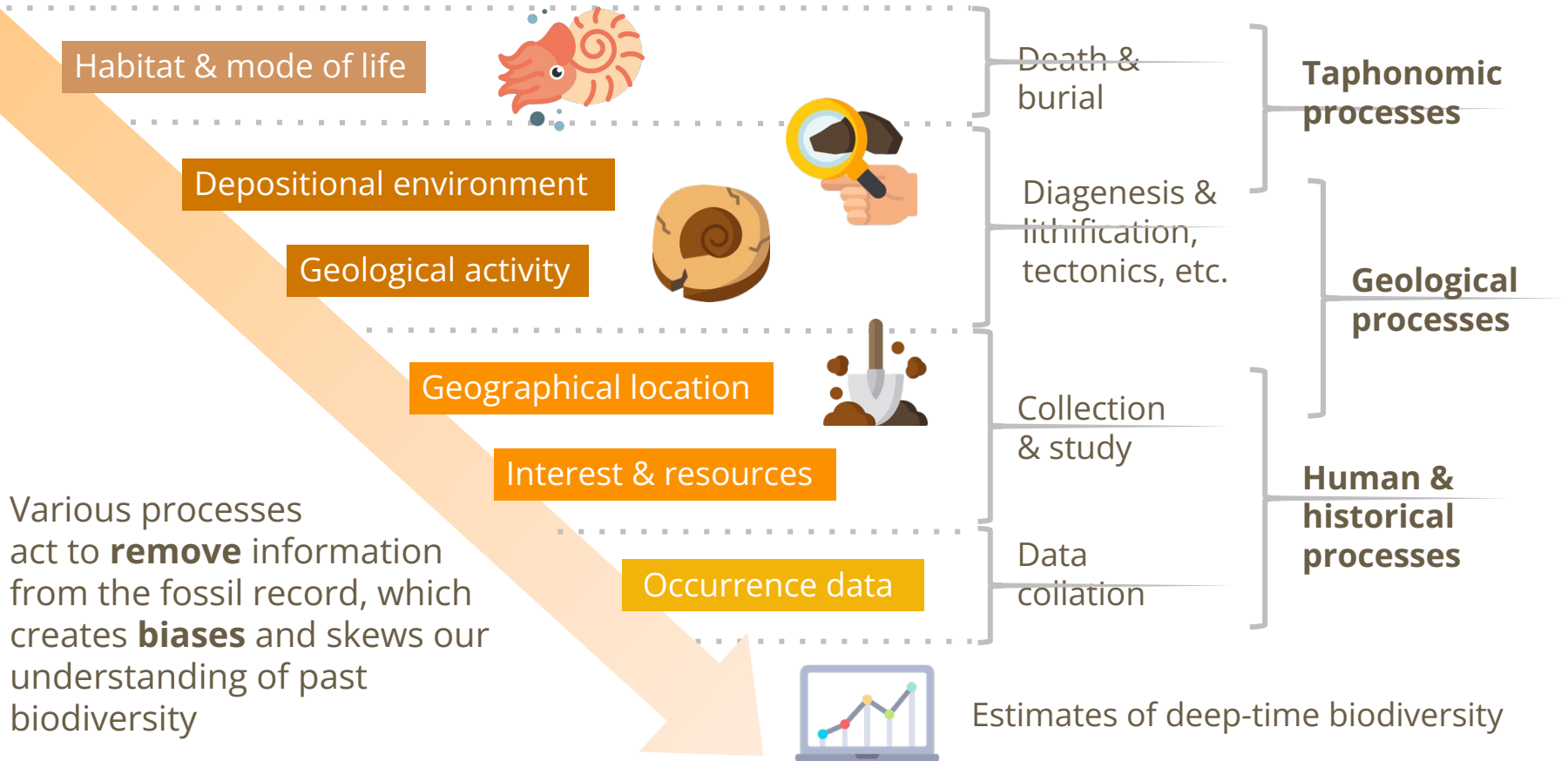

Sampling standardisation

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

FROM DEATH TO DATABASE

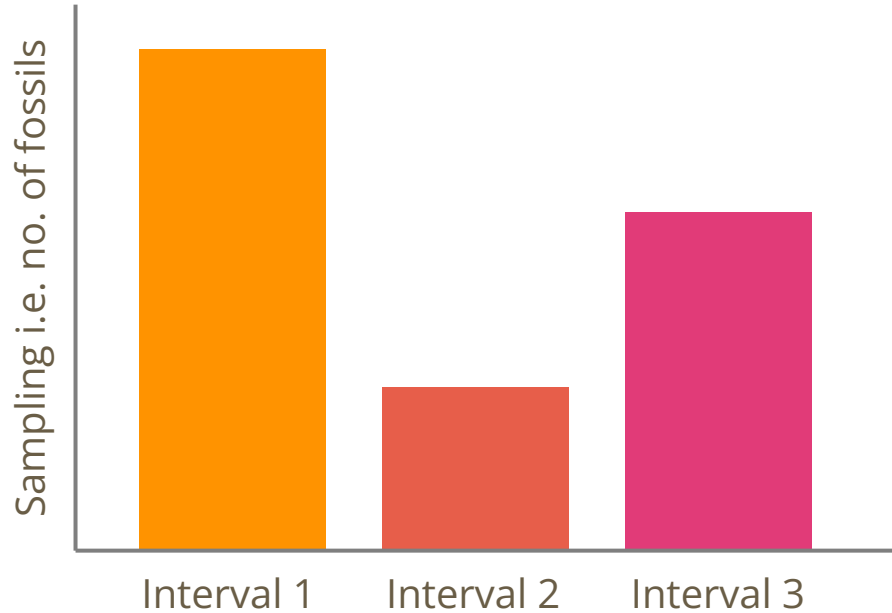


