Modeling changes to the functional composition of North American mammal diveristy

multi-level dynamics of a regional species pool

Peter D Smits

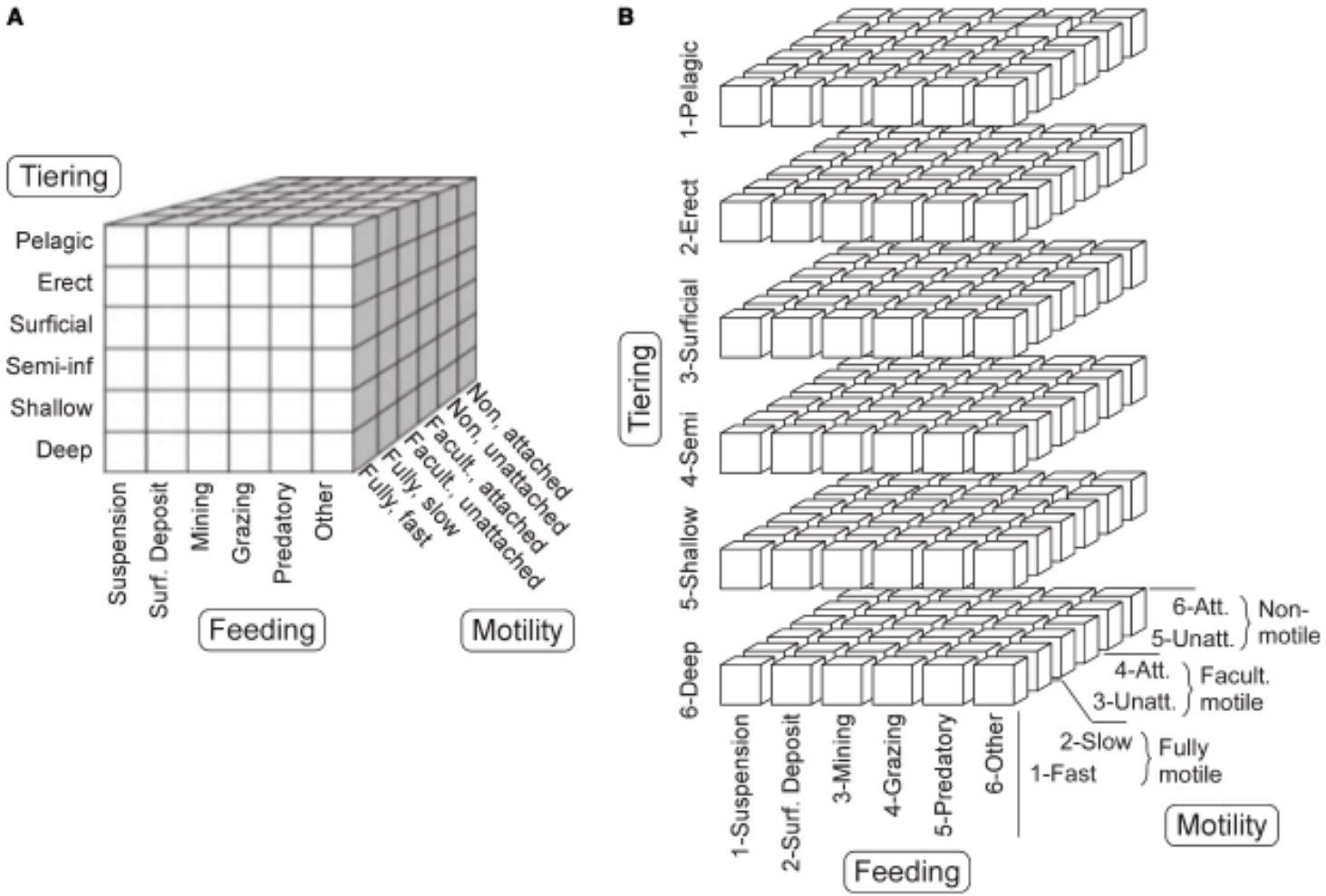
Department of Integrative Biology, University of California – Berkeley

Question

When and why are certain ecologies or functional groups enriched or depleted in a species pool?

