Sampling standardisation

Emma Dunne

FAU Erlangen-Nürnberg

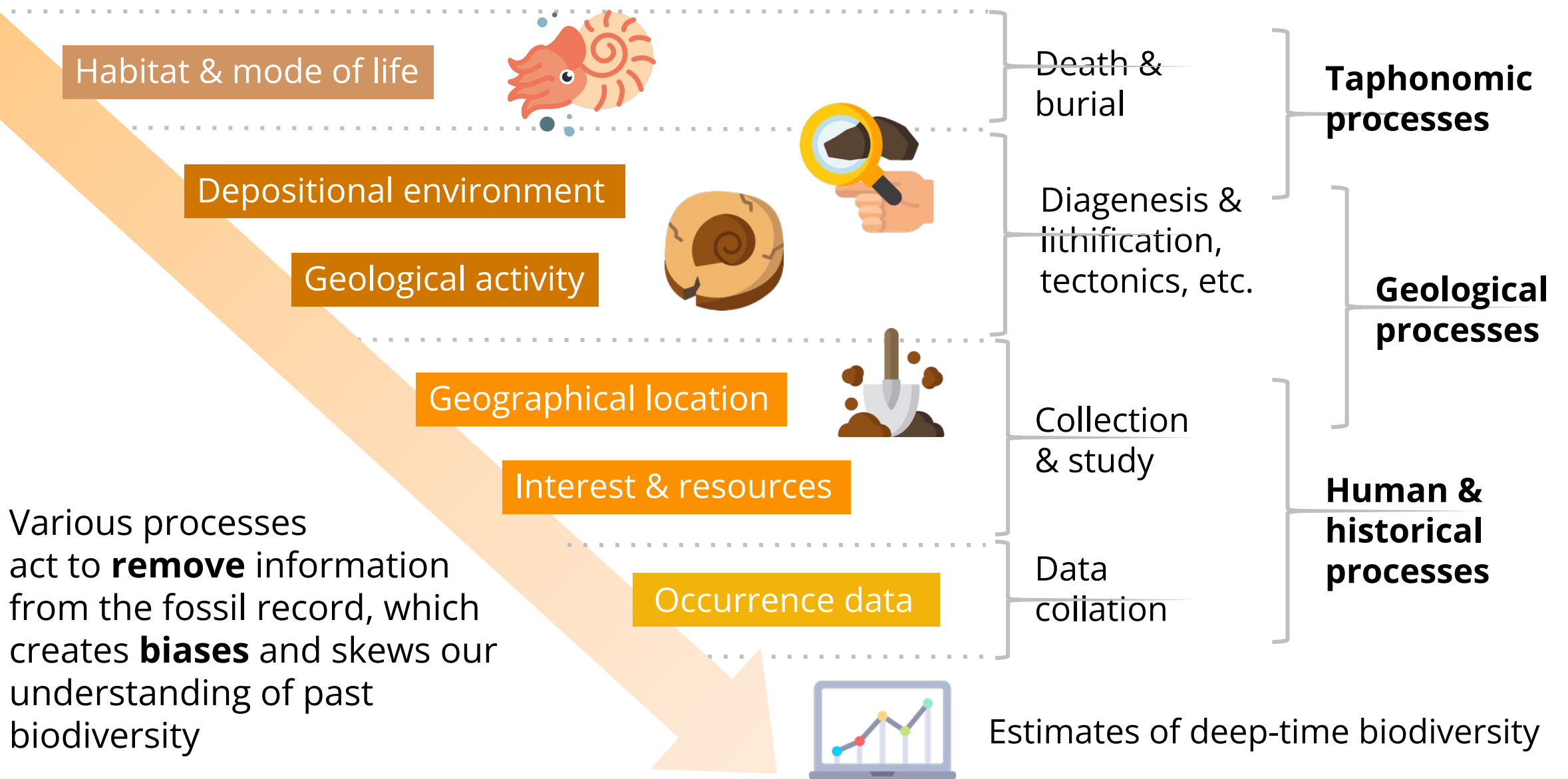Tuesday 29/08/2022

FROM DEATH TO DATABASE

Habitat & mode of life

Depositional environment

Geological activity

Geographical location

Interest & resources

Occurrence data

Death & burial

Diagenesis & lithification, tectonics, etc.

Collection & study

Data collation

Taphonomic processes

Geological processes

Human & historical processes

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

Estimates of deep-time biodiversity

Refresher: Scales of biodiversity

α

β

γ

Alpha diversity

Beta diversity

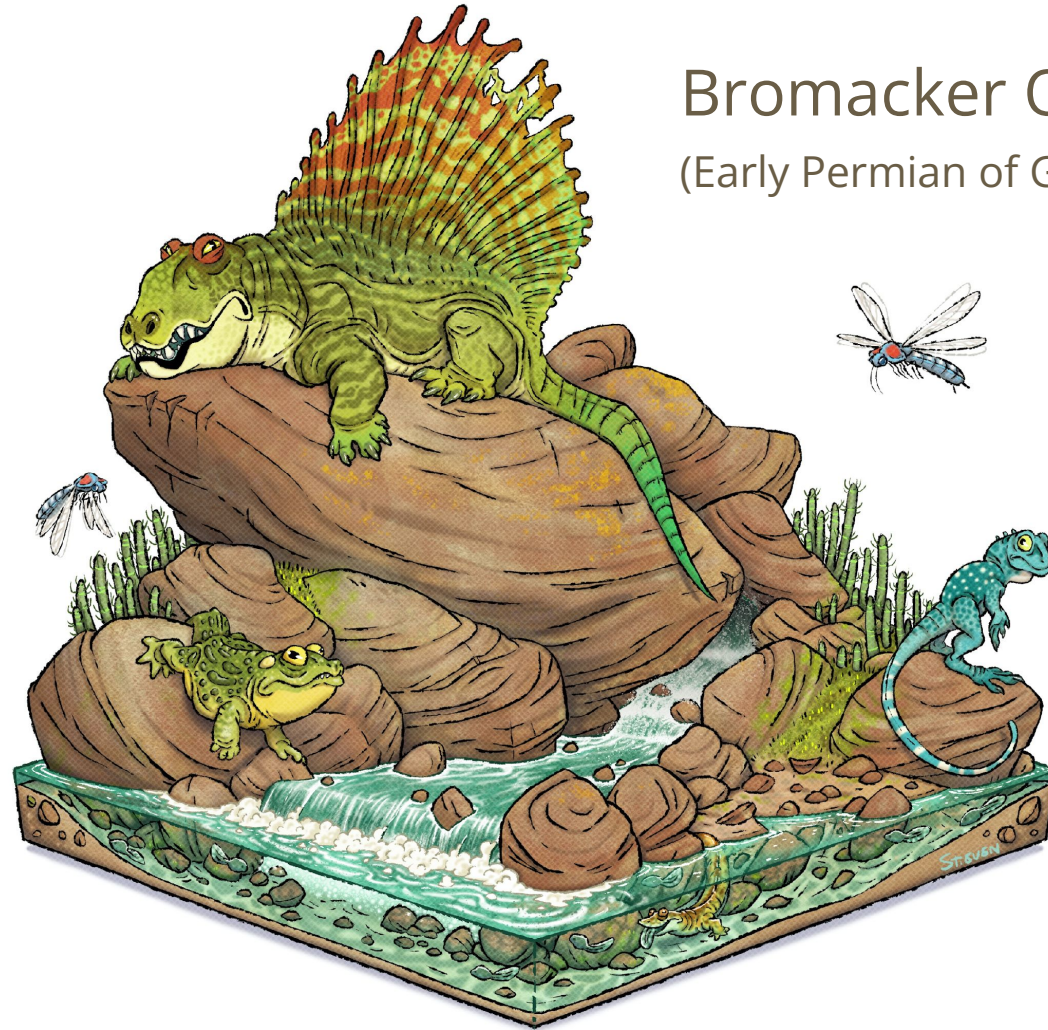Gamma diversity

"local richness"
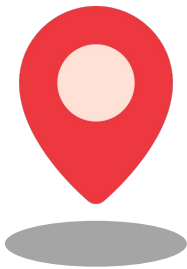
between two areas/regions

"global" diversity

α

Alpha diversity

How many species in the cartoon?

Bromacker Quarry
(Early Permian of Germany)

