3. Fossil record biases

Emma Dunne

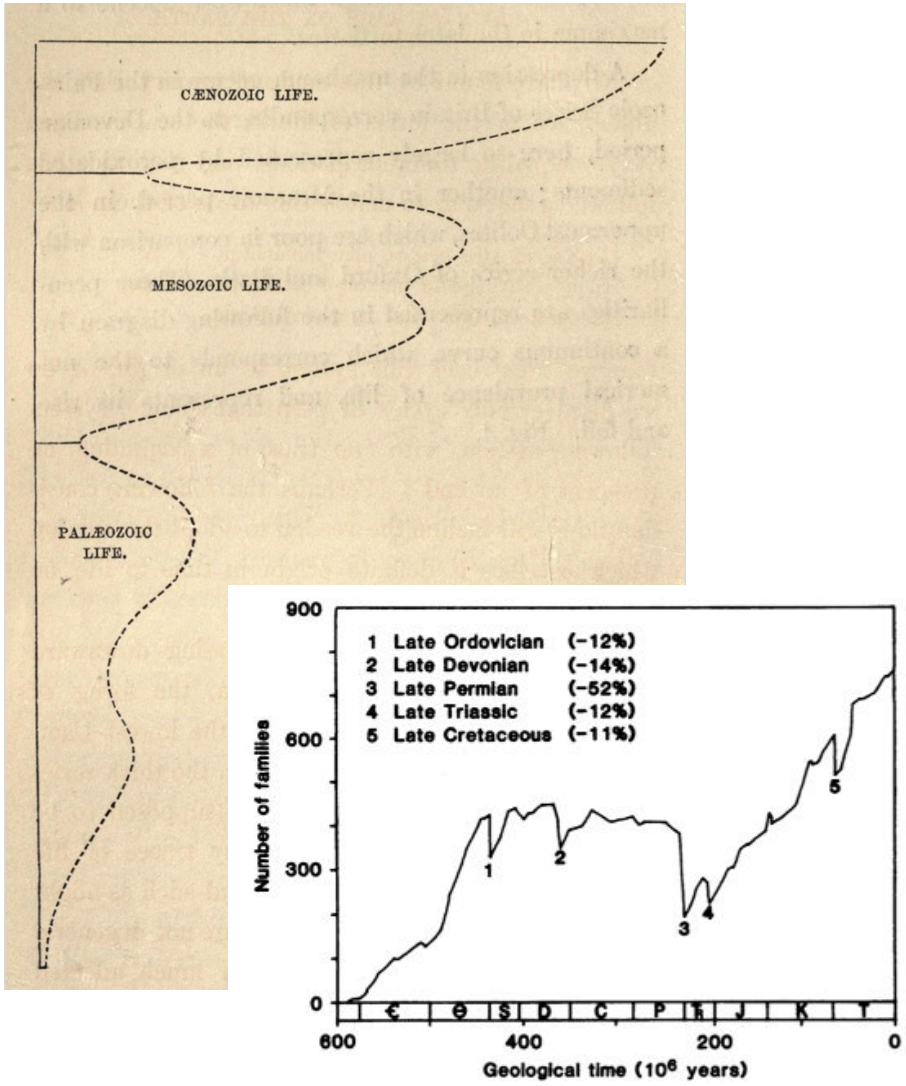
FAU Erlangen-Nürnberg

Thursday 25/08/2022

A. Understanding fossil record biases

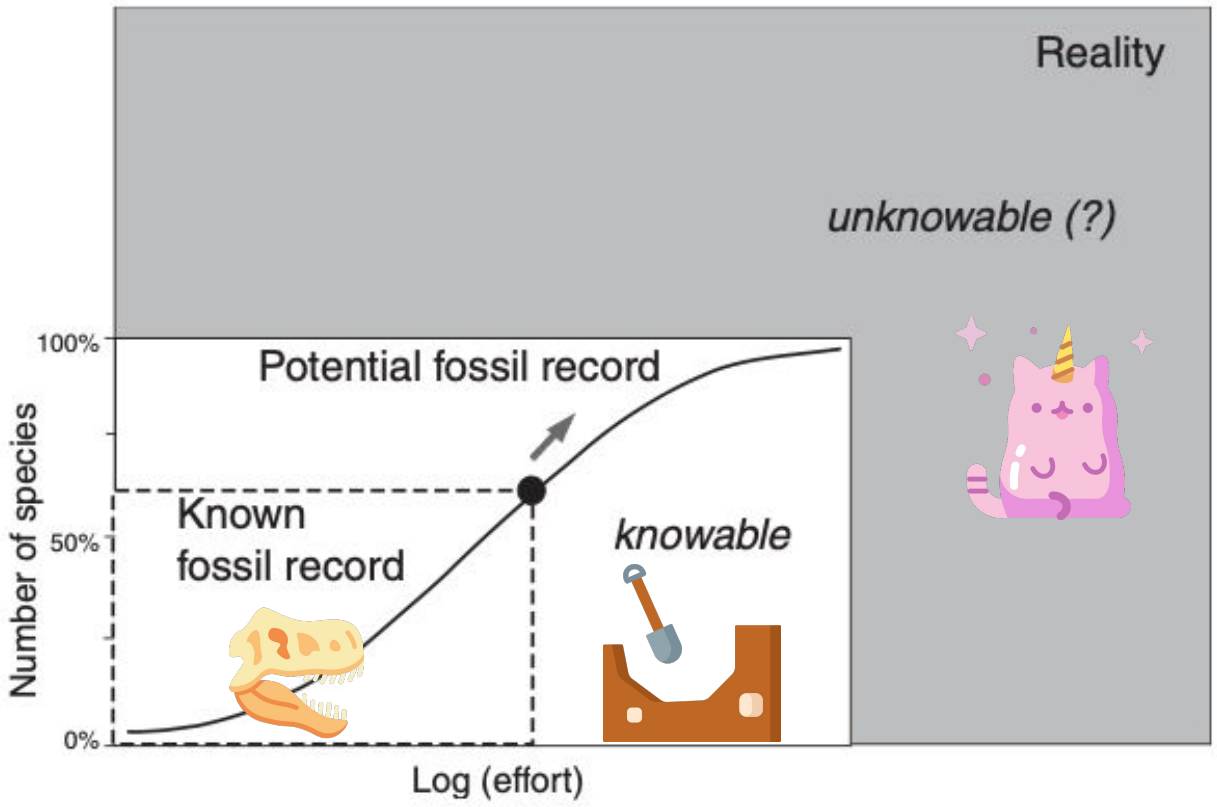
Early paleodiversity studies

- Early palaeodiversity studies took the fossil record at face value
- Even Darwin back in 1959 noted that sampling was uneven and incomplete
- It wasn't until the last half century that we started to appreciate the impacts of fossil record biases



How much of the fossil record do we even know?

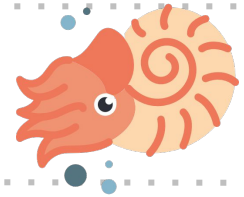
- The known fossil record is barely a fraction of what actually exists
- Even the potential fossil record contains only a tiny fraction of life that has ever lived!



Benton *et al.* (2011)

FROM DEATH TO DATABASE

Habitat & mode of life



Depositional environment

Geological activity

Geographical location

Interest & resources

Occurrence data

Various processes act to **remove** information from the fossil record, which creates **biases** and skews our understanding of past biodiversity

PROCESSES

Death & burial

Diagenesis & lithification, tectonics, etc.

Collection & study

Data collation

CATEGORY

Taphonomic processes

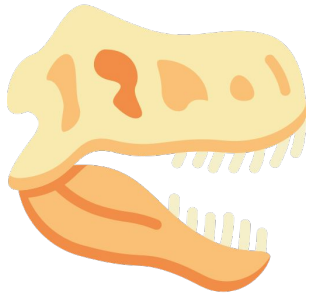
Geological processes

Human & historical processes

Estimates of deep-time biodiversity

Raup's "7 Sources of Error"

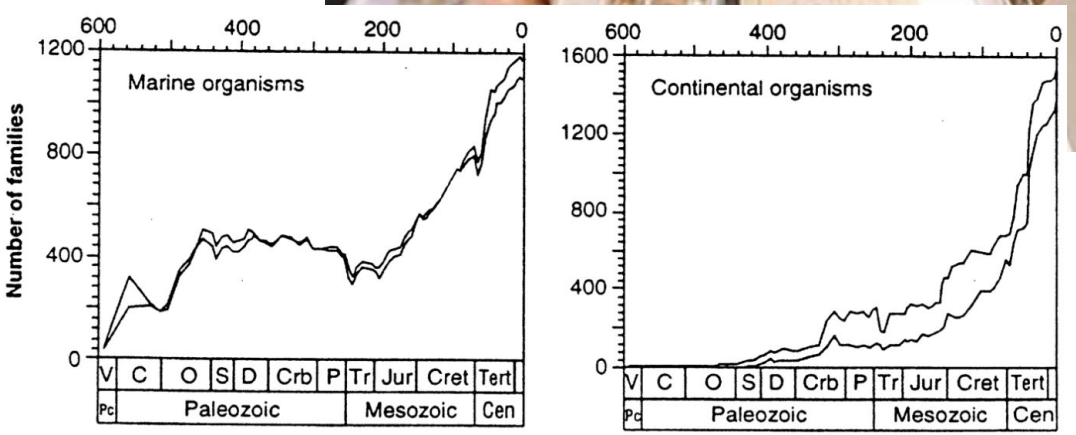
1. Range charts
2. The 'Pull of the Recent'
3. **Durations of geological units**
4. Monographic effects
5. **Lagerstätten**
6. **Area-diversity relationships**
7. Sediment volume



