Paleozoic-Mesozoic Terrestrial Total Organic Carbon and Organic Carbon Isotope Database

Yaokai Tian[1], Daoliang Chu[1]*, Jiankang Lai[1], Xiang Shu[1], Cidong Zhang[1], Fangyu Cui[1], Liangyu Lou[1], Xiaokang Liu[1], Yuyang Wu[2], Qingzhong Liang[3], Xinchuan Li[3], Hong Yao[3] and Haijun Song[1]*

[1]State Key Laboratory of Geomicrobiology and Environmental Changes, School of Earth Sciences, China University of Geosciences, Wuhan 430074, China

[2]College of Marine Science and Technology, China University of Geosciences, Wuhan 430074, China

[3]School of Computer Science, China University of Geosciences, Wuhan 430074, China

Correspondence: Daoliang Chu (chudl@cug.edu.cn); Haijun Song (haijunsong@cug.edu.cn)

Abstract. Studies of the terrestrial carbon cycle commonly rely on geochemical proxies such as total organic carbon (TOC) and the organic carbon isotopic composition ($\delta^{13}C_{org}$). However, terrestrial TOC and $\delta^{13}C_{org}$ data are widely dispersed across the literature and lack a unified compilation, limiting large-scale synthesis and cross-comparison. Here, we present a global, standardized dataset of TOC and $\delta^{13}C_{org}$ measurements derived primarily~~exclusively~~ from terrestrial sedimentary facies: Paleozoic-Mesozoic Terrestrial Total Organic Carbon and Organic Carbon Isotope Database (PM-TOCI). The dataset compiles ~~66,587~~66,583 individual data points (~~49,016~~49,014 TOC and ~~17,571~~17,569 $\delta^{13}C_{org}$) from 619 publications, spanning the Devonian to Cretaceous (419-66 Ma). Each entry is accompanied by ~~34~~37 standardized metadata fields, covering geographic information, stratigraphic age, lithology, and depositional facies, thereby enabling consistent filtering, comparison, and reuse across spatial and temporal scales. This dataset is intended to facilitate future data-driven studies of terrestrial organic carbon accumulation, paleoclimate variability, source ~~-~~rock assessment, and long-term carbon cycle dynamics, as well as the link between carbon cycle and biotic evolution. The dataset is openly accessible at https://doi.org/10.5281/zenodo.20486580~~https://doi.org/10.5281/zenodo.18163858~~ (Tian et al., 2026).

1 Introduction

Carbon is a fundamental element of life and plays a central role in matter and energy exchange across the Earth system (Post et al., 1990; Falkowski et al., 2000; Prajapati et al., 2023). Terrestrial sedimentary records, particularly lacustrine and fluvial successions, provide important archives of continental paleoenvironmental evolution. Among available geochemical proxies, total organic carbon (TOC) content and the stable carbon isotopic composition of organic matter ($\delta^{13}C_{org}$) are widely used indicators. These parameters record variations in organic productivity, preservation conditions, and carbon cycle dynamics, thus providing constraints on ancient terrestrial climate and environmental conditions, including temperature, hydrological regimes, redox states, and atmospheric CO_2 levels.

TOC, commonly expressed as weight percent (wt%), quantifies the abundance of organic carbon in sedimentary rocks (Jarvie, 1991). TOC content reflects both primary productivity and the efficiency of organic matter preservation, and it is therefore a fundamental parameter for assessing the hydrocarbon generation potential of source rocks (Akande, 2012; El Nady et al., 2015). Based on TOC abundance, source rocks are conventionally classified as fair (0.5–1 wt%), good (1–2 wt%), or very good (>2 wt%) (Peters and Cassa, 1994; Beaumont and Foster, 1999). Accordingly, vertical variations in TOC within stratigraphic successions are widely used to identify organic-rich intervals and delineate high-quality source rocks. Beyond its significance in petroleum system analysis, TOC is also a widely applied proxy for paleoenvironmental reconstruction. Because sedimentary organic matter is largely derived from aquatic primary producers (e.g., plankton and algae), elevated TOC values generally indicate enhanced paleoproductivity and increased organic carbon burial rates, though this relationship is strongly modulated

by depositional setting (e.g., anoxia, restricted circulation, upwelling) and preservation conditions (Berger et al., 1989; Tyson, 2001, 2005). In addition, TOC content is sensitive to depositional redox conditions. In semi-deep to deep lacustrine settings, water-column stratification and anoxic bottom waters suppress organic matter degradation, promoting organic carbon preservation and resulting in high TOC values—conditions that are also conducive to the development of high-quality source rocks (Chen et al., 2019; Qu et al., 2019; Hou et al., 2022).

Carbon has two stable isotopes, ^{12}C and ^{13}C, and its isotopic composition is conventionally expressed as $\delta^{13}C$ (Craig, 1957). Isotopic fractionation during biological and geological processes generates systematic variations in $\delta^{13}C$ values among different carbon reservoirs. Consequently, stratigraphic records of organic carbon isotopes ($\delta^{13}C_{org}$) provide important constraints on the evolution of the ancient carbon cycle (Ahm and Husson, 2021; Canfield, 2021). Variations in $\delta^{13}C_{org}$ reflect a combination of factors, including changes in paleo-atmospheric CO_2 concentrations, primary productivity, and carbon cycling dynamics, and are widely applied in chemostratigraphic correlation and in identifying major carbon-cycle perturbations associated with Earth-system transitions and environmental crises (Meyer et al., 2011; Li and Elderfield, 2013; Bauska et al., 2015; Ji et al., 2018; Wu et al., 2021, 2023; Ansari et al., 2023). When integrated with TOC data, $\delta^{13}C_{org}$ provides a more robust framework for paleoenvironmental interpretation. The coupled behavior of TOC and $\delta^{13}C_{org}$ helps distinguish changes in productivity, organic matter preservation, and external carbon-cycle forcing, thereby improving reconstructions of terrestrial environmental evolution and carbon-climate feedbacks.

Recent years have seen increasing emphasis on data standardization, transparency, and reproducibility in Earth system science, formalized by the FAIR principles (Findable, Accessible, Interoperable, Reusable; Wilkinson et al., 2016; "FAIR Play in geoscience data", 2019). In response, numerous curated databases have been developed, including The Data Publisher for Earth & Environmental Science (PANGAEA)(https://pangaea.de/, last access: 6 January 2026), paleoclimate compilations such as the Phanerozoic sea surface temperature database (Song et al., 2019; Judd et al., 2022), paleobiological resources such as the Paleobiology Database (https://paleobiodb.org/, last access: 30 December 2025) and the Global Acritarch Database (Shu et al., 2025), sedimentary geochemistry databases including the Sedimentary Geochemistry and Paleoenvironments Project (SGP)(Farrel et al., 2021), the international study of the marine biogeochemical cycles of trace elements and isotopes (GEOTRACES)(SCOR Working Group, 2007), the Deep-Time Marine Sedimentary Element Database (DM-SED)(Lai et al., 2025), and the Deep-time Sediment Nitrogen Isotopes in Marine Systems (DSMS-NI)(Du et al., 2025). These databases have substantially advanced data-driven research on Earth system evolution, particularly in marine settings.

In contrast, terrestrial stratigraphic geochemical data remain poorly integrated at a global scale. Existing databases are predominantly marine-focused, while terrestrial efforts are limited and restricted exclusively to stable carbon isotopes ($\delta^{13}C_{org}$), with relatively modest data volumes—e.g., the ISOORG database with ~6,900 records from ten types of terrestrial organic matter (primarily sediment, coal, and wood; Nordt et al., 2016), a Paleozoic compilation of ~1,000 records from C_3 plant remains (Wang et al., 2021), and an earlier dataset of 2,148 records from terrestrial organic matter (mainly plant fossils, coal, and bulk organics; Strauss and Peters-Kottig, 2003). Critically, no global databases currently exist for terrestrial total organic carbon (TOC), leaving a major gap in comprehensive geochemical synthesis. Terrestrial records are otherwise scattered across basin-scale studies, individual publications, and unpublished reports, further complicated by strong local and regional controls. This fragmentation hampers large-scale comparisons, synthesis of terrestrial paleoenvironments, and evaluation of terrestrial source rock development. To address this gap, we present the Paleozoic-Mesozoic Terrestrial TOC and δ^{13}Corg Database, which compiles ~~66,587~~66,583 TOC and $\delta^{13}C_{org}$ measurements from global terrestrial strata. This dataset provides a standardized foundation for studies of terrestrial paleoenvironmental evolution, paleoclimate change, carbon cycling, and marine–terrestrial system interactions. In the following sections, we describe the data sources, screening criteria, field definitions, and spatiotemporal coverage of the dataset, and discuss its potential applications and limitations.

2 Data selection and quality control

This section details the criteria applied for screening and processing sample information, including geographic location, age, lithology, depositional facies, geochemical parameters, and data sources. Each field was evaluated against a set of standardized criteria to ensure consistency in data quality (Fig. 1).