α

Alpha diversity
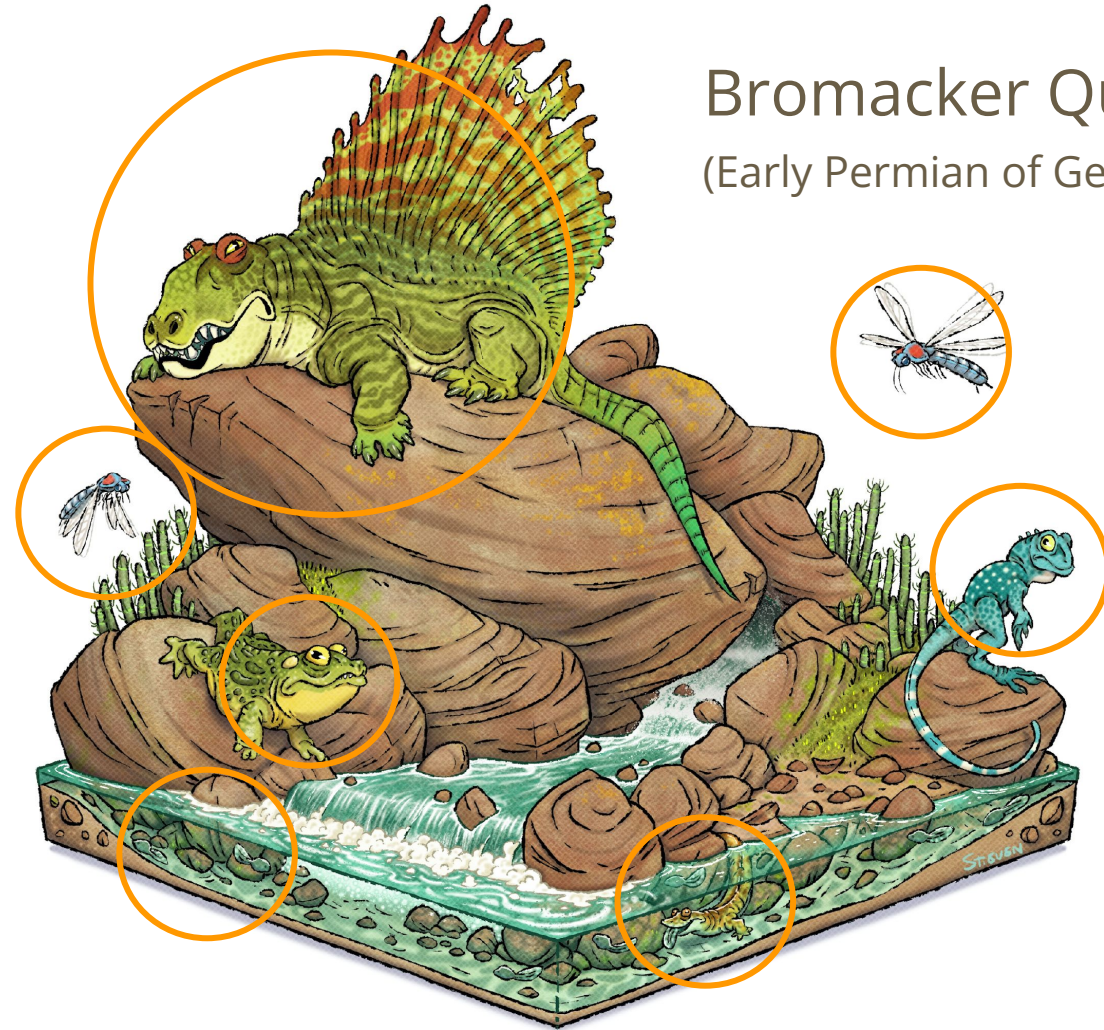
Bromacker Quarry
(Early Permian of Germany)

1 insect species
1 fish? species
4 tetrapod species

Local richness = 6 species

Art: Stieven Van der Poorten (@StievenVdP)
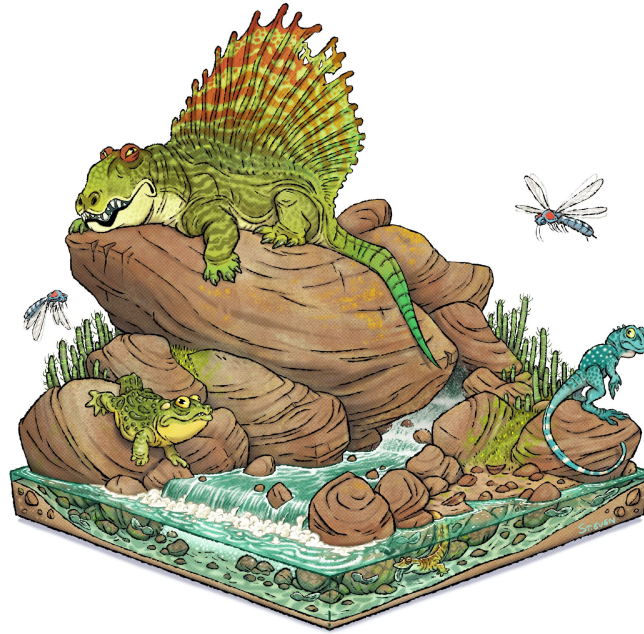
β

Beta diversity

Comparison between this sample and another sample e.g. two different regions

Bromacker Quarry
(Early Permian of **Germany**)

Arroyo Formation
(Early Permian of **USA**)

γ

Gamma diversity

"Global" diversity

e.g. species diversity during the early Permian across the globe
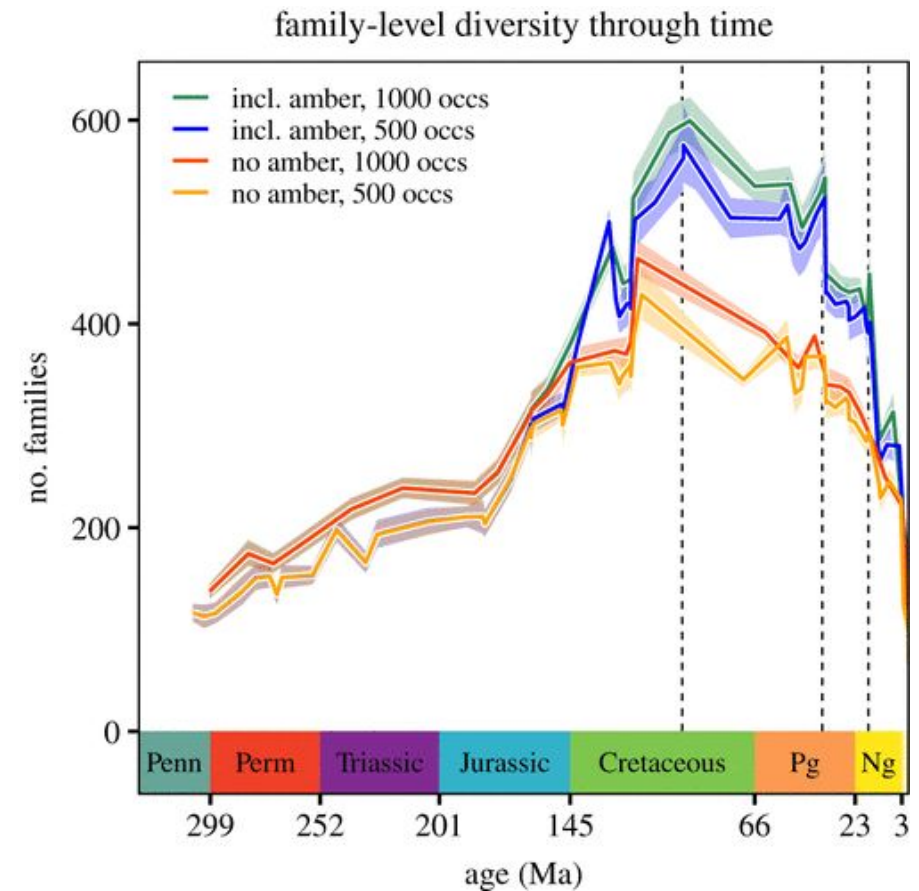
(NB: never truly global...)

Permian fauna

Methods covered in other sections

Capture-mark-recapture

- **Lee Hsiang Liow** later this week

- Estimating rates of origination and extinction and diversity-through-time

- Typically requires high resolution temporal data with abundant groups (e.g. marine invertebrates)

family-level diversity through time

Schachat *et al.* (2019) *Proc Royl. Soc. B*

Fossilized birth–death process

- We won't be covering this during this course

- If interested, speak to **Rachel Warnock** & **Laura Mulvey** next week

- Also check out the Phyloseminar recordings:

    - Dr April Wright

    - Prof Rachel Warnock

    - Dr Joëlle Barido-Sottani

Paleobiology, 46(2), 2020, pp. 137–157
DOI: 10.1017/pab.2020.12

PALEOBIOLOGY | Paleontological
A PUBLICATION OF THE | S O C I E T Y

Article

Assessing the impact of incomplete species sampling on estimates of speciation and extinction rates

Rachel C. M. Warnock, Tracy A. Heath, and Tanja Stadler

Warnock *et al.* (2020) *Paleobiology*

Taxic diversity-through-time
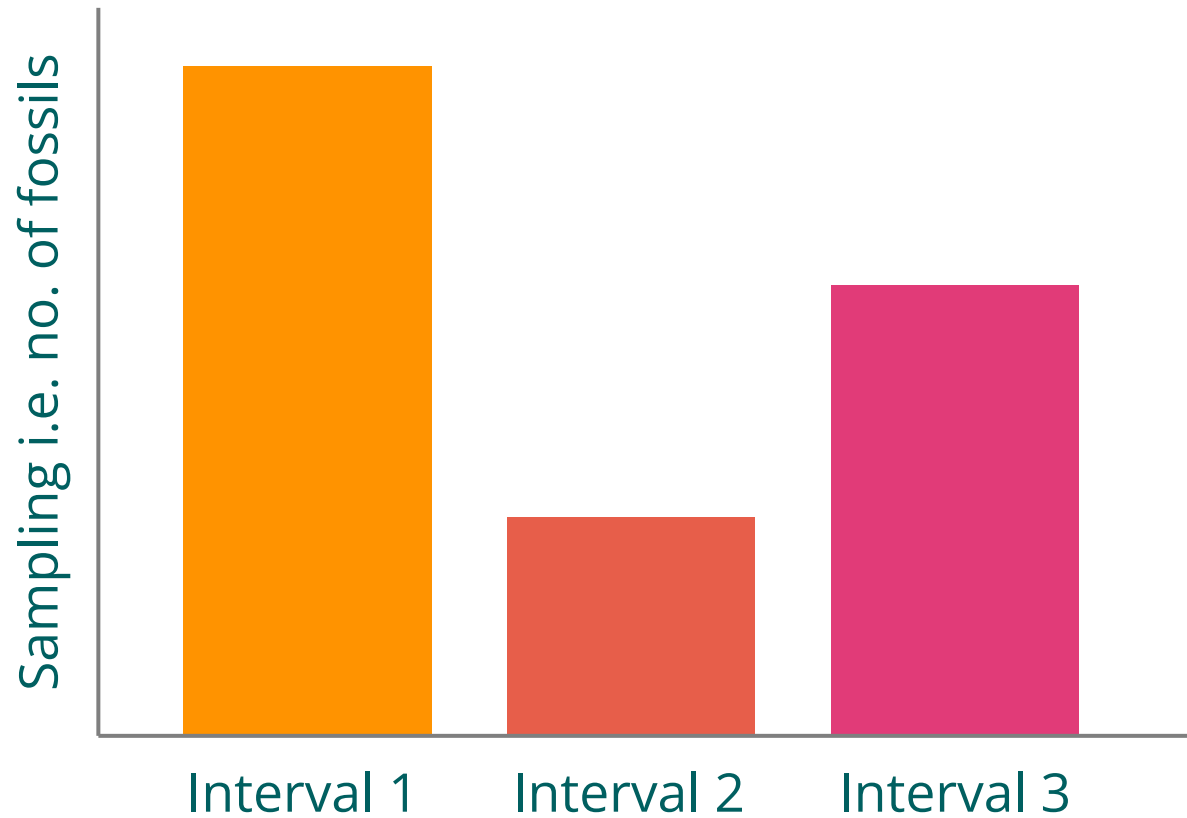
Rarefaction