- relative abundance
- species traits
- environmental context

Eco-cube and functional groups



TEXT-FIG. 1. Ecospace as defined by the three axes of tiering, motility level and feeding strategy. A, the ecospace cube with categories on each axis labelled. B, the ecospace cube 'exploded', showing 216 'bins' or modes of life specified by the combination of the categories on each ecospace axis.

(Bambach *et al.*, *2007*, Palaeontology)

Species pool concept

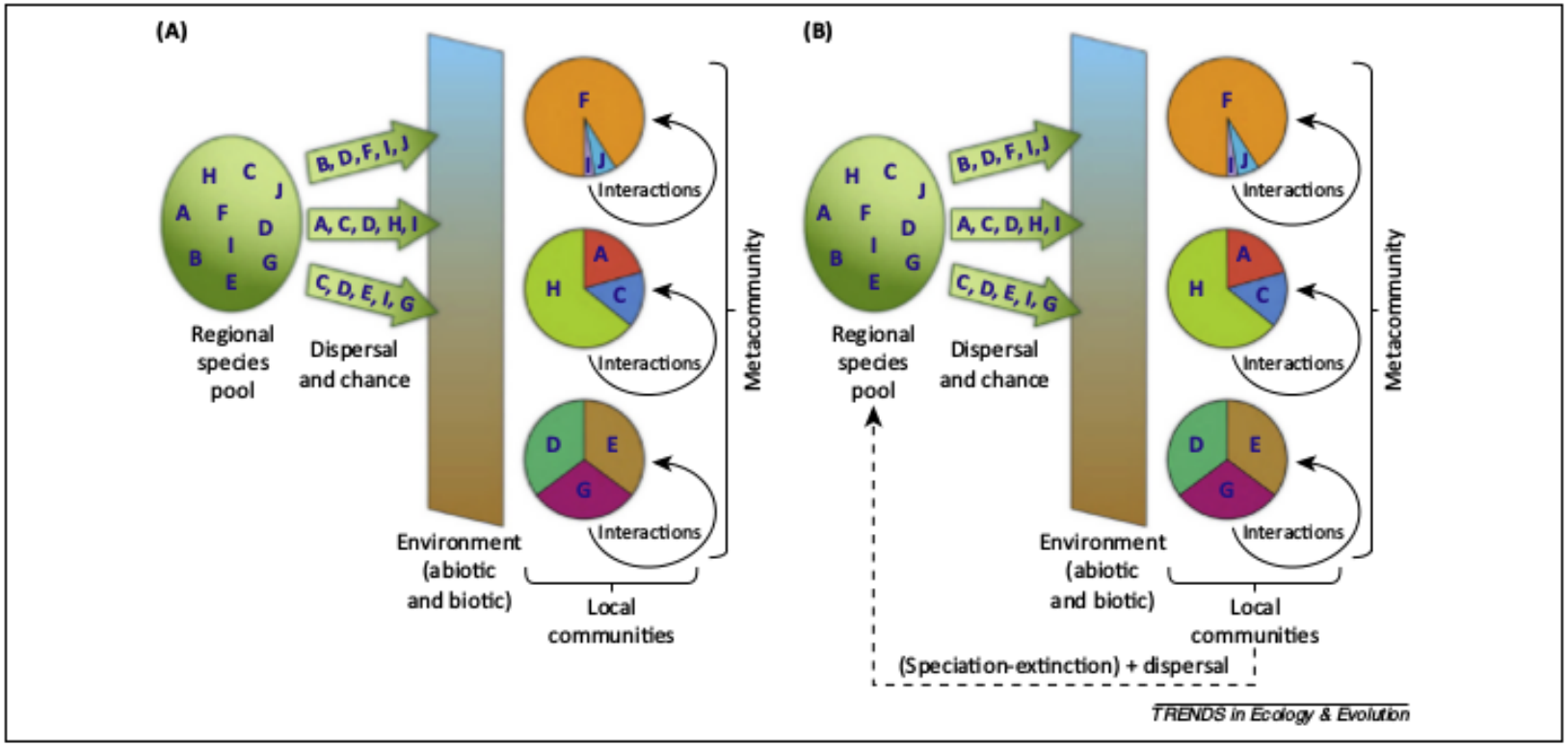
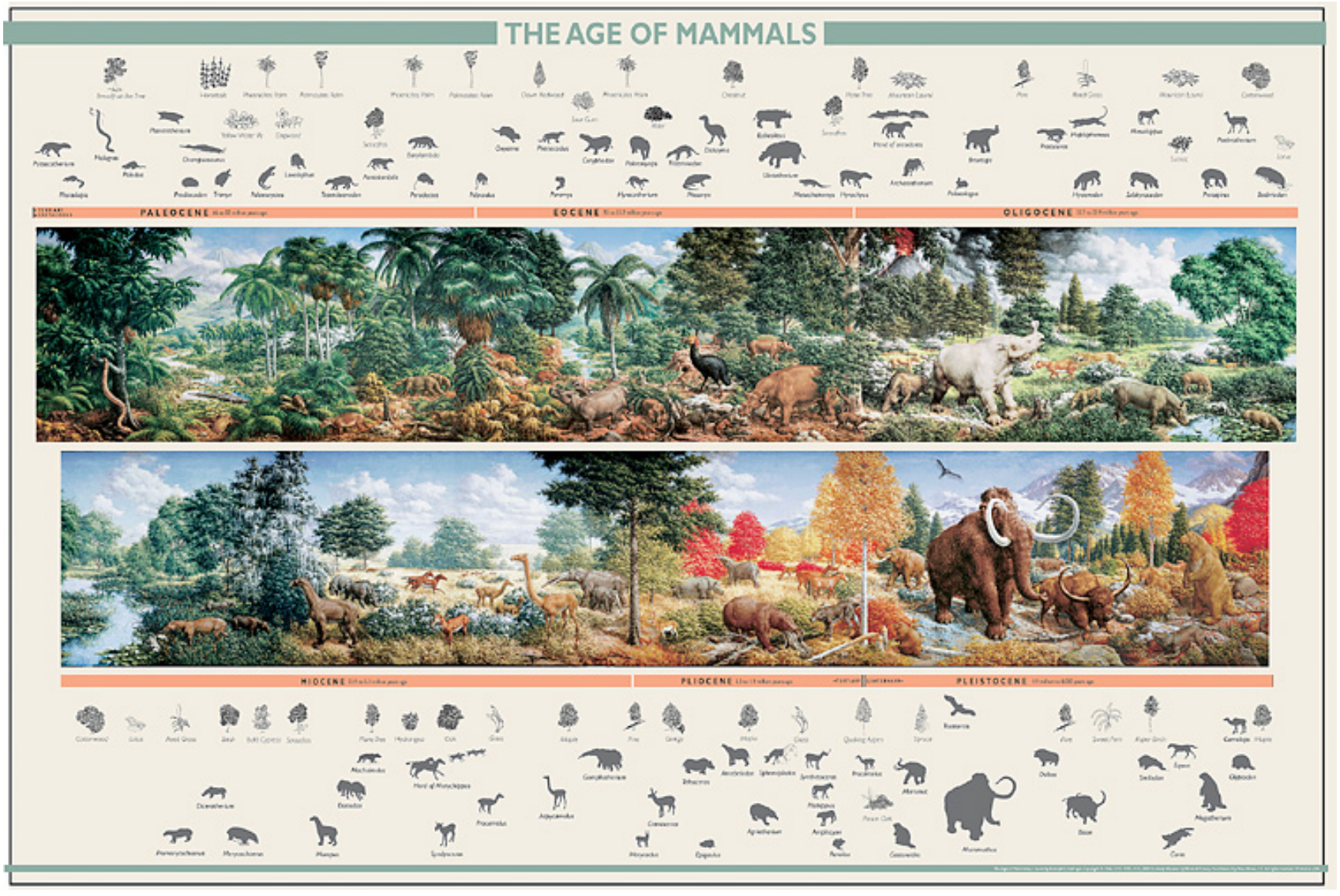


