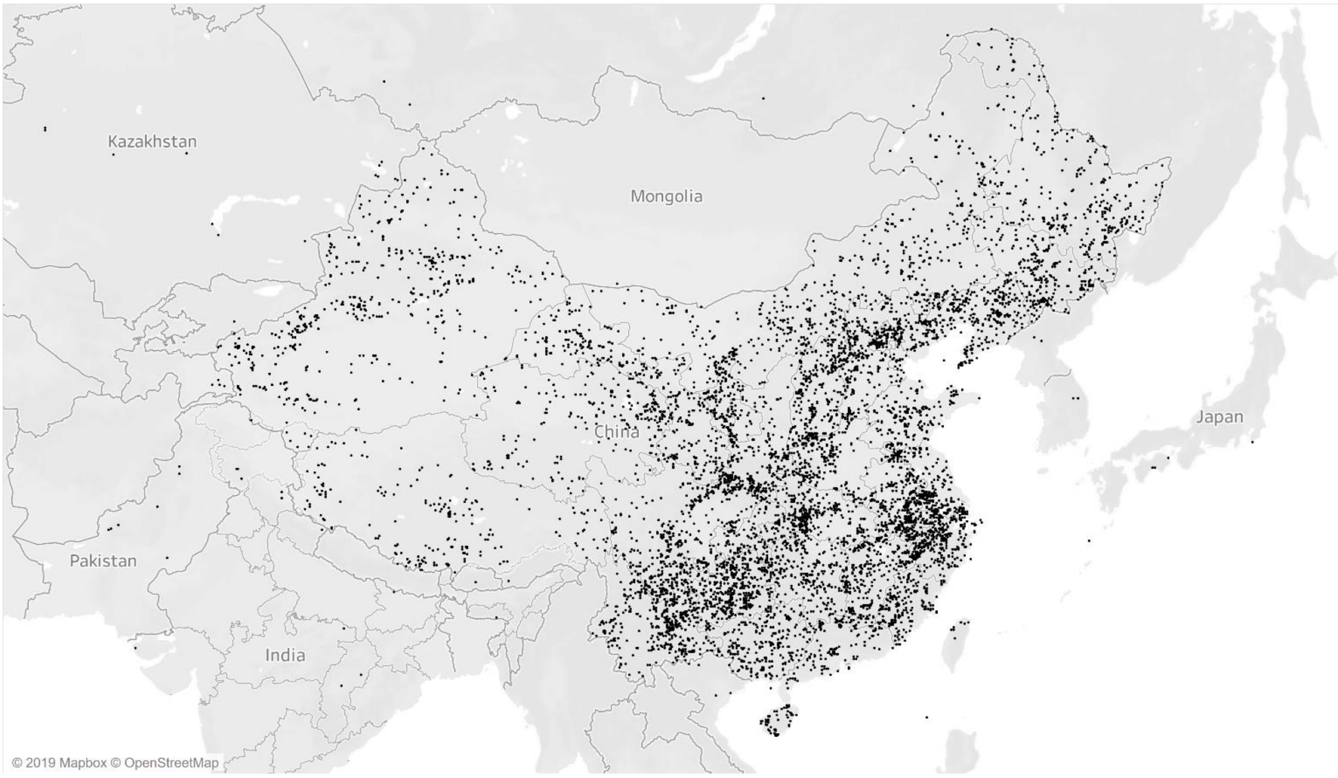

Figure 2. Regional (China-East Asia) distribution of stratigraphic and palaeontological data (2007-2018) of the Geobiodiversity Database (GBDB) (Xu, 2020). Every black dot corresponds a stratigraphic or palaeontological record of the GBDB. The map: © OpenStreetMap contributors 2020. Distributed under a Creative Commons BY-SA License.

