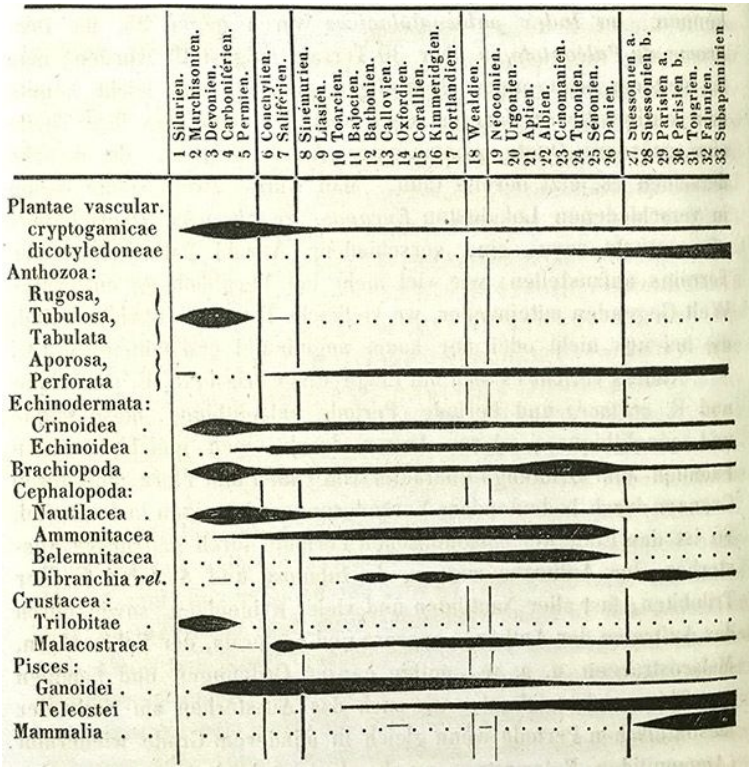
Databases in Paleobiology

Emma Dunne | APW 2024 | Wed. Aug. 7th

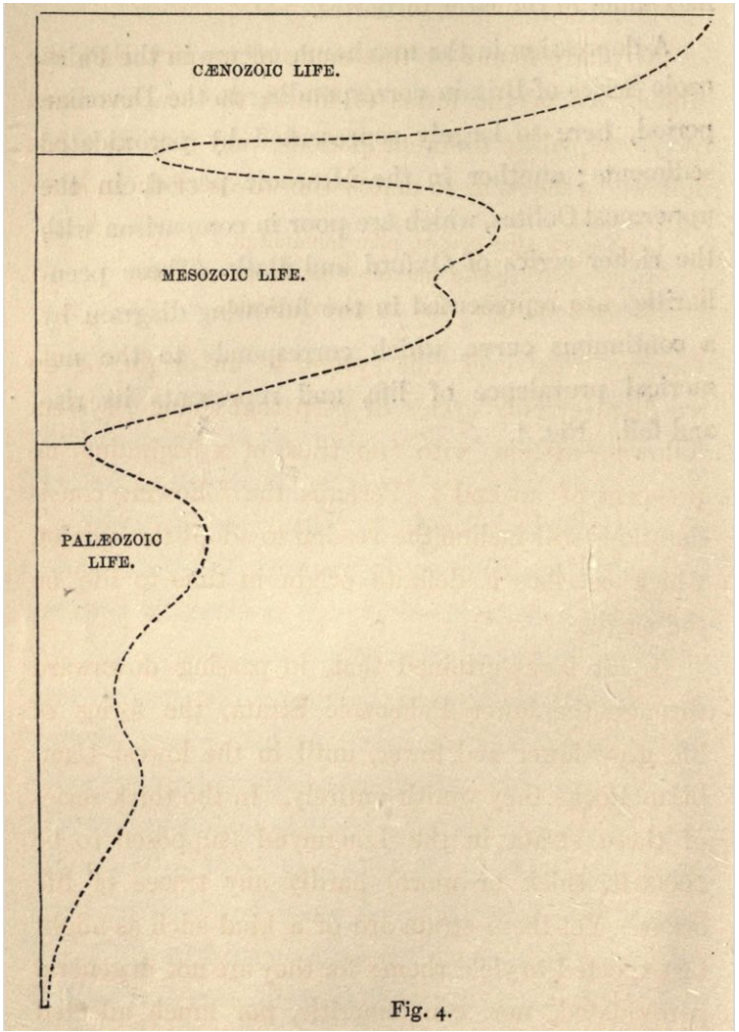
Databases in paleobiology

- Heinrich Georg Bronn (1800s), German geologist and paleontologist
    - Spindle diagrams - document originations (beginning of line), extinctions (end of line), and abundance (thickness)