Covered by Wolfgang last week

- Fixes to a sample size

- Can result in information being lost

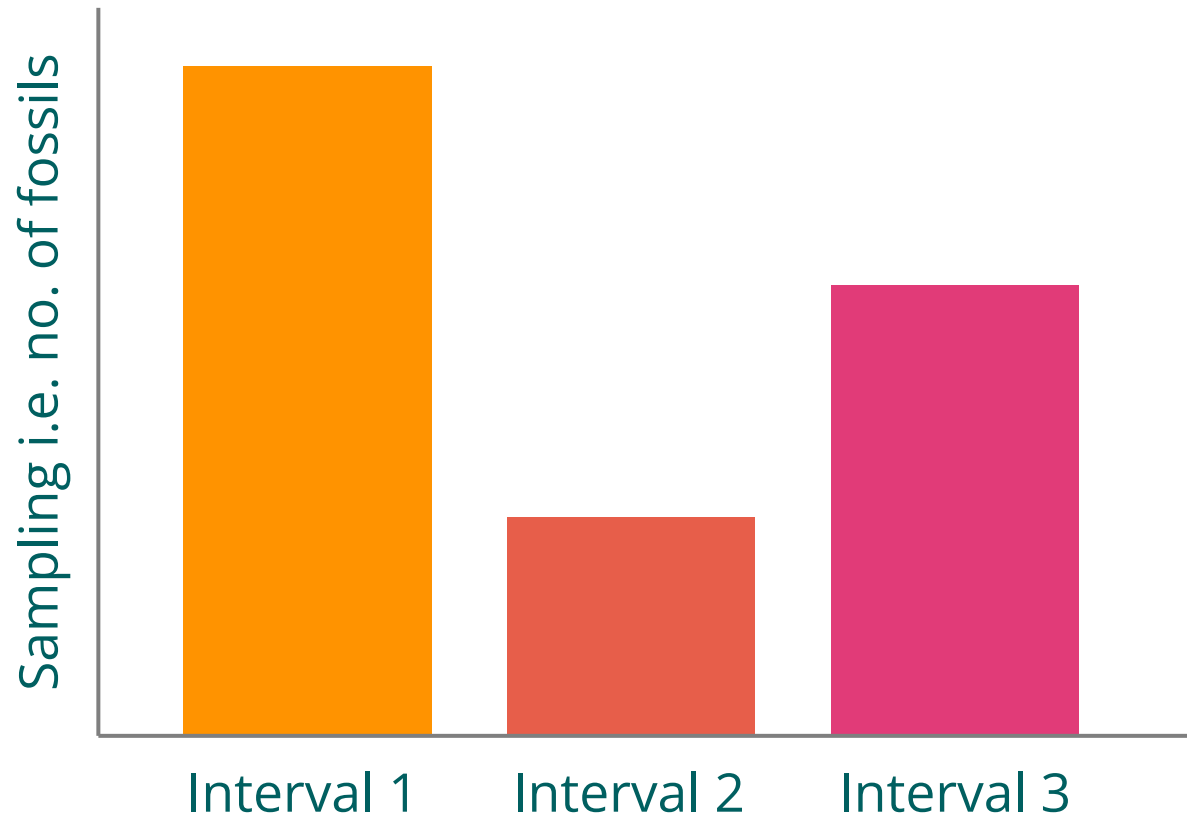- Rare taxa can be underrepresented in estimates when evenness is low
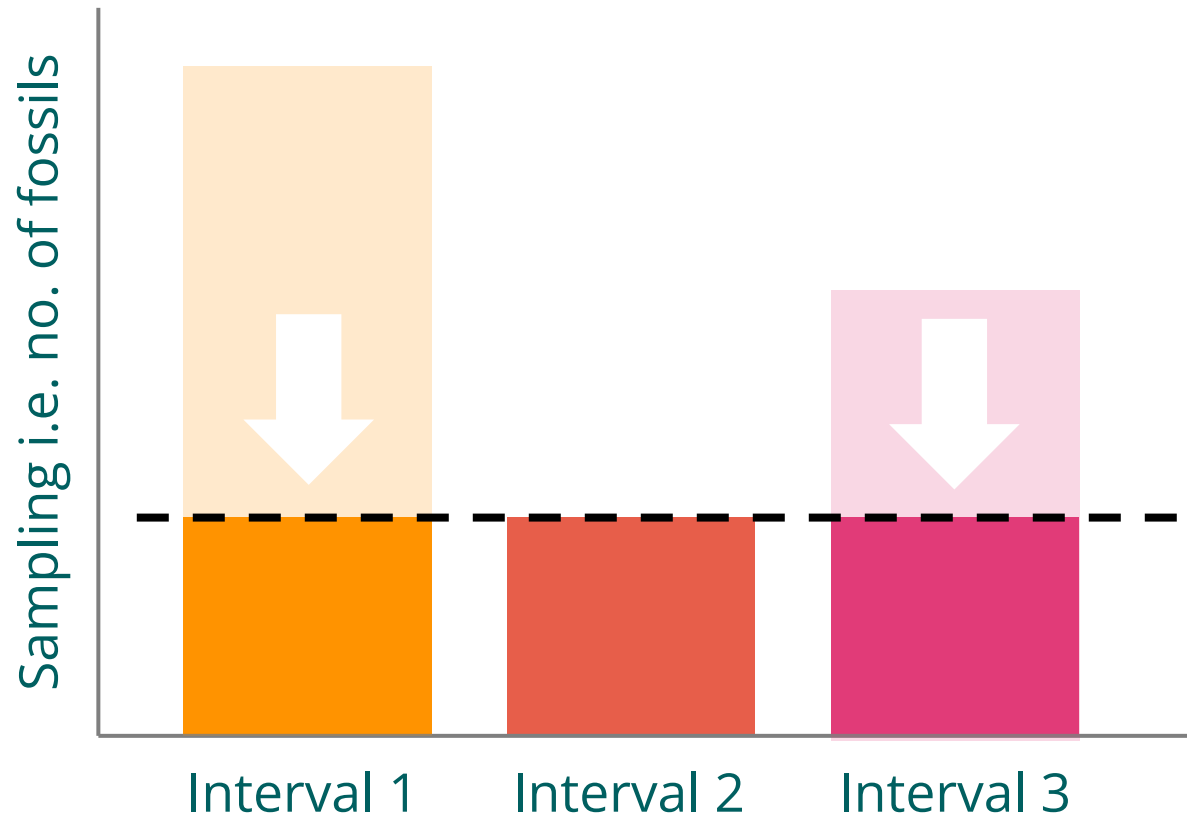
Hypothetical sampling scenario

Hypothetical sampling scenario

Classical rarefaction

- draws samples DOWN to the level of the least-well-sampled bin

Hypothetical sampling scenario

Classical rarefaction

- draws samples DOWN to the level of the least-well-sampled bin

Coverage-based methods

- Subsample to a fixed coverage
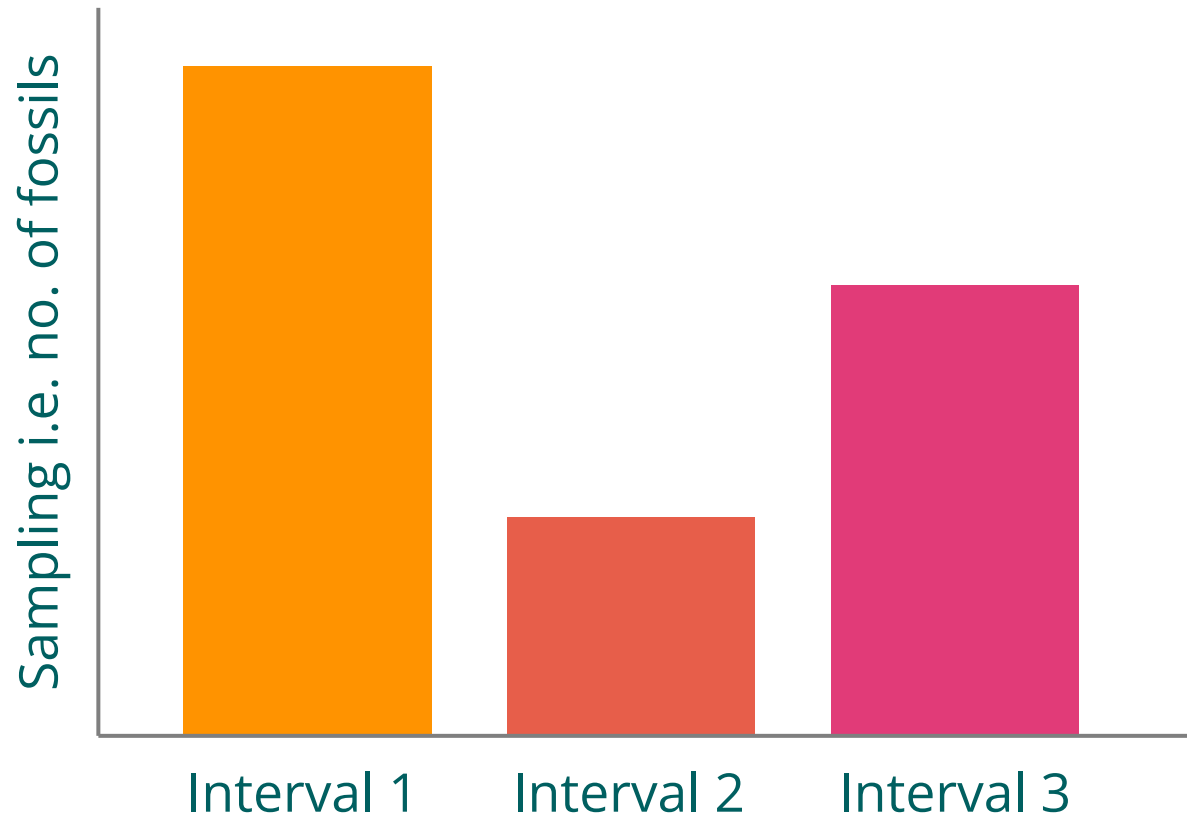
- Better at dealing with rare taxa than rarefaction

