Palaeogeography, Palaeoclimatology, Palaeoecology 679 (2025) 113296

Contents lists available at ScienceDirect

Palaeogeography, Palaeoclimatology, Palaeoecology

journal homepage: www.elsevier.com/locate/palaeo

Silent past: Biogeographic gaps in the Cenozoic fossil archive

Marta Matamala-Pagès [a,*], Oskar Hagen [b,c,d], Adrián Castro-Insua [a], Adriana Oliver [a], Eduardo Méndez-Quintas [a,e,f], Graciela Sotelo [a], Iván Rey-Rodríguez [a], Sara Gamboa [a], Sofía Galván [a], Sara Varela [a,g]

[a] MAPAS Lab, Centro de Investigación Mariña, Departamento de Ecoloxía e Bioloxía Animal, Universidade de Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Spain
[b] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[c] Senckenberg Biodiversity and Climate Research Centre (SBiK-F), 60325 Frankfurt am Main-Georg-Voigt-Straße 14–16, Germany
[d] Center for Critical Computational Studies (C^3S), Goethe-Universität Frankfurt, Theodor-W.-Adorno-Platz 1, 60629 Frankfurt am Main, Germany
[e] UNIARQ – Centre for Archaeology, School of Arts and Humanities, University of Lisbon, Alameda da Universidade, 1600-214 Lisbon, Portugal
[f] IDEA, Instituto de Evolución en África, University of Alcalá de Henares, Covarrubias 36, 28010 Madrid, Spain
[g] Oportunius. Axencia Galega de Innovación, 15702 Santiago de Compostela, Spain

ARTICLE INFO

Editor: Howard Falcon-Lang

Keywords:
Fossil record
Information loss
Cenozoic
Climate zones
Fossilisation biases
Palaeobiogeography

ABSTRACT

Exposed sedimentary rocks are unevenly distributed across space, interfering in our ability to reconstruct the spatial and temporal dynamics of past climates and ecosystems. In this study, we quantified the extent of “lost” informative areas—regions where sedimentary layers from the past 66 million years (Ma) are not presently exposed—and assessed gaps across climate zones and geological time bins. Using Chorlton’s global geological map, we reconstructed the distribution and depositional ages of exposed sedimentary rocks, and linked these with palaeoclimate simulations from the HadCM3 model, classified under the Köppen–Geiger system. This framework allowed us to compare the distribution of palaeoclimates predicted by HadCM3 for each time bin with the distribution of climates represented only by currently exposed rocks of that interval. We further integrated fossil occurrence data from the Paleobiology Database (PBDB) to evaluate mismatches between climatic coverage in the geological and fossil records. Our results indicate that more than 72 % of past continental areas across the Cenozoic lack accessible sedimentary rocks, implying a substantial loss of information about past biodiversity. Of the remaining 28 %, exposed sediments are disproportionately derived from regions that experienced tropical, temperate, and arid climates, while deposits from cold and polar climates are underrepresented. The fossil record, in contrast, shows distinct biases, with a particularly large proportion of fossils originating from past temperate environments, likely reflecting excavation effort. These findings underscore the need to account for geographic and climatic biases when interpreting macroevolutionary trends, biodiversity patterns, and species’ responses to climate change.

1. Introduction

The fossil record is the primary source of information on Earth’s history, providing direct evidence of past organisms and ecosystems (Marshall, 2017; Benson et al., 2021). Fossils allow the reconstruction of evolutionary pathways, diversification and extinction patterns, and environmental changes across geological time. The fossil record has documented key events in the history of life, such as adaptive radiations, mass extinctions that redefined the composition of ecosystems, and climatic changes that shuffled the distribution of species (Jablonski and Chaloner, 1994; Peters and Foote, 2002).

However, fossilisation is a rare event shaped by a variety of biotic and abiotic factors (see below). Only a small fraction of species and individuals are preserved in the geological record (an estimated 85 % of biological information is presumed lost and never recorded; Selden and Nudds, 2008). Fossilisation primarily occurs in sedimentary rocks, which form under relatively low temperatures and pressures, conditions that facilitate the preservation of organic material (Boggs, 2009;

* Corresponding author.
E-mail addresses: marta.matamala@uvigo.es (M. Matamala-Pagès), oskar@hagen.bio (O. Hagen), adrian.castro.insua@uvigo.es (A. Castro-Insua), adriana.oliver@uvigo.es (A. Oliver), eduardo.mendez.quintas@uvigo.es (E. Méndez-Quintas), graciela.sotelo@uvigo.es (G. Sotelo), ivan.rey.rodriguez@uvigo.es (I. Rey-Rodríguez), sara.gamboa@uvigo.es (S. Gamboa), sofia.galvan@uvigo.es (S. Galván), sara.varela@uvigo.es (S. Varela).

https://doi.org/10.1016/j.palaeo.2025.113296
Received 20 August 2025; Received in revised form 17 September 2025; Accepted 17 September 2025
Available online 18 September 2025

Lieberman and Kaesler, 2010). The fossil record is inherently biased, as certain forms of life are preserved more frequently than others (e.g. those having mineralised parts), presenting a significant challenge for the interpretation of past evolutionary trajectories and climatic changes (Mannion et al., 2013; Jones et al., 2021; Nanglu and Cullen, 2023).

Biases arise from differences in fossil preservation (Smith, 2001) due to physical and chemical properties (Starrfelt and Liow, 2016), geological features of the environment (Crampton et al., 2003; Starrfelt and Liow, 2016), biological factors (Starrfelt and Liow, 2016), taphonomic processes (Kidwell, 2001; Plotnick et al., 2016; Woolley et al., 2022), and differences in sampling efforts (Starrfelt and Liow, 2016; Prevosti and Forasiepi, 2018; Woolley et al., 2022), among others (see Benson and Upchurch, 2013; Jones et al., 2021). These biases can result in a skewed understanding of species diversity and the ecological dynamics of past environments (Alroy and Hunt, 2010; Newham et al., 2014; Silvestro et al., 2019; Nanglu and Cullen, 2023). The patchy and fragmentary nature of the fossil record often obscures critical periods of evolutionary change, and the absence of certain fossil evidence may lead to erroneous conclusions, such as the non-existence of particular species or lineages or the misinterpretation of their adaptive responses and lifespans (Sakamoto et al., 2017; Silvestro et al., 2019; Dean et al., 2025), hindering our ability to fully comprehend the environmental conditions under which species evolved.

Recent methods allow accounting for biases in the fossil record, such as rarefaction and extrapolation techniques (Alroy, 2010; Chao and Jost, 2012; Starrfelt and Liow, 2016) or probabilistic models such as PyRate, which estimate diversification rates considering the incompleteness and uncertainties of fossil data (Silvestro et al., 2014, 2016; Morlon et al., 2024). These tools have improved the accuracy of macroevolutionary inferences, but a key challenge remains largely unaddressed: the loss of fossil potential across different climate zones. The shift of climatic and sedimentary conditions over time has led to varying preservation potential across regions, meaning that areas less favourable for fossilisation are vastly under-represented in the fossil record. As a result, the links between environmental changes and evolutionary processes may remain poorly documented or entirely undocumented.

