Sampling standardisation

Emma Dunne | APW 2023 | Thurs. Aug. 24th

Fossil record biases
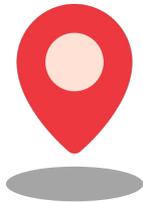
PROCESSES **CATEGORY**

Habitat & mode of life

Death & burial — **Taphonomic processes**

Depositional environment

Geological activity

Diagenesis & lithification, tectonics, etc. — **Geological processes**

Geographical location

Interest & resources

Collection & study

Occurrence data

Data collation — **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

Scales of biodiversity

α

β

γ

Alpha diversity

Beta diversity

Gamma diversity

"local richness"

between two areas/regions

"global" diversity

Scales of biodiversity

α

Alpha diversity

"local richness"

β
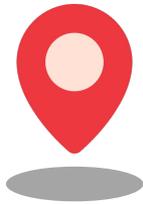
Beta diversity

between two areas/regions

γ

Gamma diversity

"global" diversity

Global vs. regional diversity

- Some authors argue that 'global' diversity is meaningless

    - **Spatial sampling biases** are pervasive and must be corrected for

- In this session, we will focus on temporal global diversity to better understand this argument and spatial diversity will be covered in other modules

Maastrichtian

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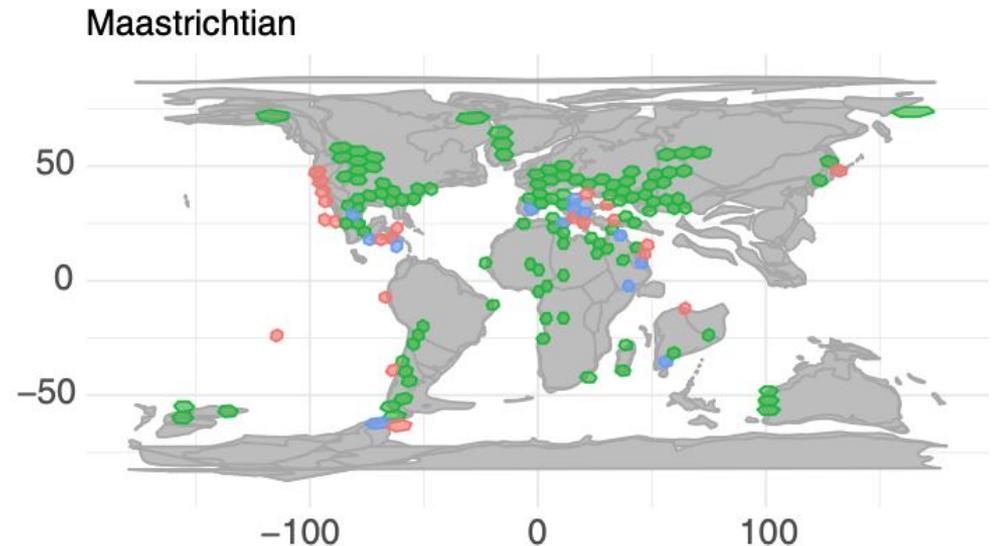
Sampling standardisation

- To infer genuine patterns of deep-time biodiversity, we need methods that both:

  - successfully standardise samples of unequal sizes

  - permit direct comparisons of richness among assemblages

- Several different approaches have been developed - each have various (dis)advantages