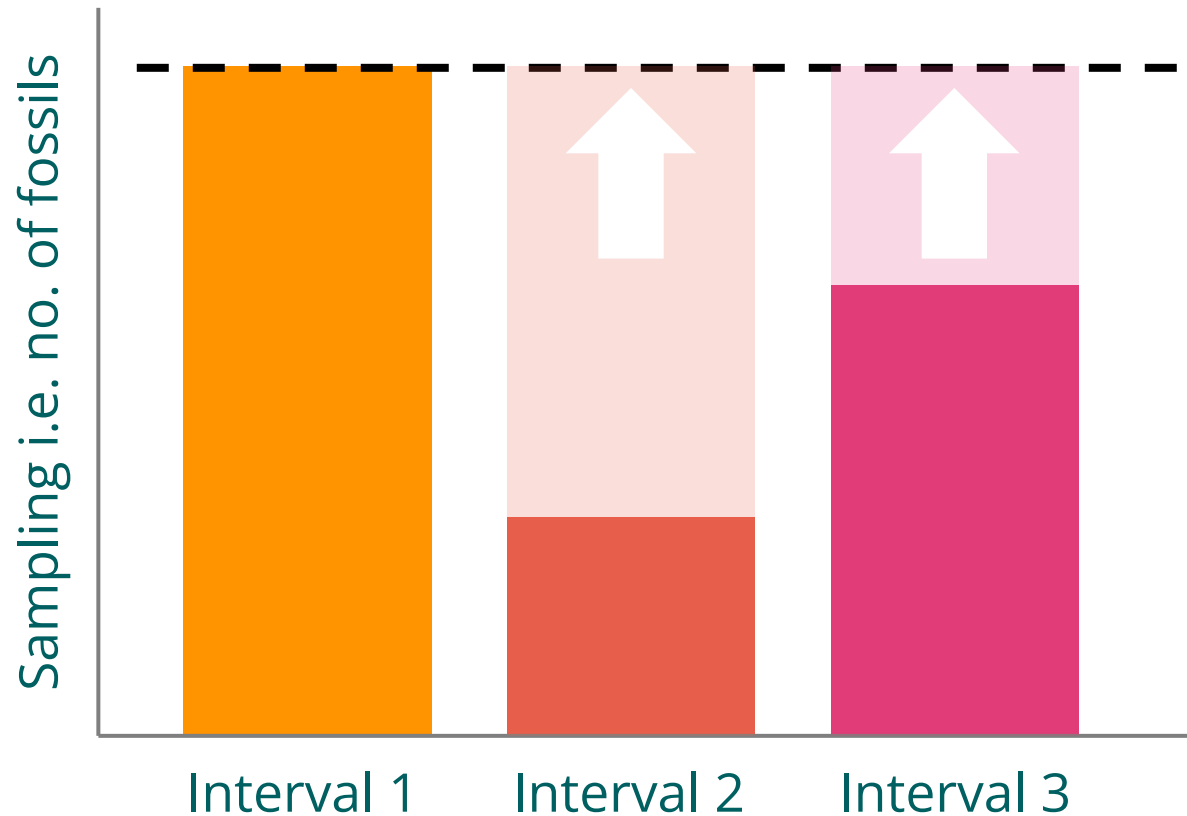# Coverage-based methods

Shareholder Quorum Subsampling (SQS)

- Developed by John Alroy in 2010

- Interpolation method (i.e. estimates from known data)

- Known among ecologists as coverage-based rarefaction (`iNEXT` R package - **in**terpolation and **ext**rapolation)
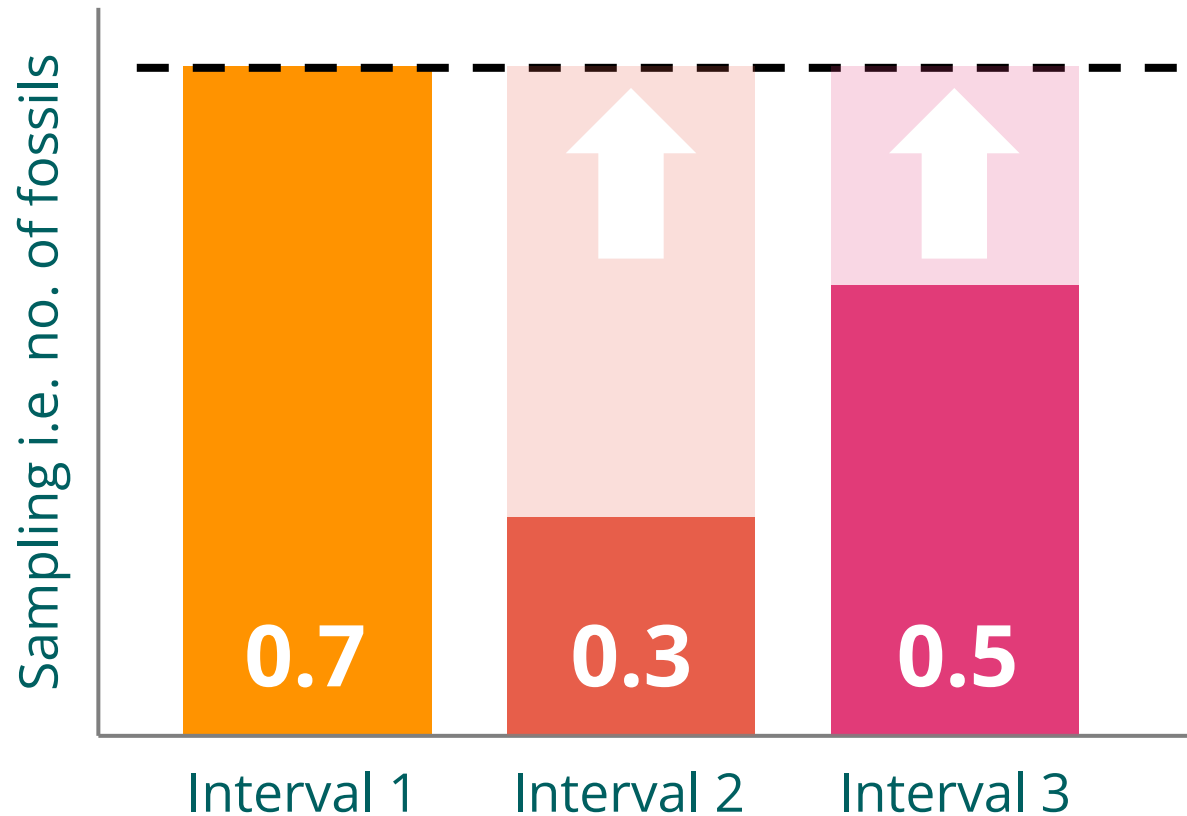
Hypothetical sampling scenario

Sampling i.e. no. of fossils

Interval 1   Interval 2   Interval 3

Hypothetical sampling scenario

Sampling i.e. no. of fossils

Interval 1   Interval 2   Interval 3

Coverage-based methods

- draw samples UP based on an inference from the known data

Hypothetical sampling scenario

SQS

- Based on sample 'coverage' as measured by Good's u

- This is based on species frequencies to give an estimate of how 'complete' the bin is

Example: Early tetrapod diversity

- Was early tetrapod (terrestrial vertebrate) diversity impacted by the Carboniferous 'rainforest collapse' (CRC)?

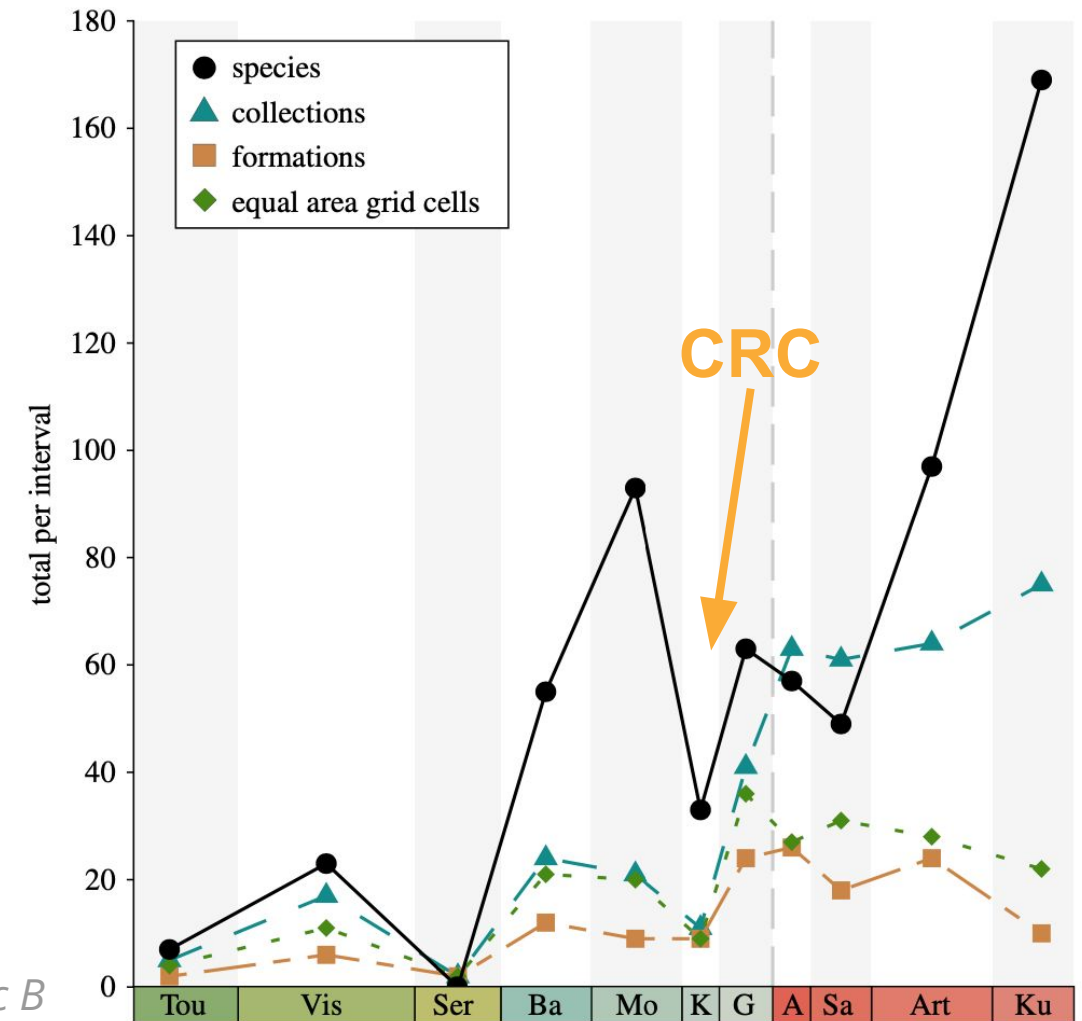- How does sampling impact our understanding of this?