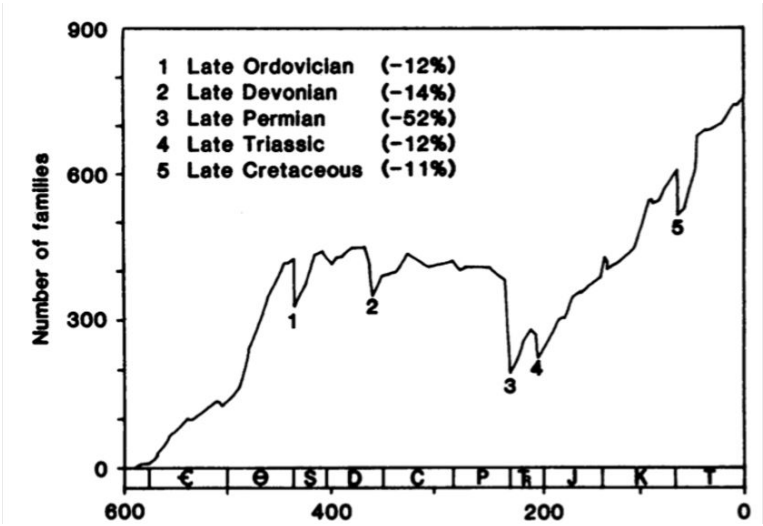
Databases in paleobiology

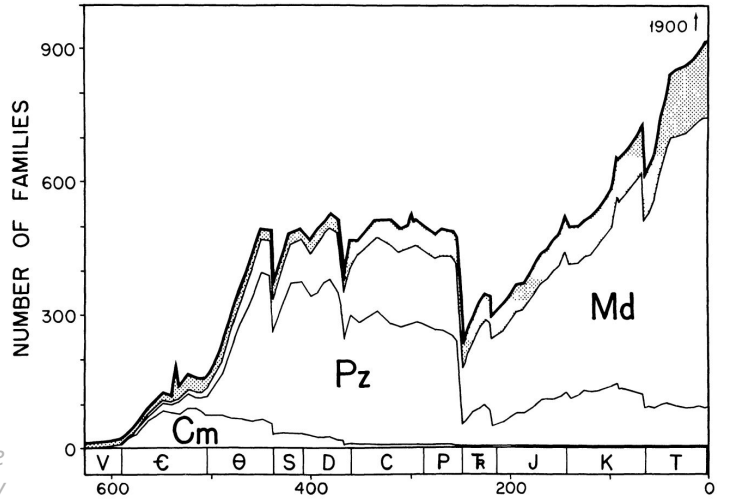
- Heinrich Georg Bronn (1800s), German geologist and paleontologist
    - Spindle diagrams - document originations (beginning of line), extinctions (end of line), and abundance (thickness)
- John Phillips (1860), British geologist
    - First ‘diversity curve’
    - Based on compilation of the British fossil record



Databases in paleobiology

- James Valentine, Elisabeth Vrba, Jack Sepkoski, David Raup, Michael Benton
- Sepkoski's compendium (1970-1980s)
  - First digital database
  - Phanerozoic global marine invertebrate fossil record
  - The "Big Five" mass extinctions
  - See the online genus database and `sepkoski` R package