Rarefaction

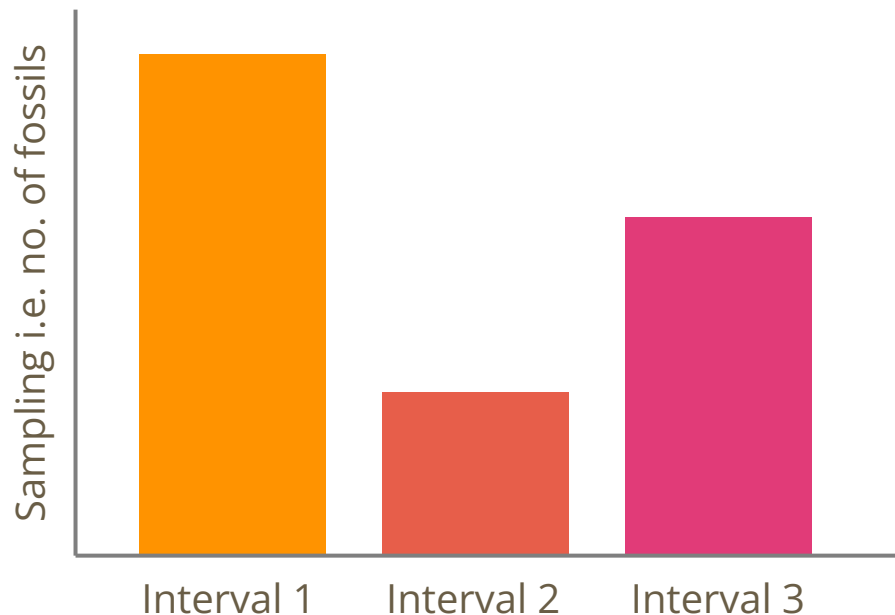
- Fixes to a sample **size**
- Can result in information being lost
- Rare taxa can be underrepresented in estimates when **evenness** is **low**