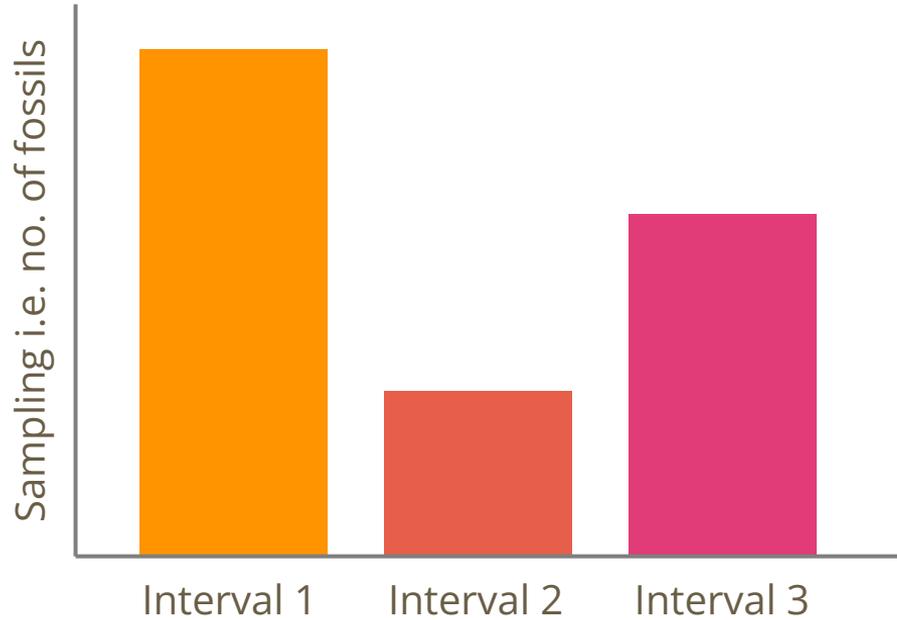Close *et al.* (2018) *Methods Ecol. Evol.*
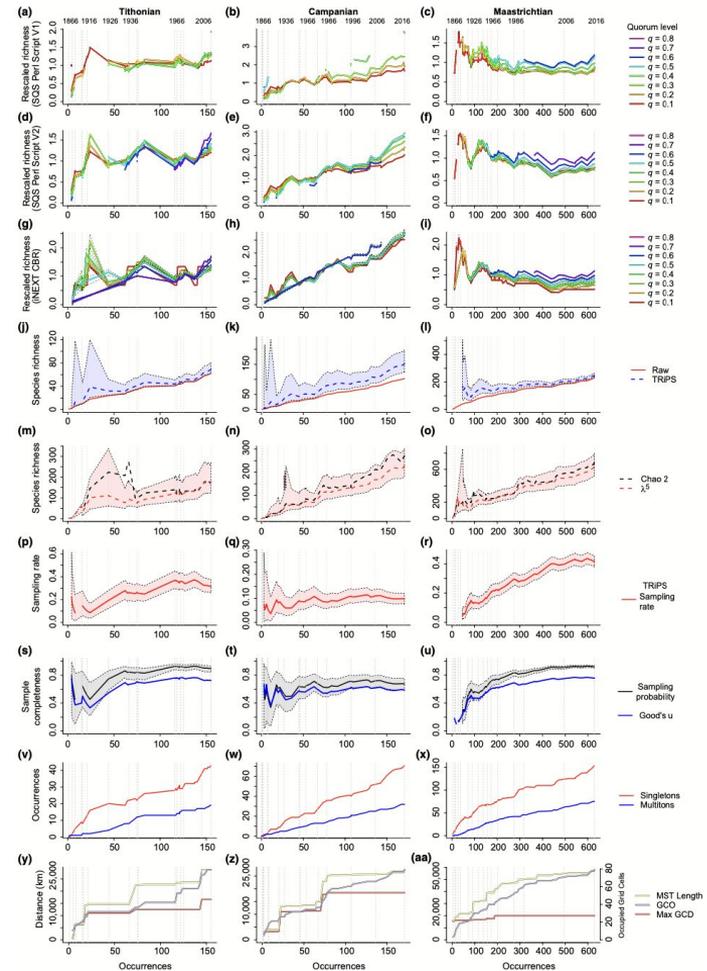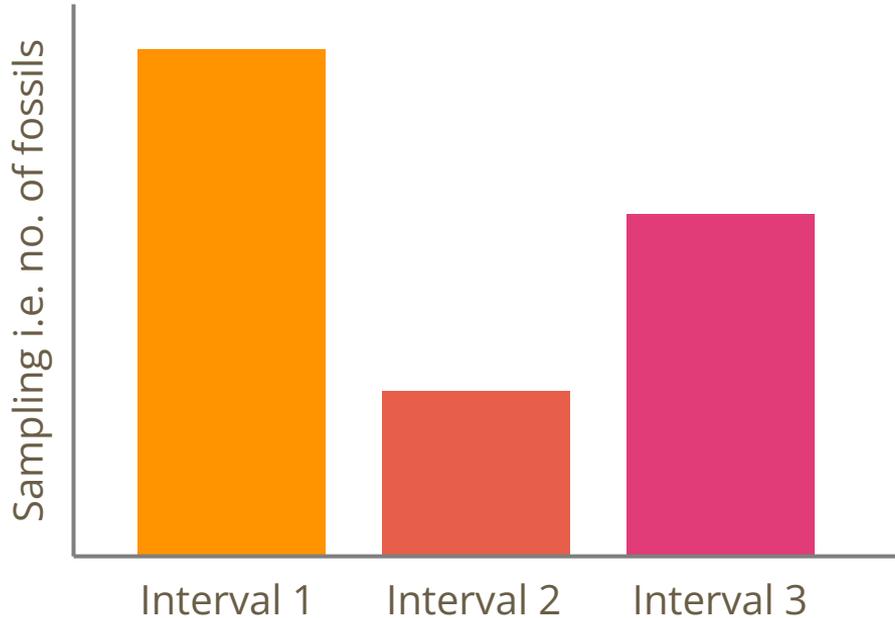
Sampling standardisation