Raup (1972)

Taphonomic biases

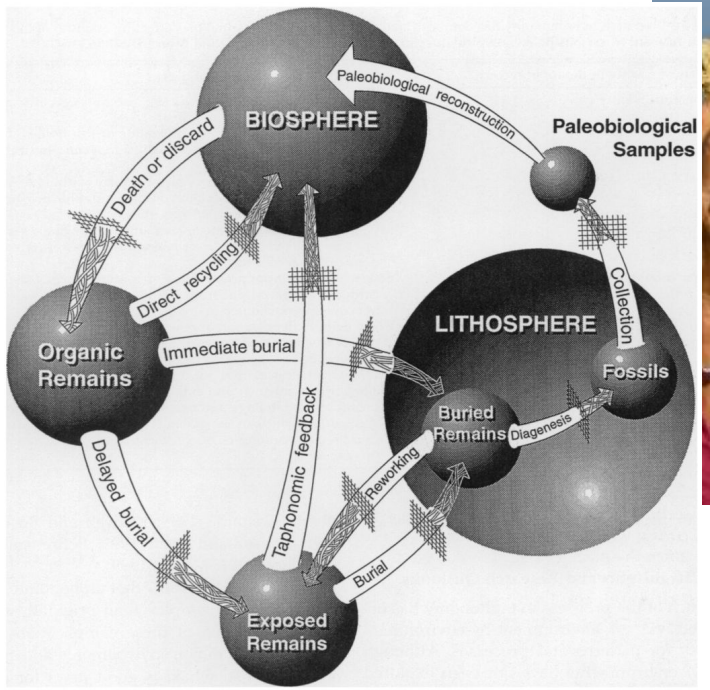
- **Susan Kidwell** (Chicago)
- Relationship between fossil diversity and sedimentation rates
- More sedimentary rock means more opportunities to find fossils
- Also, experimental taphonomy - how do fossils fossilise?



Kidwell, & Sepkoski (1999)

Taphonomic biases

- **Anna ‘Kay’ Behrensmeyer** (Smithsonian NMNH)
- Composition of fossil faunas vary with sedimentary environment (channel, floodplain, lake margin)
- Importance of accounting for taphonomic biases in paleoecological studies



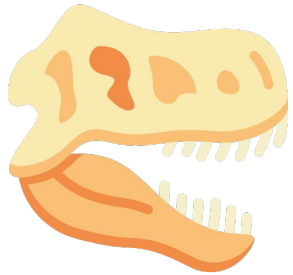
Behrensmeyer et al. (2000)

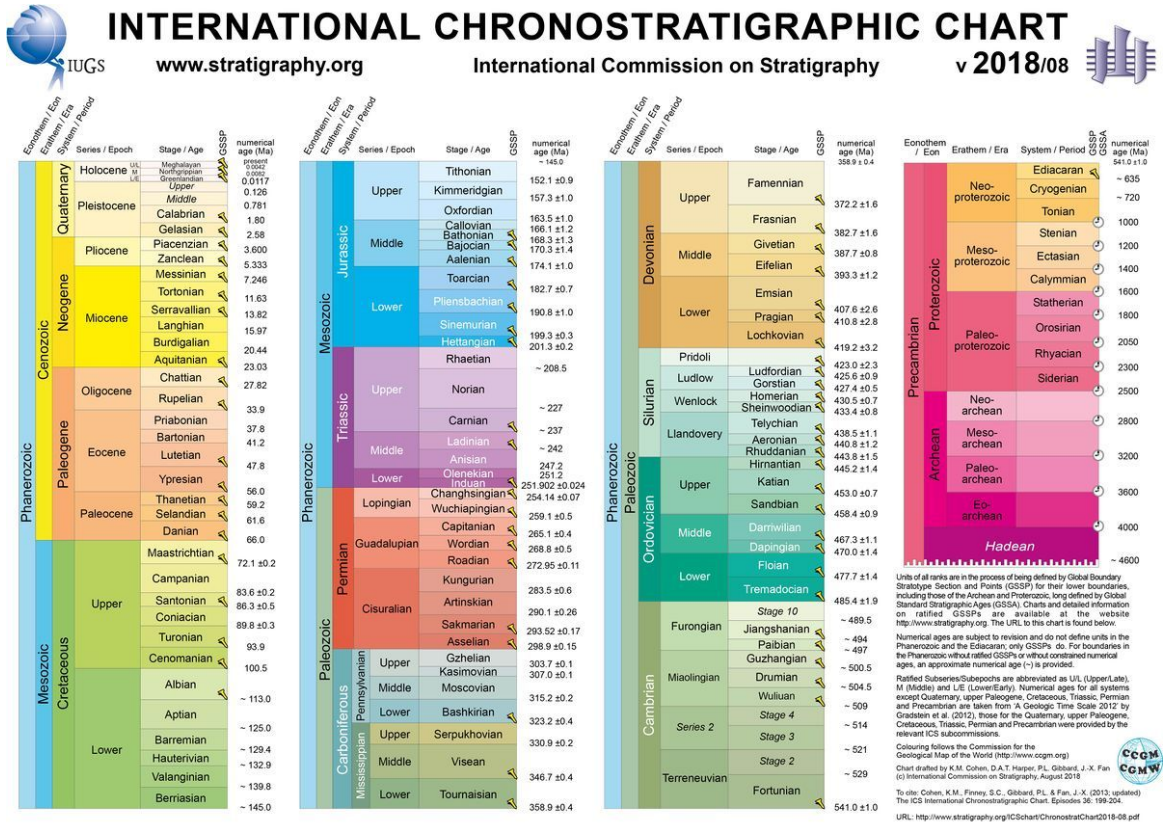
Temporal resolution in the fossil record

Geological time intervals are not equal in length

Example: Late Triassic epochs:

- Rhaetian ~8 Ma
- Norian ~20 Ma
- Carnian ~10 Ma



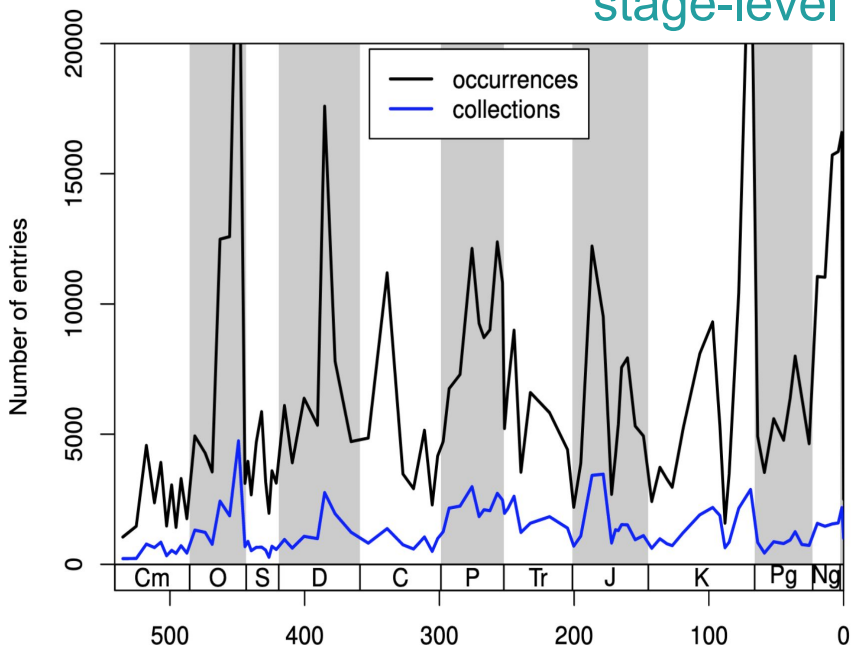


Temporal resolution in the fossil record

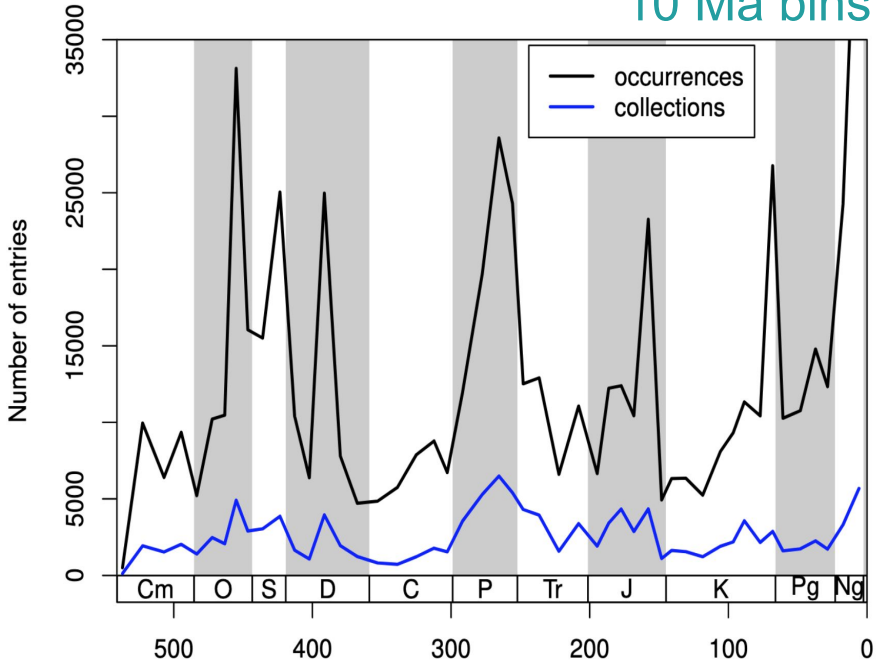
- Many studies focus on stage-level or equal-interval ages