Figure 1. Two models of community assembly. **(A)** Local communities comprise a subset of species from the regional species pool that have passed through environmental filters. There is no feedback from the metacommunity (collection of local communities) to the regional species pool. Adapted from [5]. **(B)** Local communities are assembled as in (A), but speciation adds new species to the pool, extinction removes others, and dispersal allows the persistence of species that might otherwise go extinct.

(Mittelbach and Schemske, 2015, *TREE*)

Cenozoic mammals of North America



Structured, multi-level data in biology

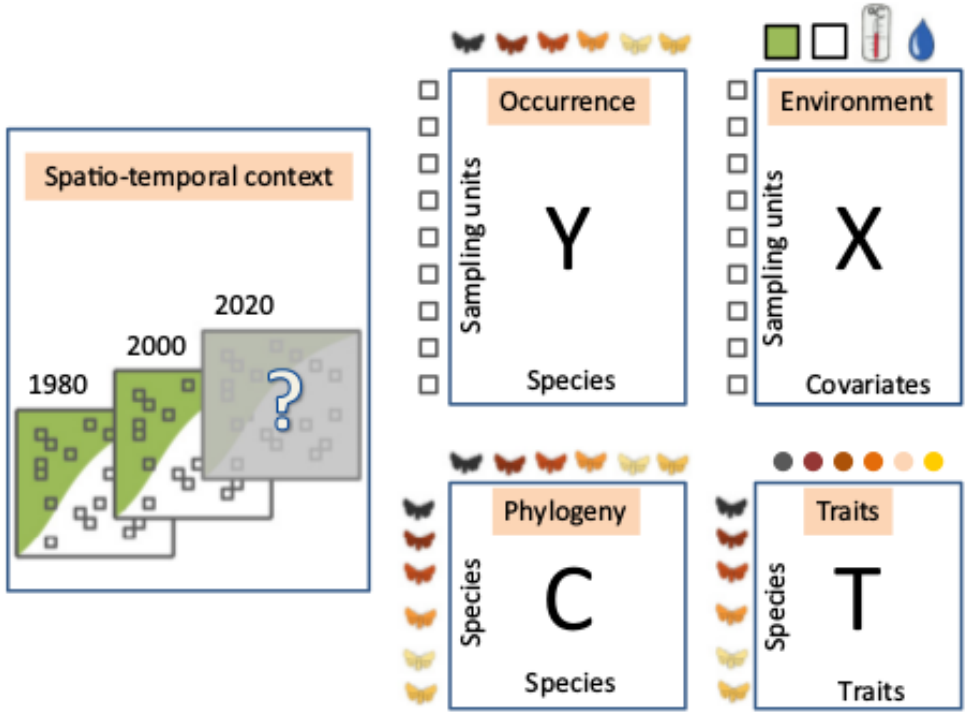
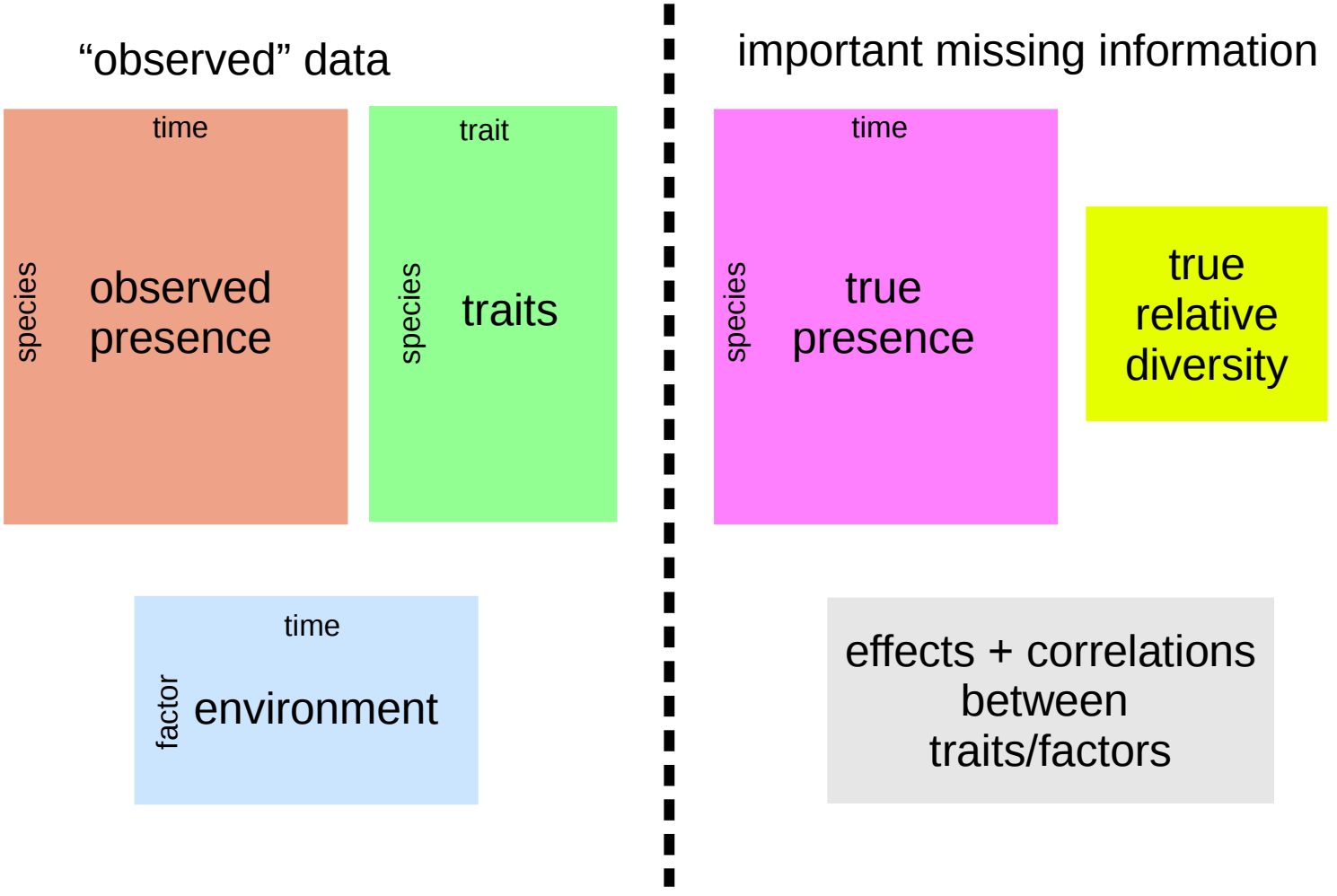