- An early approach was to standardise samples by **size**

- e.g. **Classical rarefaction**

  - Fixes estimates to sample size

  - Can result in information being lost

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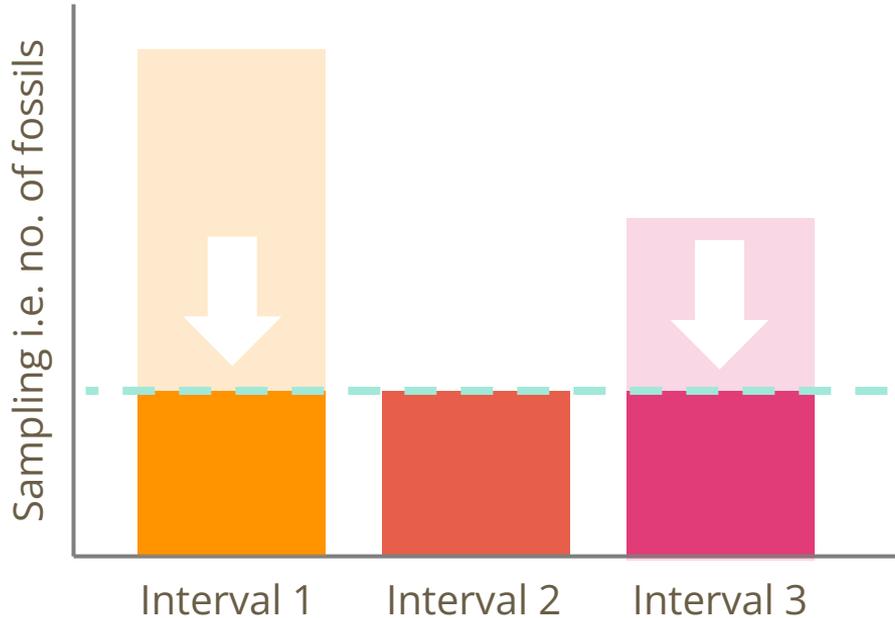
Hypothetical sampling scenario