Kocsis *et al.* (2019)
divDyn R package

stage-level

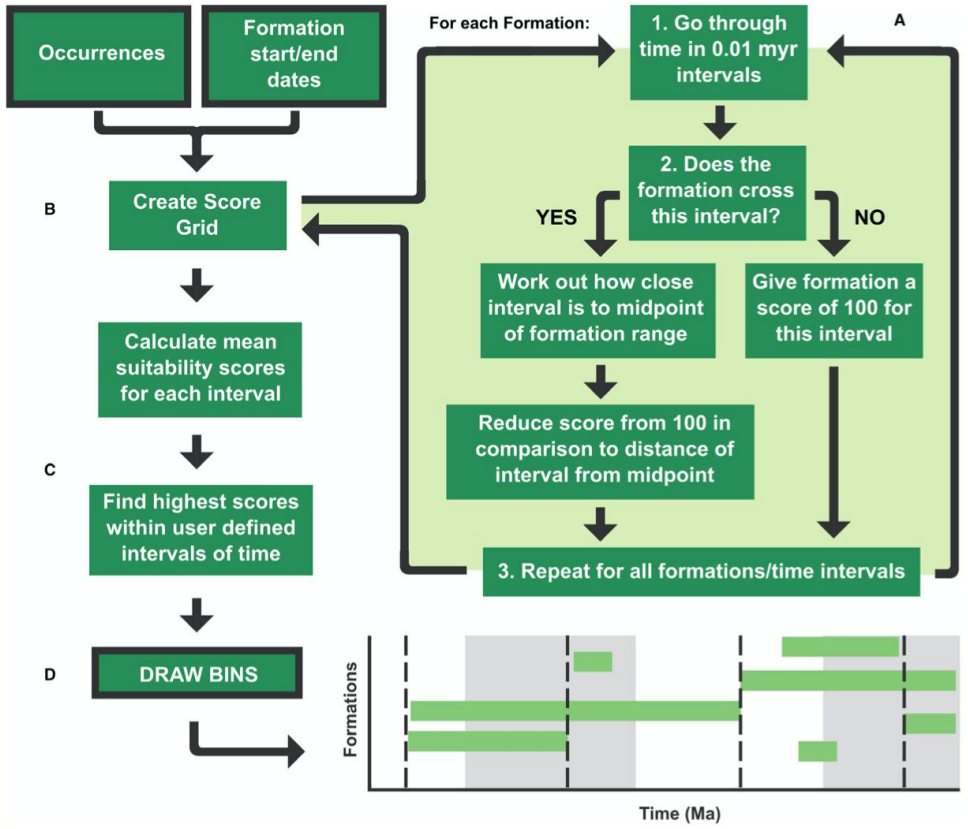


10 Ma bins



Temporal resolution in the fossil record

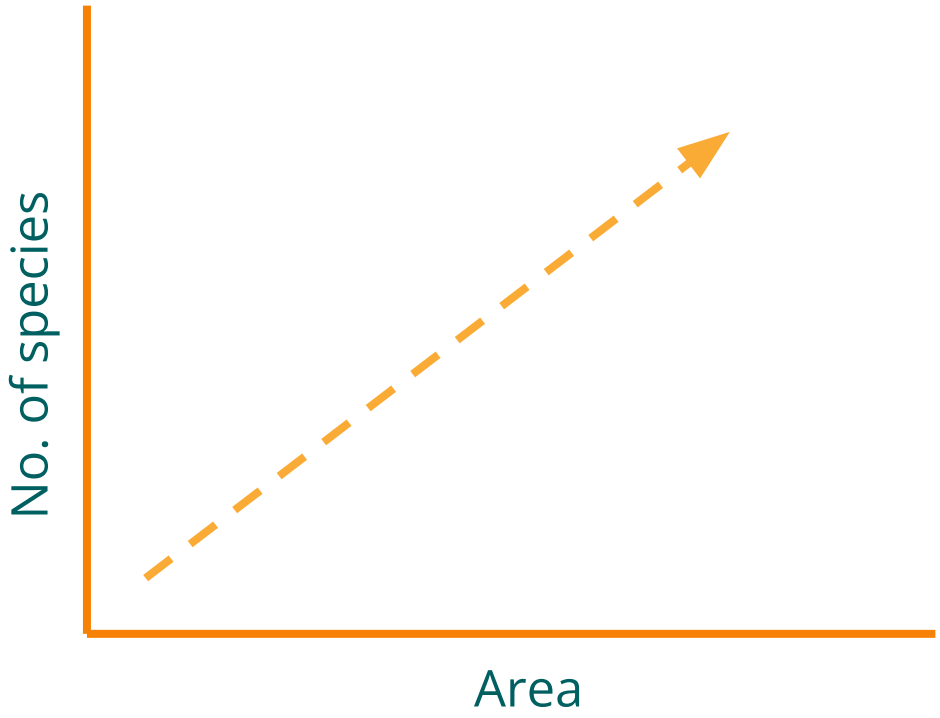
- Many studies focus on stage-level or equal-interval ages
- Not suitable for regional studies
- More recently, statistical methods have been developed to bin data (e.g. using regional stratigraphy)



Dean *et al.* (2020) *Palaeontology*

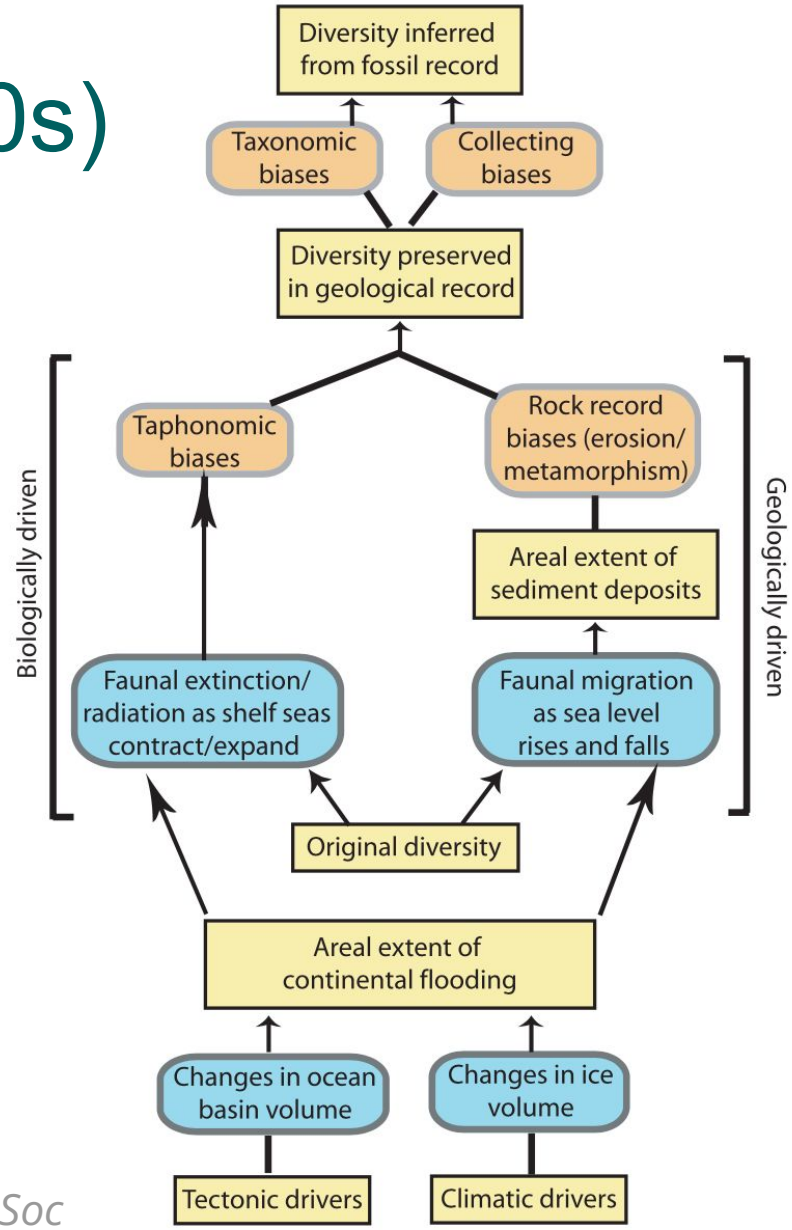
Spatial resolution in the fossil record

- Species-area effect
- Larger areas tend to contain larger numbers of species
- Opening of new areas to search for fossils inevitably leads to new species being described



Renewed interest in biases (2000-10s)

