

Obtaining data for palaeobiological analysis



Bethany Allen

What types of data?

- Occurrences
- Taxonomy
- Traits and/or morphology
- Stratigraphic information
- Geological data
- Palaeoenvironmental information
- Phylogenies
- ...

Paleobiology Database

paleobiodb.org

Over 20 years old, mostly funded by NSF

Coverage is global, good for macrofossils throughout geological time

Can be explored using the Navigator, but data can also be downloaded

All fossils entered by palaeontologists from the published literature



The Paleobiology Database
revealing the history of life

Paleobiology Database



The Paleobiology Database
revealing the history of life

Newly published user guide gives more information on

- Data structure
- How to enter data
- Caveats and common problems with data

Geobiodiversity Database

geobiodiversity.com



Over 25 years old, mostly funded by NSFC

Focus is on a stratigraphic approach, has links with the International Commission on Stratigraphy

Coverage is global but focused on China and the UK (due to links with British Geological Survey)

Fossils entered by a team of paid staff from published literature

Neotoma

neotomadb.org



Focuses on relatively recent fossils (up to 5Ma), good coverage for mammals and pollen, especially in North America

Has an explorer to browse data, but data can also be downloaded

All fossils entered by specialists for that clade, termed “investigators”

Has a dedicated R package

GBIF

gbif.org



Global Biodiversity Information Facility

Compiles modern and fossil occurrence data from a wide variety of sources, including direct reporting from museum collections

Also has extensive taxonomic records

iDigBio

idigbio.org



Integrated Digitized Biocollections

Specimen-level data, mostly compiled by museums across the US

Funded by NSF

Neptune

nsb.mfn-berlin.de



Specifically for marine planktonic microfossils

Most data comes from deep sea ocean drilling, and is dated using integrated stratigraphic age models

Triton is a cleaned and curated version of the foraminifera in this dataset

BioDeepTime

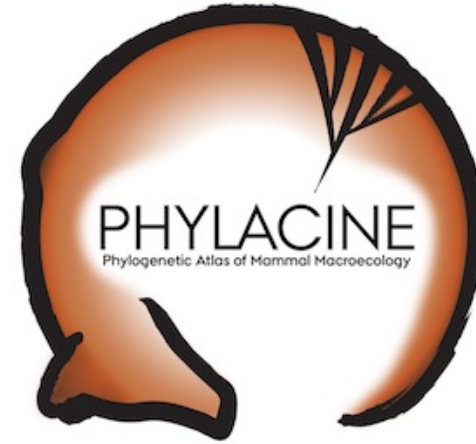
doi.org/10.1111/geb.13735

Curated dataset of high-resolution time series across a variety of temporal scales and resolutions (recent to deep time)

Includes terrestrial, marine and freshwater environments

PHYLACINE

megapast2future.github.io/PHYLACINE_1.2/



Global database for late Quaternary mammals, both extant and recently extinct

Includes a variety of data including traits, IUCN status, past and present ranges, and phylogenetic trees

Other useful resources

- IUCN: iucnredlist.org
- TreeBASE: treebase.org
- MorphoBank: morphobank.org
- Fossil Calibration Database: fossilcalibrations.org
- MorphoSource: morphosource.org
- Phenome10k: phenome10k.org
- Macrostrat: macrostrat.org
- EarthByte: earthbyte.org/category/resources/data-models/
- BRIDGE palaeoclimate models: bristol.ac.uk/geography/research/bridge
- CHELSA: chelsa-climate.org

Top tips

Databases only show presence, they don't tell you anything about absence, and abundance data is uncommon

Understand the underlying literature and construction processes

Make sure you take uncertainty into account

Check and clean the data – if it doesn't make sense, it's probably a mistake