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

Example: Early tetrapod diversity

Raw (face-value) richness

- Raw (face-value) species richness tracks proxies for sampling

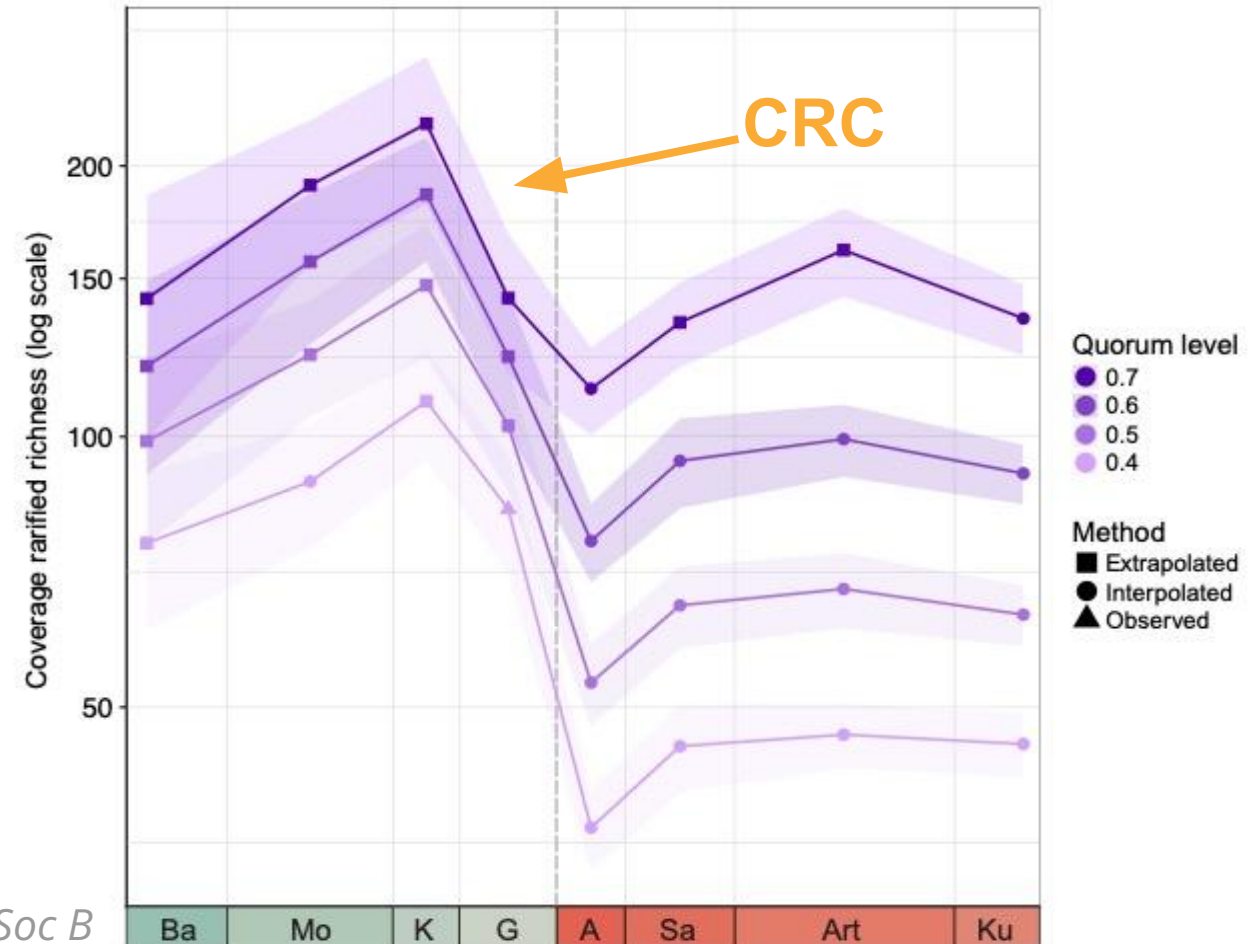- More sampling = greater species richness

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

Example: Early tetrapod diversity

Coverage-rarifed richness

- Implemented through
  `iNEXT` R package

- Shows a more 'intuitive'
  pattern given what we
  know about the CRC

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

SQS: Criticisms

- Found to track evenness i.e. if sample is dominated by common species, SQS might miss true richness changes (Hannisadal *et al.* 2012)

- Still needs relatively high levels of sampling to compute estimates
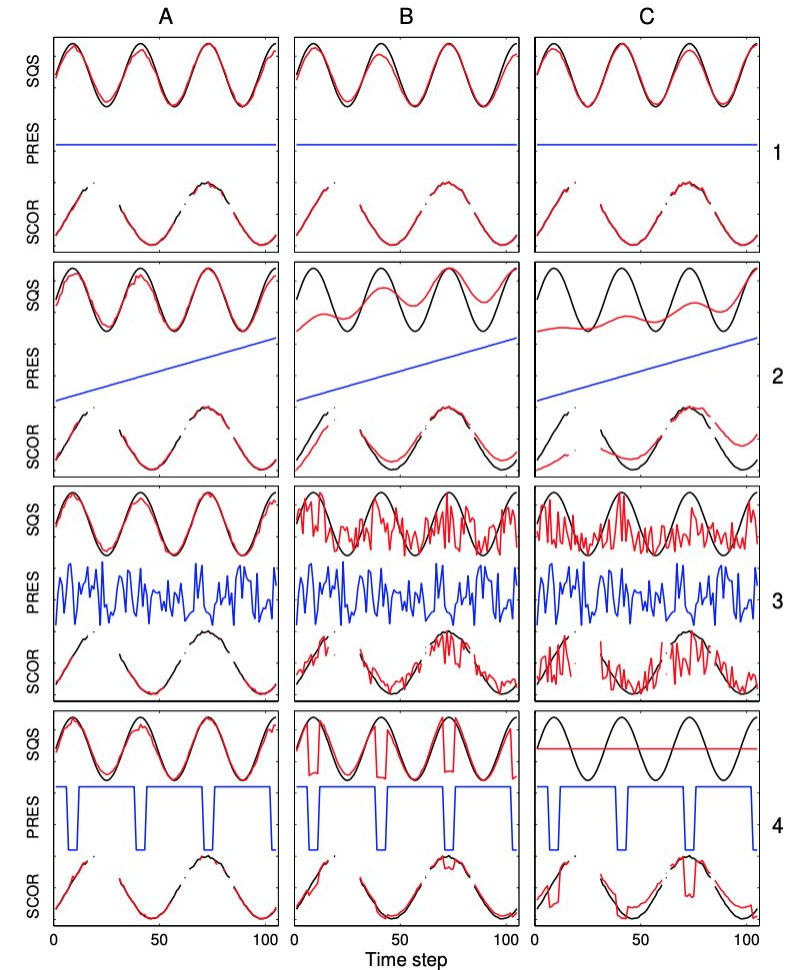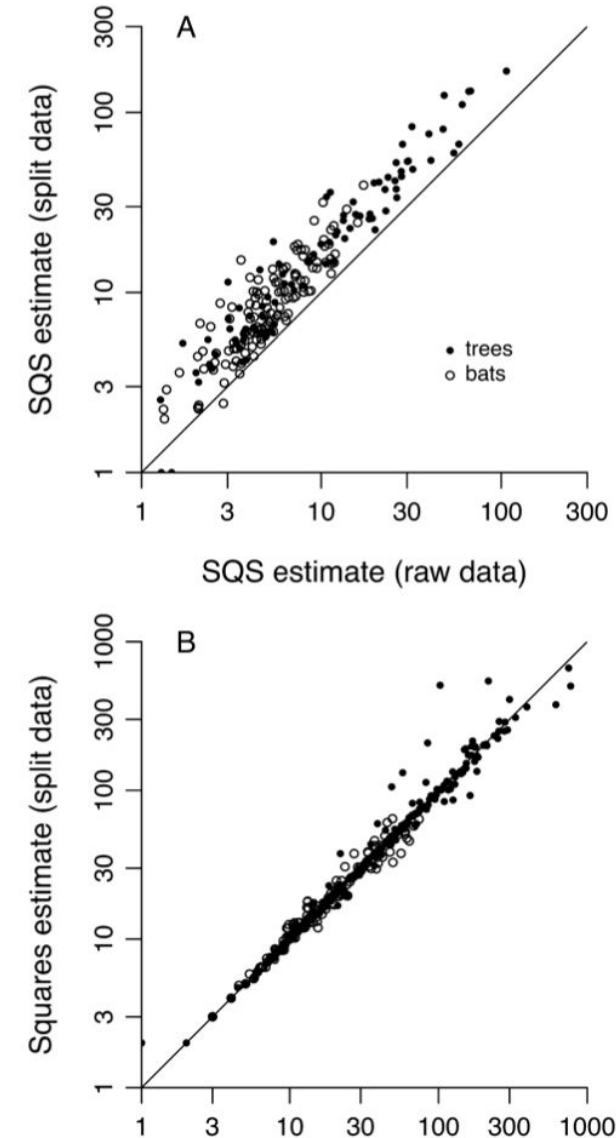