Hypothetical sampling scenario



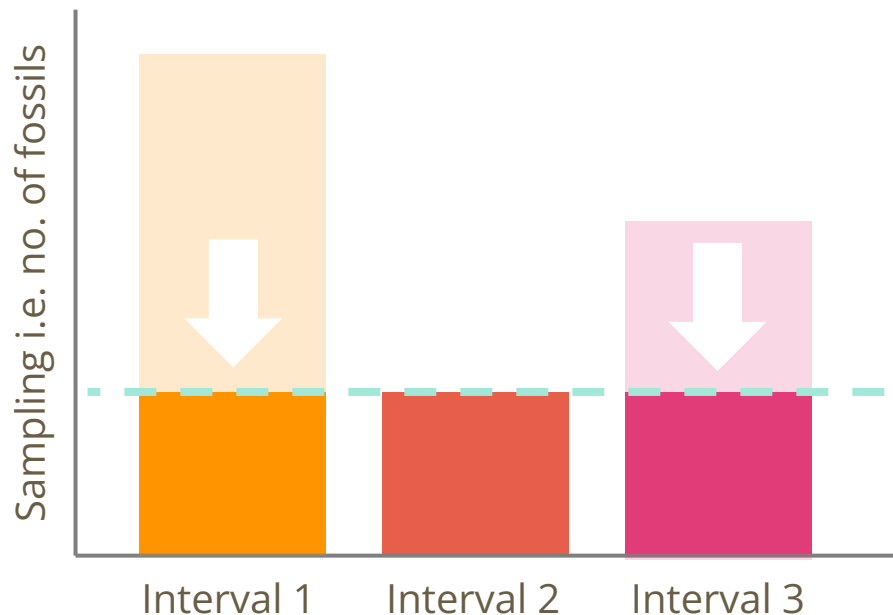
Hypothetical sampling scenario



Classical rarefaction

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

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

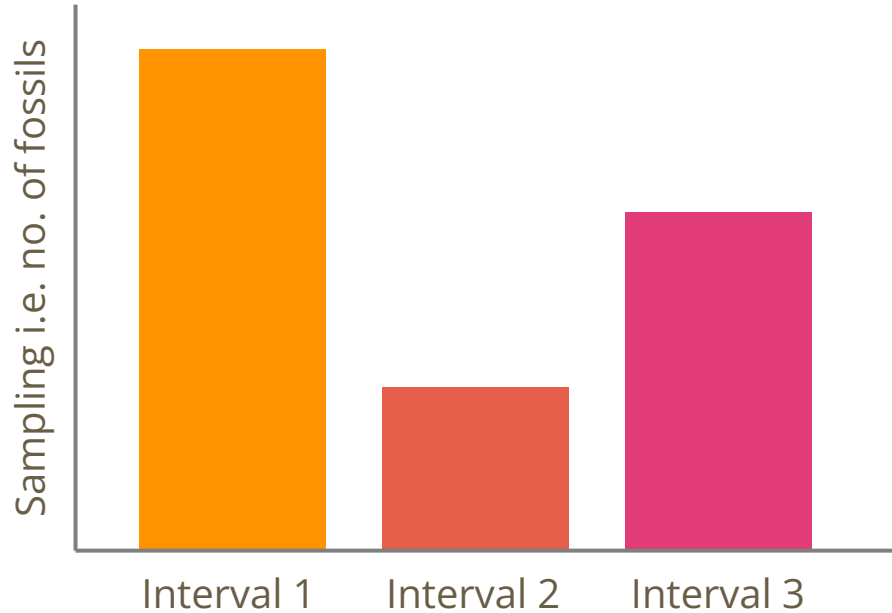
- Subsample to a fixed coverage
- Better at dealing with rare taxa than rarefaction



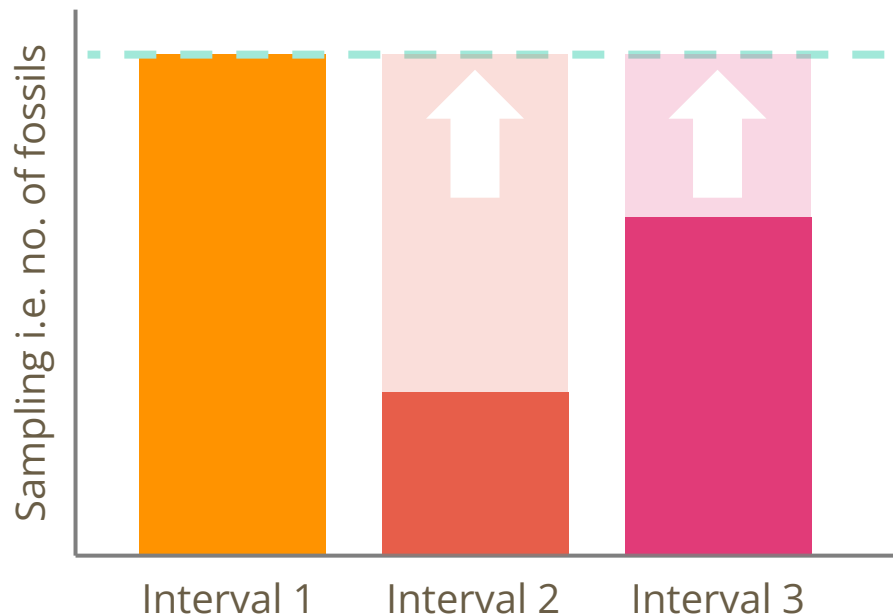
Shareholder Quorum Subsampling (SQS)