Recent studies also show that broad geographic regions can follow distinct macroevolutionary trajectories. Global compilations of palaeodiversity may therefore be biased if the composition of fossil assemblages shifts from one basin or region to another across time bins, since such shifts can conflate regional signals into apparent global patterns (Close et al., 2018; Benson et al., 2021; Flannery-Sutherland et al., 2022). This highlights the need to consider not only sampling biases but also gaps in the fossil record across climate zones.

In this study, using a Geographic Information System (GIS) layer of currently exposed rocks with potential fossil-forming capacity (Chorlton, 2007), we reprojected the geological record into its corresponding time intervals and extracted the climatic conditions under which rocks were deposited. By comparing the total extent of each climate zone (derived from the HadCM3 model for the last 66 Ma) with the area that could have preserved fossils (currently exposed sedimentary rocks and fossil data from the Paleobiology Database (PBDB; Paleobiology Database, 2024)), we calculated the percentage of area completely lost to fossilisation or currently unreachable (buried sedimentary rocks). Thus, we quantified how much area per time bin and past climatic zone is not currently accessible for research on Cenozoic, shedding light on factors that remain largely unaccounted for in current macroevolutionary and biogeographic studies.

2. Methods

To quantify fossil information loss throughout the Cenozoic, we integrated palaeoclimatic, lithological, and fossil occurrence data.

2.1. Palaeoclimatic data

We obtained climatic data from the HadCM3 model (Valdes et al., 2017), covering the entire Cenozoic. HadCM3 can perform consistent long-term simulations at defined geological intervals that enable the study of climate change over time. Although alternative climate models are available, such as the ecoClimate for the Quaternary (Lima-Ribeiro et al., 2015), PALEO-PGEM emulator for the last 5 Ma (Holden et al., 2019), or Climber-3α with a low resolution for the continents (Montoya et al., 2005), HadCM3 provides a spatial resolution, 3.75° longitude × 2.5° latitude, suitable for detailed reconstructions across the studied time span.

Climatic reconstructions included temperature and precipitation data for the following geological intervals: 0, 3, 11, 15, 20, 26, 31, 36, 40, 45, 52, 56, 61, and 66 Ma. These intervals reflect key climatic transitions rather than uniform divisions to effectively capture critical palaeoclimatic events.

Climate zones were classified according to Köppen-Geiger (Beck et al., 2018) into tropical, arid, temperate, cold, and polar, using the method described by Galván et al. (2023). Unlike classifications based on species occurrences and expert range mapping (e.g., Olson and Dinerstein, 2002), this method exclusively relies on temperature and precipitation patterns, ensuring consistent palaeoclimatic categorisation across geological periods.

2.2. Lithological data

We used the generalised global geological map from Chorlton (2007), which includes global geological formations classified into igneous, metamorphic, sedimentary, tectonic and volcanic rock domains (GIS-shapefiles). From lithology, we extracted all sediments with the capacity to accumulate biological remains over time (Tyson, 1995; Haldar and Tišljar, 2014). We overlaid climatic and palaeogeographic maps with lithological polygons from Chorlton (2007) to further extract average temperature and precipitation values for each area in the different geological periods. To enable palaeogeographically consistent analyses, all sediment polygons were rotated back to their original palaeopositions using the «**rgplates**» package in R (Müller et al., 2018; Kocsis et al., 2025). A significant methodological challenge arose from the broad geological age and range assigned to sediment polygons in Chorlton's database. To address this, we followed a conservative approach, allocating each sediment polygon to every temporal interval it overlapped with. As a result, the same polygon appears in multiple time periods, leading to a potential overestimation of sediment presence. Additionally, for periods 0 and 3 Ma, we considered Antarctica and Greenland as polygons with fossiliferous potential since Bierman et al. (2024) discovered fossils in these regions.

2.3. Fossil data

We combined fossil data from the PBDB data with HadCM3 palaeoclimatic layers to obtain the distribution of past climates within collected fossil data. The PBDB is an online platform that provides global palaeontological data on the occurrence and taxonomy of organisms from all geological epochs. It is organised and operated collaboratively, allowing access to data on fossil organisms from most of the Phanerozoic Eon. In our study, fossil occurrence records were downloaded directly from the PBDB web interface (https://paleobiodb.org) on the 27th February 2025. We retrieved all available taxonomic for the interval 70–0 Ma, resulting in 613,334 occurrences (see Supplementary Table S1).

2.4. Analysis

For each geological interval, we integrated three GIS data layers: palaeoclimate (raster files), lithological distributions (polygons), and fossil occurrences (points). Our workflow involved extracting information from climatic raster layers using the vector layers and then plotting the results to examine their temporal dynamics. Specifically, we extracted three variables for each lithological polygon (representing exposed sedimentary rocks) and for each fossil data point: climate type (tropical, temperate, arid, cold, or polar), annual mean temperature (°C), and annual precipitation (mm). In addition, we extracted equivalent information from all continental areas in each time bin to establish a baseline for comparison with the climatic model predictions, which served as our ground truth.

The resulting subsamples of continental climates were visualised using boxplots and bar charts to assess discrepancies across time intervals. To illustrate the data integration process, we also prepared an schematic figure summarising pre-processing steps: Fig. 1 presents an example at 45 Ma, showing (A) Köppen–Geiger climate zones derived from the HadCM3 model, (B) lithological polygons representing sedimentary rock exposures and fossil preservation potential, and (C) fossil occurrences obtained from the PBDB. This figure highlights the three key datasets and how they were subsequently overlaid in our analyses.

Data processing and analyses were performed using R version 4.2.2 for Windows (R Core Team, 2021). All data integration, extraction, and visualization were conducted using a custom R script, and packages used include «terra» version 1.7–83 (Hijmans et al., 2024), «rgplates» version 0.5.0 (Müller et al., 2018; Kocsis et al., 2025), «sf» version 1.0.14 (Pebesma, 2018; Pebesma and Bivand, 2023), «maps» version 3.4.1 (Becker et al., 2025), «dplyr» version 1.1.3 (Wickham et al., 2023), «tidyr» version 1.3.0 (Wickham et al., 2024), «lepidochroma» version 0.1.0 (Gamboa et al., 2025).

3. Results

3.1. Climate zones across time

The resulting Köppen-Geiger maps for the Cenozoic show that, globally, arid and cold zones occupy large parts of the continents (Fig. 2). From the Oligocene (31–26 Ma) onwards, tropical zones decrease while polar regions begin to establish in Antarctica. In the Miocene (20–11 Ma), temperate and arid regions expanded in the northern hemisphere, while tropical zones continued to shrink. This pattern becomes more pronounced during the Pliocene (3 Ma), when cold and polar regions consolidate in both extreme latitudes. Finally, at present, polar zones reach their largest historical extension, especially in Antarctica and Greenland, while temperate and arid regions occupy most of the mid-latitudes and tropical zones are confined to a more restricted equatorial range compared to earlier geological times.

3.2. Area, temporal dynamics, and information loss of each climate zone

Arid and cold climate zones remain the largest across time periods (Fig. 3). The arid zone peaks around 36 and 15 Ma (with maximum values of 55.2e+15 km^2 and an average value of 44.8e+13 km^2) and the cold, at 40–36 Ma (reaching maximum values of 54.6e+13 km^2 and an average value of 47.7e+13 km^2).