The dataset includes the following categories of information: basic sample details (SampleID, SectionName, SampleDepth), spatial information (Region, ModLat, ModLon, PalaeoLat, PalaeoLon), chronostratigraphic constraints (Age, Period, Epoch, Stage), stratigraphic and lithological attributes (Lithology, Stratigraphic Unit~~Strata~~, Facies), geochemical data (total organic carbon (TOC) and organic carbon isotope composition ($\delta^{13}C_{org}$)), and data sources (References). Detailed field names and descriptions for each column in the dataset are provided in Table 1.

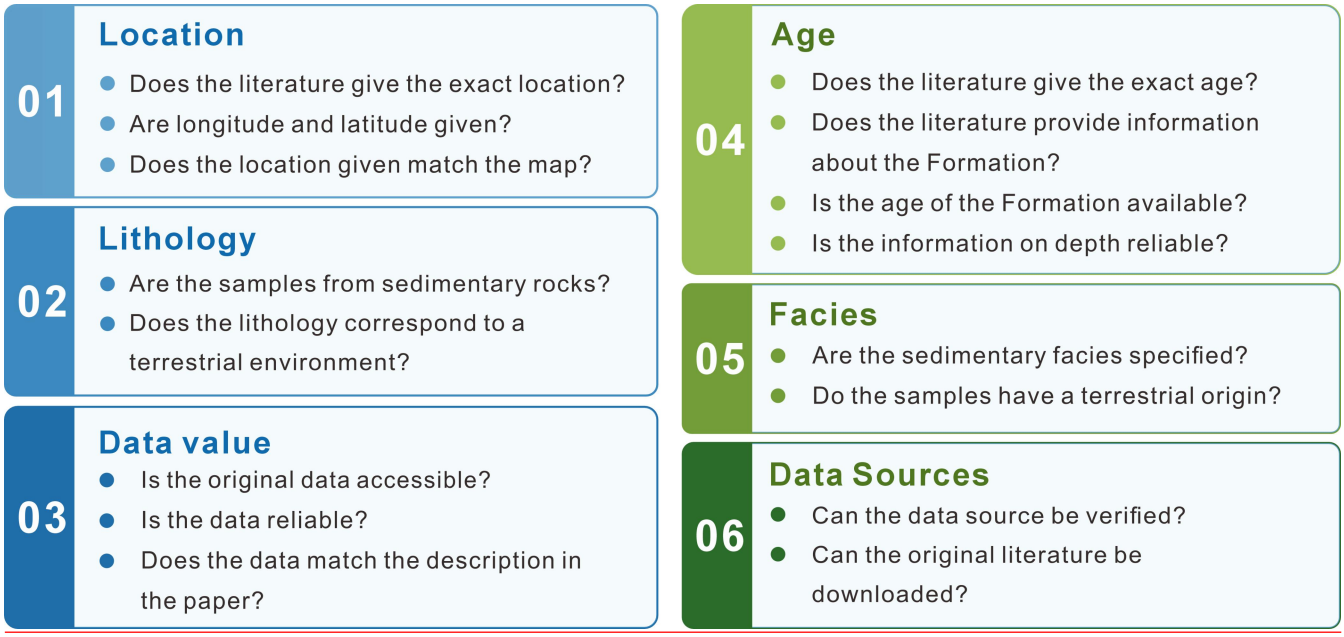

Figure 1. Data filtering and screening workflow

2.1 Location fields

Regarding the location fields, the dataset includes the SampleID, SampleName, SampleDepth, Region, SectionName, ModLat, ModLon, PalaeoLat, and PalaeoLon fields. In the dataset, each sample is assigned a unique SampleID which serves as its identification code. SampleName corresponds to the sample name or number given in each reference, used for cross-referencing with the original published data; since many publications do not provide sample numbers, this field is not populated for a large portion of the data. SampleDepth refers to the height or depth of the sample within an outcrop or drill hole, uniformly recorded in this field. In many studies, sampling depths are not explicitly reported but can be inferred from lithological columns. In such cases, depths were manually digitized from published figures using WebPlotDigitizer, and the extracted values were recorded to two decimal places. Whenever heights (or depths) were reported in feet or centimetres, we converted them to metres. Region denotes the sampling location. Because most samples are from onshore sites, with only a few derived from ocean drilling programs, the vast majority of regions can be assigned to a specific country. SectionName is the name of the sampling section or borehole. Modern latitude and longitude (ModLat and ModLon) are crucial for recovering the sample's location information. Although some publications provide exact coordinates, most do not directly give coordinate locations but instead mark the section position on a map or merely provide the section name. For publications that mark the section location on a map, we used Google Maps to determine the precise coordinates. For those that only provided the section name, we referred to other studies conducted in the same section to determine its location. Data for which coordinates could not be reliably determined were not included in the dataset. Latitude and longitude coordinates are in decimal degrees, rounded to

two decimal places, with positive values indicating north latitude and east longitude and negative values indicating south latitude and west longitude. The coordinate reference system is WGS84 (World Geodetic System). Paleolatitude and paleolongitude (PalaeoLat and PalaeoLon) were reconstructed from sample ages and present-day coordinates using Time Machine (https://deeptime.world/TimeMachine/, last access: 30 December 2025).

2.2 Age field

Only a small fraction of source publications report absolute numerical ages for individual samples. To ensure consistency and reproducibility in the Age field, we implemented a tiered and traceable age-assignment workflow that integrates rule-based automated calculations with documented expert curation. Where absolute ages are available for multiple horizons, sample ages were calculated by linear interpolation assuming a constant sedimentation rate. However, this simplification ignores natural sedimentation variability and introduces some age uncertainty. Users should be aware of potential misplacement into adjacent time bins, especially in high-resolution applications. Otherwise, samples were assigned to formation-level or finer stratigraphic units, and approximate ages were determined using published stratigraphic frameworks, prioritizing high-precision astrochronological and zircon U–Pb constraints as well as authoritative regional syntheses. For samples lacking depth information, ages were estimated by evenly distributing them within the temporal bounds of the corresponding stratigraphic unit following a predefined rule. The Age accuracy field documents the method used for age assignment, distinguishing between linearly interpolated ages and expert-estimated ages based on published stratigraphic constraints. Age assignments are resolved at the geological stage level, and samples that could not meet this minimum temporal resolution were excluded. All ages are standardized to the International Chronostratigraphic Chart (v2024-12) (https://stratigraphy.org/, last access: 30 December 2025), with the specific version documented in the dataset metadata to ensure reproducibility and future compatibility.

2.3 Lithology field

In the Lithology field, recorded rock types are dominated by shale, mudstone, siltstone, sandstone, and other clastic sediments. Although many source publications provide more detailed lithological descriptions, we prioritized fidelity to the original literature by retaining the reported terminology rather than simplifying it into broader categories. Where lithology was not explicitly stated, rock types were typically determined from accompanying stratigraphic columns. In rare cases where no lithological information was available, lithology was inferred by cross-referencing other studies from the same section; records for which lithology could not be confidently established were excluded from the dataset.

For stratigraphic units, the formation or group name was recorded as reported in the source literature. Where the stratigraphic unit did not correspond to a formal formation or group, the original stratigraphic term was retained. Depositional facies are predominantly lacustrine, fluvial, and deltaic, with minor occurrences of coastal transitional, swamp, and lagoonal facies. Because organic matter in these transitional environments is likely largely terrestrially derived, data from such facies were also included in the dataset.

2.4 Data field

For the geochemical data compiled in the dataset, TOC and $\delta^{13}C_{org}$ values were primarily obtained from tables or supplementary materials in the original publications. When numerical data were presented exclusively in graphical form (e.g., scatter plots or line charts), values were manually extracted using WebPlotDigitizer and subsequently rounded to two decimal places. To ensure data reliability, extracted values were cross-checked against axis scales and accompanying textual descriptions where available. Data points that were clearly erroneous, inconsistent with the reported trends, or contradicted the descriptions in the source publications were excluded from the dataset. This data acquisition and screening procedure was applied consistently

across all sources to maximize accuracy, transparency, and reproducibility.

2.5 Data source

All data were compiled from publicly accessible sources, including peer-reviewed journal articles, books, and academic theses. To distinguish between peer-reviewed and non-peer-reviewed datasets, a filter flag "Peer-reviewed (yes/no)" has been added to the database. Users may optionally exclude non-peer- reviewed entries. It should be noted that theses and conference proceedings constitute only a small fraction (<5%) of the total entries; they have been retained for completeness, but the new filter flag enables their straightforward removal when desired. For each record, the reference field documents the first author, publication title, publication year, and journal name, or the corresponding bibliographic information for books and theses. To ensure machine readability and interoperability, special characters, non-standard symbols, and garbled text were systematically removed during data processing, while preserving the original bibliographic content.