- Developed by John Alroy in 2010
- Interpolation method (i.e. estimates from known data)
- Known among ecologists as coverage-based rarefaction

Hypothetical sampling scenario



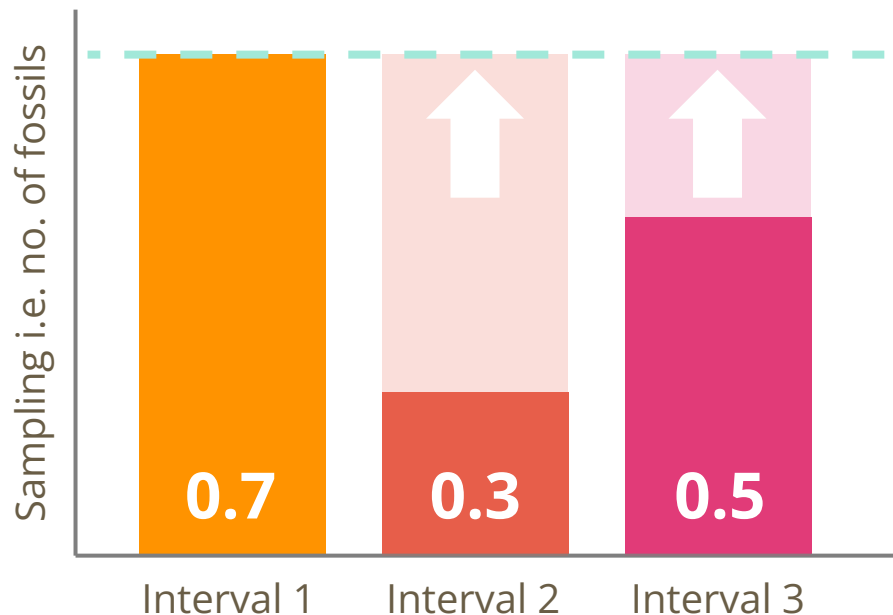
Hypothetical sampling scenario



Coverage-based methods

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

Hypothetical sampling scenario



SQS

- Based on sample 'coverage' as measured by **Good's u**
- This is based on **species frequencies** to give an estimate of how 'complete' a bin is

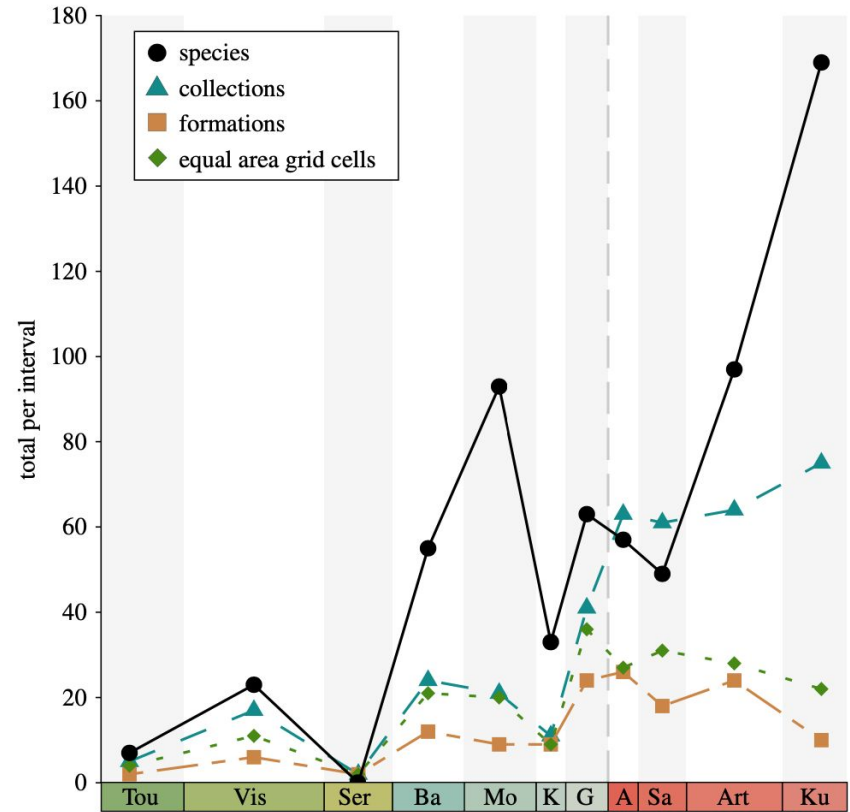
Case study: Early tetrapods

- Was early tetrapod (terrestrial vertebrate) diversity impacted by the Carboniferous 'rainforest collapse' (CRC)?
- How does sampling impact our understanding of this?