Hypothetical sampling scenario

Classical rarefaction

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

Hypothetical sampling scenario

Classical rarefaction

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

- Lots of potentially useful information may be lost

Coverage-based methods

- Subsample to a fixed **coverage**

- Better at dealing with rare taxa than rarefaction
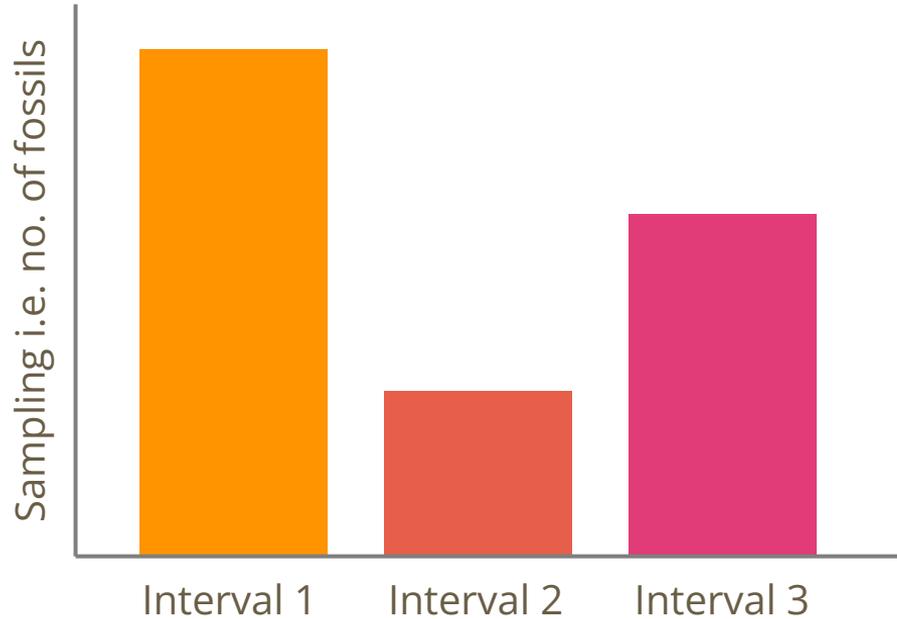
Coverage-based methods

- Subsample to a fixed **coverage**

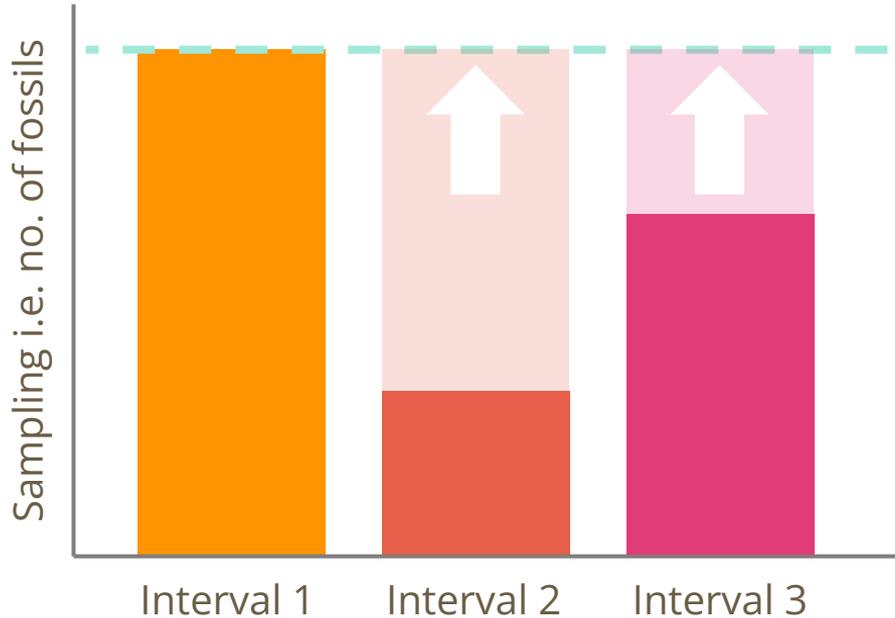- Better at dealing with rare taxa than rarefaction

Shareholder Quorum Subsampling (SQS)

- Developed by John Alroy ([2010a](#), [2010b](#))