Hannisadal *et al.* 2012 *Glob. Change Biol.*

Squares

- Also developed by John Alroy

- Another coverage-based method

- Performs well when the rank abundance of distributions of samples is particularly skewed (i.e. there are many rare taxa)
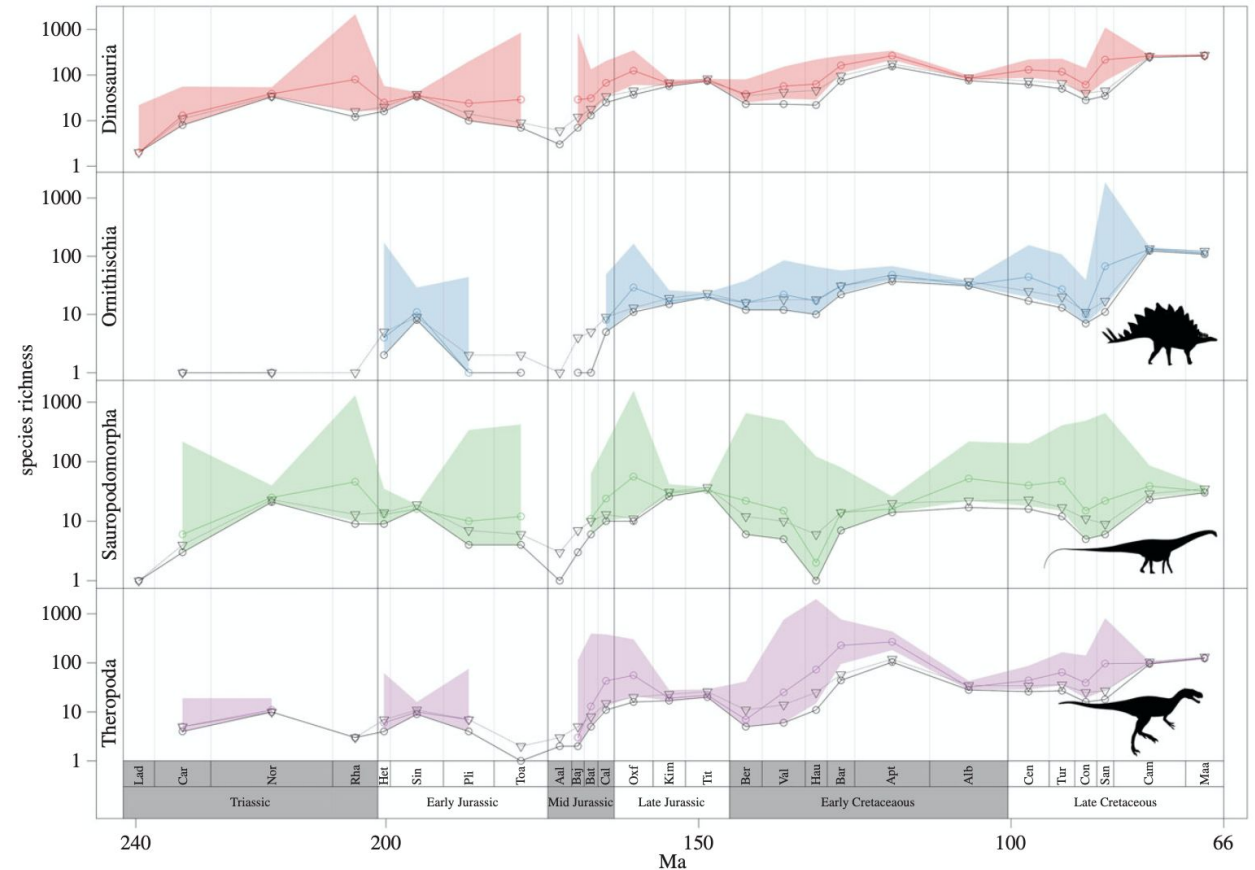

Alroy (2020) *Paleobiology*
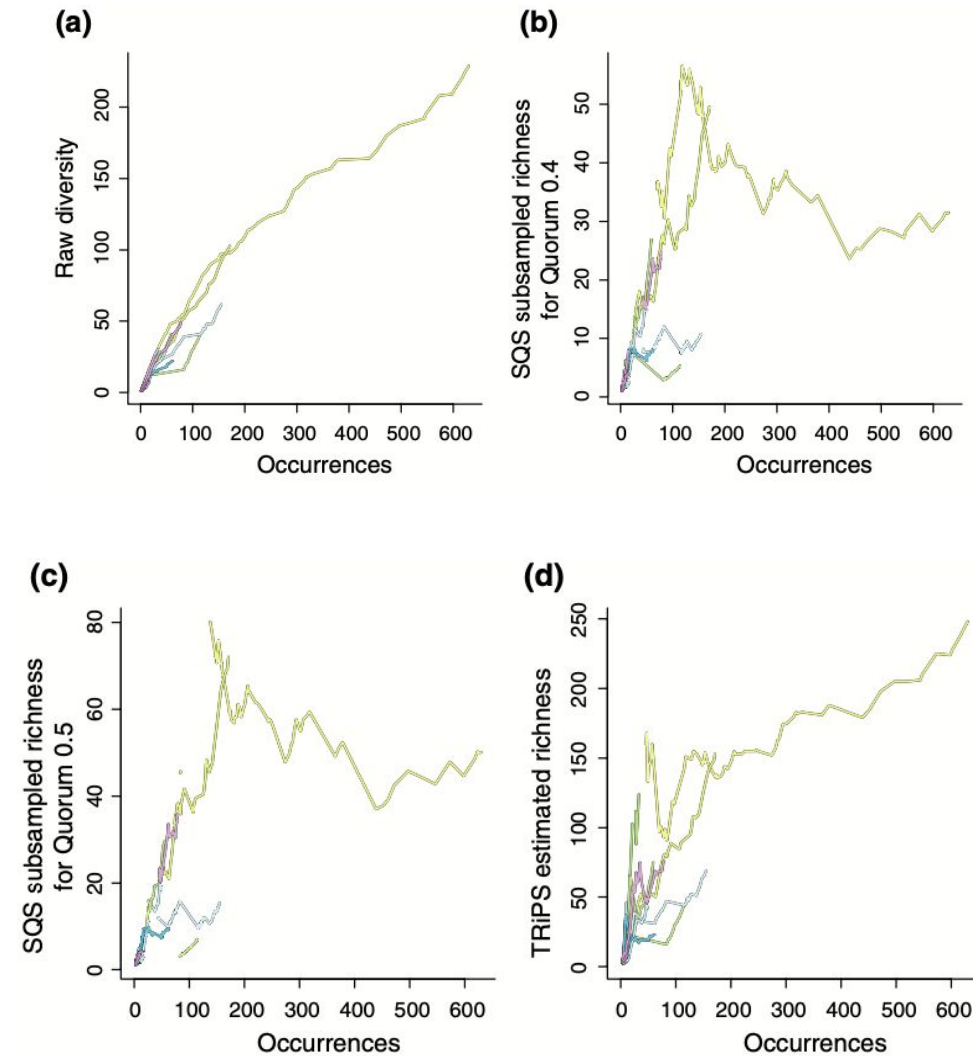
Other subsampling methods

TRiPS

- **T**rue **Ri**chness estimated using a **P**oisson **S**ampling model

- Extrapolation method

- Available via a Shiny app: starrfelt.shinyapps.io/TRiPS/

Starrfelt & Liow (2016) *Phil. Trans.*

TRiPS: Criticism

- Tracks unstandardised diversity, particularly in well-sampled bins

- Doesn't deal well when dataset contains many common species

Close *et al.*(2018) *Method Ecol. Evol.*

Residual modelling

- Looks at the relationship between sampling and diversity, then plots the residuals to identify instances where diversity is lower or higher than expected

- Introduced by Smith & McGowan, developed by Graeme Lloyd
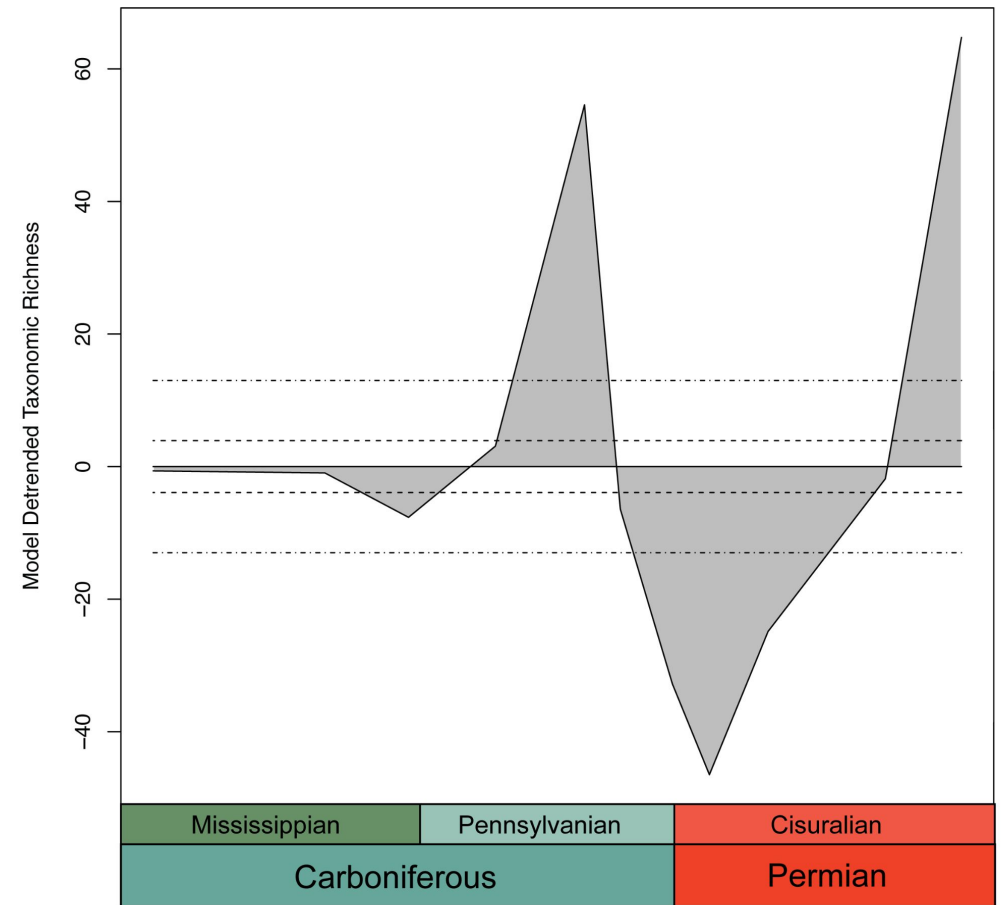
E. Dunne (unpublished data)