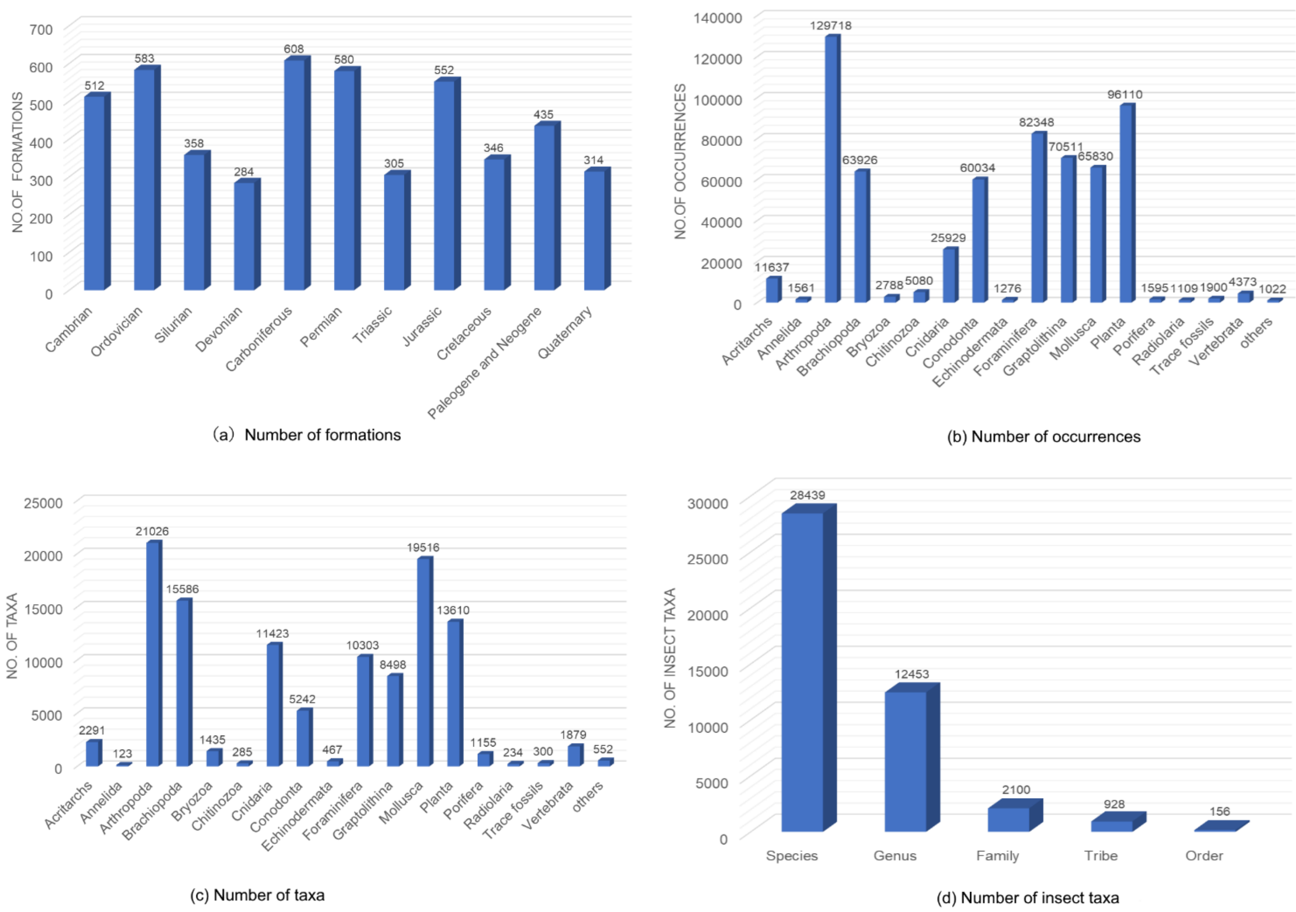

Figure 3. Histograms showing the statistic outcome of the data in Geobiodiversity Database (GBDB), the detailed numbers are shown

on every item. (a) Stratigraphic formations of different ages from China. (b, c) Fossil taxa and occurrences of different groups. (d) Newly-added taxa of the Class Insecta, these taxa are not included in the statistic outcome of the Table 1.