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

- Ecologists implement a very similar method: **coverage-based rarefaction**
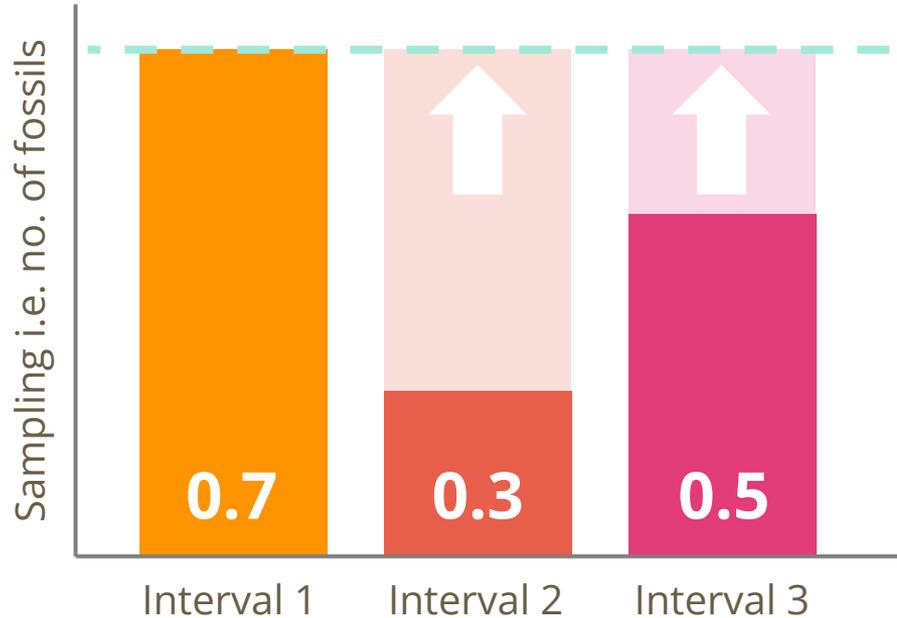
Hypothetical sampling scenario

Hypothetical sampling scenario

Coverage-based methods

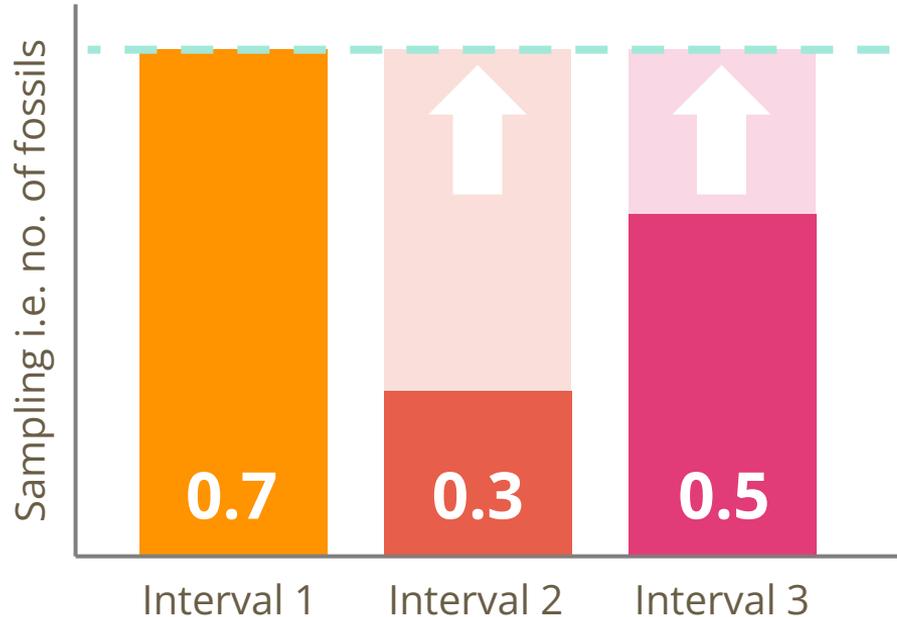- draw samples **UP** based on an inference from the known data

Hypothetical sampling scenario

SQS

- Computes sample 'coverage' using **Good's u**

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

Hypothetical sampling scenario

Coverage-based rarefaction

- Computed using the **equations of Chao & Jost** (2012) (analogous to SQS)

- and extrapolation based on the **Chao1 estimator** (Chao 1984)