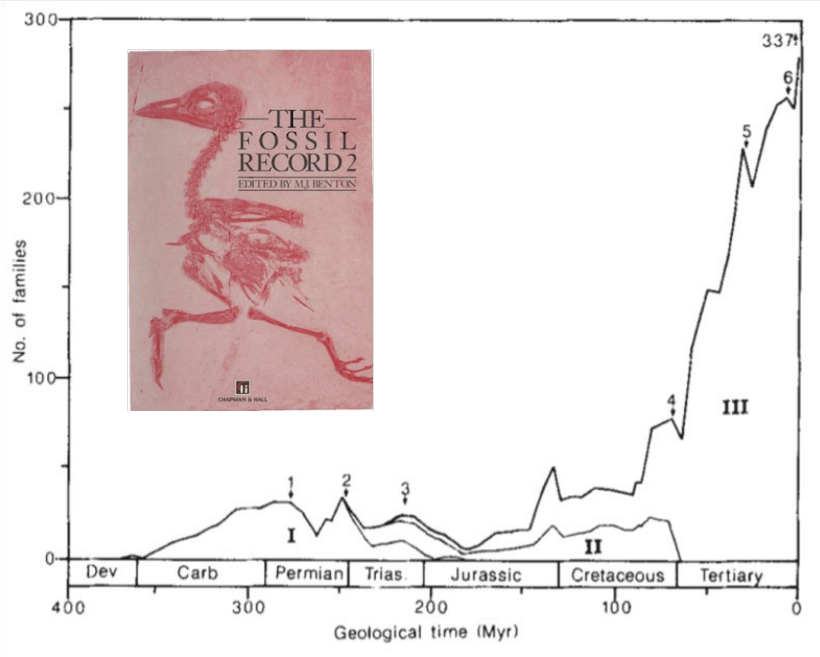




Raup & Sepkoski (1982) *Science*
Sepkoski (1984) *Paleobiology*

Databases in paleobiology

- James Valentine, Elisabeth Vrba, Jack Sepkoski, David Raup, Michael Benton
- Sepkoski's compendium (1970-1980s)
  - First digital database
  - Phanerozoic global marine invertebrate fossil record
  - The "Big Five" mass extinctions
- Benton's The Fossil Record dataset (1990s) - vertebrates