Table 1. Database fields and descriptions

Field name	Description of the field (units)
Location Fields	
Sample-ID	Unique sample identification code
Sample-Name	Sample name as given in the reference
Section-Name	Well name (for boreholes) or section name
Region	Country or region where the sample was collected
Location	Specific sampling locality
Sample-Depth	Stratigraphic depth or height (m)
ModLat	Modern latitude of the sampling site, rounded to two decimal places; negative values indicate the southern hemisphere (decimal degrees)
ModLon	Modern longitude of the sampling site, rounded to two decimal places; negative values indicate the western hemisphere (decimal degrees)
Paleolat	Paleolatitude of the sampling site, rounded to two decimal places; negative values indicate the southern hemisphere (decimal degrees)
Paleolon	Paleolongitude of the sampling site, rounded to two decimal places; negative values indicate the western hemisphere (decimal degrees)
Age Fields	
Age	Absolute age (if available)
Period	Geological period, following the International Chronostratigraphic Chart (v2024-12)
Stage	Geological stage, following the International Chronostratigraphic Chart (v2024-12)
Stratigraphy	
Lithology	Lithological name of the sample
Stratigraphic Unit	Name of the stratigraphic unit
Facies	Depositional environment (e.g., lacustrine or fluvial)
Data Fields	
Total Organic Carbon	Total organic carbon content, reported as mass fraction and rounded to two decimal places (wt%)
Carbon Isotope	$\delta^{13}C_{org}$ value, rounded to two decimal places (‰)
Data Source	

Reference	Source publication of the data, including peer-reviewed journals, theses, books, and other publicly available materials

3 Data overview

3.1 Data distribution

This dataset focuses on terrestrial organic carbon content and carbon isotope geochemistry. It integrates data from 619 publicly available sources, primarily peer-reviewed journal articles, with additional contributions from monographs, conference proceedings, and academic theses. The current version comprises ~~66,587~~66,583 data points, including ~~49,016~~49,014 TOC measurements and ~~17,571~~17,569 $\delta^{13}C_{org}$ values, spanning more than 340 million years from the Devonian to the Cretaceous. These data are primarily derived from bulk samples, with a minor subset of 1,572 data points originating from terrestrial organic matter (e.g., wood, coal, charcoal, cuticles), all of which are explicitly annotated within the database. The data have global geographic coverage, encompassing North America, South America, Europe, Asia, Africa, Oceania, and Antarctica. The underlying literature was published between 1988 and 2025, reflecting a substantial increase in research output over time. Both the annual number of publications and the volume of compiled geochemical data exhibit pronounced growth, particularly during the past decade (Fig. 2).

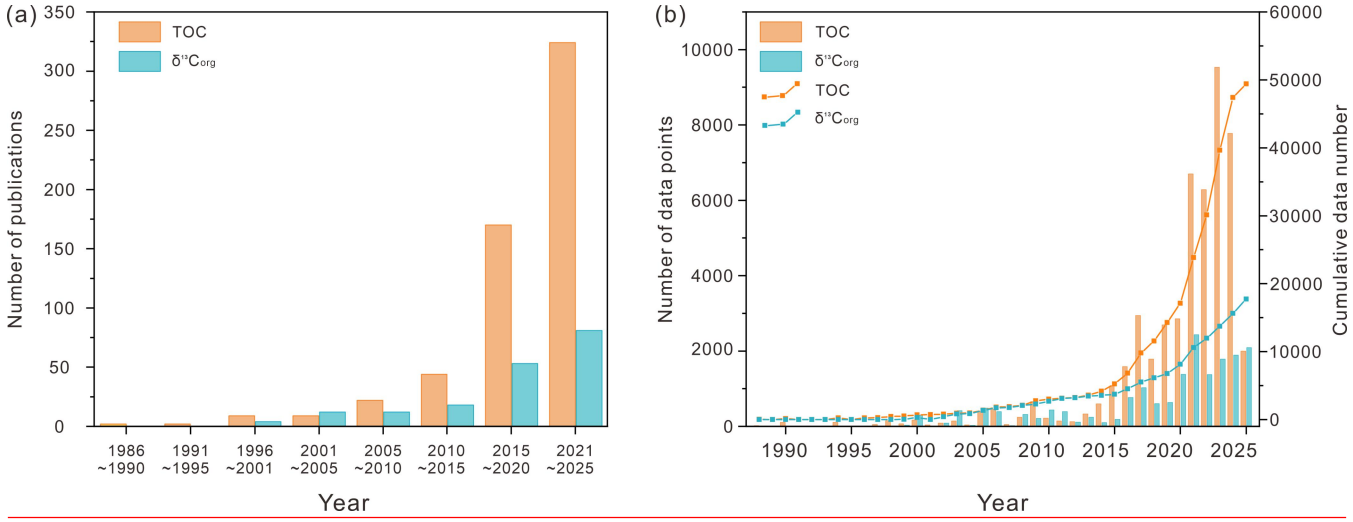

Figure 2. Distribution of publication year of source references and statistics of compiled data records. (a) Distribution of the source literature by publication year. (b) Annual and cumulative numbers of data records included in the dataset; bars indicate the number of records compiled per year, and the line shows the cumulative total.

From a geographic perspective, TOC data exhibit a global distribution, with the highest densities in Asia and Europe, moderate coverage in North and South America, relatively sparse records in Africa and Australia, and the fewest data points in Antarctica. The spatial distribution of $\delta^{13}C_{org}$ data broadly mirrors that of TOC, with records concentrated in Europe, Asia, Australia, and North America, but with notably limited coverage in South America and Africa (Fig. 3). Overall, the compiled dataset shows clear spatial heterogeneity, with pronounced differences in data density among regions.

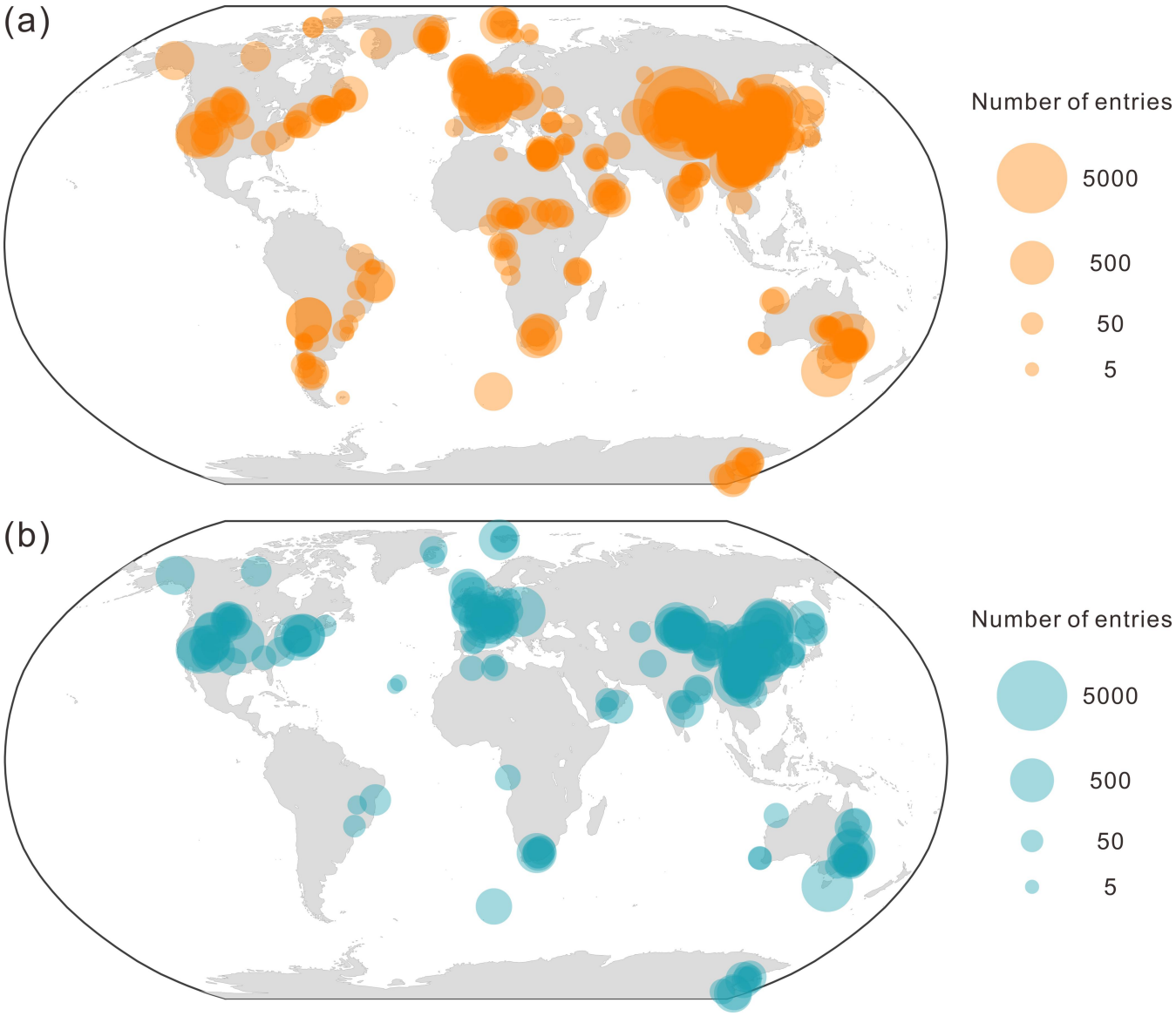

Figure 3. Bubble chart showing the modern geographic distribution and sample abundance of the compiled dataset. Each bubble represents a unique section or borehole. (a) Global distribution of TOC data. (b) Global distribution of $\delta^{13}C_{org}$ data.