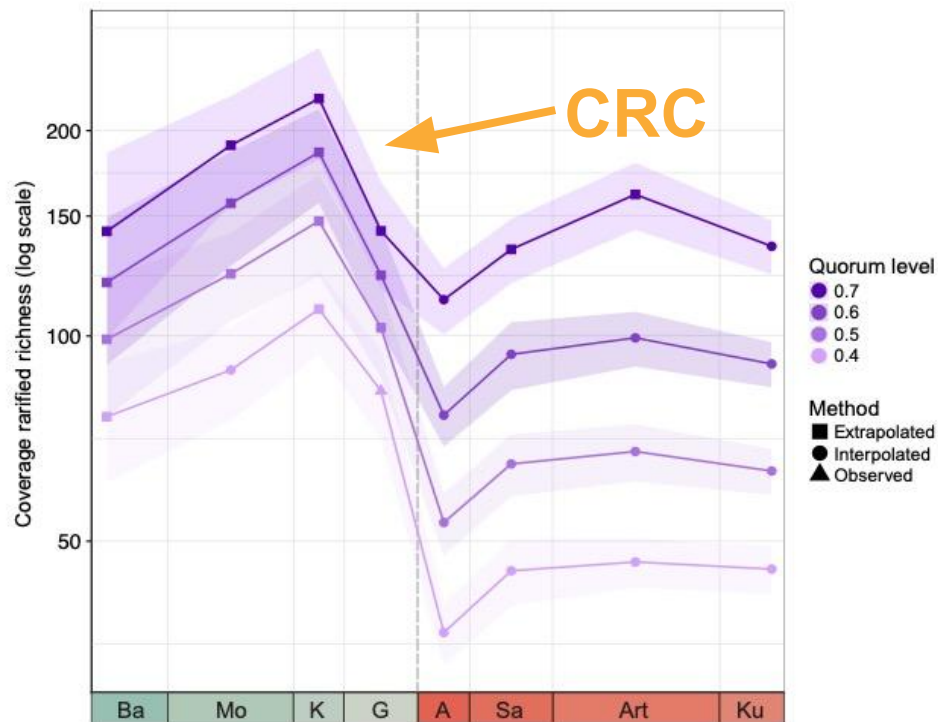
Case study: Early tetrapods

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



Case study: Early tetrapods

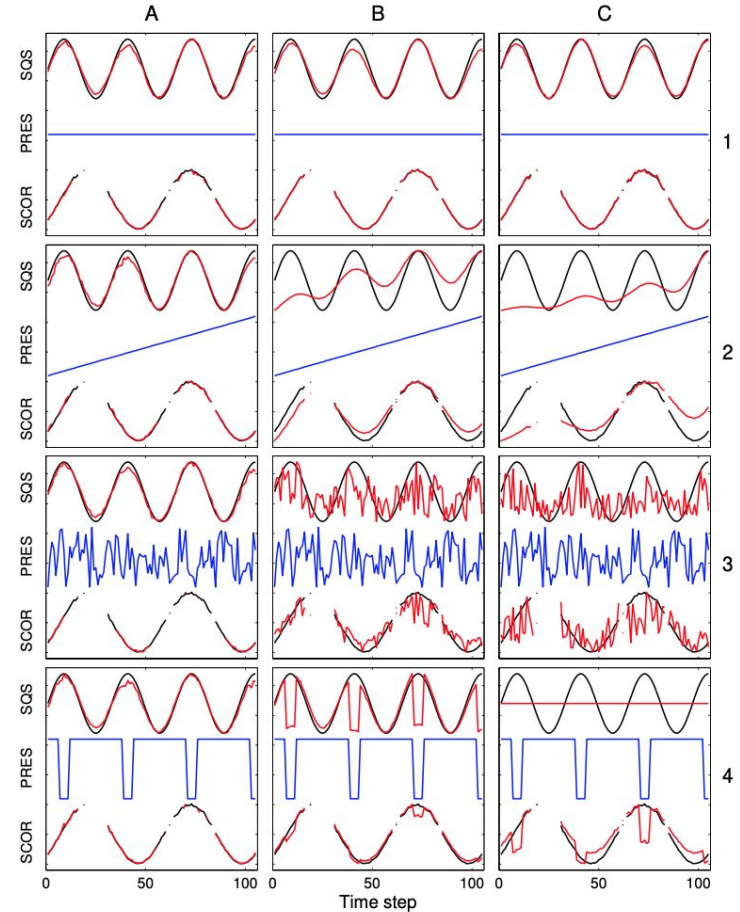
- Coverage-rarified richness
- Implemented through iNEXT R package
- Shows a more 'intuitive' pattern given what we know about the CRC