The temporal dynamics of the expanse differ between climate zones. While tropical areas tend to decrease towards the present, polar ones generally increase since their appearance 36 Ma. Temperate areas, and cold ones to a lesser extent, remain quite stable over time, whereas the arid areas show the largest fluctuations.

Remarkably, potentially fossiliferous rocks cover less than 25 % of the total area of each climate zone. The relative area covered by sediments is larger in tropical and arid areas (14 % and 13 % respectively)

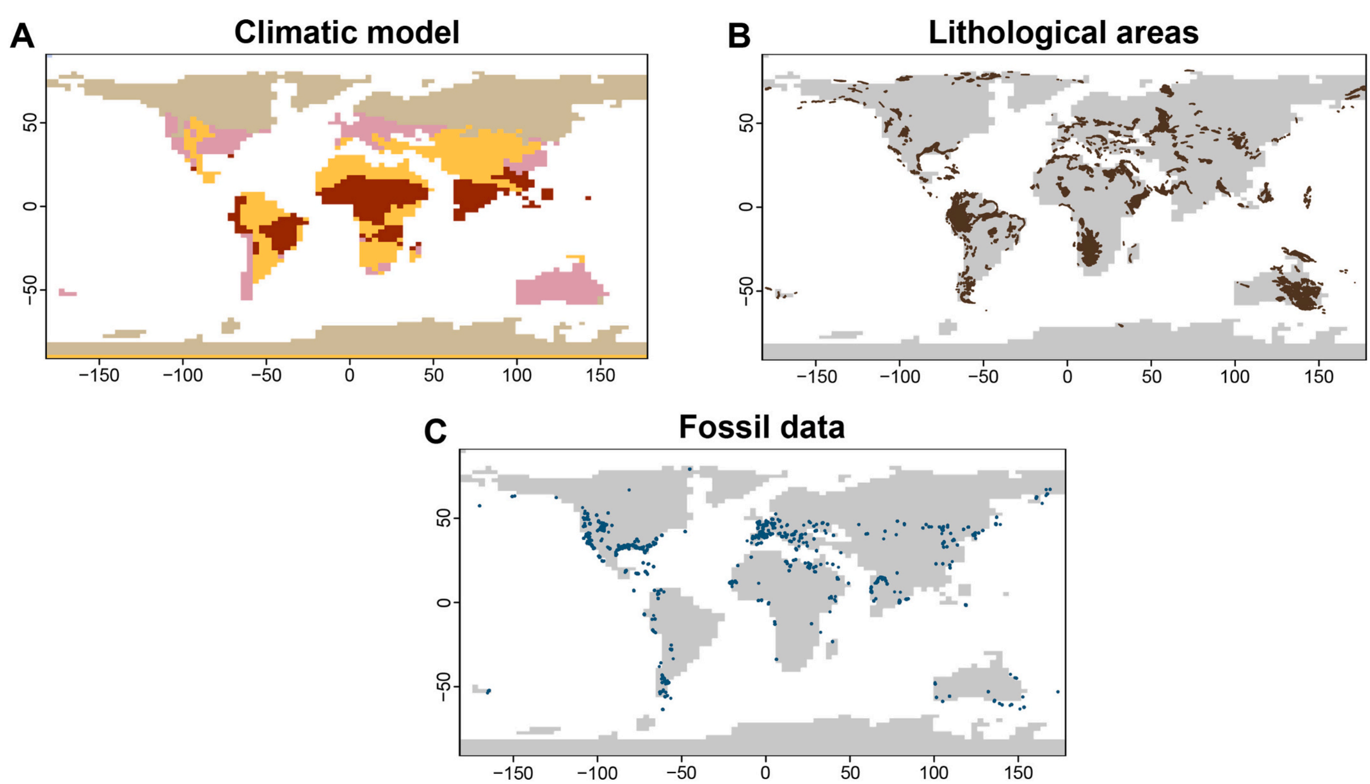

Fig. 1. Graphical representation of the 3 overlapping data layers at 45 Ma (shown as an example). A) Climate zones obtained from the reclassification of the climatic model. Climate zones are colour-coded as follows: tropical (red), arid (orange), temperate (pink), cold (light brown) and polar (light blue). B) Delimitation of lithological areas with fossilisation potential shown in brown. C) Distribution of fossil occurrences for this time interval, shown in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

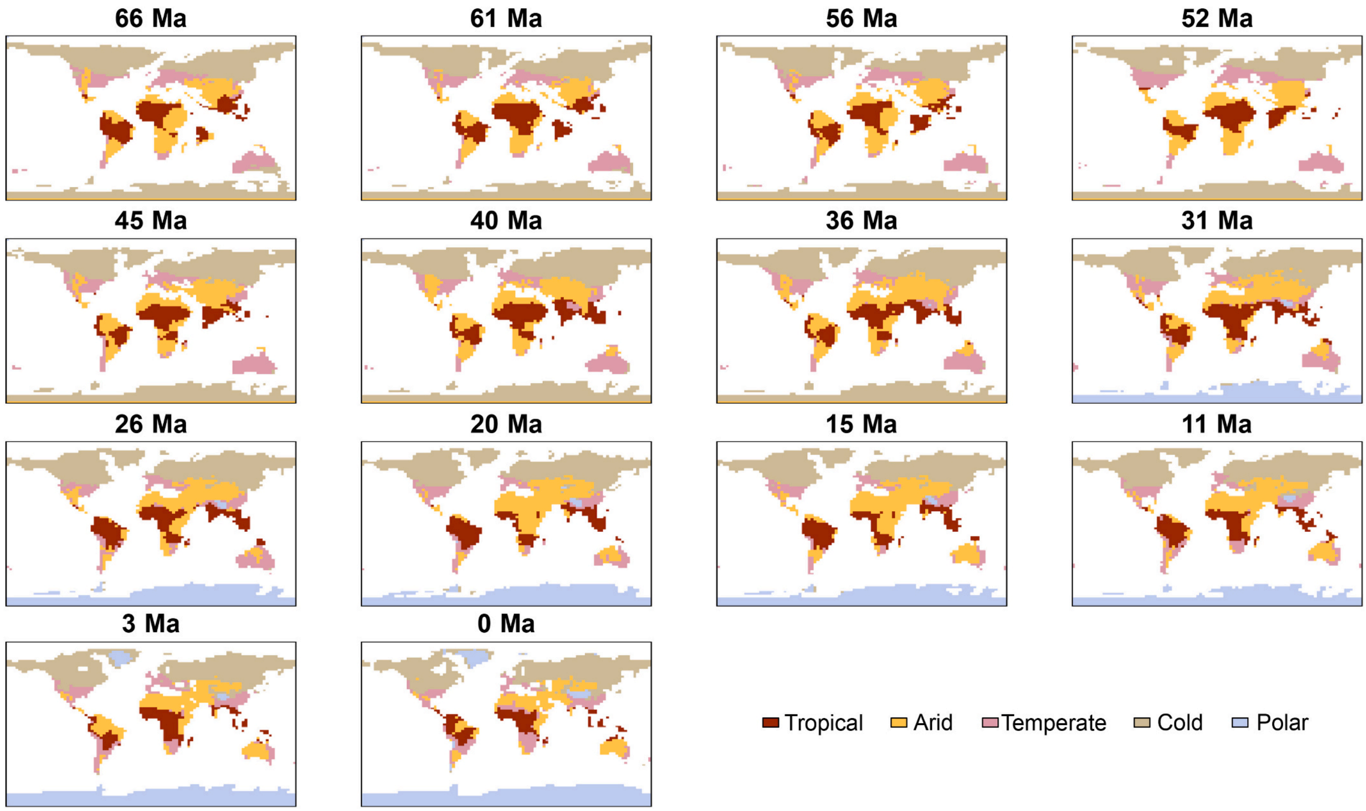

Fig. 2. Evolution of the extent and distribution of the five main Köppen-Geiger climate zones through the Cenozoic. The tropical zone is shown in red, arid in orange, temperate in pink, cold in light brown and polar in light blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