Case study: Early tetrapods

- 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*

Case study: Early tetrapods

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

- Clearly evident sampling biases

  - More sampling = greater species richness


Dunne *et al.* (2018) *Proc Royl Soc B*
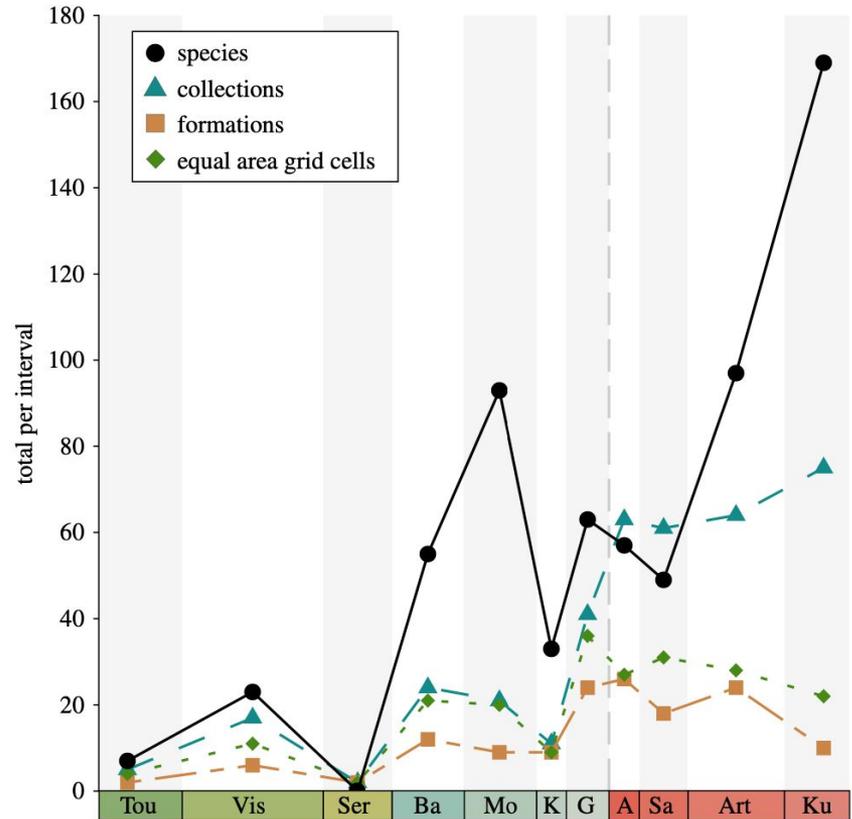
Case study: Early tetrapods

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

- Clearly evident sampling biases
  - More sampling = greater species richness

- Coverage-rarifed richness implemented through `iNEXT` R package (Hsieh *et al.* 2016)

- Standardises diversity across time bins as described previously


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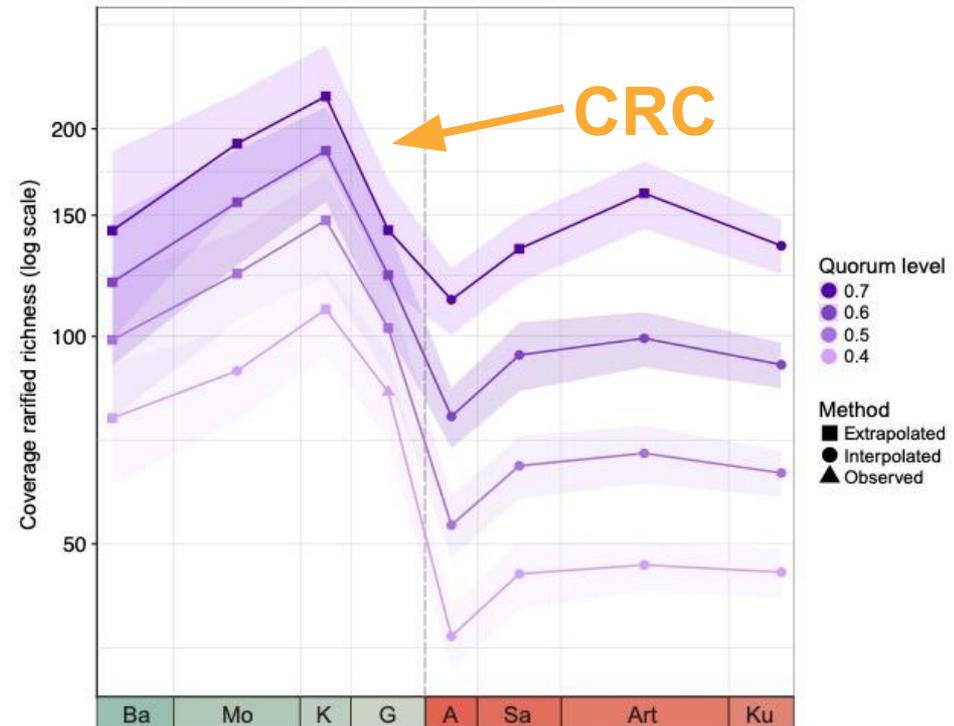
SQS