3.2 Temporal distribution

Although vascular plants began to colonize land during the Silurian, the earliest records included in this dataset date to the Devonian (~419 Ma). The absence of Silurian data likely reflects a combination of geological and preservational factors rather than a true lack of terrestrial organic matter. First, continental land area during the Silurian was relatively limited, and well-developed terrestrial depositional environments (e.g., fluvial, lacustrine, and swamp systems) were less widespread (Ziegler et al., 1977; Golonka et al., 2023), reducing the availability of suitable sedimentary archives. Second, early terrestrial strata are more susceptible to later tectonic modifications, including burial, metamorphism, and erosion, which may have substantially reduced their preservation and accessibility (Peters and Husson, 2017). Third, sparse vegetation cover and limited soil development on early land surfaces likely enhanced surface erosion and hindered the accumulation and long-term preservation of terrestrial sediments containing organic carbon (Algeo and Scheckler, 1998; Xue et al., 2022). Together, these factors likely contribute to the absence of Silurian records in the compiled dataset. A different set of considerations applies to the choice of the upper boundary at 66 Ma. First, the Cenozoic Era has already been extensively studied, with numerous well-established, publicly available compilations covering its organic geochemical and isotopic records. Second, our primary objective is to focus on the Paleozoic and Mesozoic intervals, which remain comparatively under-explored in terms of

integrated TOC-carbon isotope databases for terrestrial sections. Third, including Cenozoic data would dramatically increase the total number of entries (given the much larger volume of published Cenozoic studies), potentially overwhelming the current scope and requiring disproportionate curation efforts. Fourth, the Cenozoic benefits from a wider array of high-resolution proxy indicators that are not available for older periods; integrating these would introduce methodological heterogeneity that could compromise the internal consistency of the database. Therefore, we deliberately restricted the temporal coverage to the 419–66 Ma.

Figure 4 illustrates the temporal distribution of data records by geological period. For TOC data, Paleozoic records account for approximately 43.2% of the total. Devonian data are limited, followed by a moderate increase in the Carboniferous, although overall data volumes remain relatively low. A notable peak occurs from the late Carboniferous to the early Permian, culminating in a pronounced increase in the Permian, which contributes the largest proportion (30.6%). Mesozoic records comprise 56.8% of the dataset and exhibit generally higher data abundance than Paleozoic records, with a more even distribution among periods despite some inter-period variability, including several prominent peaks in the Ladinian (Triassic), the Toarcian–Aalenian (Jurassic), and the Aptian and Turonian (Cretaceous). For $\delta^{13}C_{org}$ data, Paleozoic records represent 28.5% of the total. Data from the Devonian and Carboniferous are sparse to nearly absent, with a marked increase beginning in the late Carboniferous to early Permian. Mesozoic records dominate the dataset (71.5%), exhibiting both higher overall abundance and a relatively more uniform temporal distribution, although substantial variability is still evident among individual periods. Overall, data density varies considerably over time, reflecting pronounced temporal heterogeneity in the compiled dataset.

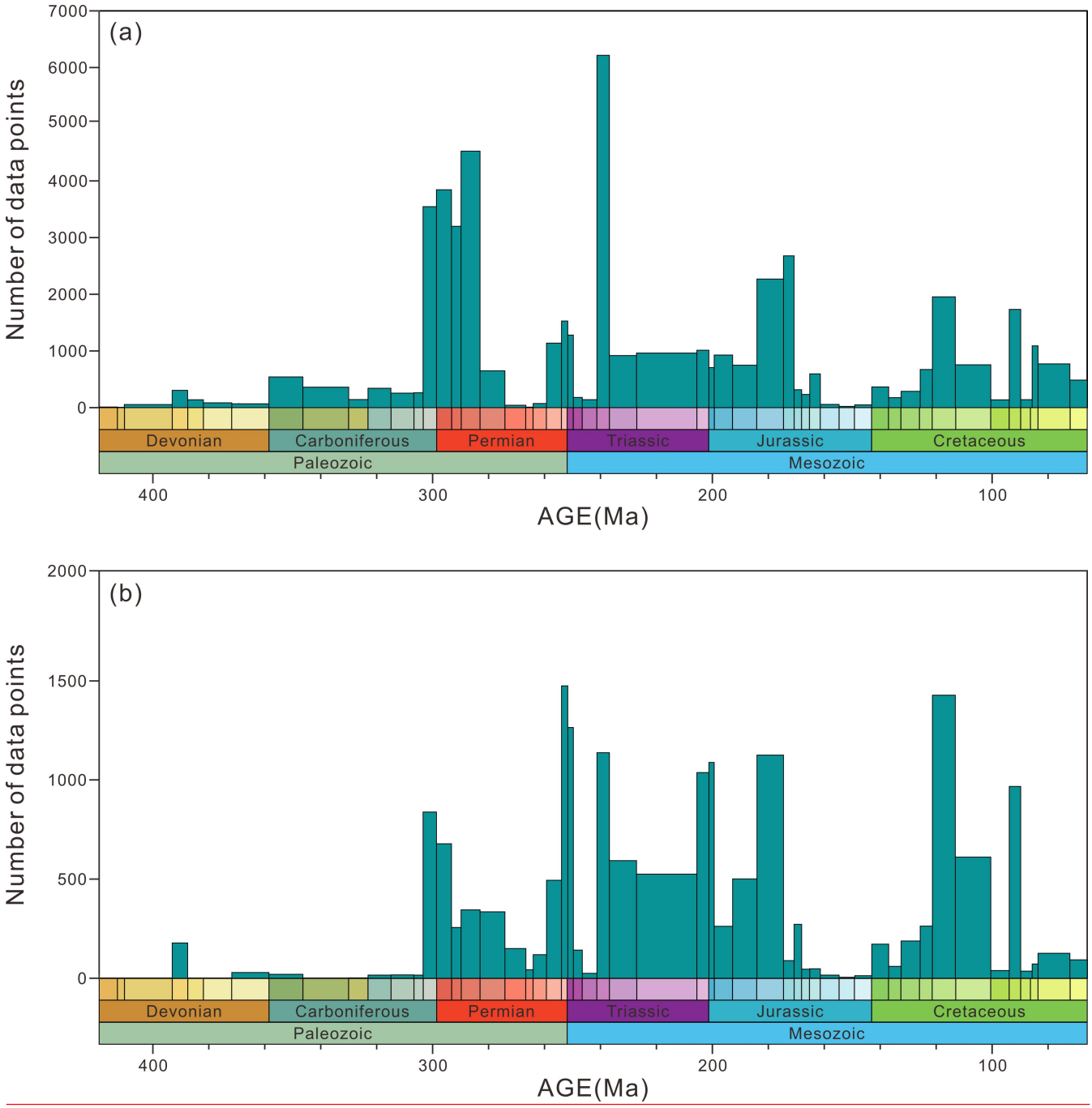

Figure 4. Temporal distribution of data by geological period. (a) Age distribution of TOC data. (b) Age distribution of $\delta^{13}C_{org}$ data. The colorful boxes on the x-axis represent the stratigraphic framework; see details in the International Chronostratigraphic Chart (v2024-12), (https://stratigraphy.org/, last access: 30 December 2025).

The temporal distribution of $\delta^{13}C_{org}$ values from the Devonian to the Cretaceous is shown in Figure 5. Overall, $\delta^{13}C_{org}$ values range from approximately −45‰ to −15‰ (VPDB), with most data concentrated between −30‰ and −20‰. The LOWESS smoothing curve indicates long-term fluctuations through the Paleozoic and Mesozoic, including a gradual positive shift from the Devonian to Carboniferous, a pronounced negative excursion during the Permian, and relatively stable but variable values throughout the Triassic to Cretaceous. Several short-term negative excursions are also evident, particularly during the Permian and Early Cretaceous intervals, reflecting major perturbations in the global carbon cycle.