than in other climates. Globally, the proportion of information preserved in sedimentary rocks increases towards the present, especially if we assume ice caps are able to bear fossils (Table 1). Without ice, recovery rates drop to 7–10 % recovery per time period. In general, currently exposed sedimentary rocks do not capture most areas from past climate zones, implying a strong under-representation of the past in the fossil record and a substantial information loss for science.

3.3. *Distortion in climatic coverage between lithological areas and fossil samples*

The representation of Köppen-Geiger climate zones varies significantly depending on the data source: the full extent of the continents, the currently exposed lithological rocks, or the fossil occurrences recorded in the PBDB (Fig. 4). According to the HadCM3 palaeoclimate model, which we take as a reference, arid and cold zones have dominated the continental surface through time, while tropical and temperate zones together account for less than 50 % of the area. This climatic distribution remains relatively stable until the emergence of polar zones in more recent periods.

However, when using only currently exposed lithological rocks as the sampling base, cold zones become strongly under-represented. Instead, tropical, temperate, and to a lesser extent, arid zones are over-represented. Additionally, polar regions appear over-represented in the most recent periods, likely due to our method of classifying modern ice caps as potentially fossiliferous.

Fossil data show the most pronounced distortion: temperate zones are consistently more heavily sampled across time, whereas cold and arid zones are poorly sampled, and tropical zones are extremely under-sampled for fossils. This skew is especially evident at specific time points, such as at 52 Ma, when nearly all recorded fossils correspond to temperate climates.

These patterns are mirrored in temperature and precipitation data derived from each source. While the palaeoclimatic model indicates a long-term trend of global cooling and drying—from stable, warm conditions until around 31 Ma to the progressive establishment of polar climates—fossil data systematically overestimate both temperature and precipitation. In particular, fossils reflect higher and less variable temperatures, especially in warmer periods and the recent past (3–0 Ma). Precipitation values from fossil data are also more variable and often exaggerated compared to the model and lithological data, especially at certain intervals (e.g., 56 Ma, 45 Ma). By contrast, lithological data generally align with the model but occasionally overestimate precipitation (e.g., at 52 Ma and between 20 and 11 Ma).

In summary, both the distribution of climate zones and the estimates of temperature and precipitation reveal similar biases: fossil data disproportionately reflect temperate conditions, lithological data under-represent cold climates, and neither serves as a fully accurate proxy for global climate distributions through time.

4. Discussion

Here, we explore gaps in the amount of information that we have from past climate zones across time. Our results have strong implications on our inferences on how we measure past species adaptation and trait evolution and in our current understanding on how climate shapes species life history traits and the evolutionary fate of lineages.

Our results show that less than 28 % of past climate zone area are recoverable at any time interval we evaluated, from 66 Ma to the present. On average, only 14 % of tropical zones can be accessed, 13 % of arid, 13 % of temperate, 4 % of cold, and 16 % of polar ones (Fig. 3). When ice in excluded, potentially fossiliferous rocks cover merely 7–10 % of the total area per time period. This means that over 70 % of continental information is permanently lost, even under conservative

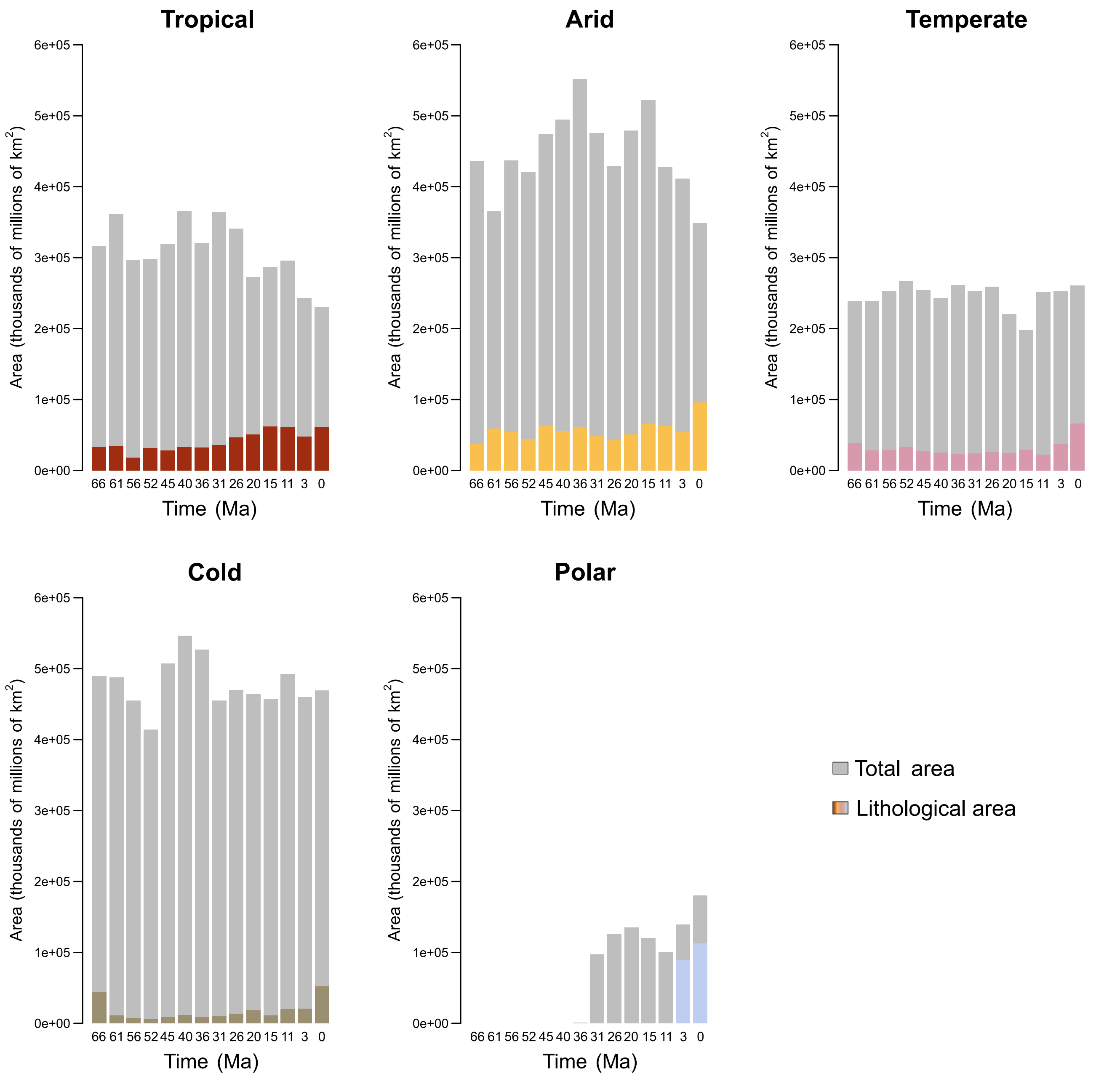

Fig. 3. Extent of each main Köppen-Geiger climate zone (shown in gray) and corresponding lithological areas with fossilisation potential per time interval (shown in the same colour previously assigned to each climate zone: tropical (red), arid (orange), temperate (pink), cold (light brown), and polar (light blue)). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Percentage of continental area with fossilisation potential across time (Ma), with ice caps considered fossil-bearing substrates.