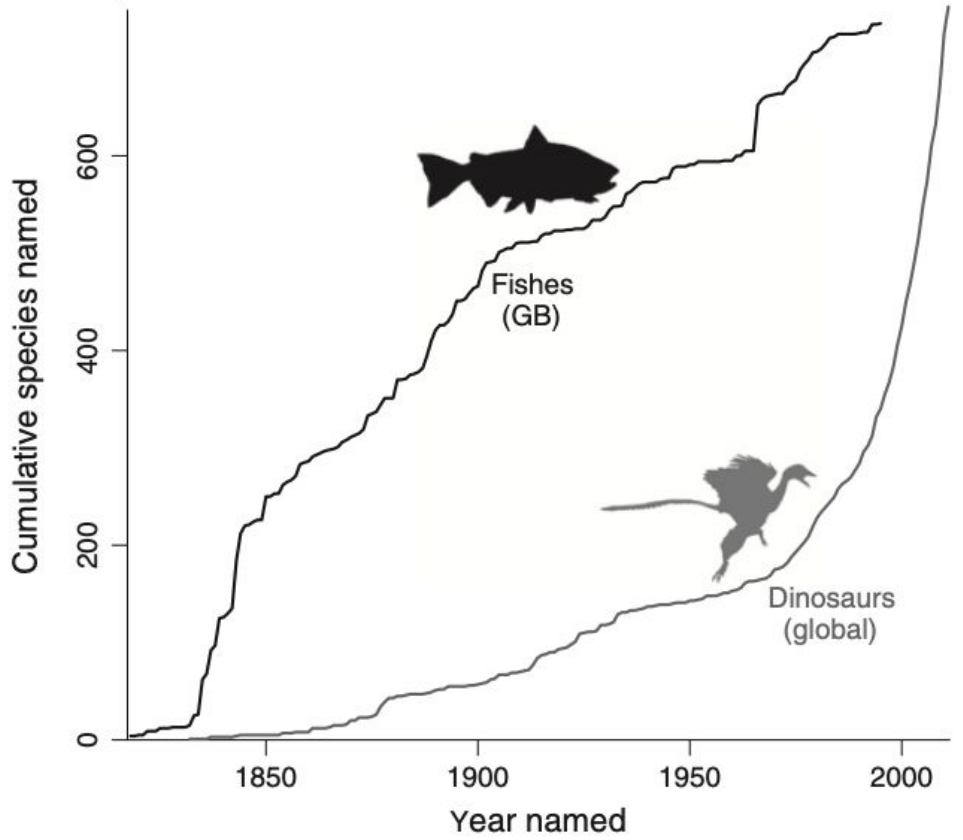
- Resulted in an even greater exploration of the factors that influence analyses of the fossil record
- Sampling standardisation methods begin to be developed
- Also, large compilations of data (e.g. PBDB) and data sharing/open data have been making even more analyses possible



Smith (2007) *Geol Soc*

Research interest over time

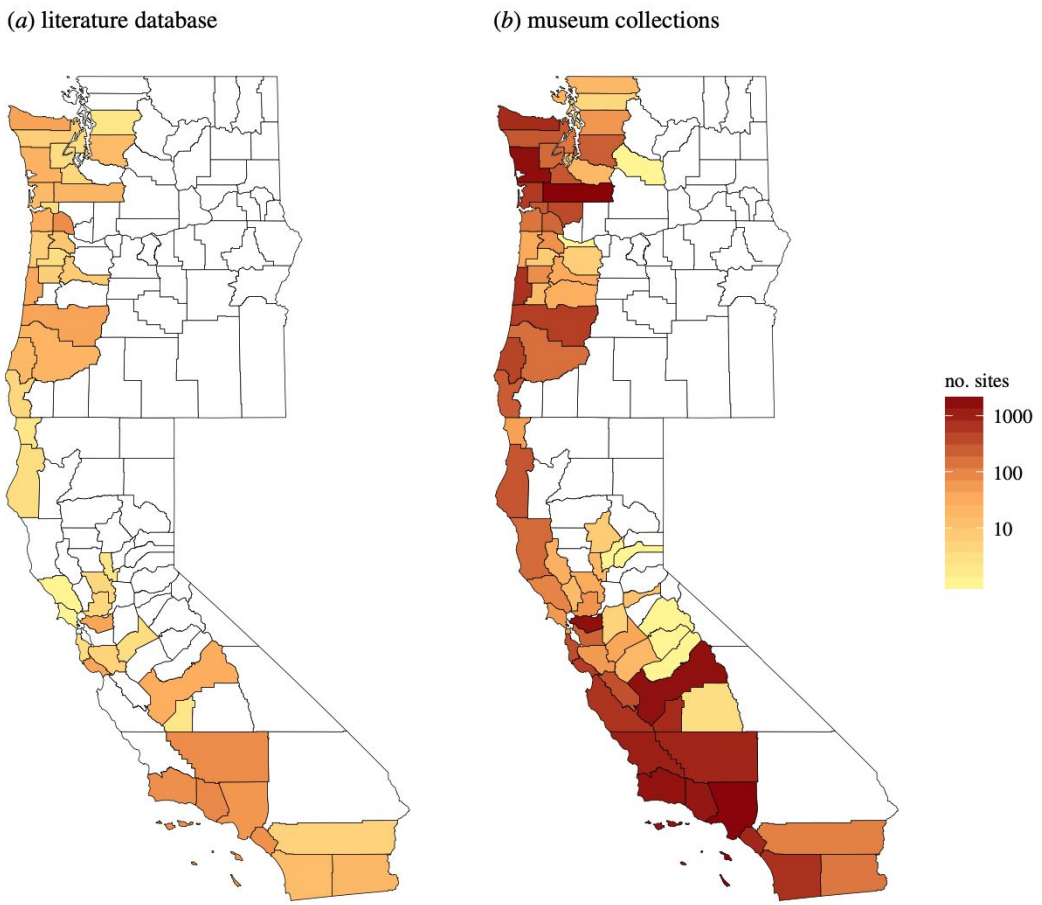
- Some groups are well understood and well sampled e.g. British fossil fishes
- Other groups have only become popular relatively recently (e.g. dinosaurs), or some have stalled in the number of new taxa named each year



Lloyd & Friedman (2013) *Palaeo3*

Insights into collecting biases

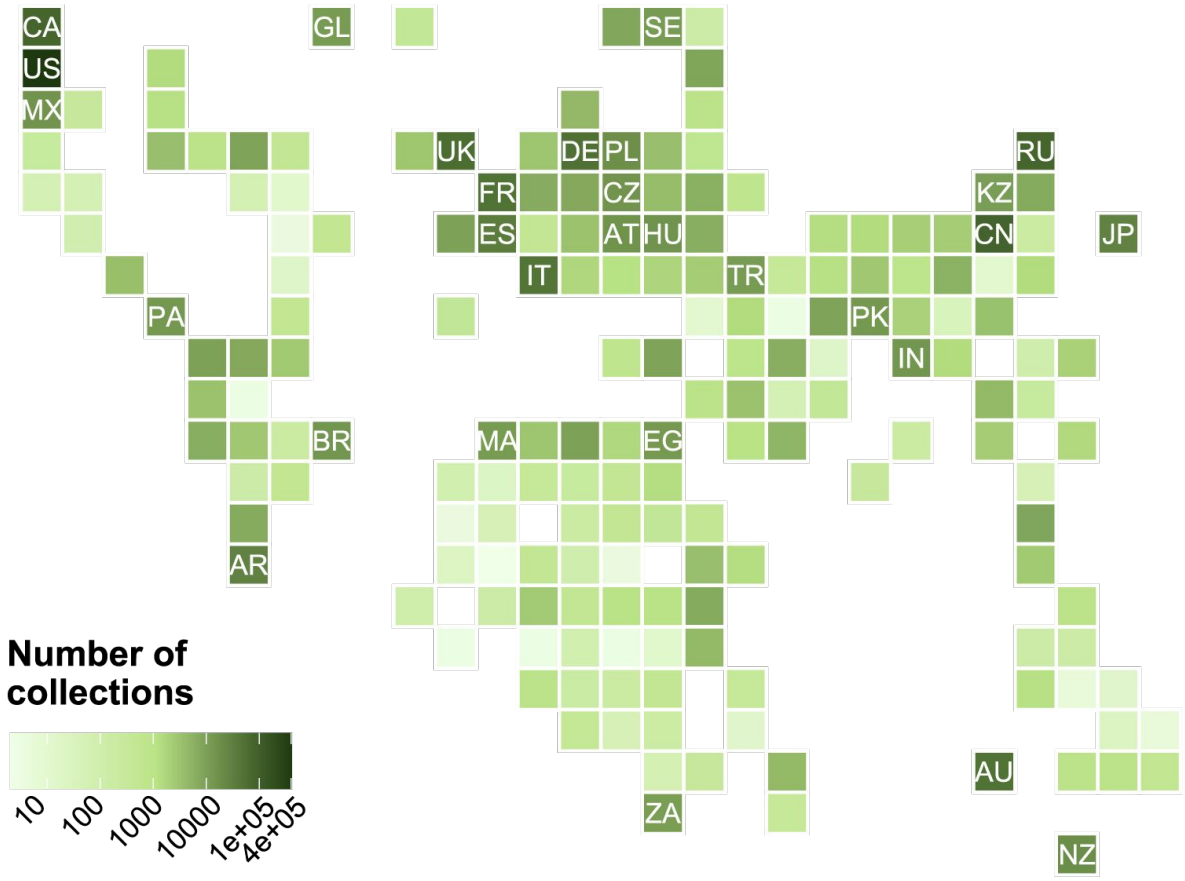
- Online occurrence databases rely on data from published literature
- Museum collection could hold up to 23 times more data (at least for marine invertebrates on the west coast of America)