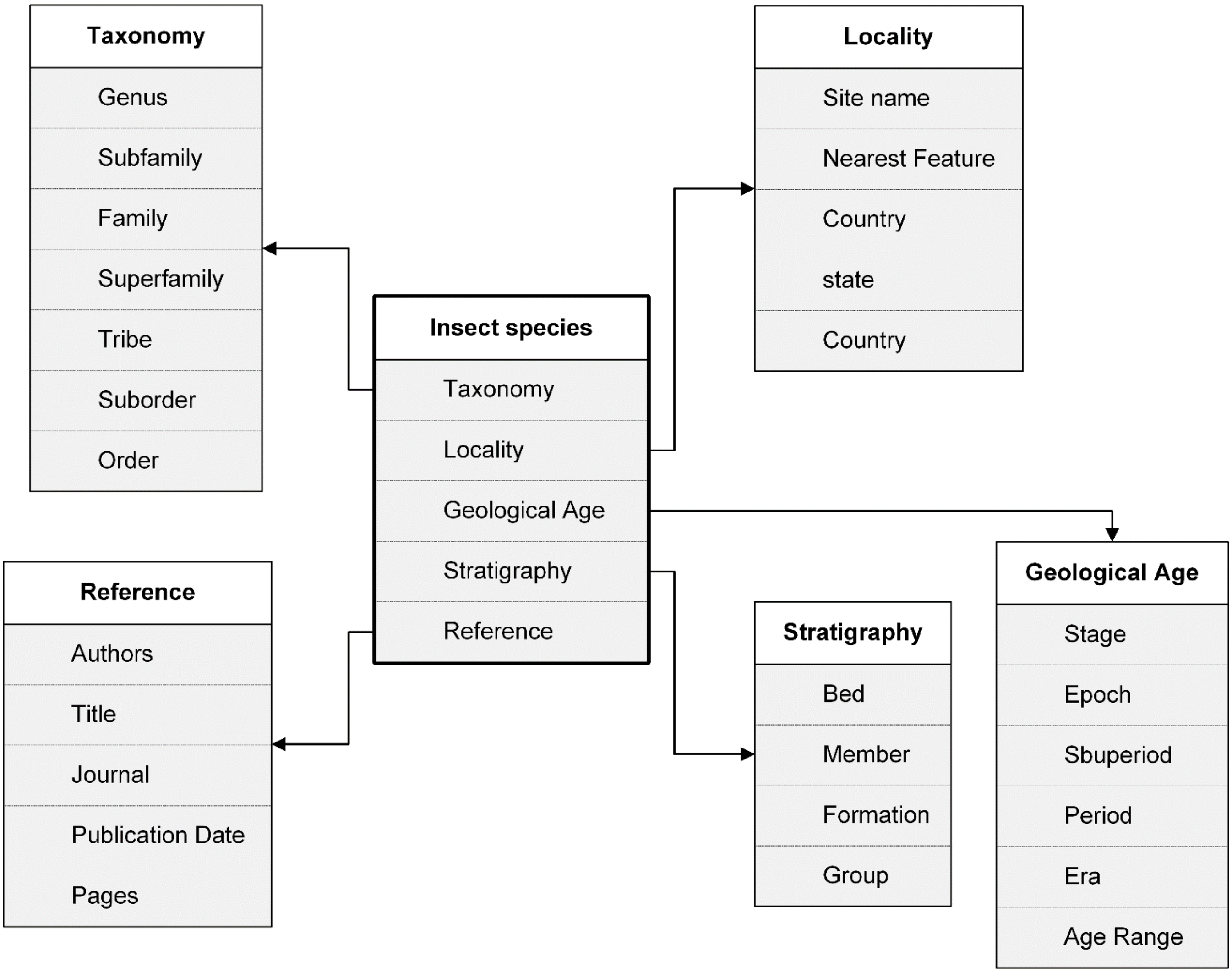

Figure 4. The data structure of insect species name dataset of the Geobiodiversity Database (GBDB) (Xu, 2020).