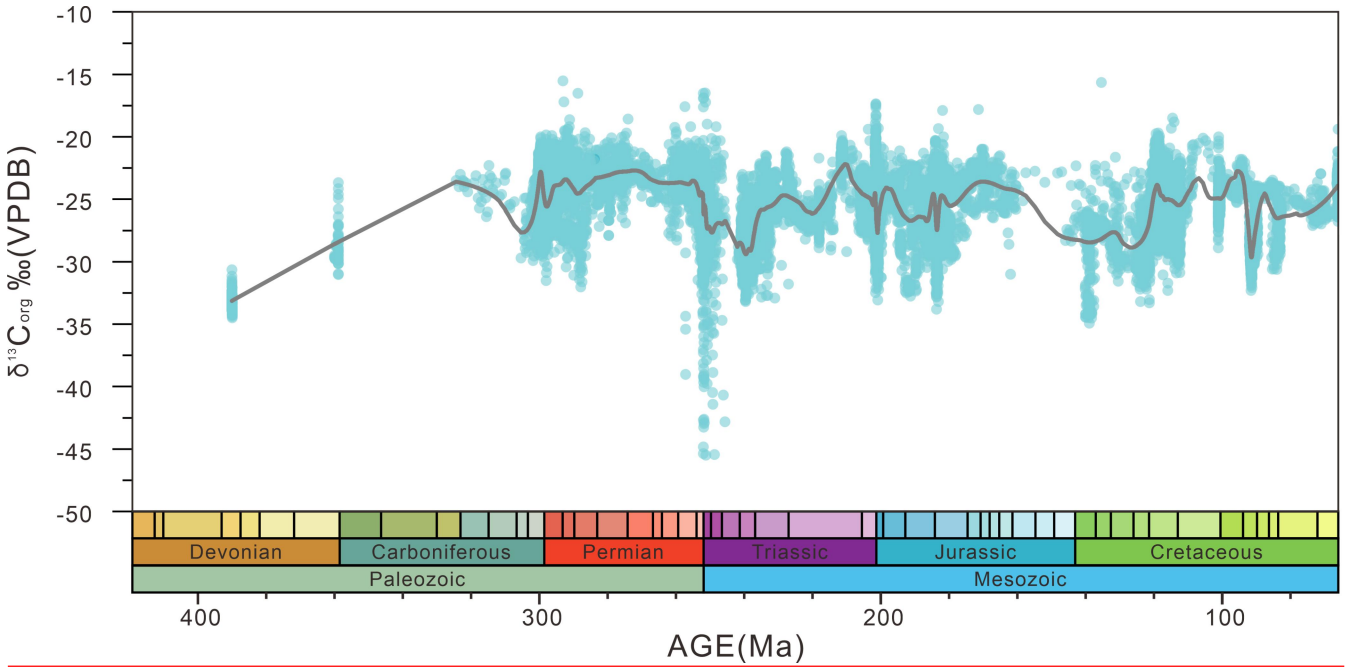

Figure 5. Temporal distribution of $\delta^{13}C_{org}$ values from the Devonian to the Cretaceous. Blue dots represent individual measurements, and the purple line shows the LOWESS smoothing curve illustrating long-term secular trends and short-term carbon isotope excursions through the Paleozoic and Mesozoic.

3.3 Depositional facies characteristics

Regarding the depositional facies, lacustrine facies dominate the terrestrial TOC dataset across all geological periods, accounting for approximately 73.6% of the total records. The number of lacustrine TOC records increases from the Devonian through the Carboniferous and peaks in the Permian. Although absolute record numbers decrease during the Triassic, Jurassic, and Cretaceous, the relative contribution of lacustrine facies continues to increase. Among the remaining facies, deltaic and fluvial environments are consistently represented, whereas the remaining~~other~~ facies types contribute only minor proportions. For the $\delta^{13}C_{org}$ dataset, lacustrine facies likewise constitute the dominant category, representing approximately 56.3% of the records and showing an overall increasing trend through time. Fluvial facies (including alluvial plain settings) account for relatively higher proportions during the Permian and Triassic but contribute less in other periods. Records from all remaining~~other~~ depositional facies are generally limited (Fig. 5).

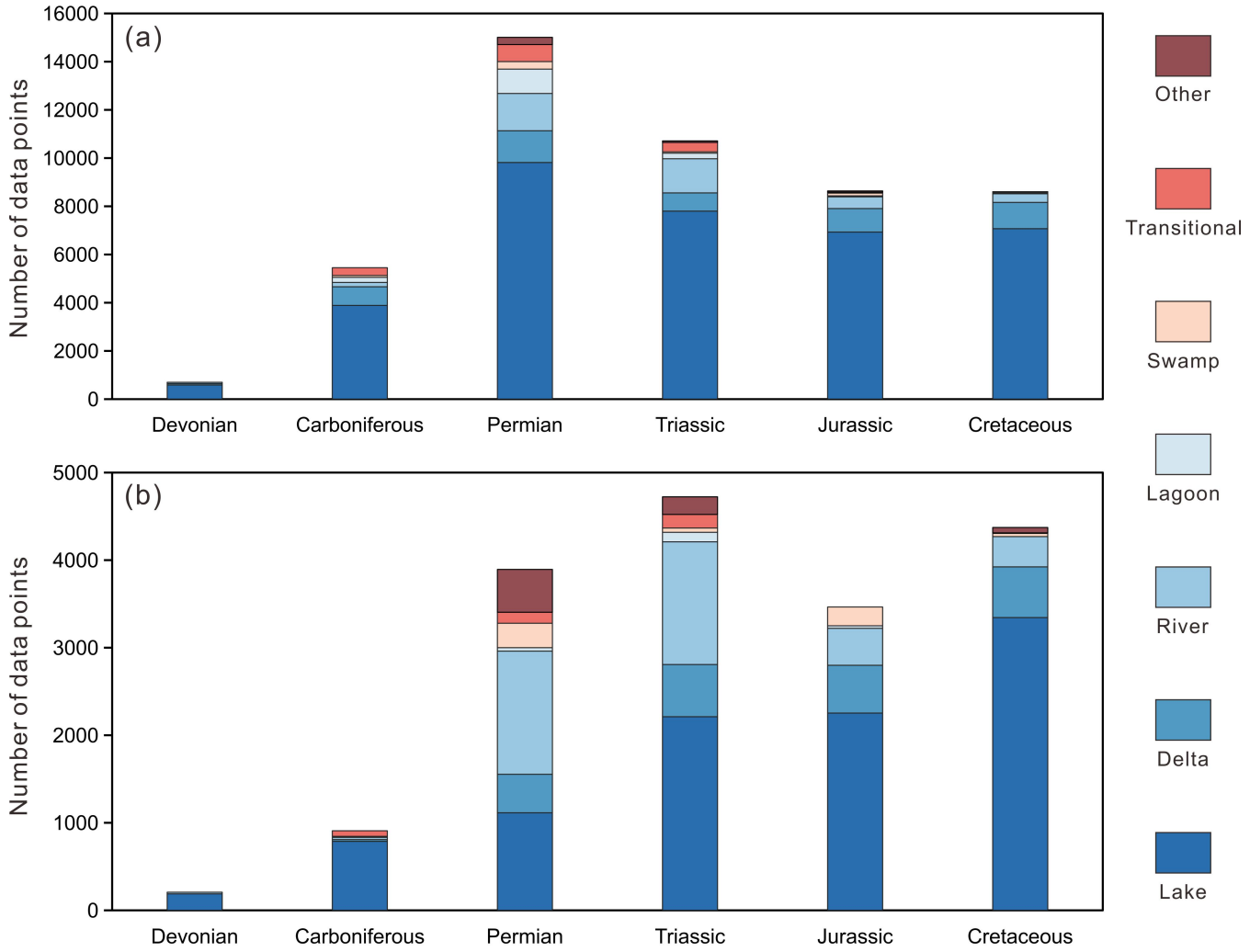

Figure 56. Depositional facies distribution of the compiled dataset. (a) Distribution of TOC data across depositional facies by geological period. (b) Distribution of $\delta^{13}C_{org}$ data across depositional facies by geological period. The 'Other' facies category includes paleosol, glacial, and other continental settings that are only described as 'continental' in the text.

3.4 Paleogeographic distribution

As shown in Fig. 6, terrestrial TOC records from the Devonian are mainly concentrated in low paleolatitudes (30° N–30° S), although the total number of records is limited. During the Carboniferous, TOC records remain largely confined to low-paleolatitude regions but become more widely distributed across both intracontinental and coastal settings of paleocontinents, accompanied by a clear increase in data abundance relative to the Devonian. In the Permian, records are distributed across low- to mid-paleolatitudes (45° N–45° S), covering paleocontinental interiors and adjacent marine–continental transitional zones, with additional occurrences in high-paleolatitude regions of the Southern Hemisphere (60° S–90° S). During the Triassic, records from mid-paleolatitudes become more prominent, and data from intracontinental settings account for a larger proportion of the dataset. Jurassic records are predominantly located in mid- to high-paleolatitude intracontinental and coastal regions. By the Cretaceous, TOC data are widely distributed across low- and mid-paleolatitude zones, with a marked increase in records from intracontinental environments. The paleogeographic distribution of $\delta^{13}C_{org}$ data broadly mirrors the latitudinal pattern observed for TOC, but is characterized by a substantially lower number of records overall (Fig. 6).