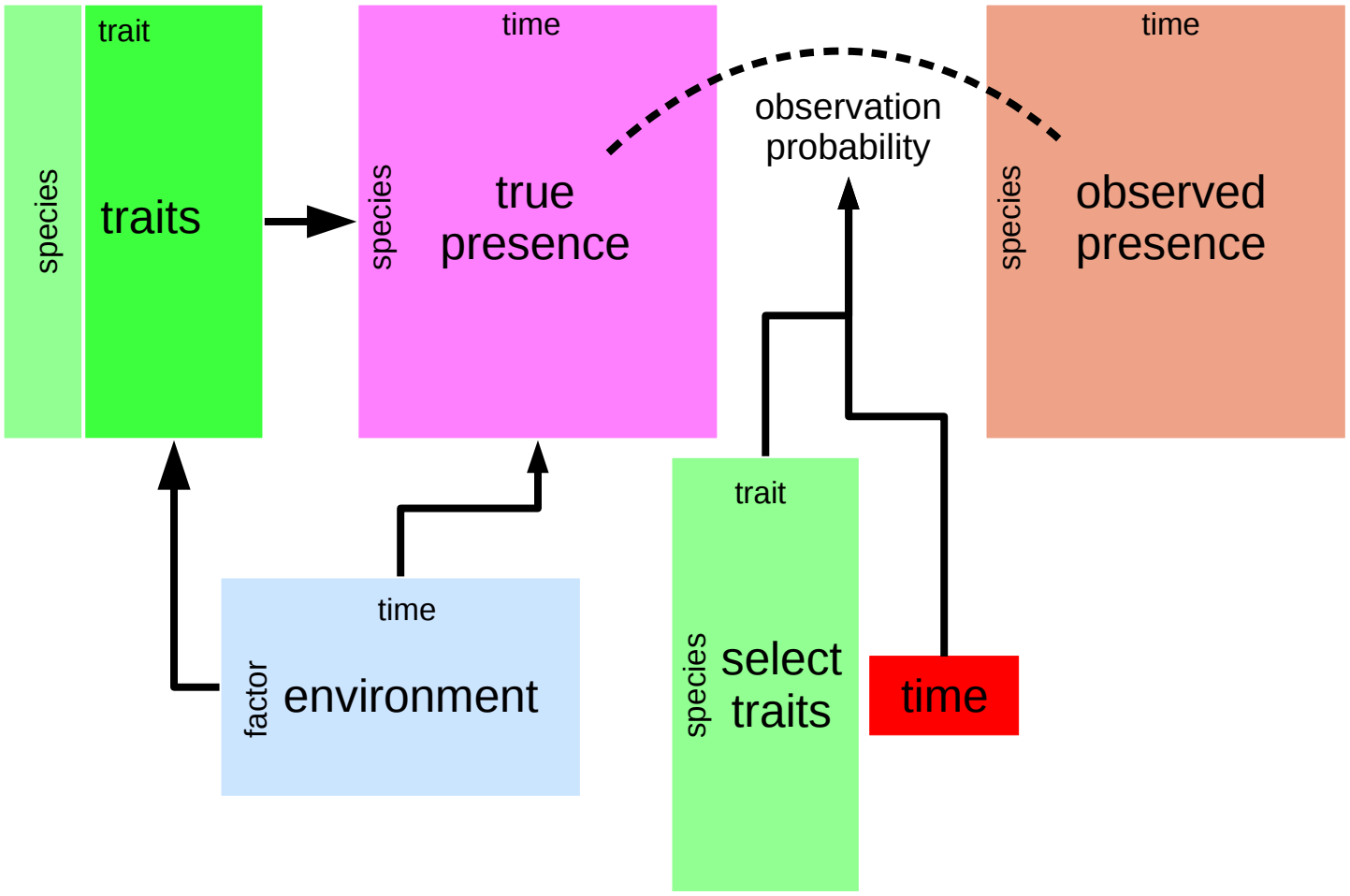
Figure 3 Data typically collected in community ecology. The occurrence data (denoted as the **Y** matrix) includes the occurrences of the species recorded in a set of temporal and/or spatial sampling units. The environmental data (denoted as the **X** matrix) consists of the environmental covariates measured over the sampling units. The traits data (denoted as the **T** matrix) consists of a set of traits measured for the species present in the **Y** matrix. To account for the phylogenetic dependencies among the species, we can include a fourth matrix consisting of the phylogenetic correlations among the species (denoted as the **C** matrix). The spatiotemporal context includes location and time information about the samples.