Time (Ma)	0	3	11	15	20	26	31	36	40	45	52	56	61	66
% area potentially fossiliferous	26	16	11	11	9	8	7	8	8	8	8	8	9	10

estimates. The magnitude of this loss implies a severe under-representation of past biodiversity and environmental conditions in the fossil record. However, a recent study brings optimistic results about the use of biased information in palaeomacroecology, showing that some large and conspicuous macroecological patterns might be robust to these strong biases, such as the Latitudinal Diversity Gradient (Galván et al., 2025).

The biases of currently exposed rocks have long been recognised, especially with regard to their uneven preservation and exposure in different environments and time periods (Purnell and Donoghue, 2005; Garzanti et al., 2010; Dunhill et al., 2014). From a geological perspective, it is well known that tectonic activity, erosion, sedimentation rates and sea level changes influence the preservation potential of rocks (Bluck, 1969), often leading to differential preservation of certain depositional environments: for example, arid continental environments tend to retain more fossiliferous sediments (Rieke et al., 2003; Lancaster, 2009), whereas humid tropical areas are poorly preserved in terms of rock exposure due to high weathering rates (Morton, 2002; Gupta,

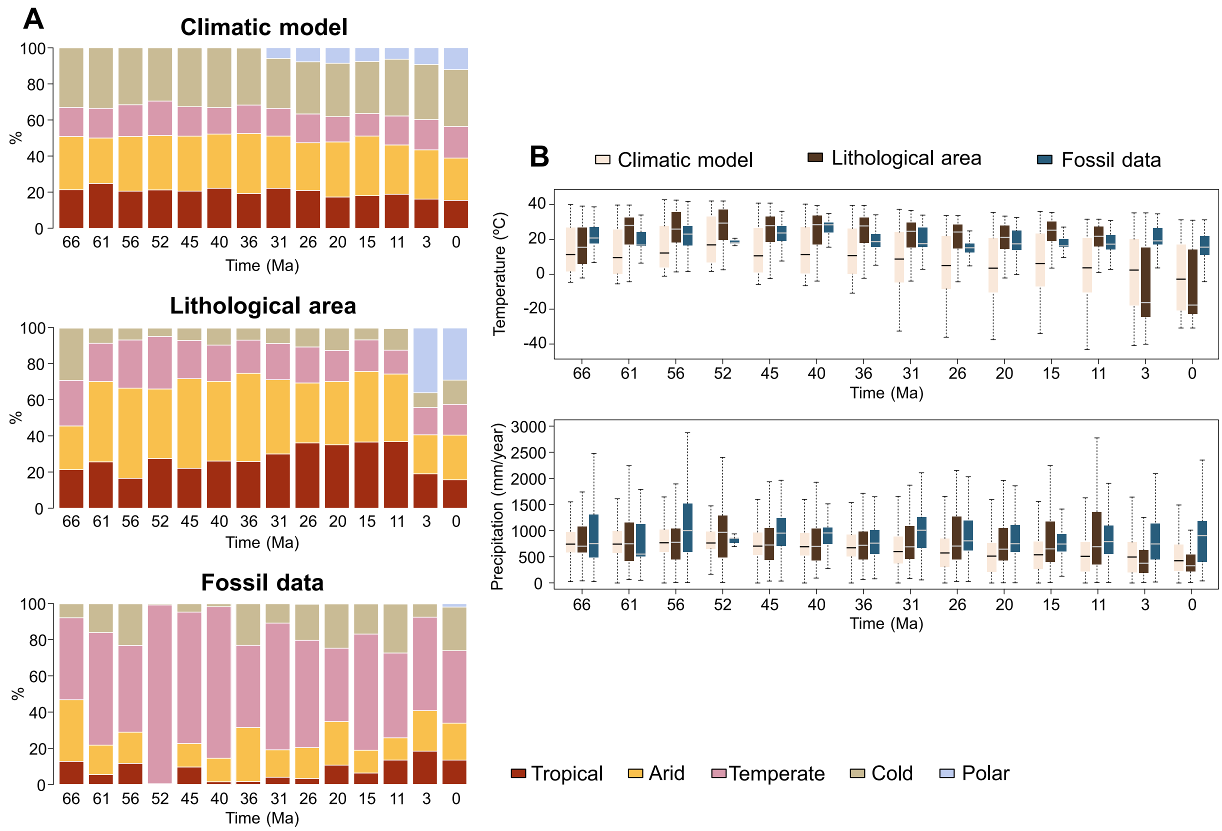

Fig. 4. A) Relative extent of climate zones and B) distribution of climatic conditions over time, based on three different data sources: climatic model, lithological area, and fossil data.

2011). From a macroecological and biogeographical point of view, this uneven preservation translates into an incomplete and biased view of species distributions, trait evolution, and lineage dynamics.

The implications of the paucity of fossil data might be far more serious than is often recognised (Starrfelt and Liow, 2016). Although many researchers in macroecology and evolutionary biology are aware of taphonomic biases, such as which organisms are most likely to fossilise and how fossilisation occurs, they tend to focus on filters operating within already fossiliferous areas (Foote and Sepkoski, 1999; Behrensmeyer et al., 2000). What is often overlooked is that these zones themselves represent only a small and unequal fraction of past environments. This means that large regions of Earth's historical climatic and ecological diversity are completely missing from the fossil record, not because fossils did not form but because the rocks that might contain them were already erased by geologic processes or are not currently exposed. Importantly, our findings underscore the striking disparity between the area occupied by each climate zone in the past and the limited area that remains available within exposed rocks. This discrepancy is particularly pronounced in colder regions (Fig. 3), where extensive historical climatic coverage sharply contrasts with minimal (less than 10 %) rocks with potential fossil-forming capacity exposed nowadays. Several factors likely contribute to poor fossil preservation in these regions, such as low sedimentation rates characteristic of cold climates (Gray and Bart, 2007; Wolfe, 2013) and permafrost dynamics disrupting sediment accumulation. Although reduced microbial activity in cold environments tends to slow organic matter decomposition (Friedmann and Weed, 1987; Dudko et al., 2024), other taphonomic constraints—such as physical disturbance by freeze-thaw cycles and limited mineralisation—may counteract this potential for preservation. A significant part of the global biological record has simply not been preserved, not because of sampling failure but because the physical conditions were never suitable for fossilisation to occur (Krone et al., 2024). These insights highlight the need for careful consideration when interpreting the evolutionary history and biodiversity of higher-latitude ecosystems.