Overall, the compiled dataset exhibits pronounced spatial heterogeneity in paleogeographic coverage. Records are unevenly distributed among latitudinal bands, paleocontinental settings, and geological periods, with a clear sampling bias toward the

Northern Hemisphere. Mid-latitude regions of the Northern Hemisphere (30° N–60° N) are the most densely sampled, accounting for 54.4% of all records (Fig. 7), whereas the Southern Hemisphere and high-latitude regions are comparatively underrepresented. These characteristics indicate that, while the dataset provides robust coverage for specific regions and intervals, spatial biases should be considered when interpreting large-scale paleogeographic patterns.

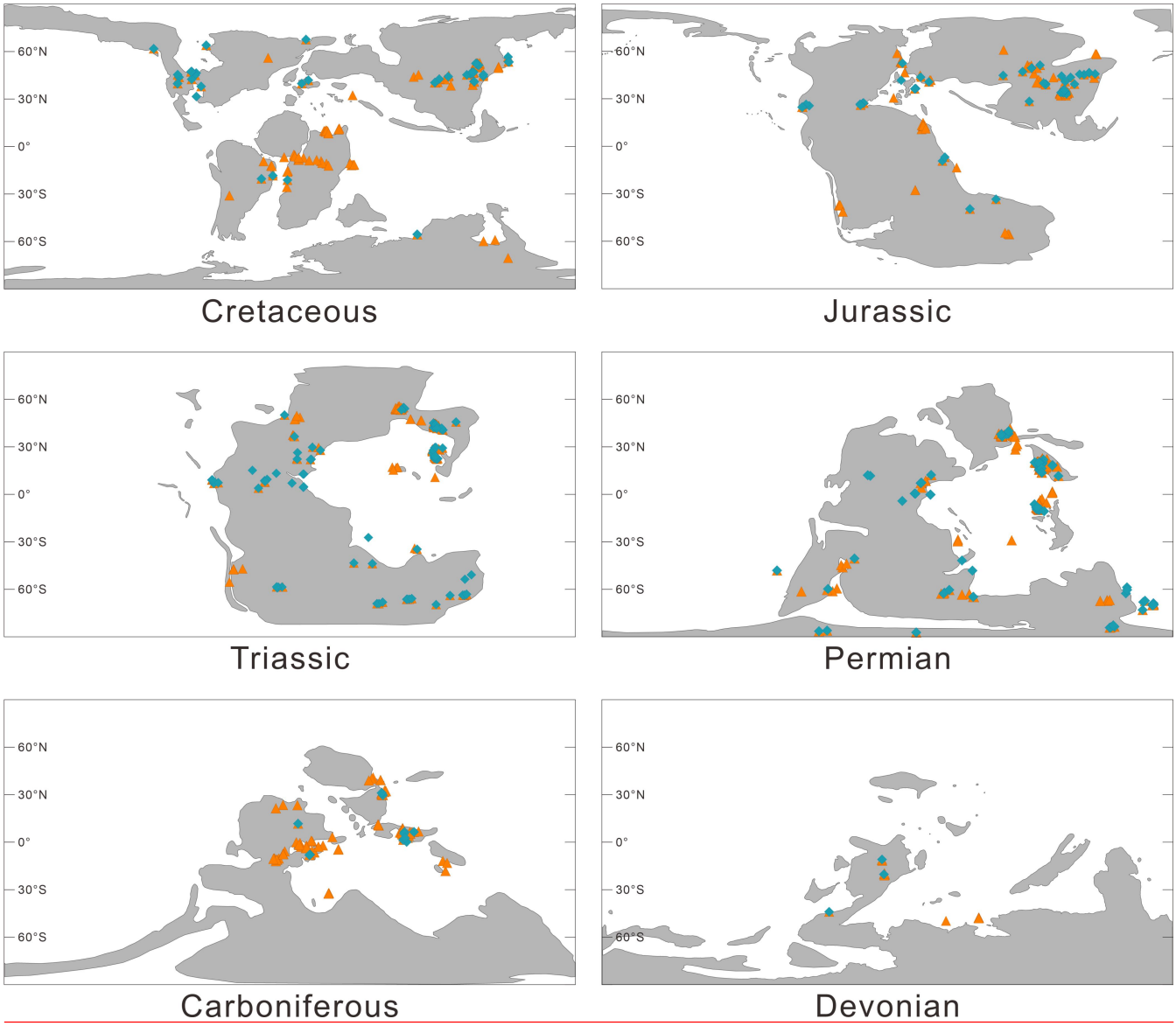

Figure 67. Paleogeographic distribution of the compiled dataset. Orange triangles denote the paleogeographic locations of TOC records, while ~~pink~~teal diamonds represent the paleogeographic locations of $\delta^{13}C_{org}$ records. The base paleogeographic map is adapted from Scotese and Wright (2018).

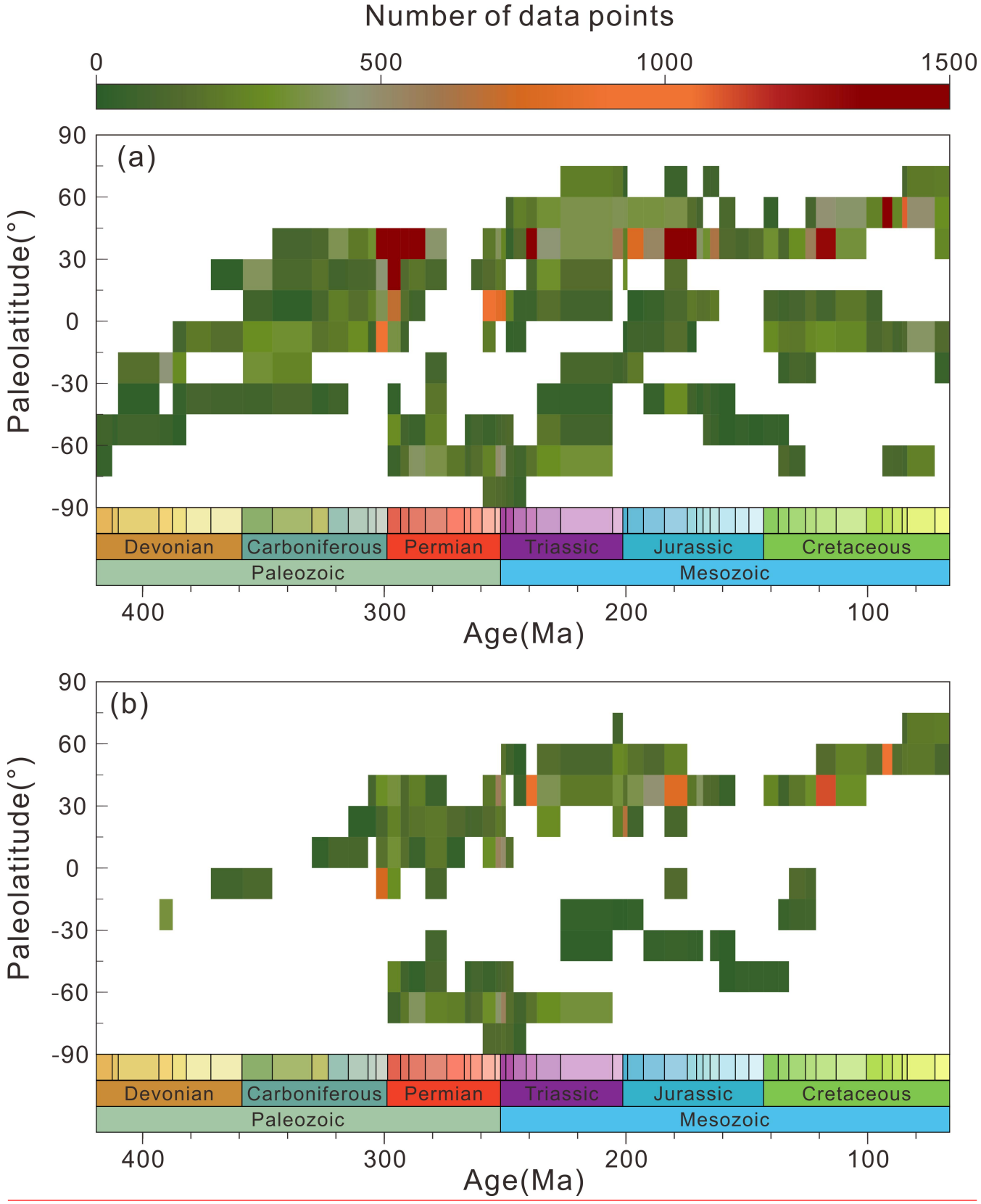

Figure ~~7~~8. Spatiotemporal distribution of the compiled data. (a) Distribution of TOC records over time and across latitudinal bands. (b) Distribution of $\delta^{13}C_{org}$ records over time and across latitudinal bands. The colorful boxes on the x-axis represent the stratigraphic framework; see details in the International Chronostratigraphic Chart (v2024-12), (https://stratigraphy.org/, last access: 30 December 2025).