SQS

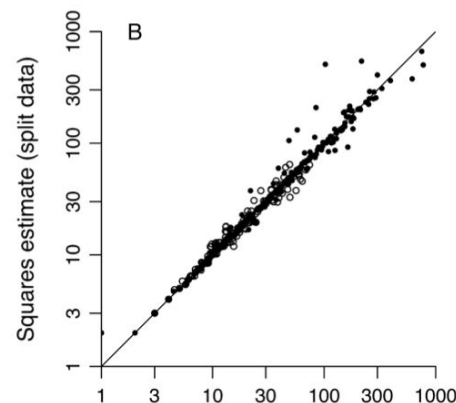
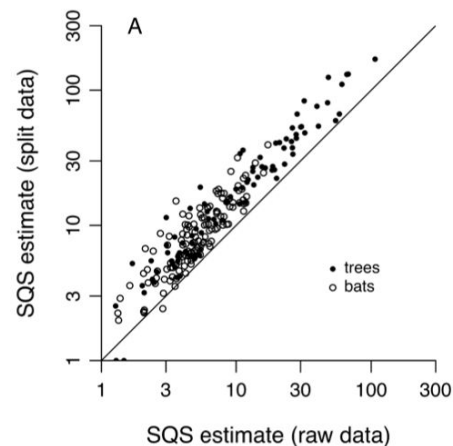
Criticism

- 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



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)

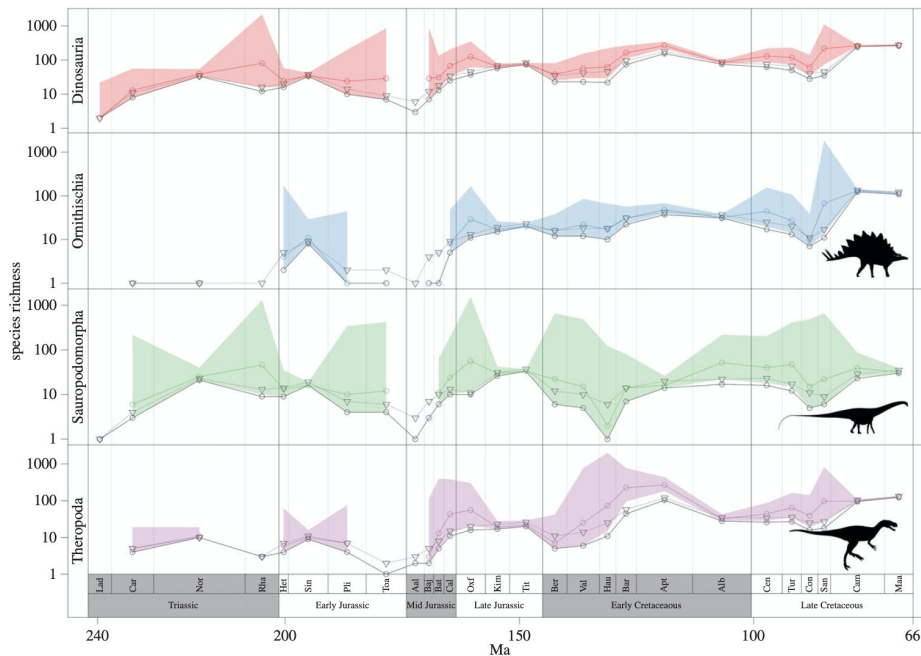


TRiPS

- True **R**ichness estimated using a **P**oisson **S**ampling model
- Extrapolation method ([Shiny app](#))