Benton (1985) *Nature*

Databases in paleobiology

Databases in paleobiology

WoRMS
World Register of Marine Species

DigPaleo

GBIF
Global Biodiversity Information Facility

NEOTOMA PALEOECOLOGY DATABASE

MORPHO SOURCE

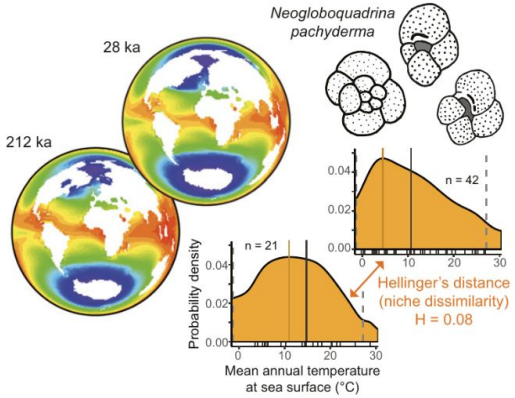
MORPHOBANK
HOMOLOGY OF PHENOTYPES OVER THE WEB

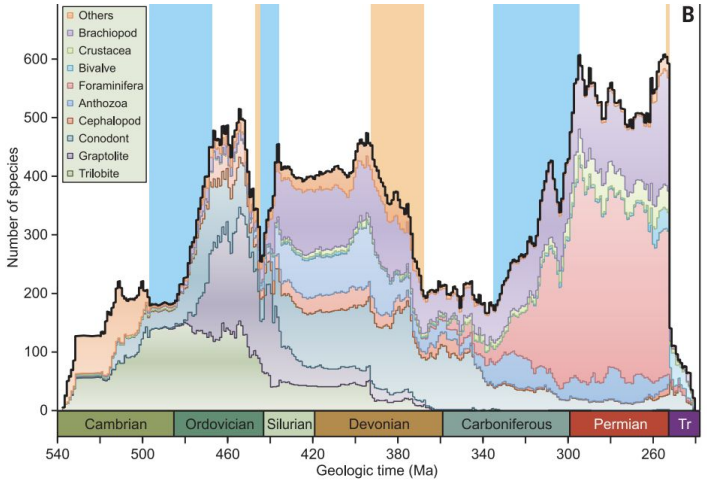
phenome 10k

Neptune SB ANDBOX Berlin

Geobiodiversity Database

Database positives

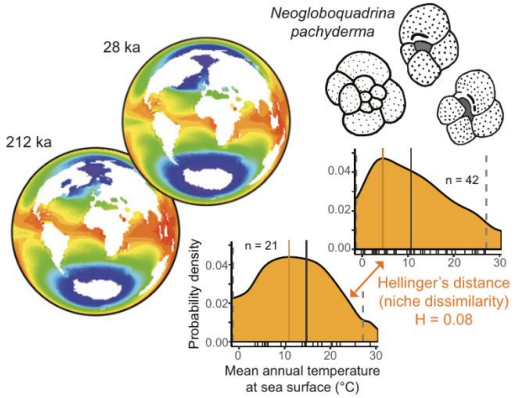


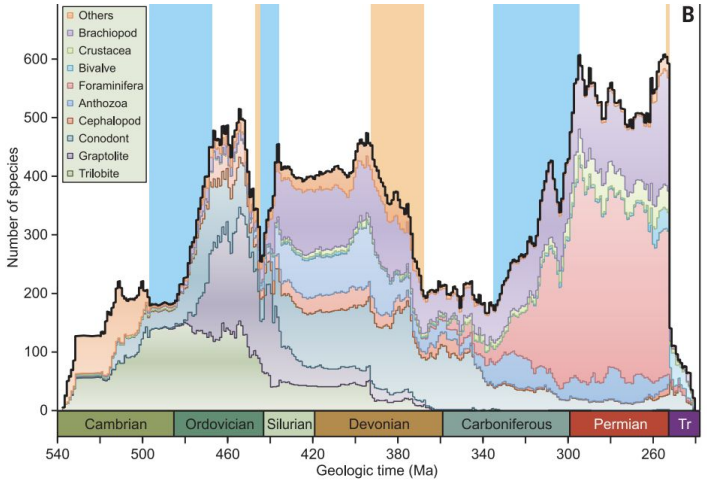


Antell *et al.* (2021) *PNAS*, Fan *et al.* (2020) *Science*

Database positives

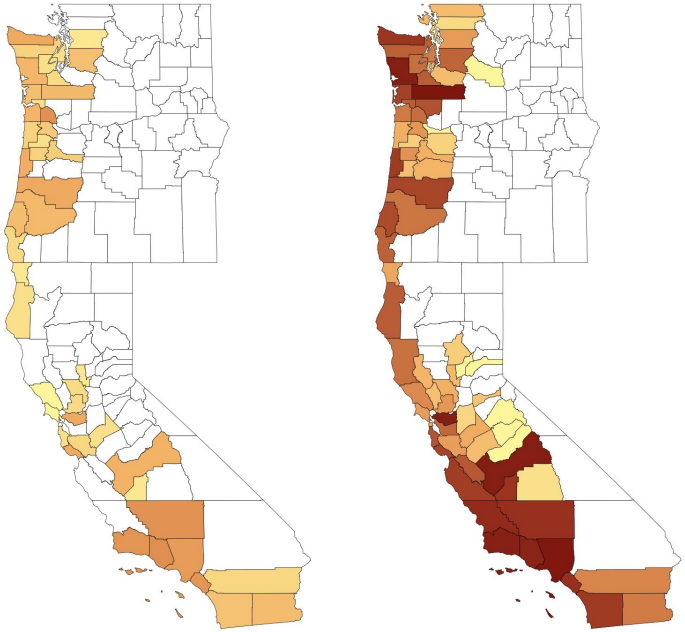
- Deeper insights past biodiversity, evolutionary patterns, and extinction
- Continually open up new research avenues
    - Promote methods development
- Increase transdisciplinary opportunities
- Improve data access and sharing
- Promotes data consistency
- Greater data security