On the other hand, temperate areas are more heavily sampled for fossils (Fig. 4), in line with previous studies linking fossil recovery to temperate and, interestingly, also wealthy countries (Raja et al., 2022; Krone et al., 2024). Further, palaeoclimatic data overlapping sedimentary records and fossil occurrences led to very different inferences on palaeoclimatic distributions, and none of them captured the complete image provided by the palaeoclimatic model for any of the studied temporal intervals, revealing pronounced differences in the inferred palaeoclimatic distribution among these data sources, with none of these data showing a complete image of the modelled palaeoclimatic distribution of continents of each time bin (Fig. 4). Notably, the model occasionally assigns polar climate zones to regions such as South-Central Asia across multiple time periods, including the present. This arises from the Köppen-Geiger definition of polar regions and highlights how the resolution and accuracy of the palaeoclimate model can influence the spatial assignment of climate zones. Such patterns may also result from coarse model resolution, borderline classification thresholds, or local palaeotopographic effects, and should be interpreted cautiously. While the climatic model and lithological sedimentary data tend to agree on a large scale in their estimates of temperature and precipitation over time, the fossil data show a higher deviation, particularly in recent periods, where they tend to overestimate temperatures and precipitation levels. This suggests that the distribution of fossilisable sediments is a better proxy indicator of palaeoclimatic conditions than the fossil record itself, especially in recent times where preservation and sampling bias are greater. This difference can be explained, in part, by a higher representation of fossils from temperate environments, where preservation conditions are more favourable (Andrews and Cook, 1985). Additionally, current fossil distributions are strongly influenced by human accessibility and habitability in these areas, as well as the bias towards developed countries, which enhances spatial and climatic biases (Raja et al., 2022).

Our findings stress the necessity of explicitly incorporating lithological and climatic biases into palaeontological and macroecological studies. Methodological frameworks that mitigate these biases (e.g., through sampling standardisation techniques, explicit palaeoenvironmental context integration, and quantitative models accounting for preservation potential) will significantly enhance the accuracy and reliability of their findings (Silvestro et al., 2016, 2019; Starrfelt and Liow, 2016). Moreover, future studies may benefit from integrating additional proxies, such as geochemical signals, isotopic records, molecular phylogenetics, and palaeoecological niche modelling, to reconstruct ecological histories independently of fossil presence. By triangulating multiple data sources, researchers can partially circumvent inherent biases in the fossil record, achieving a more comprehensive and nuanced view of past biodiversity and ecological dynamics (Close et al., 2018; Benson et al., 2021; Hagen et al., 2021; Jones et al., 2021; Hagen, 2023). Modelling how much biodiversity the Earth sustained/held at each past time interval, and where this biodiversity lived, is one of the most vibrant research lines in palaeontology nowadays. Combining GIS geologic layers, new palaeoclimatic layers from Earth System Models, big data (e.g., PBDB), and cutting-edge techniques (e.g., Deep Learning; Cooper et al., 2024), will allow us to set novel ways to approach the long-standing question of how many species have inhabited the planet through time.

5. Conclusions

Our quantification of fossil information loss across climate zones provides essential context for understanding the structural biases inherent in the palaeontological record. We found that, regardless of temporal interval, approximately 72 % of the land once occupied by arid, tropical, temperate, cold, and polar zones is permanently inaccessible in the sedimentary record, with the exception of recent polar deposits. This indicates that vast portions of Earth's past biodiversity are beyond scientific reach. Of the remaining 28 % of land area that may preserve fossils, sedimentary rocks are disproportionately derived from regions formerly characterized by tropical, temperate, and arid climates, while cold and polar environments are underrepresented. Geological processes alone therefore provide only a fragmented and limited archive of Cenozoic life and its relationship with climate.

In addition, we observed that fossil sampling is strongly biased towards temperate zones across all intervals, resulting in distorted representations of global climatic conditions. Such biases reflect not only the legacy of geological preservation but also geographic disparities in human sampling effort.

These combined biases have significant implications for macroecological and evolutionary inference. If not explicitly accounted for, they may lead to inaccurate reconstructions of species diversity, extinction rates, and adaptive dynamics The overrepresentation of temperate regions, alongside the underrepresentation of cold and arid zones, risks obscuring key evolutionary events occurring in that climatic conditions, such as adaptive radiations, mass extinctions, and biotic responses to climate change. Recognising and addressing these information gaps is thus critical for developing bias-aware methodologies, improving palaeobiological analyses, and refining our understanding of the climatic and geographic drivers of biodiversity through Earth's history.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.palaeo.2025.113296.

CRediT authorship contribution statement

Marta Matamala-Pagès: Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing, Writing – original draft. **Oskar Hagen:** Visualization, Validation, Formal analysis, Data curation, Conceptualization, Writing – review & editing. **Adrián Castro-Insua:** Visualization, Validation, Methodology, Writing – review & editing. **Adriana Oliver:** Visualization, Validation, Conceptualization, Writing – original draft. **Eduardo Méndez-Quintas:** Visualization, Validation, Conceptualization, Writing – review & editing. **Graciela Sotelo:** Visualization, Validation, Conceptualization, Writing – review & editing. **Iván Rey-Rodríguez:** Visualization, Validation, Conceptualization, Writing – review & editing. **Sara Gamboa:** Visualization, Validation, Conceptualization, Writing – review & editing. **Sofía Galván:** Visualization, Validation, Conceptualization, Writing – review & editing. **Sara Varela:** Visualization, Validation, Supervision, Methodology, Investigation, Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements and funding statement

We thank all developers and contributors of the implemented public data that allowed us to carry out this work. This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement 947921; Sara Varela) as part of the MAPAS project. This paper is part of LOST project (PID2021-123202NA-I00) funded by the MCIN/AEI/10.13039/501100011033/ and by "ERDF A way of making Europe". This research was also funded by an "Axuda Complementaria beneficiario axuda StG do ERC" from Xunta de Galicia GAIN Oportunius programme and Consellería de Educación (Galicia, Spain). Oskar Hagen is supported by the Swiss National Science Foundation (SNSF) (P500PB_206815). Iván Rey-Rodríguez is supported by the "Axudas de apoio á etapa de formación posdoutoral grant" (ED481B-2023-040), from the Consellería de Cultura, Educación, Formación Profesional e Universidades, Xunta de Galicia. Sofía Galván was funded by the Universidade de Vigo through a predoctoral fellowship PREUVIGO-2022 (00VI 131H 6410211).

Funding for open access charge: Universidade de Vigo/CISUG.