Marshall *et al.* (2018) *Biol. Lett*

Insights into socio-economic biases

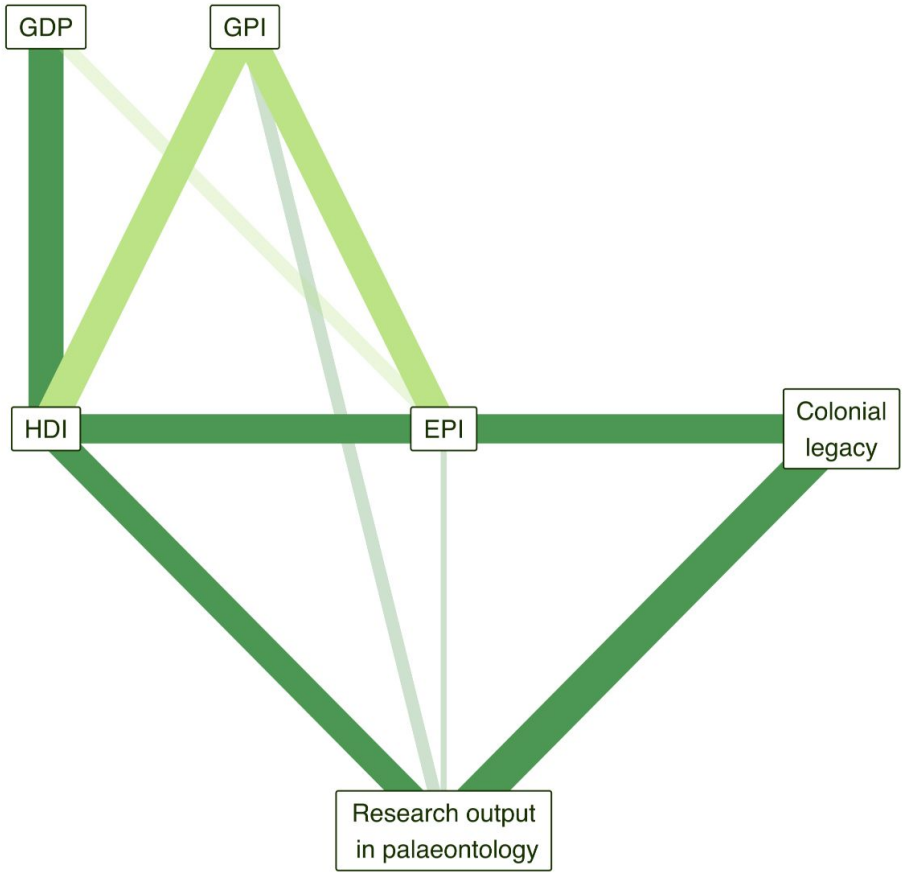
- Palaeontological research is dominated by researchers based in middle- and high-income countries
- 97% of fossil occurrence data in the PBDB were generated by researchers in northern America and western Europe



Raja & Dunne *et al.* (2022) *Nat. Ecol Evol.*

Insights into socio-economic biases

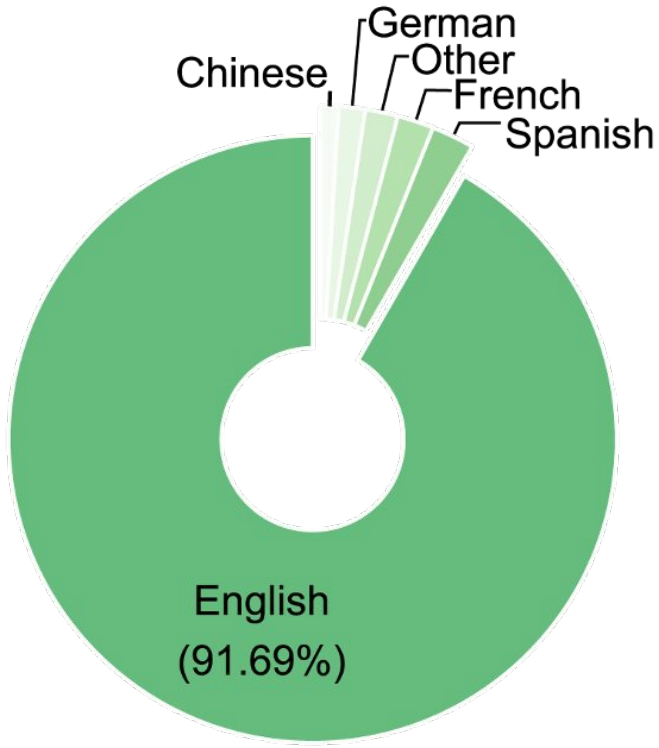
- Greater research output in palaeo (counted as co-authored publications) is linked to:
  - Higher GDP
  - Higher HDI
  - Greater security (GPI)
  - A history of, or profiting from, colonialism



Raja & Dunne *et al.* (2022) *Nat. Ecol Evol.*

Insights into socio-economic biases

- English is the dominant language in palaeodiversity studies
- Knowledge in other languages is overlooked – this has been shown to bias outcomes of meta- analyses (see Konno *et al.* 2020 *Ecol. Evol.*)
- Impedes the communication of science



Languages of references in the PBDB

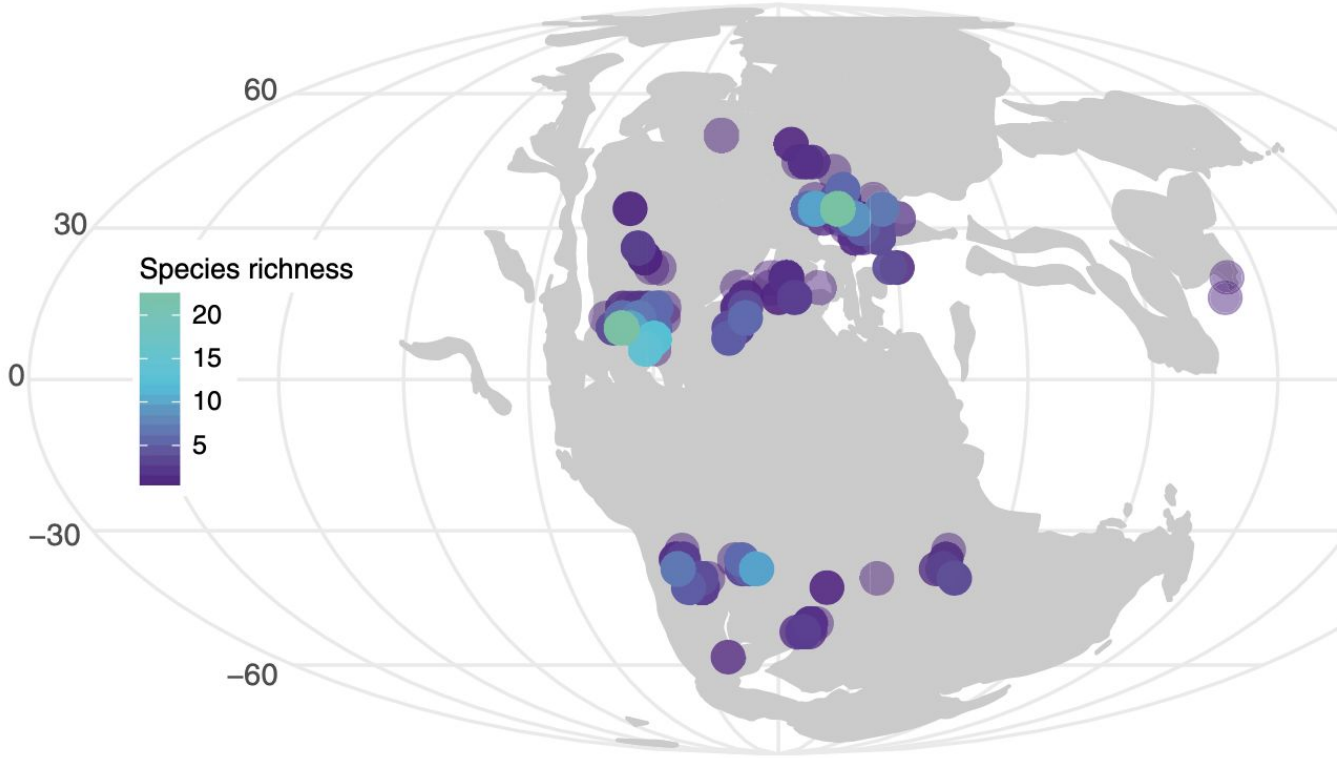
Raja & Dunne *et al.* (2022) *Nat. Ecol Evol.*

B. Detecting and quantifying fossil record biases

Data visualisations

An important step in any research project!

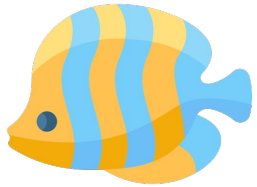
- What do your data look like?
- What patterns jump out at you?
- What happens when you use a different method/scale/colour scheme?