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 (as does coverage-based rarefaction)

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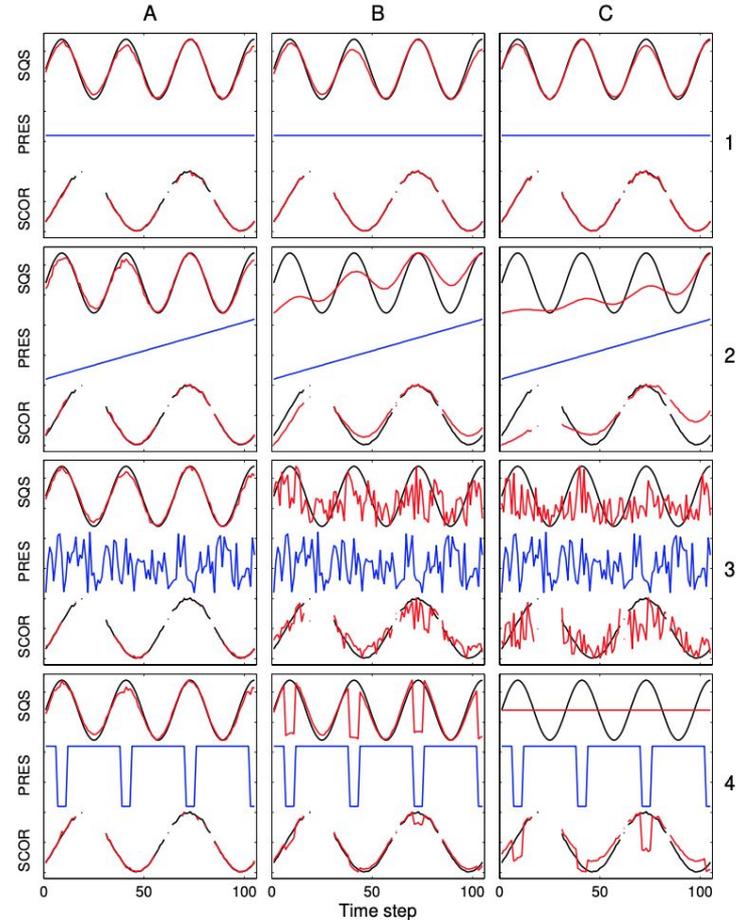
Squares

- Another coverage-based method

- Also developed by John Alroy ([2020](#))

- Performs well when the rank abundance of distributions of samples is particularly skewed