Data availability

The data and code that supports the findings of this study are openly available in Dryad at Reviewer URL: http://datadryad.org/share/Od-SotdTAIA-riq18p0VwwMBM8OZy-qqL5elcRsItHA

Forthcoming on Dryad. https://doi.org/10.5061/dryad.34tmpg4wk

References

Alroy, J., 2010. The shifting balance of diversity among major marine animal groups. Science 329, 1191–1194.
Alroy, J., Hunt, G., 2010. Quantitative methods in paleobiology, 16. Paleontol. Soc, pp. 55–80.
Andrews, P., Cook, J., 1985. Natural modifications to bones in a temperate setting. Man 20, 675–691.
Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data 5, 180214.
Becker, R., Wilks, A.R., Brownrigg, R., Minka, T.P., Deckmyn, A., 2025. maps: Draw Geographical Maps.
Behrensmeyer, A.K., Kidwell, S.M., Gastaldo, R.A., 2000. Taphonomy and paleobiology. Paleobiology 26, 103–147.
Benson, R.B.J., Upchurch, P., 2013. Diversity trends in the establishment of terrestrial vertebrate ecosystems: Interactions between spatial and temporal sampling biases. Geology 41, 43–46.
Benson, R.B.J., Butler, R., Close, R.A., Saupe, E., Rabosky, D.L., 2021. Biodiversity across space and time in the fossil record. Curr. Biol. 31, R1225–R1236.
Bierman, P.R., Mastro, H.M., Peteet, D.M., Corbett, L.B., Steig, E.J., Halsted, C.T., Caffee, M.M., Hidy, A.J., Balco, G., Bennike, O., Rock, B., 2024. Plant, insect, and