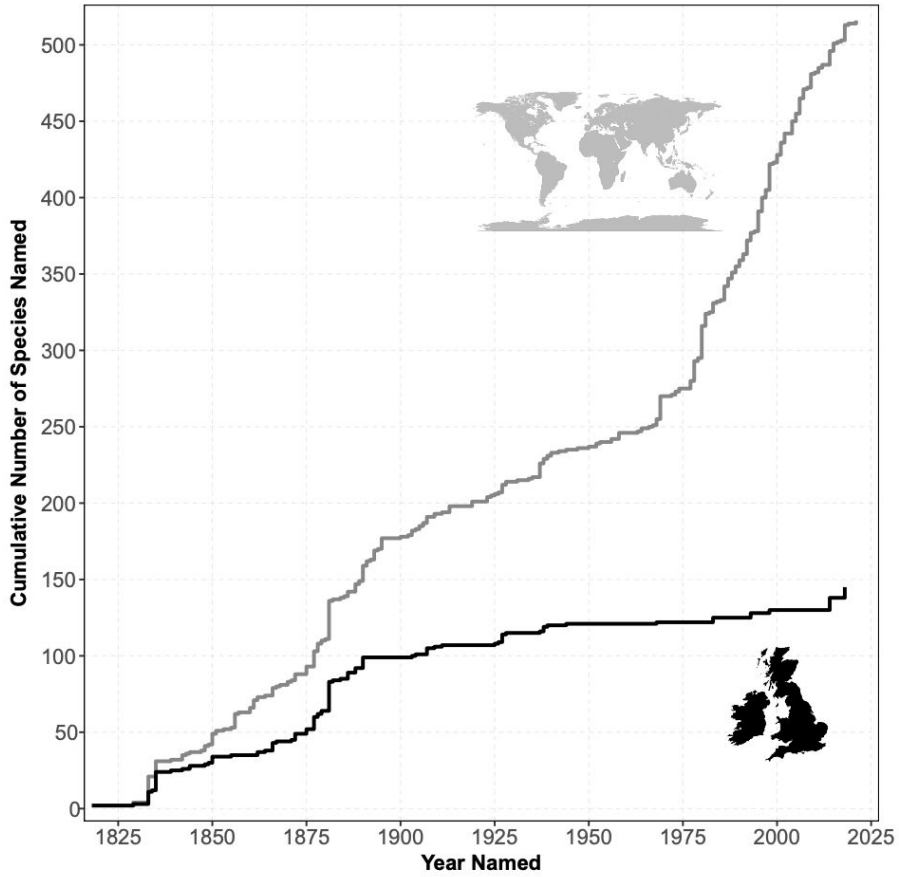


Dunne *et al.* (2020) *Paleontology*

Collector's / species accumulation curves

- Allow comparison of diversity across assemblages or to evaluate the benefits of additional sampling
- **Example**: Species named each year in different regions
- As research interest grows (and new areas open up) - expect pattern to increase and eventually level off

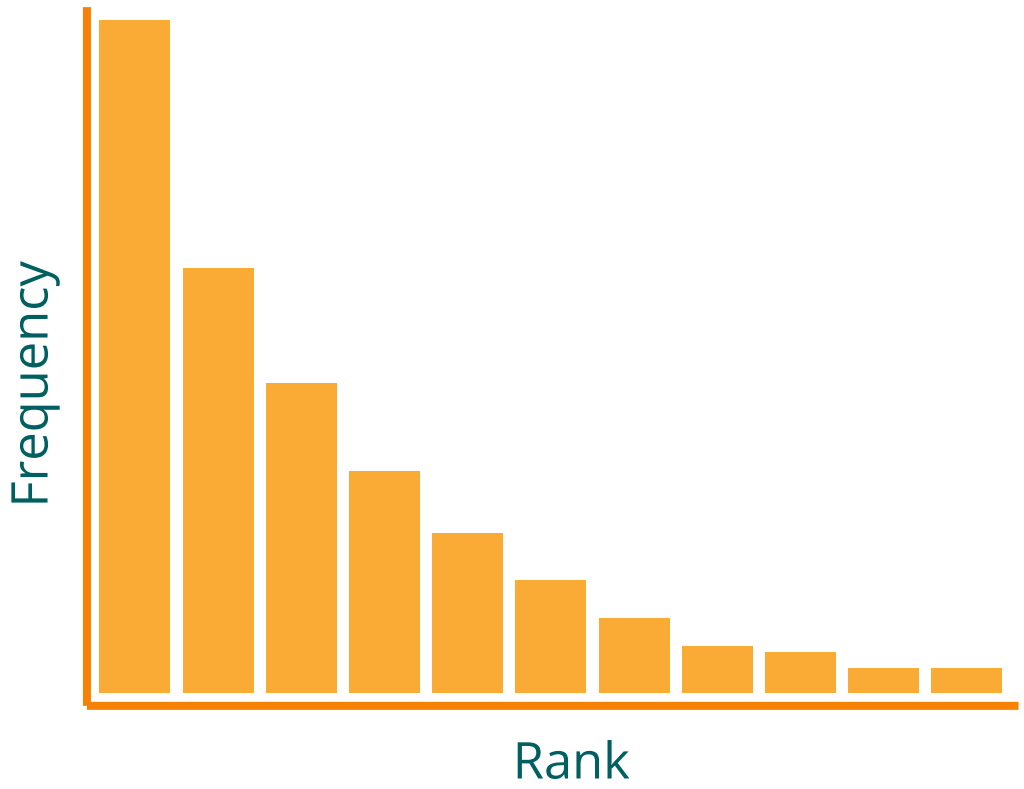




Henderson *et al.* (2021) *EarthArXiv*

Rank order abundance

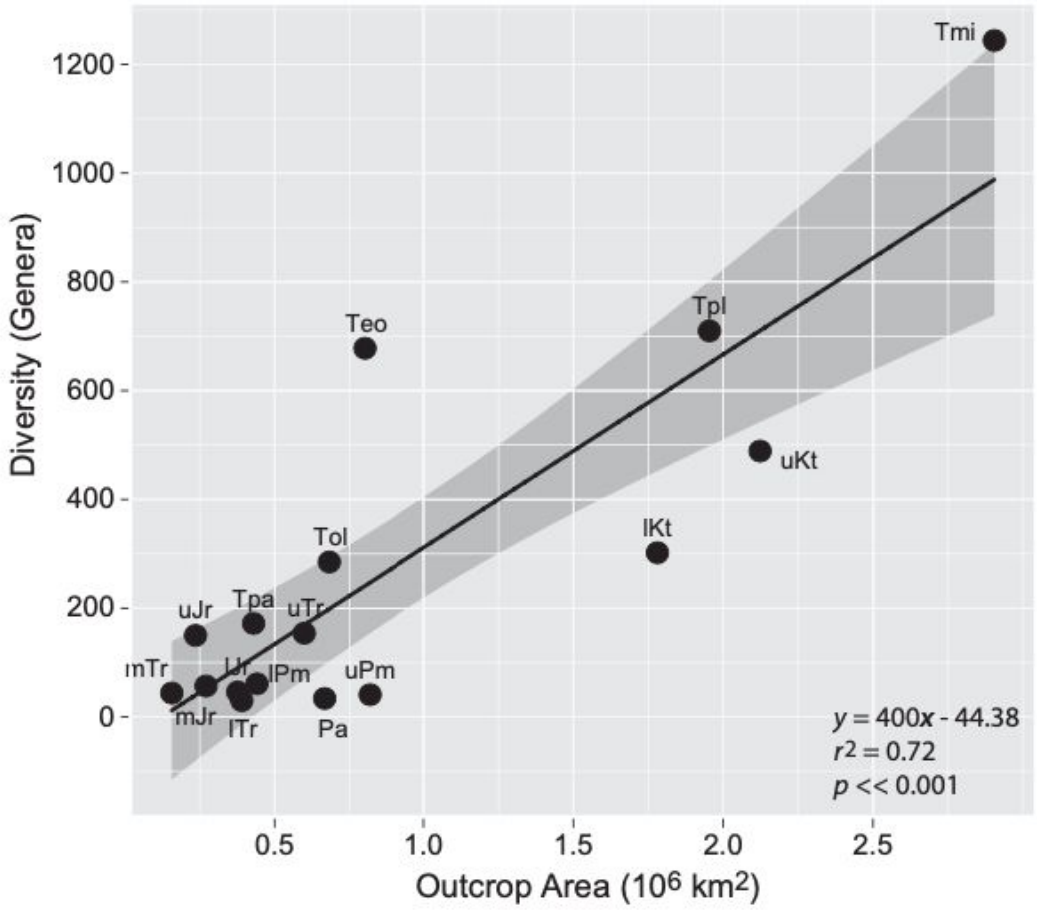
- Visualise species richness and species evenness
- Most abundant species = rank 1
- Least abundant species will be recorded less often



Simple correlations

- Use regression plots and simple statistics to quantify the strength of the correlation between proxies for sampling effort and fossil diversity

- But what to use as a proxy for sampling?