4 Data availability

The dataset is openly accessible at Zenodo (https://doi.org/10.5281/zenodo.20486580,~~https://doi.org/10.5281/zenodo.18163858,~~ Tian et al., 2026). A static version of the dataset is archived in the Geobiology Database (http://geobiologydata.cug.edu.cn/, last access: 30 December 2025). The database will be continuously updated and expanded, and we welcome collaboration with authors of existing and future data compilations to facilitate the incorporation of newly published data.

5 Code availability

The software tools used in this study are available from the following sources: WebPlotDigitizer can be downloaded from https://github.com/ankerl/WebPlotDigitizer/releases (last access: 30 December 2025). Time Machine can be downloaded from https://github.com/sc66cc/TimeMachine/releases (last access: 30 December 2025).

6 Usage instructions

This dataset establishes the first systematic data platform for global terrestrial organic carbon research. When using it, attention should be paid to its inherent limitations and scope of application. The data are primarily derived from published literature, with spatial coordinates, stratigraphic ages, and geochemical indicators (TOC and $\delta^{13}C_{org}$) for some samples obtained through literature review and digital extraction. Their accuracy depends on the completeness of the original sources and the extraction methods, requiring careful evaluation in high-resolution comparisons. The data exhibit significant heterogeneity in both temporal and spatial distribution, reflecting, to some extent, current research focus, outcrop availability, and preservation conditions. Caution should be exercised regarding potential sampling bias when conducting global or cross-epoch integrated analyses. This dataset is primarily suited for macro-scale studies such as large-scale paleoenvironmental reconstruction, cross-basin source rock potential assessment, and long-term carbon cycle evolution. If used for high-resolution stratigraphic correlation or short-term paleoenvironmental event identification, users are advised to verify and calibrate data for key intervals by consulting the original references. Despite systematic quality control during data collection, extraction, and entry, occasional errors or data gaps may still exist in a dataset of this scale. We welcome and appreciate feedback from the academic community to continuously improve these data resources through collaborative maintenance.

Author contributions

YKT: writing – original draft, visualization, data collection, investigation. DLC: writing – review & editing, supervision, investigation, funding acquisition. HJS, QZL, XCL, HY: writing – review & editing, investigation. JKL, XKL, YYW, X~~C~~S: writing – review & editing. CDZ, FYC, LYL: data collection.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgments

The authors would like to acknowledge the valuable assistance provided by Jincheng Zhang in literature research and Yue Su in figure preparation. We also extend our sincere gratitude to other members of the research group for their support and contributions.

Financial support

This study was supported by the State Key R&D Project of China (grant no. 2024YFF0808100), the National Natural Science Foundation of China (grant no. 4472021), and the Natural Science Foundation of Hubei (2023AFA006).

References

Ahm, A.-S. and Husson, J. M.: Local and global controls on carbon isotope chemostratigraphy. Cambridge University Press, https://doi.org/10.1017/9781009028882, 2022.

Akande, W. G.: Evaluation of hydrocarbon generation potential of the Mesozoic organic-rich rocks using TOC content and Rock-Eval pyrolysis techniques, Geosciences, 2, 164–169, https://doi.org/10.5923/j.geo.20120206.03, 2012.

Algeo, T. J. and Scheckler, S. E.: Terrestrial-marine teleconnections in the Devonian: links between the evolution of land plants, weathering processes, and marine anoxic events, Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 353, 113–130, https://doi.org/10.1098/rstb.1998.0194, 1998.

Ansari, A. H., Pandey, S. K., Ahmad, S., Sharma, M., Govil, P., Chaddha, A. S., and Sharma, A.: High primary productivity in an Ediacaran shallow marine basin influenced by strong seasonal to perennial upwelling, Geological Magazine, 160, 1607–1623, https://doi.org/10.1017/S0016756823000614, 2023.

Bauska, T. K., Joos, F., Mix, A. C., Roth, R., Ahn, J., and Brook, E. J.: Links between atmospheric carbon dioxide, the land carbon reservoir and climate over the past millennium, Nature Geoscience, 8, 383–387, https://doi.org/10.1038/ngeo2422, 2015.

Beaumont, E. A. and Foster, N. H. (Eds.): Exploring for Oil and Gas Traps, American Association of Petroleum Geologists, https://doi.org/10.1306/TrHbk624, 1999.

Berger, W. H., Smetacek, V. S., and Wefer, G.: Ocean productivity and paleoproductivity—an overview, Productivity in the Ocean - Present and Past, 44: 1-34, 1989.

Canfield, D. E.: Carbon cycle evolution before and after the great oxidation of the atmosphere, American Journal of Science, 321, 297–331, https://doi.org/10.2475/03.2021.01, 2021.

Chen, Y., Zhu, Z., and Zhang, L.: Control actions of sedimentary environments and sedimentation rates on lacustrine oil shale distribution, an example of the oil shale in the Upper Triassic Yanchang Formation, southeastern Ordos Basin (NW China), Marine and Petroleum Geology, 102, 508–520, https://doi.org/10.1016/j.marpetgeo.2019.01.006, 2019.

Craig, H.: Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide, Geochimica et Cosmochimica Acta, 12, 133–149, https://doi.org/10.1016/0016-7037(57)90024-8, 1957.

Dong, Y., Cui, Y., Wang, J., Chen, H., Zhang, F., Wu, Y., Li, Z., Zhu, P., and Jiang, S.: Paleozoic carbon cycle dynamics: Insights from stable carbon isotopes in marine carbonates and C_3 land plants. Earth-Science Reviews, 222, 103813, https://doi.org/10.1016/j.earscirev.2021.103813, 2021.

Du, Y., Song, H., Algeo, T. J., Zhang, H., Peng, J., Wu,Y., Lai, J., Shu, X, Song, H., Wei, L., Zhang, J., Stüeken, E. E., Grasby, S. E., Dal Corso, J., Liu, X., Chu, D., Tian, L., Liang, Q., Li, X., Yao, H., and Song, H.: The global database of deep-time marine nitrogen isotope data, Earth System Science Data Discussions, https://doi.org/10.5194/essd-2025-377, 2025.

El Nady, M. M., Ramadan, F. S., Hammad, M. M., and Lotfy, N. M.: Evaluation of organic matters, hydrocarbon potential and thermal maturity of source rocks based on geochemical and statistical methods: Case study of source

rocks in Ras Gharib oilfield, central Gulf of Suez, Egypt, Egyptian Journal of Petroleum, 24, 203–211, https://doi.org/10.1016/j.ejpe.2015.05.012, 2015.

FAIR: FAIR Play in geoscience data, Nature Geoscience, 12, 961, https://doi.org/10.1038/s41561-019-0506-4, 2019.

Falkowski, P., Scholes, R. J., Boyle, E. E. A., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Högberg, P., Linder, S., Mackenzie, F. T., Moore III, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., and Steffen, W. : The global carbon cycle: a test of our knowledge of earth as a system, Science, 290, 291–296, https://doi.org/10.1126/science.290.5490.291, 2000.

Farrell, U. C., Samawi, R., Anjanappa, S., Klykov, R., Adeboye, O. O., Agic, H., Ahm, A. C., Boag, T. H., Bowyer, F., Brocks, J. J., Brunoir, T. N., Caneld, D. E., Chen, X., Cheng, M., Clarkson, M. O., Cole, D. B., Cordie, D. R., Crockford, P. W., Cui, H., Dahl, T. W., Mouro, L. D., Dewing, K., Dornbos, S. Q., Drabon, N., Dumoulin, J. A., Emmings, J. F., Endriga, C. R., Fraser, T. A., Gaines, R. R., Gaschnig, R. M., Gibson, T. M., Gilleaudeau, G. J., Gill, B. C., Goldberg, K., Guilbaud, R., Halverson, G. P., Hammarlund, E. U., Hantsoo, K. G., Henderson, M. A., Hodgskiss, M. S. W., Horner, T. J., Husson, J. M., Johnson, B., Kabanov, P., Brenhin Keller, C., Kimmig, J., Kipp, M. A., Knoll, A. H., Kreitsmann, T., Kunzmann, M., Kurzweil, F., LeRoy, M. A., Li, C., Lipp, A. G., Loydell, D. K., Lu, X., Macdonald, F. A., Magnall, J. M., Mand, K., Mehra, A., Melchin, M. J., Miller, A. J., Mills, N. T., Mwinde, C. N., O'Connell, B., Och, L. M., Ossa Ossa, F., Pages, A., Paiste, K., Partin, C. A., Peters, S. E., Petrov, P., Playter, T. L., Plaza-Torres, S., Porter, S. M., Poulton, S. W., Pruss, S. B., Richoz, S., Ritzer, S. R., Rooney, A. D., Sahoo, S. K., Schoepfer, S. D., Sclafani, J. A., Shen, Y., Shorttle, O., Slotznick, S. P., Smith, E. F., Spinks, S., Stockey, R. G., Strauss, J. V., Stueken, E. E., Tecklenburg, S., Thomson, D., Tosca, N. J., Uhlein, G. J., Vizcaino, M. N., Wang, H., White, T., Wilby, P. R., Woltz, C. R., Wood, R. A., Xiang, L., Yurchenko, I. A., Zhang, T., Planavsky, N. J., Lau, K. V., Johnston, D. T., and Sperling, E. A.: The Sedimentary Geochemistry and Paleoenvironments Project, Geobiology, 19, 545–556, https://doi.org/10.1111/gbi.12462, 2021.