Criticism

- Tracks unstandardised diversity, particularly in well-sampled bins
- Doesn't deal well when dataset contains many common species



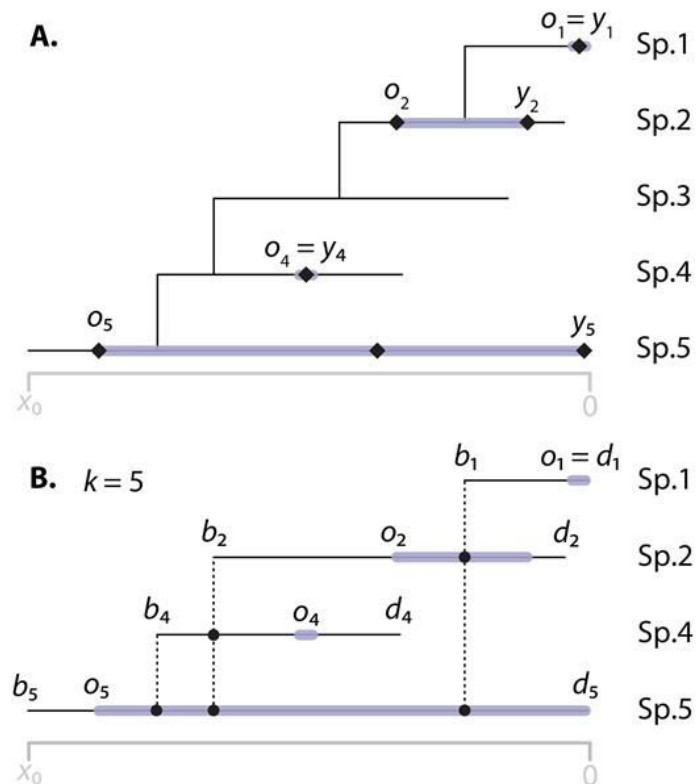
Other methods

Capture-mark-recapture

- Estimating rates of origination and extinction and diversity- through-time
- Next week: **Isaiah Smith & Lee Hsiang Liow**

Fossilized birth-death process

- 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**

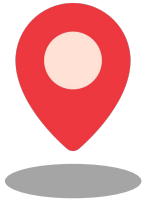


Scales of biodiversity

 α

Alpha diversity

“local richness”