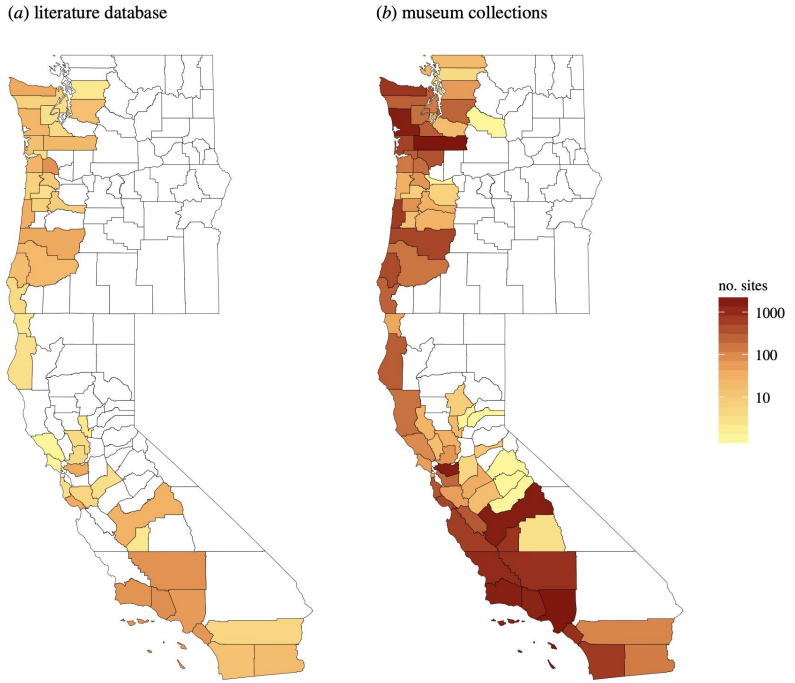
Antell *et al.* (2021) *PNAS,* Fan *et al.* (2020) *Science*

Database challenges



Database challenges

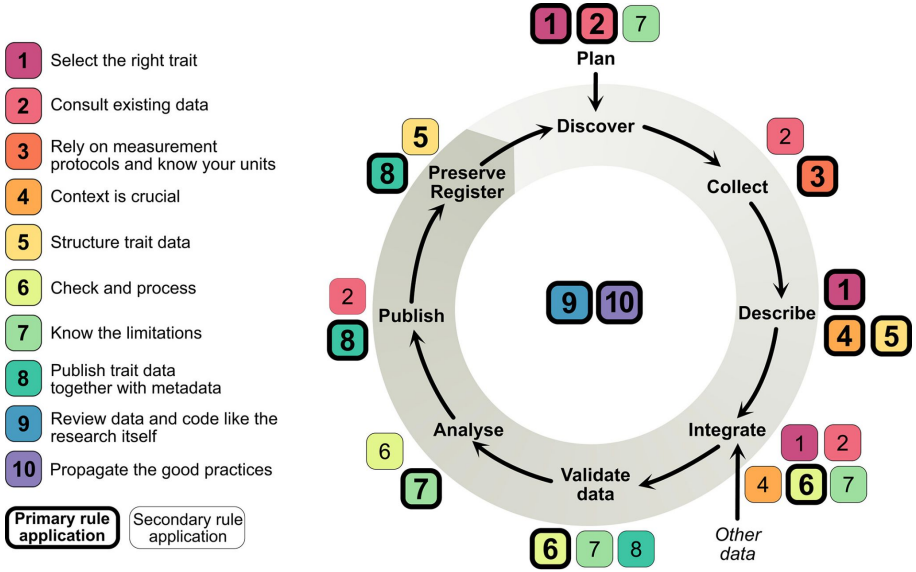
- Funding is short-term, scattered and sporadic - maintenance issues
- Data coverage still poor (e.g. museum specimens, geographical/economic biases)
- Integration across platforms and data types is challenging
- Redundancy across databases
- Data collation/curation is not adequately credited (nor is data sharing, etc.)



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

Database initiatives

- New coordination networks to communicate, innovate, and standardize research practices, training, and educational activities
    - e.g. for paleoclimate (NSF)
- Establishment of new data lakes to bridge existing data and disciplines
    - e.g. IRAL (Paleosynthesis)
- Community-led initiatives for future-proofing (e.g. Keller *et al.* 2022)