   - i.e. when there are many rare taxa

Alroy (2020) *Paleobiology.*
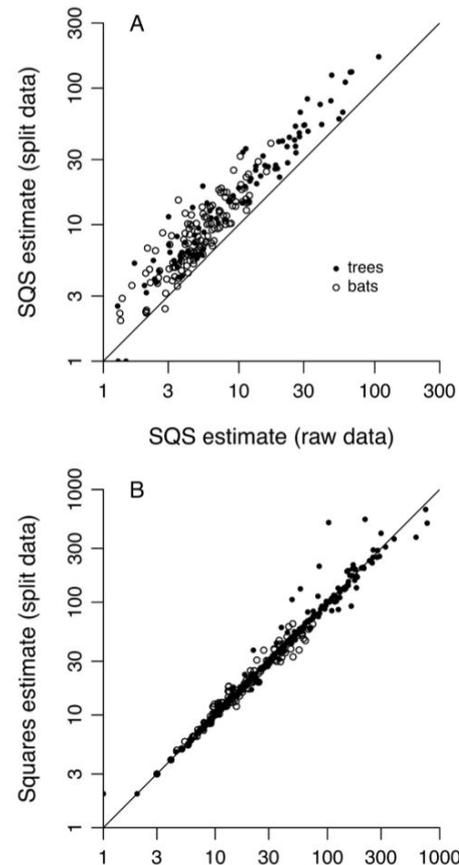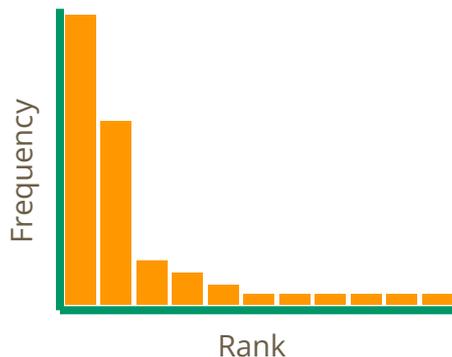
Other approaches

Capture-mark-recapture (ecology)

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

- Next week: **Isaiah Smith** & **Lee Hsiang Liow**

Fossilized birth–death process (phylogenetics)

- A model that explicitly recognizes that the branching events in a phylogenetic tree and sampled fossils were generated by the same underlying diversification process

- Speak to **Rachel Warnock**

Warnock *et al.* (2020) *Paleobiology*
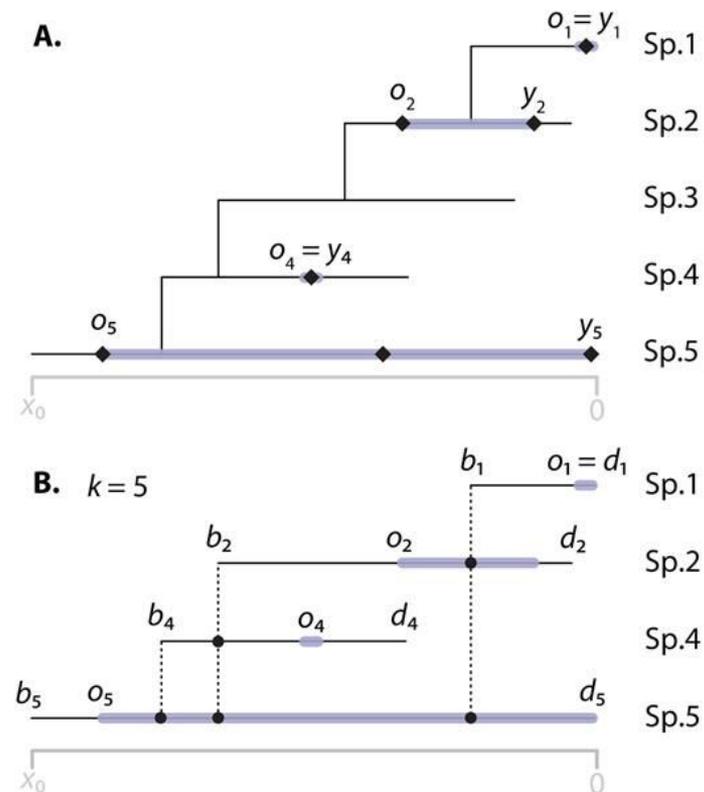
Which method is the best?

- Most popular for temporal diversity = SQS / coverage-based rarefaction

- Can use more than one method (e.g. Allen et al. 2020; Henderson *et al.* 2022; Schnetz *et al. in review*)

- Most important step is exploring and understanding your data

- Remember that 'standardised' estimates are not automatically 'correct'!

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