Residual modelling: Criticism

- Reliant on the sampling proxy used - which can mean different things

- Modelling is a generally an underdeveloped standardisation method
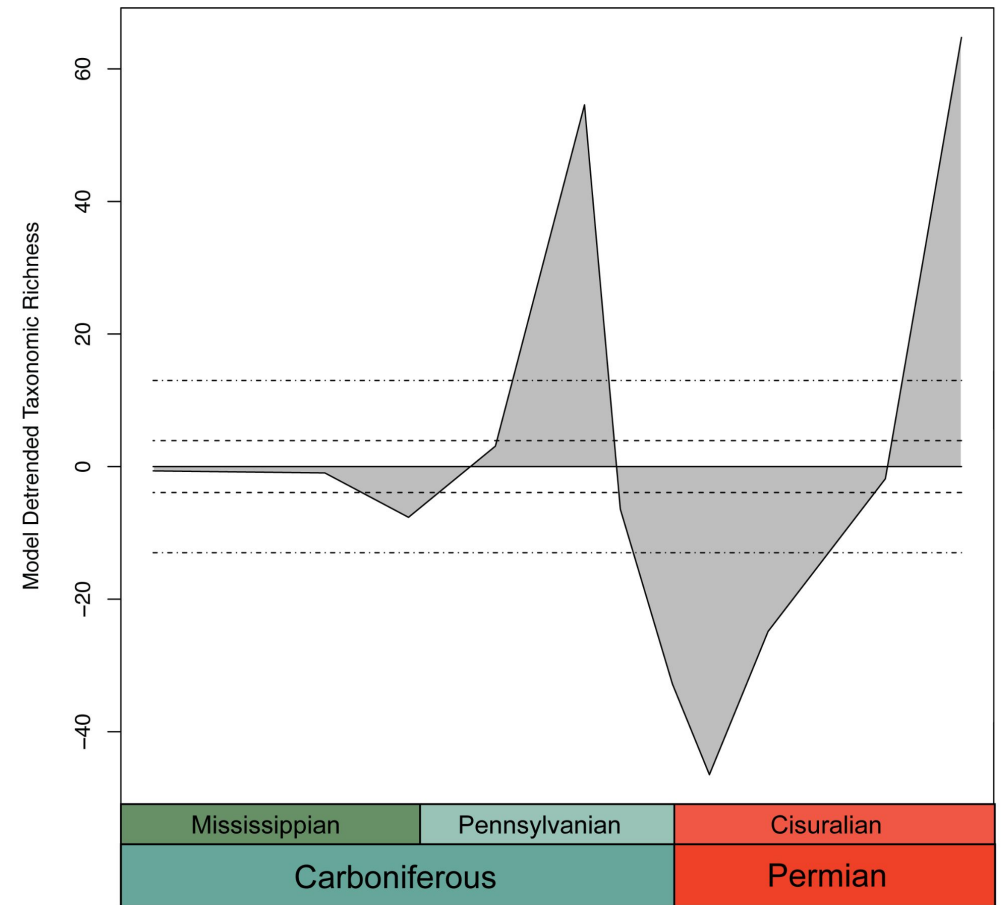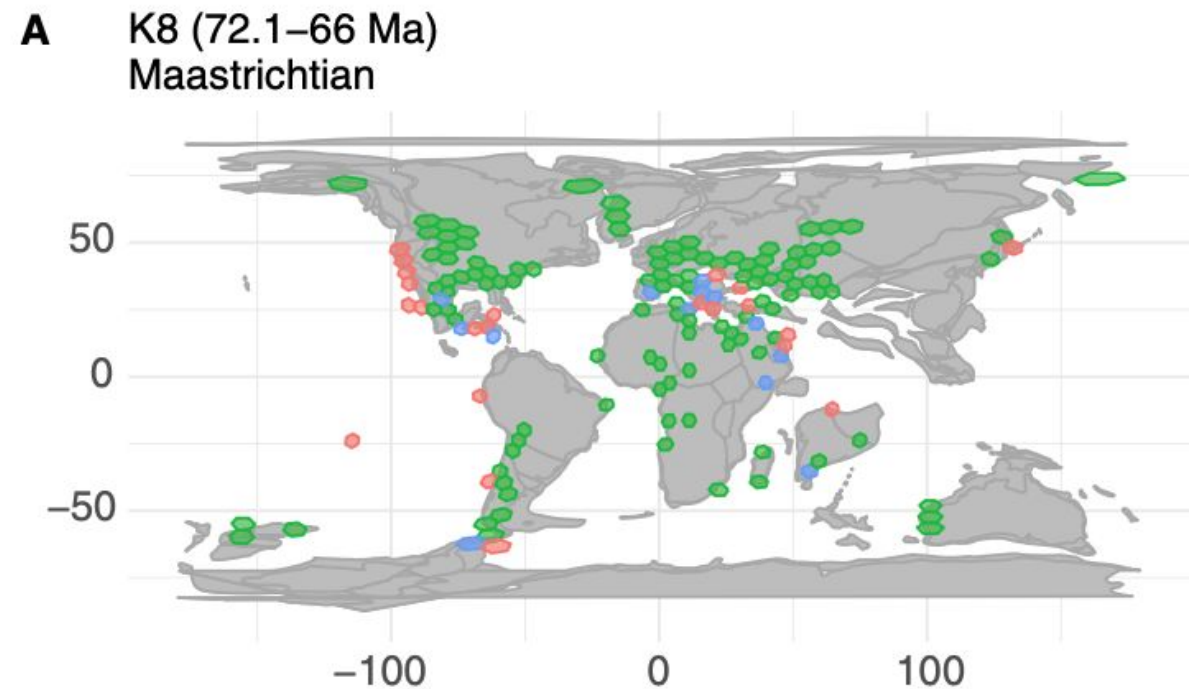

E. Dunne (unpublished data)

Correcting for spatial sampling

Global vs. regional diversity

- Some authors argue that 'global' diversity is meaningless

- Spatial sampling biases are pervasive and must be corrected for

- Same authors argue for greater focus on regional diversity

A   K8 (72.1–66 Ma)
     Maastrichtian

Benson *et al* (2021) *Curr. Biol.*

Minimum-spanning tree length

- Attempting to control for the species-area effect

- MST length draws spatial samples of fixed extent

- MST example: a cable company wanting to lay line to multiple towns - minimizing amount of cable saves money
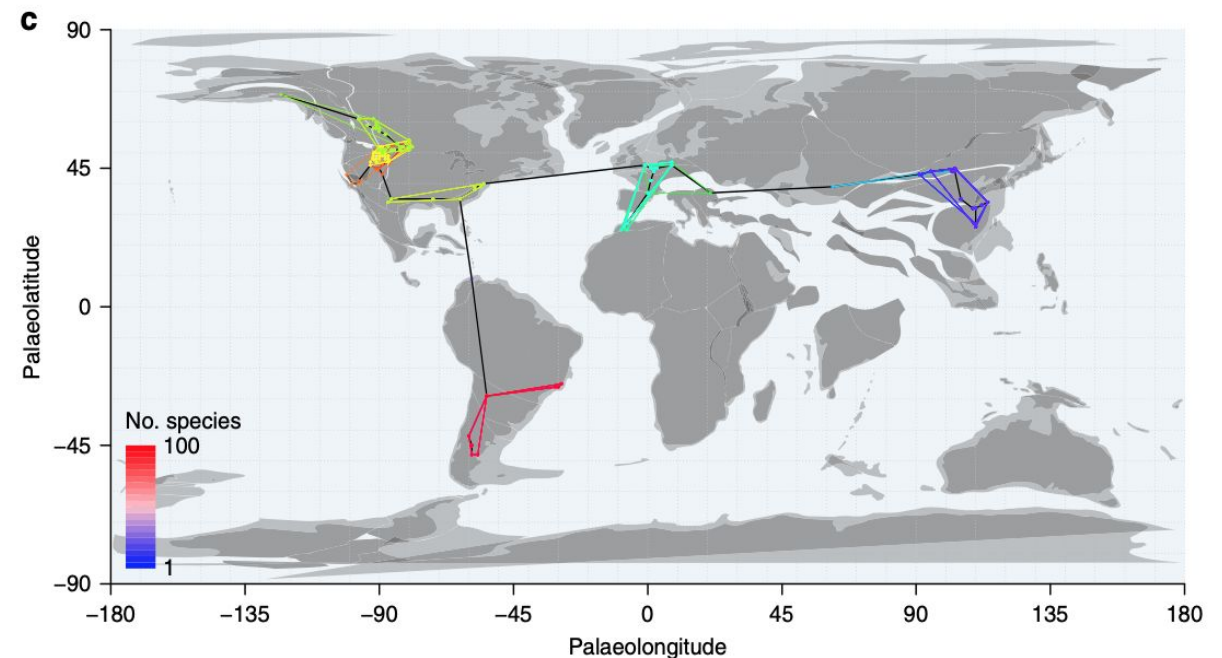
Close *et al* (2017) *Nat. Commun.*