fungi fossils under the center of Greenland's ice sheet are evidence of ice-free times. Proc. Natl. Acad. Sci. 121, e2407465121.
Bluck, B.J., 1969. Introduction to sedimentology. In: Eglinton, G., Murphy, M.T.J. (Eds.), Organic Geochemistry: Methods and Results. Springer, Berlin, Heidelberg, pp. 245–261.
Boggs, J., 2009. In: Sam (Ed.), *Origin, Classification, and Occurrence of Sedimentary Rocks. Petrology of Sedimentary Rocks*. Cambridge University Press, Cambridge, pp. 3–18.
Chao, A., Jost, L., 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology 93, 2533–2547.
Chorlton, L.B., 2007. Generalized Geology of the World: Bedrock Domains and Major Faults in GIS Format. Geological Survey of Canada, p. 5529. Open File.
Close, R.A., Evers, S.W., Alroy, J., Butler, R.J., 2018. How should we estimate diversity in the fossil record? Testing richness estimators using sampling-standardised discovery curves. Methods Ecol. Evol. 9, 1386–1400.
Cooper, R.B., Flannery-Sutherland, J.T., Silvestro, D., 2024. DeepDive: estimating global biodiversity patterns through time using deep learning. Nat. Commun. 15, 4199.
Crampton, J.S., Beu, A.G., Cooper, R.A., Jones, C.M., Marshall, B., Maxwell, P.A., 2003. Estimating the Rock volume Bias in Paleobiodiversity Studies. Science 301, 358–360.
Dean, C.D., Chiarenza, A.A., Doser, J.W., Farnsworth, A., Jones, L.A., Lyster, S.J., Outhwaite, C.L., Valdes, P.J., Butler, R.J., Mannion, P.D., 2025. The structure of the end-cretaceous dinosaur fossil record in North America. Curr. Biol. 0.
Dudko, R.Y., Alfimov, A.V., Gurina, A.A., Meshcheryakova, E.N., Reshetnikov, S.V., Legalov, A.A., Berman, D.I., 2024. Insufficient cold resistance as a possible reason for the absence of darkling beetles (Coleoptera, Tenebrionidae) in Pleistocene Sediments of Siberia. Insects 15, 64.
Dunhill, A.M., Hannisdal, B., Benton, M.J., 2014. Disentangling rock record bias and common-cause from redundancy in the British fossil record. Nat. Commun. 5, 4818.
Flannery-Sutherland, J.T., Silvestro, D., Benton, M.J., 2022. Global diversity dynamics in the fossil record are regionally heterogeneous. Nat. Commun. 13, 2751.
Foote, M., Sepkoski, J.J., 1999. Absolute measures of the completeness of the fossil record. Nature 398, 415–417.
Friedmann, E.I., Weed, R., 1987. Microbial trace-fossil formation, biogenous, and abiotic weathering in the Antarctic cold desert. Science 236, 703–705.
Galván, S., Gamboa, S., Varela, S., 2023. Köppen Geiger Climatic Classification for R Users.
Galván, S., Gamboa, S., Chiarenza, A.A., Rotatori, F.M., Oliver, A., Varela, S., 2025. Future palaeontologists will detect current mammal latitudinal biodiversity gradient. Glob. Ecol. Biogeogr. 34, e70075.
Gamboa, S., Insua, A.C., Galván, S., Sotelo, G., Matamala-Pagès, M., Oliver, A., Rey-Rodríguez, I., Méndez-Quintas, E., Varela, S., 2025. Lepidochroma: Colour Palettes inspired by Butterflies.
Garzanti, E., Andò, S., Malusà, M., Vezzoli, G., 2010. Understanding bias in Provenance Studies. EGU General Assembly Conference Abstracts, p. 1143.
Gray, S.W. & Bart, P.J. (2007) Sediment accumulation rates from the Ross Sea continental shelf and Deepwater sites: a physical proxy for the onset of polar conditions. Antarctica: A Keystone in a changing World–Online Proceedings of the 10th ISAES, edited by AK Cooper and CR Raymond et al., USGS Open-File Report, p. 4. Extended Abstract, USGS Open-File Report.
Gupta, A., 2011. Tropical Geomorphology. Cambridge University Press.
Hagen, O., 2023. Coupling eco-evolutionary mechanisms with deep-time environmental dynamics to understand biodiversity patterns. Ecography 2023, e06132.
Hagen, O., Skeels, A., Onstein, R.E., Jetz, W., Pellissier, L., 2021. Earth history events shaped the evolution of uneven biodiversity across tropical moist forests. Proc. Natl. Acad. Sci. 118, e2026347118.
Haldar, S.K., Tišljar, J., 2014. Chapter 5 - Sedimentary Rocks. In: Haldar, S.K., Tišljar, J. (Eds.), Introduction to Mineralogy and Petrology. Elsevier, Oxford, pp. 121–212.
Hijmans, R.J., Bivand, R., Dyba, K., Pebesma, E., Sumner, M.D., 2024. Terra: Spatial Data Analysis.
Holden, P.B., Edwards, N.R., Rangel, T.F., Pereira, E.B., Tran, G.T., Wilkinson, R.D., 2019. PALEO-PGEM v1.0: a statistical emulator of Pliocene–Pleistocene climate. Geosci. Model Dev. 12, 5137–5155.
Jablonski, D., Chaloner, W.G., 1994. Extinctions in the Fossil Record [and Discussion]. Philos. Trans.: Biol. Sci. 344, 11–17.
Jones, L., Dean, C., Mannion, P., Farnsworth, A., Allison, P., 2021. Spatial sampling heterogeneity limits the detectability of deep time latitudinal biodiversity gradients. Proc. R. Soc. B Biol. Sci. 288, 1–10.
Kidwell, S.M., 2001. Major Biases in the Fossil Record. Palaeobiology II. Blackwell Science, pp. 297–303.
Kocsis, A.T., Raja, N.B., Williams, S., Dowding, E.M., 2025. rgplates: R Interface for the GPlates Web Service and Desktop Application.
Krone, I.W., Magoulick, K.M., Yohler, R.M., 2024. All the Earth will not remember: how geographic gaps structure the record of diversity and extinction. Paleobiology 50, 214–225.
Lancaster, N., 2009. In: Young, R., Norby, L. (Eds.), Aeolian Features and Processes. Geological Monitoring. Geological Society of America, p. 0.
Lieberman, B.S., Kaesler, R.L., 2010. Prehistoric Life: Evolution and the Fossil Record. John Wiley & Sons.
Lima-Ribeiro, M.S., Varela, S., González-Hernández, J., Oliveira, G. de, Diniz-Filho, J.A. F. & Terribile, L.C. (2015) EcoClimate: a database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. Biodivers. Inform., 10.
Mannion, P.D., Benson, R.B.J., Butler, R.J., 2013. Vertebrate palaeobiodiversity patterns and the impact of sampling bias. Palaeogeogr. Palaeoclimatol. Palaeoecol. 372, 1–4.
Marshall, C.R., 2017. Five palaeobiological laws needed to understand the evolution of the living biota. Nat. Ecol. Evol. 1, 1–6.
Montoya, M., Griesel, A., Levermann, A., Mignot, J., Hofmann, M., Ganopolski, A., Rahmstorf, S., 2005. The earth system model of intermediate complexity CLIMBER-3α. Part I: description and performance for present-day conditions. Clim. Dyn. 25, 237–263.
Morlon, H., Andréoletti, J., Barido-Sottani, J., Lambert, S., Perez-Lamarque, B., Quintero, I., Senderov, V., Veron, P., 2024. Phylogenetic insights into diversification. Annu. Rev. Ecol. Evol. Syst. 55, 1–21.
Morton, R., 2002. Coastal geoindicators of environmental change in the humid tropics. Environ. Geol. 42, 711–724.
Müller, R.D., Cannon, J., Qin, X., Watson, R.J., Gurnis, M., Williams, S., Pfaffelmoser, T., Seton, M., Russell, S.H.J., Zahirovic, S., 2018. GPlates: building a virtual Earth through deep time. Geochem. Geophys. Geosyst. 19, 2243–2261.
Nanglu, K., Cullen, T.M., 2023. Across space and time: a review of sampling, preservational, analytical, and anthropogenic biases in fossil data across macroecological scales. Earth Sci. Rev. 244, 104537.
Newham, E., Benson, R., Upchurch, P., Goswami, A., 2014. Mesozoic mammaliaform diversity: the effect of sampling corrections on reconstructions of evolutionary dynamics. Palaeogeogr. Palaeoclimatol. Palaeoecol. 412, 32–44.
Olson, D.M., Dinerstein, E., 2002. The global 200: priority ecoregions for global conservation. Ann. Mo. Bot. Gard. 89, 199–224.
Paleobiology Database, 2024. The Paleobiology Database.
Pebesma, E., 2018. Simple features for R: standardized support for Spatial Vector Data. The R J. 10, 439–446.
Pebesma, E., Bivand, R., 2023. Spatial Data Science: With Applications in R. Chapman and Hall/CRC, New York.
Peters, S.E., Foote, M., 2002. Determinants of extinction in the fossil record. Nature 416, 420–424.
Plotnick, R.E., Smith, F.A., Lyons, S.K., 2016. The fossil record of the sixth extinction. Ecol. Lett. 19, 546–553.
Prevosti, F.J., Forasiepi, A.M., 2018. The Fossil Record of Mammalian Carnivores in South America: Bias and Limitations. In: Prevosti, F.J., Forasiepi, A.M. (Eds.), Evolution of South American Mammalian Predators during the Cenozoic: Paleobiogeographic and Paleoenvironmental Contingencies. Springer International Publishing, Cham, pp. 137–154.
Purnell, M.A., Donoghue, P.C., 2005. Between death and data: biases in interpretation of the fossil record of conodonts. Spec. Pap. Palaeontol. 73, 7–25.
R Core Team, 2021. R: The R Project for Statistical Computing.
Raja, N.B., Dunne, E.M., Matiwane, A., Khan, T.M., Nätscher, P.S., Ghilardi, A.M., Chattopadhyay, D., 2022. Colonial history and global economics distort our understanding of deep-time biodiversity. Nat. Ecol. Evol. 6, 145–154.
Rieke, H., McCann, T., Krawczyk, C.M., Negendank, J.F.W., 2003. In: McCann, T., Saintot, A. (Eds.), Evaluation of Controlling Factors on Facies Distribution and Evolution in an Arid Continental Environment: An Example from the Rotliegend of the NE German Basin. Tracing Tectonic Deformation Using the Sedimentary Record. Geological Society of London, p. 0.
Sakamoto, M., Venditti, C., Benton, M.J., 2017. 'Residual diversity estimates' do not correct for sampling bias in palaeodiversity data. Methods Ecol. Evol. 8, 453–459.
Selden, P., Nudds, J., 2008. Fossil Ecosystems of North America: A Guide to the Sites and their Extraordinary Biotas. CRC Press.
Silvestro, D., Salamin, N., Schnitzler, J., 2014. PyRate: a new program to estimate speciation and extinction rates from incomplete fossil data. Methods Ecol. Evol. 5, 1126–1131.
Silvestro, D., Zizka, A., Bacon, C.D., Cascales-Miñana, B., Salamin, N., Antonelli, A., 2016. Fossil biogeography: a new model to infer dispersal, extinction and sampling from palaeontological data. Phil. Trans. R. Soc. B 371, 20150225.
Silvestro, D., Salamin, N., Antonelli, A., Meyer, X., 2019. Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. Paleobiology 45, 546–570.
Smith, A.B., 2001. Large–scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 356, 351–367.
Starrfelt, J., Liow, L.H., 2016. How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. Philos. Trans. R. Soc. B 371, 20150219.
Tyson, R.V., 1995. Introduction: The Importance of Sedimentary Organic Matter. In: Tyson, R.V. (Ed.), Sedimentary Organic Matter: Organic Facies and Palynofacies. Springer, Netherlands, Dordrecht, pp. 1–6.
Valdes, P.J., Armstrong, E., Badger, M.P.S., Bradshaw, C.D., Bragg, F., Crucifix, M., Davies-Barnard, T., Day, J.J., Farnsworth, A., Gordon, C., Hopcroft, P.O., Kennedy, A.T., Lord, N.S., Lunt, D.J., Marzocchi, A., Parry, L.M., Pope, V., Roberts, W.H.G., Stone, E.J., Tourte, G.J.L., Williams, J.H.T., 2017. The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0. Geosci. Model Dev. 10, 3715–3743.
Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., Software, P., PBC, 2023. dplyr: A Grammar of Data Manipulation.
Wickham, H., Vaughan, D., Girlich, M., Ushey, K., Software, P., PBC, 2024. tidyr: Tidy Messy Data.
Wolfe, S., 2013. Cold-climate Aeolian Environments. Treatise on Geomorphology 11, 375–394.
Woolley, C.H., Thompson, J.R., Wu, Y.-H., Bottjer, D.J., Smith, N.D., 2022. A biased fossil record can preserve reliable phylogenetic signal. Paleobiology 48, 480–495.