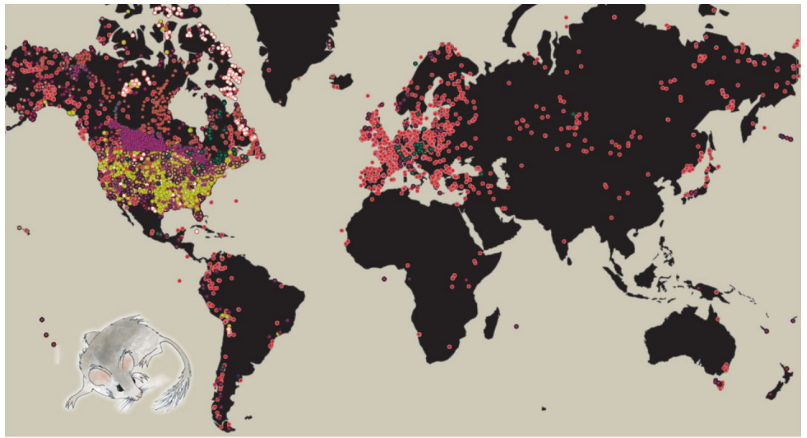
Databases used in paleo research

- No single database is perfect - each one has advantages and disadvantages
- Choose the one that works for you
- **Join in the efforts!**
    - Collect and add data to existing databases
    - Curate data in existing databases
    - Dedicate time to learning and sharing your data (and code)

See Table S1 in Dillon, Dunne *et al.* (2023) *Paleobiology*

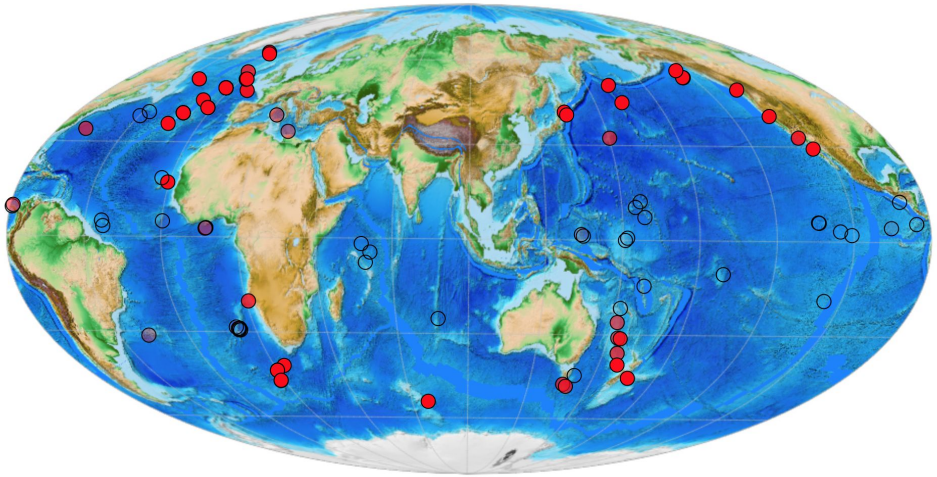
Neotoma Paleoecology Database

- Neotoma is a database of databases
  - Includes North American Pollen Database and fossil mammals (FAUNMAP)
  - Fossil pollen, vertebrates, diatoms etc. from the Pliocene-Quaternary
  - Focus on global-change research
- Publicly available (CC-BY 4.0 license)
- R package: `neotoma2` (via GitHub)