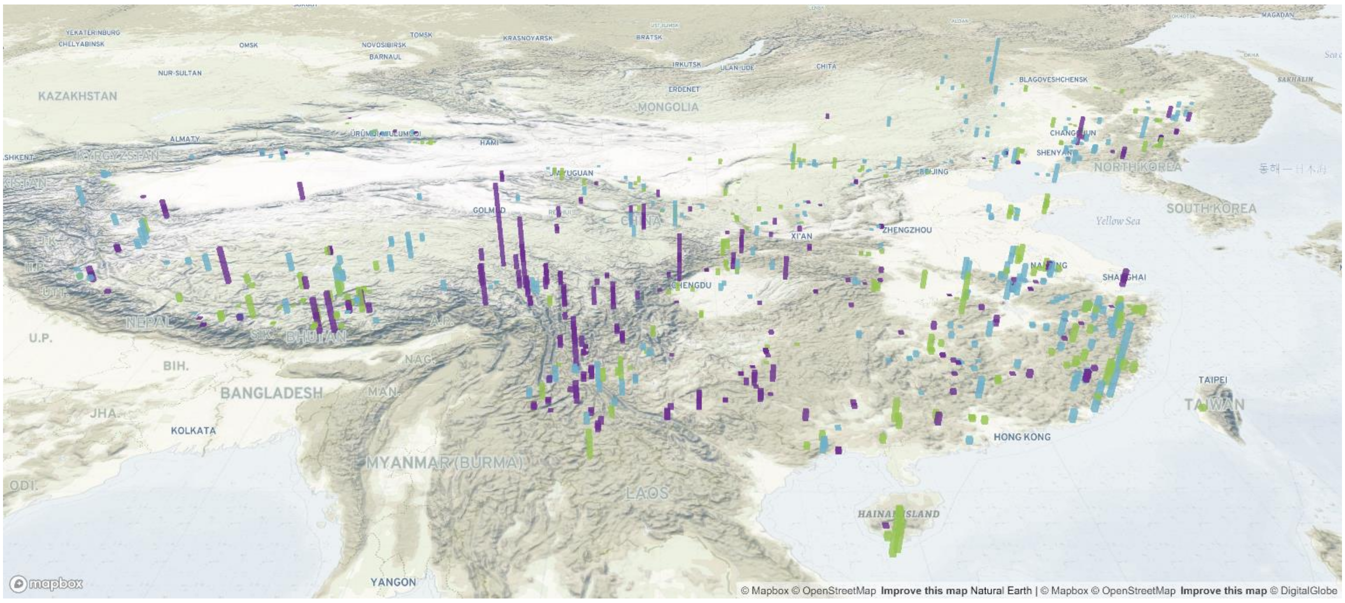

Figure 5. The screenshot of a three-dimensional bar graph visualizing the Mesozoic stratigraphic formations from China. Data are from the Geobiodiversity Database (GBDB). The colors of the bars are based on those in the International Chronostratigraphic

Chart: Triassic (pink), Jurassic (blue) and Cretaceous (light green). The map: ©OpenStreetMap contributors 2020. Distributed under a Creative Commons BY-SA License. The website of this graph is: http://167.71.205.3/3dbar/

Table 1. The comparison of the two widely-used palaeontological databases. Note that the newly-added data of terrestrial organisms, plant and insect fossil records, are not included in the GBDB statistic outcome (by November 2019).

	Paleobiology Database (PBDB)	Geobiodiversity Database (GBDB)
Type	fossil occurrence-based	section-based
No. of references	69 248	96 511
No. of taxa	388 533	113 925
No. of opinions	718 165	18 058
No. of collections	202 189	124 456
No. of occurrences	1 414 981	626 747
No. of sections	n/a	26 423
No. of formations	16 252*	4 736
No. of publications	344	45
Founded at	1998	2007
Website	https://paleobiodb.org/	http://geobiodiversity.com/

***The stratigraphic formation data of the PBDB was obtained from Prof. W. Kiessling whilst one can see these records from the portal of the PBDB.**