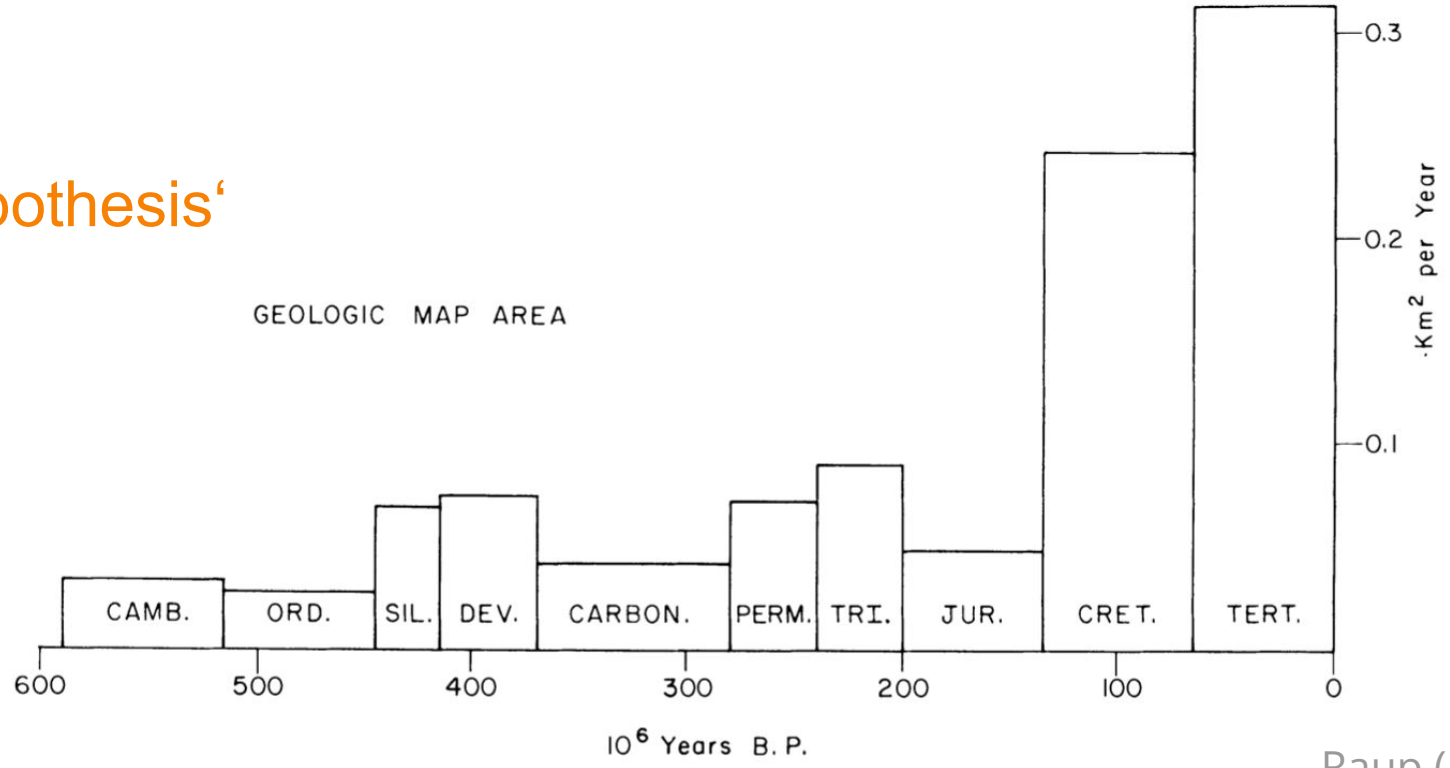


Wall *et al.* (2011) *Geol Soc*

Proxies for sampling & the ‘sampling hypothesis’

- Raup documented that species diversity correlates with sedimentary rock volume and area

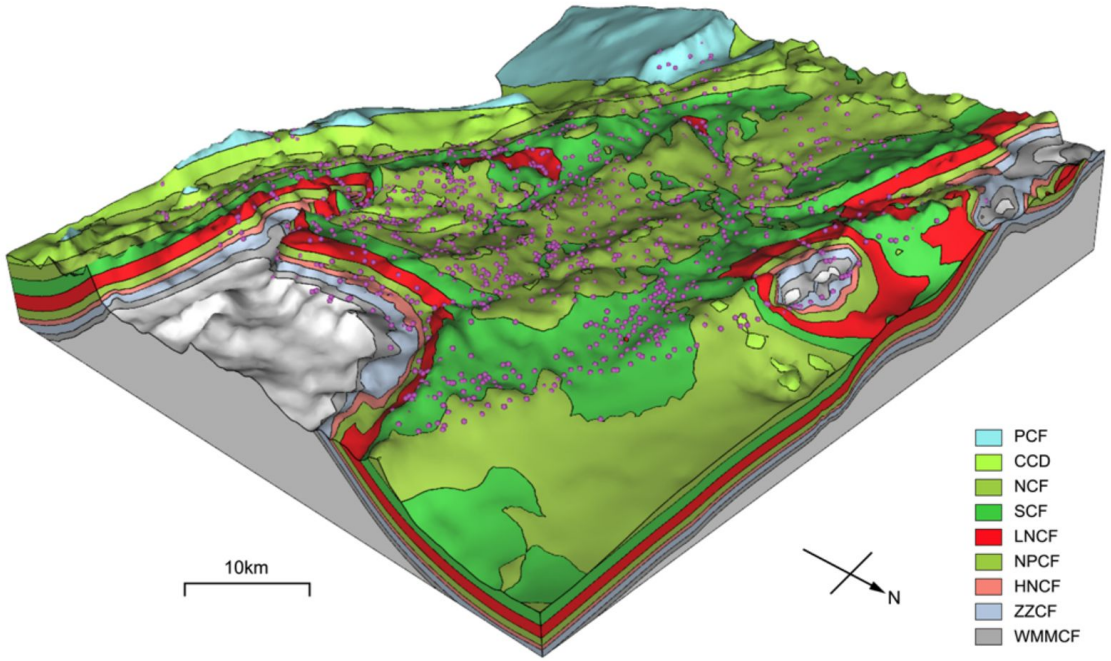
- This is the ‘sampling hypothesis‘ or ‘bias hypothesis’



Raup (1976)

Proxies for sampling - 1. Outcrop area

- **Definition**: The amount of rock available for sampling
- Good proxy for rock volume and area
- Rock must be exposed for fossils to be sampled

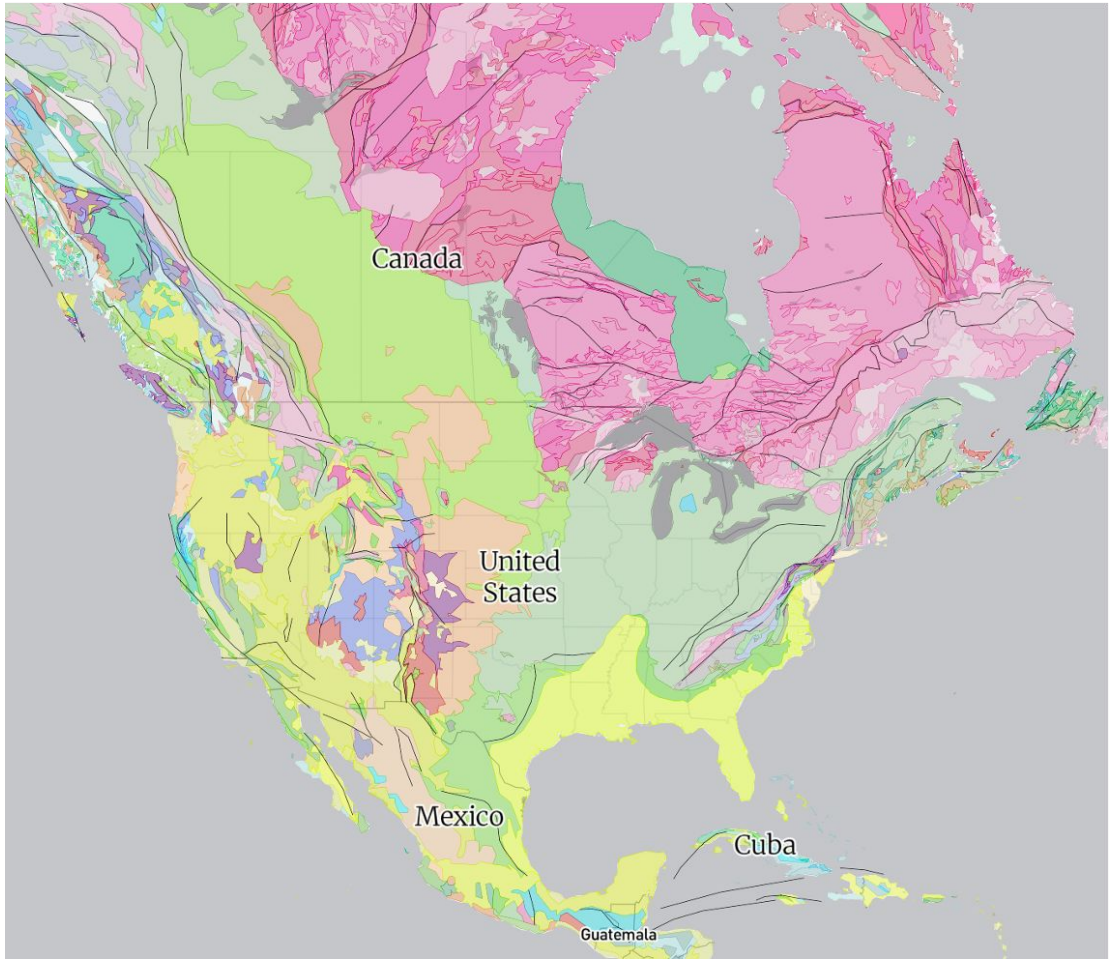


Walker *et al.* (2017)