Golonka, J., Porebski, S. J., and Waśkowska, A.: Silurian paleogeography in the framework of global plate tectonics, Palaeogeography, Palaeoclimatology, Palaeoecology, 622, 111597, https://doi.org/10.1016/j.palaeo.2023.111597, 2023.

Hou, H., Shao, L., Li, Y., Liu, L., Liang, G., Zhang, W., Wang, X., and Wang, W.: Effect of paleoclimate and paleoenvironment on organic matter accumulation in lacustrine shale: Constraints from lithofacies and element geochemistry in the northern Qaidam Basin, NW China, Journal of Petroleum Science and Engineering, 208, 109350, https://doi.org/10.1016/j.petrol.2021.109350, 2022.

Jarvie, D. M.: Total organic carbon (TOC) analysis, in: Source and Migration Processes and Evaluation Techniques, edited by: Merrill, R. K., Treatise of Petroleum Geology, 113–118, https://doi.org/10.1306/TrHbk543C11, 1991.

Ji, S., Nie, J., Lechler, A., Huntington, K. W., Heitmann, E. O., and Breecker, D. O.: A symmetrical CO_2 peak and asymmetrical climate change during the middle Miocene, Earth and Planetary Science Letters, 499, 134–144, https://doi.org/10.1016/j.epsl.2018.07.011, 2018.

Judd, E. J., Tierney, J. E., Huber, B. T., Wing, S. L., Lunt, D. J., Ford, H. L., Inglis, G. N., McClymont, E. L., O'Brien, C. L., Rattanasriampaipong, R., Si, W., Staitis, M. L., Thirumalai, K., Anagnostou, E., Cramwinckel, M. J., Dawson, R. R., Evans, D., Gray, W. R., Grossman, E. L., Henehan, M. J., Hupp, B. N., MacLeod, K. G., O'Connor, L. K., Sánchez Montes, M. L., Song, H., and Zhang, Y. G.: The PhanSST global database of Phanerozoic sea surface temperature proxy data, Scientific Data, 9, 753, https://doi.org/10.1038/s41597022-01826-0, 2022.

Lai, J., Song, H., Chu, D., Dal Corso, J., Sperling, E. A., Wu, Y., Liu, X., Wei, L., Li, M., Song, H., Du, Y., Jia, E., Feng, Y., Song, H., Yu, W., Liang, Q., Li, X., and Yao, H.: Deep-Time Marine Sedimentary Element Database, Earth System Science Data, 17, 1613–1626, https://doi.org/10.5194/essd-17-1613-2025, 2025.

Li, G., and Elderfield, H.: Evolution of carbon cycle over the past 100 million years, Geochimica et Cosmochimica

Acta, 103, 11–25, https://doi.org/10.1016/j.gca.2012.10.014, 2013.

Meyer, K. M., Yu, M., Jost, A. B., Kelley, B. M., and Payne, J. L.: $\delta^{13}C$ evidence that high primary productivity delayed recovery from end-Permian mass extinction, Earth and Planetary Science Letters, 302, 378–384, https://doi.org/10.1016/j.epsl.2010.12.033, 2011.

Nordt, L., Tubbs, J., and Dworkin, S.: Stable carbon isotope record of terrestrial organic materials for the last 450 Ma yr. Earth-science reviews, 159: 103-117, https://doi.org/10.1016/j.earscirev.2016.05.007, 2016.

Peters, K. E. and Cassa, M. R.: Applied source rock geochemistry, in: The Petroleum System—From Source to Trap, edited by: Magoon, L. B. and Dow, W. G., AAPG Memoir 60, 93–120, https://doi.org/10.1306/M60585C5, 1994.

Peters, S. E., and Husson, J. M.: Sediment cycling on continental and oceanic crust. Geology, 45(4): 323-326, https://doi.org/10.1130/G38861.1, 2017.

Post, W. M., Peng, T. H., Emanuel, W. R., King, A. W., Dale, V. H., and DeAngelis, D. L.: The global carbon cycle, American Scientist, 78, 310–326, http://www.jstor.org/stable/29774118, 1990.

Prajapati, S. K., Kumar, V., Dayal, P., Gairola, A., Borate, R. B., and Srivastava, R.: The Role of Carbon in Life's Blueprint and Carbon Cycle under-Standing Earth's Essential Cycling System: Benefits and Harms to Our Planet, International Journal, 1, 21–32, https://doi.org/10.5281/zenodo.8385431, 2023.

Qu, C. S., Qiu, L. W., Cao, Y. C., Yang, Y. Q., and Yu, K. H.: Sedimentary environment and the controlling factors of organic-rich rocks in the Lucaogou Formation of the Jimusar Sag, Junggar Basin, NW China, Petroleum Science, 16, 763–775, https://doi.org/10.1007/s12182-019-0353-3, 2019.

SCOR Working Group: GEOTRACES–An international study of the global marine biogeochemical cycles of trace elements and their isotopes, Geochemistry, 67, 85–131, https://doi.org/10.1016/j.chemer.2007.02.001, 2007.

Scotese, C. R., and Wright, N.: PALEOMAP paleodigital elevation models (PaleoDEMS) for the Phanerozoic. Paleomap Proj, 1-26, https://doi.org/10.5281/zenodo.5460860, 2018.

Shu, X., Song, H., Lei, Y., Chu, D., Dal Corso, J., Liu, X., Ye, Q, Song, H., Wei, L., Jia, E., Feng, Y., Du, Y., Song, H., Yu, W., Liang, Q., Li, X., Yao, H., and Wu, Y.: Global Acritarch Database (> 110 000 occurrences), Earth System Science Data, 17, 3567–3582, https://doi.org/10.5194/essd-17-3567-2025, 2025.

Song, H., Wignall, P. B., Song, H., Dai, X., and Chu, D.: Seawater temperature and dissolved oxygen over the past 500 million years, Journal of Earth Science, 30, 236–243, https://doi.org/10.1007/s12583-018-1002-2, 2019.

Strauss, H., and Peters-Kottig, W.: The Paleozoic to Mesozoic carbon cycle revisited: the carbon isotopic composition of terrestrial organic matter. Geochemistry, Geophysics, Geosystems, 4(10), https://doi.org/10.1029/2003GC000555, 2003.

Tian, Y., Chu, D., Lai, J., Shu, X., Zhang, C., Cui, F., Lou, L., Liu, X., Wu, Y., Liang, Q., Li, X., Yao, H., and Song, H.: Paleozoic-Mesozoic Terrestrial Total Organic Carbon and Organic Carbon Isotope Database [Data set]. Zenodo. https://doi.org/10.5281/zenodo.20486580,~~https://doi.org/10.5281/zenodo.18163858,~~ 2026.

Tyson, R. V.: Sedimentation rate, dilution, preservation and total organic carbon: some results of a modelling study, Organic Geochemistry, 32, 333–339, https://doi.org/10.1016/S0146-6380(00)00161-3, 2001.

Tyson, R. V.: The "productivity versus preservation" controversy: cause, flaws, and resolution, SEPM Special Publication, 82, 17–33, https://doi.org/10.2110/pec.05.82.0017, 2005.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., t Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn,

G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3, 160018, https://doi.org/10.1038/sdata.2016.18, 2016.

Wu, Y., Chu, D., Tong, J., Song, H., Dal Corso, J., Wignall, P. B., Song, H., Du, Y., and Cui, Y.: Six-fold increase of atmospheric pCO_2 during the Permian–Triassic mass extinction, Nature Communications, 12, 2137, https://doi.org/10.1038/s41467-021-22298-7, 2021.

Wu, Y., Cui, Y., Chu, D., Song, H., Tong, J., Dal Corso, J., and Ridgwell, A.: Volcanic CO2 degassing postdates thermogenic carbon emission during the end-Permian mass extinction, Science Advances, 9, eabq4082, https://doi.org/10.1126/sciadv.abq4082, 2023.

Xue, J. Z., Wang, J. S., Li, B. X., Huang, P., and Liu, L.: Origin and early evolution of land plants and the effects on Earth's environments, Earth Science, 47, 36483664, in Chinese, http://www.earth-science.net/article/doi/10.3799/dqkx.2022.332, 2022.

Ziegler, A. M., Hansen, K. S., Johnson, M. E., Kelly, M. A., Scotese, C. R., and Van Der Voo, R.: Silurian continental distributions, paleogeography, climatology, and biogeography, Tectonophysics, 40, 13–51, https://doi.org/10.1016/0040-1951(77)90028-2, 1977.