 β

Beta diversity

between two
areas/regions

 γ

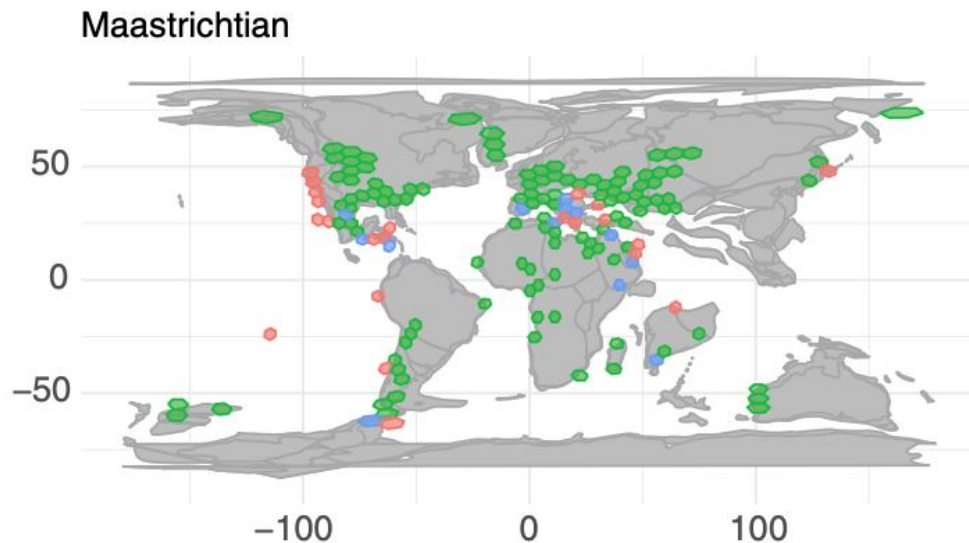
Gamma diversity

“global” diversity



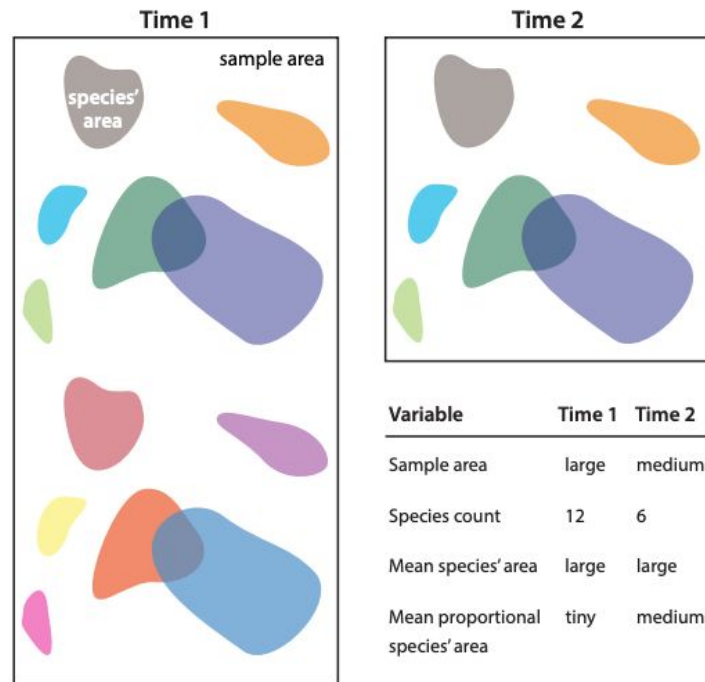
Global vs. regional diversity

- Some authors argue that 'global' diversity is meaningless
- Spatial sampling biases are pervasive and must be corrected for
- Greater focus needed on regional diversity?



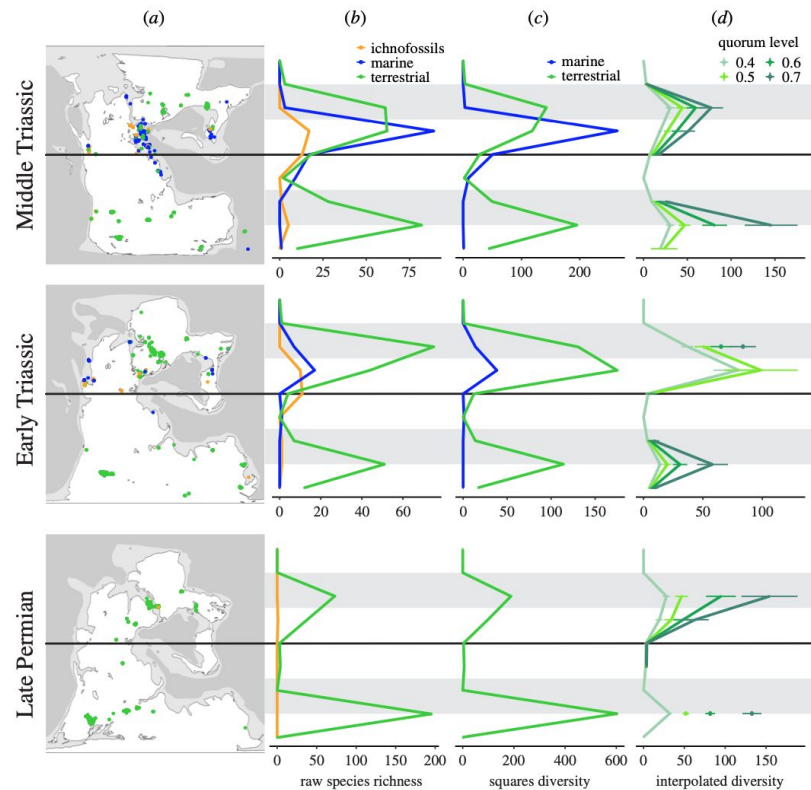
Correcting for spatial sampling

1. **Minimum spanning tree length (MST):** Attempting to control for the species-area effect. Draws spatial samples of fixed extent
2. **Spatial subsampling:** Control both total sampling area (no. of occupied grid cells) and dispersion of sampling sites through time (MSTs) - R package coming soon!



Which method is the best?

- Most popular = 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!



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(s) best suit your data and your questions/hypotheses
5. Remember that your *corrected* estimates are not automatically 'correct'!