Proxies for sampling - 1. Outcrop area

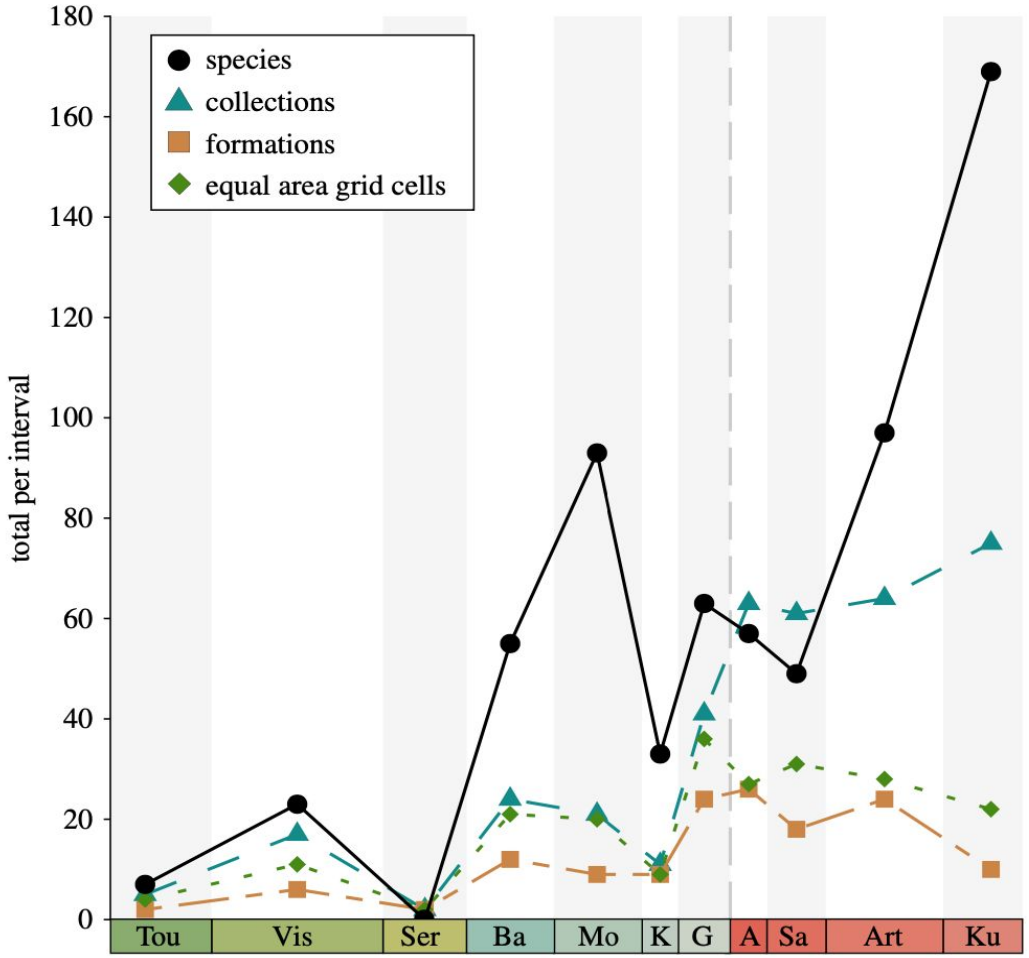
Macrostrat

- Geologic map database
- Focus on North American and other specific regions e.g. the Caribbean
- Download geological maps from various providers



Proxies for sampling - 2. Formations

- **Definition**: lithostratigraphic units that contain fossils (fossiliferous)
- Mostly well-studied, geologically constrained areas
- Good proxy for rock area and volume as well as human effort
- Vary in their geographic size



Dunne *et al.* (2018) *Proc Royl Soc B*

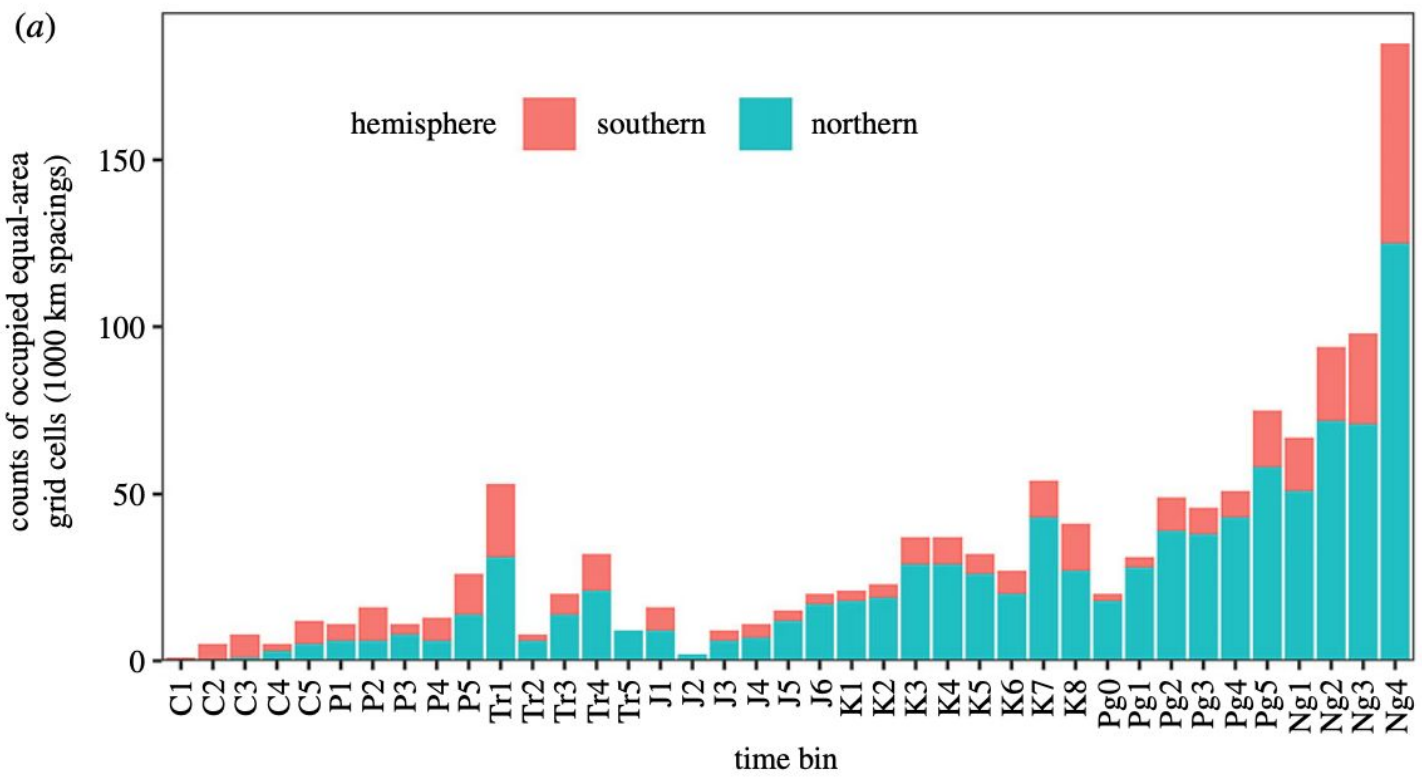
Proxies for sampling - 3. Fossil sites / collections

- **Definition**: distinct localities that contain fossils (‘collections’ in the PDBD)
- Definitions vary between research groups, taxa, etc.
- Great proxy for human effort

Image: A.A. Chiarenza

Proxies for sampling - 4. Occupied grid cells

- **Definition**: Equal-area grid cells on a map that contain fossils
- Standardised geographical areas
- Check out the R package dggridR



Close *et al.* (2020) *Proc. Royl. Soc. B.*

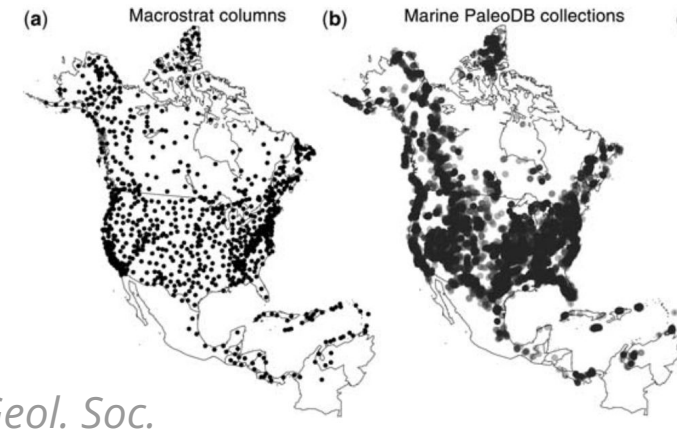
Other sampling hypotheses

Redundancy hypothesis

- **Definition**: Sampling and diversity are entirely or partially redundant with each other
- Proxies for sampling can rely on the presence of fossils (rarely their absence)
- Formations are sometimes defined by fossils

Common-cause hypothesis

- **Definition**: Both sampling and diversity are driven by some common factor(s)
- Examples: fluctuations in sea level, environmental perturbations or tectonic activity



Peters & Heim (2011) *Geol. Soc.*