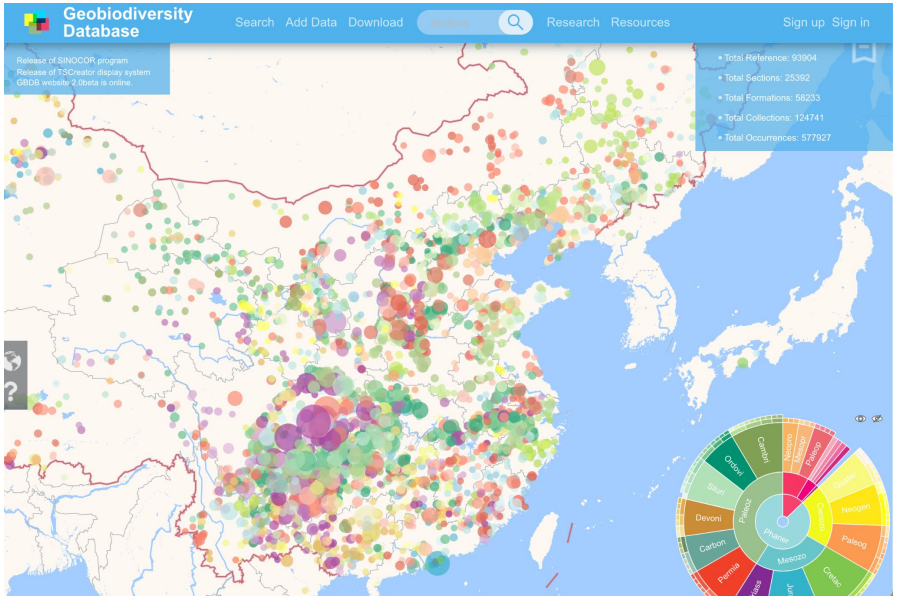
Neptune database

- Available publicly through the Neptune Sandbox Berlin
- ShareAlike 4.0 International (CC BY-SA 4.0) license
- Occurrence records for thousands of marine plankton microfossil species from hundreds of deep-sea ocean drilling sections
- Inspired by Sepkoski's database which, in turn, inspired the PBDB



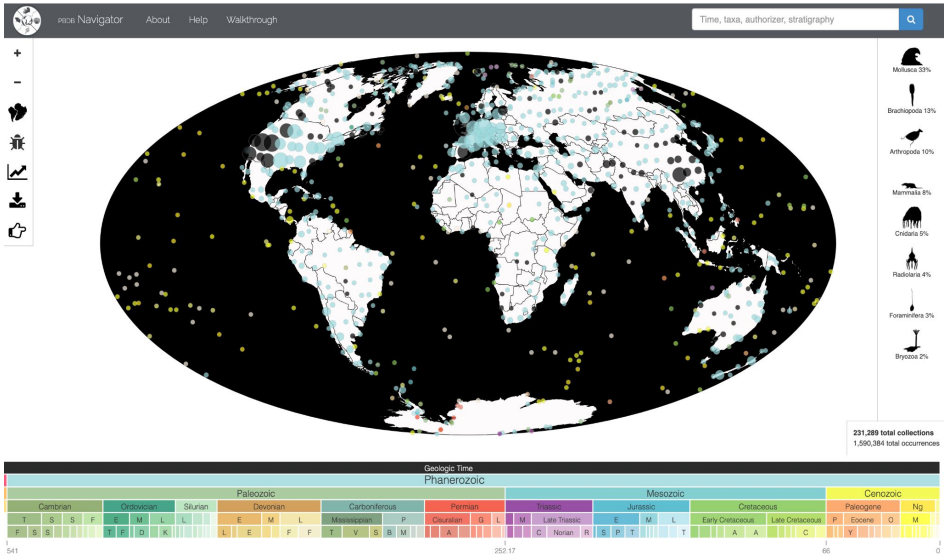
Geobiodiversity Database

- GBDB began in 2006
- Database of the International Commission on Stratigraphy
- Geological section-based system (Fan *et al.* 2013)
- Includes fossil occurrences, taxonomy, and descriptions
- Focus on invertebrates
- Currently focused on China

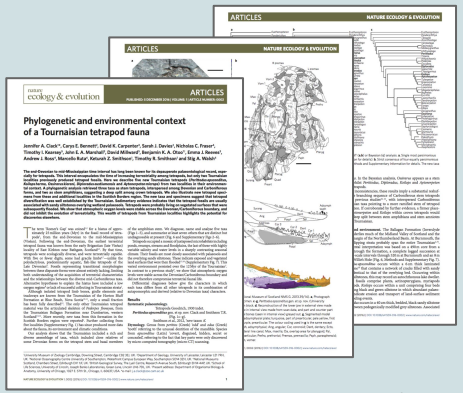


Paleobiology Database (PBDB)

- Began in 1998 as the Phanerozoic Marine Paleofaunal Database (John Alroy & Charles Marshall)
    - Contains data from the Sepkoski compendia
- Global occurrence data across Phanerozoic and beyond
- Taxonomic info and specimen data
- CC BY 4.0 license



PBDB structure & workflow



REFERENCES
Primary literature

DATA


Taxonomy


Specimen info


Location

Stratigraphy

DATA ENTRY & CURATION

STORAGE & SHARING

ANALYSES & PUBLICATION

What is a “collection” in the PBDB?

“Collection” can refer to several different things:

1. Individual localities (i.e. sites where fossils are found)
2. Small collection of fossils collected from same place
3. Corresponding to museum collection or research trip