Spatial subsampling

- Antell *et al.*'s method controls both total sampling area (no. of occupied equal-area grid cells) and dispersion of sampling sites through time (MSTs)

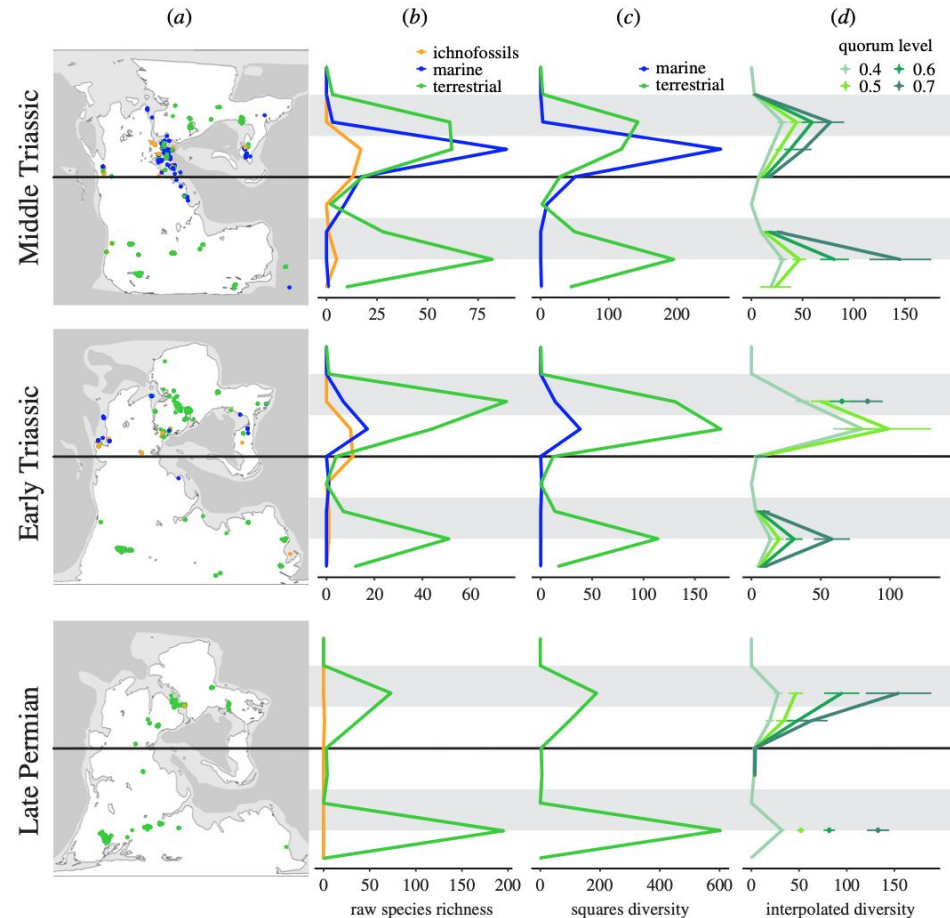- Keep an eye out for a new R package being released soon!
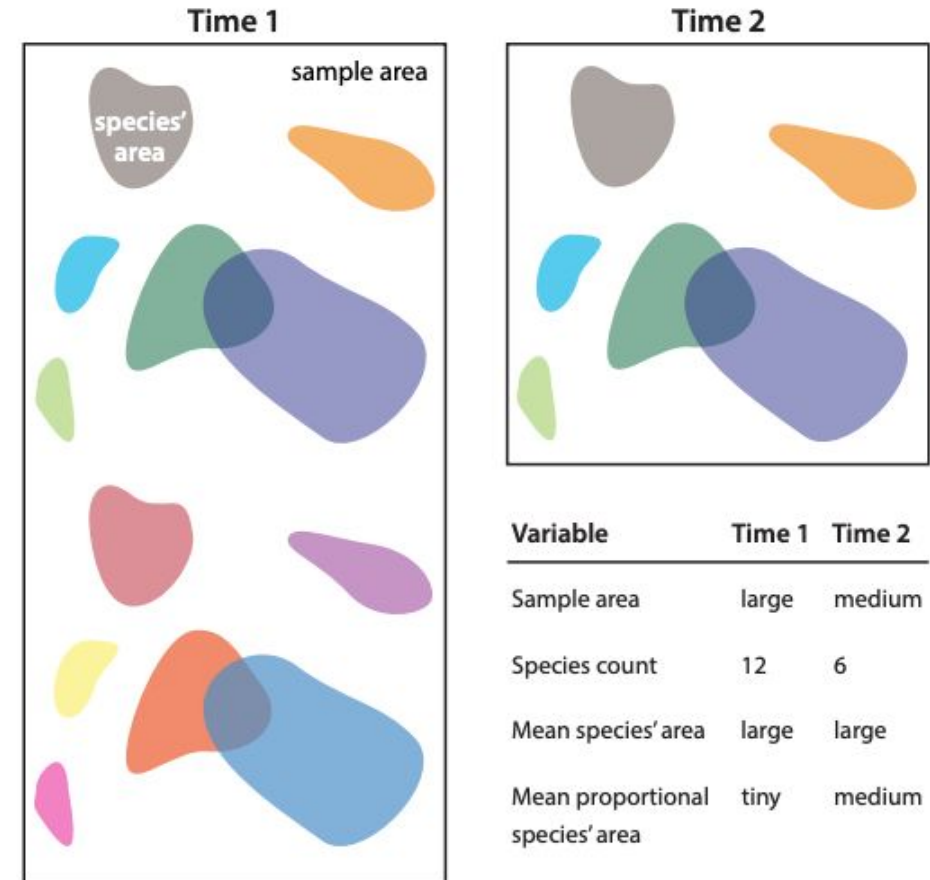
Antell *et al* (2020) *Curr. Biol.*

Which method should I use?

Which method is the 'best'?

- Current most popular for temporal trends = SQS/ coverage-based rarefaction

- Useful to use more than one method

- Most important that you explore your data

- Don't expect one single method to fix all problems!

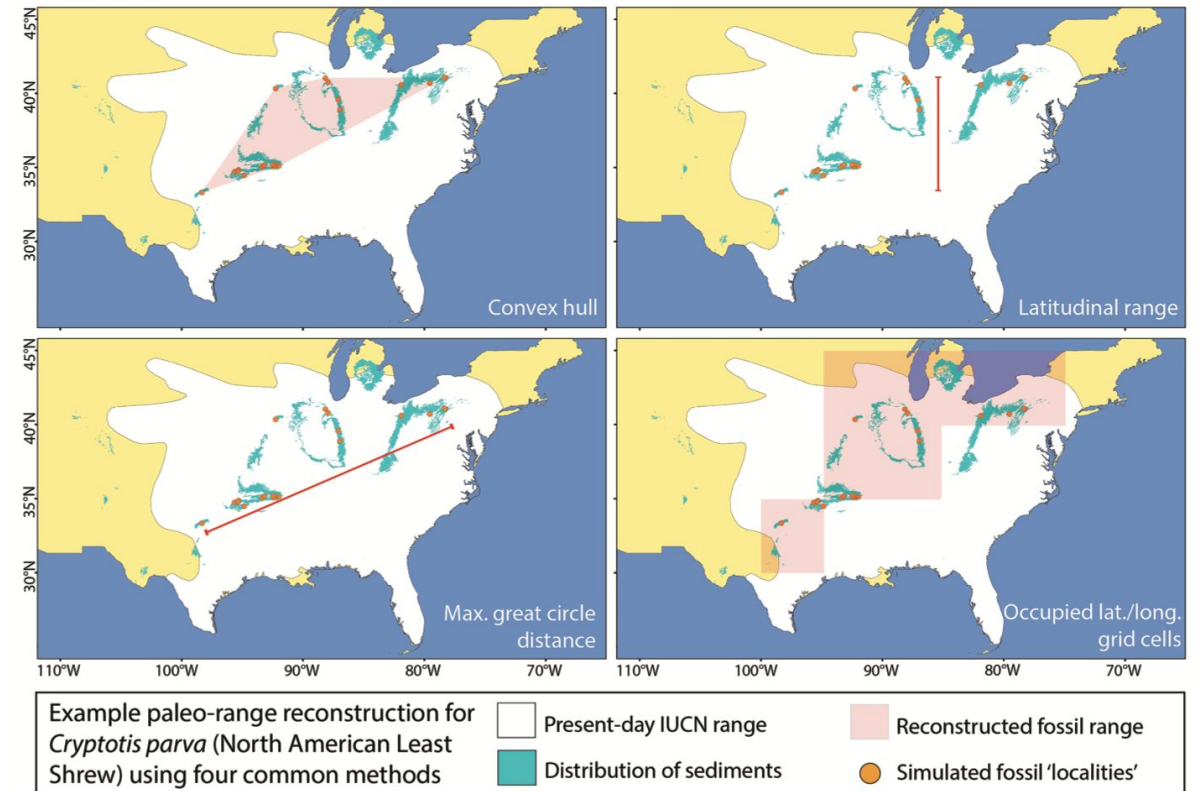Allen *et al* (2020) *Proc. Royl. Soc. B.*

Best practice

1. Always consider that there are biases lurking among your data

2. Explore your data thoroughly

3. Choose a suitable metric to quantify sampling bias

4. Consider which subsampling method best suits your data

5. Remember that your corrected estimates are not automatically 'correct'!

Can we even trust the fossil record?

- Simulations have shown that the fossil record has good potential to preserve palaeogeographic ranges (in both recent time and deep time)

- Temporal patterns are a little trickier though - they don't seem to preserve so well

Example paleo-range reconstruction for *Cryptotis parva* (North American Least Shrew) using four common methods

- Present-day IUCN range
- Reconstructed fossil range
- Distribution of sediments
- Simulated fossil 'localities'

Darroch *et al* (2020) *Am. Nat*