(Ovaskainen *et al.* 2017 *Ecology Letters*)

Conceptualizing the question and data



Conceptualizing the analysis



Hidden Markov Model with absorbing state

Jolly-Seber CMR/Restricted occupancy model

$$y_{i,t} \sim \text{Bernoulli}(z_{i,t} p_{i,t})$$

$$z_{i,t=1} \sim \text{Bernoulli}(\phi_{i,t=1})$$

$$z_{i,t} \sim \text{Bernoulli}\left(z_{i,t-1}\pi_{i,t} + \sum_{x=1}^{t}(1 - z_{i,x})\phi_{i,t}\right)$$

y observed state; z estimated state.

p observation; ϕ origination; π survival.

i in N; t in T.

Modeling the probabilities; individual-level

Multi-level logistic regression

$$p_{i,t} \sim \text{logit}^{-1}(b_t + e_{j[i]} + \beta^p mass_i)$$
$$\phi_{i,t} \sim \text{logit}^{-1}(f^{\phi}_{j[i],t} + o^{\phi}_{k[i]} + \beta^{\phi} mass_i)$$
$$\pi_{i,t} \sim \text{logit}^{-1}(f^{\pi}_{j[i],t} + o^{\pi}_{k[i]} + \beta^{\pi} mass_i)$$

observation: b_t time-varying intercept; $e_{j[i]}$ functional group eff; β^p mass eff.

origination: $f^{\phi}_{j[i],t}$ time/FG-varying intercept; $o^{\phi}_{j[i]}$ order eff; β^{ϕ} mass eff.

survival: $f^{\pi}_{j[i],t}$ time/FG-varying intercept; $o^{\pi}_{j[i]}$ order eff; β^{π} mass eff.

Modeling the probabilities; group-level

Multivariate regression of time/FG-varying intercept

$$f^{\phi} \sim \text{MVN}\left(\begin{matrix} U\gamma^{\phi}_{j=1} \\ \vdots \\ U\gamma^{\phi}_{j=J} \end{matrix}, \text{diag}(\tau_{f^{\phi}})\Omega_{f^{\phi}}\text{diag}(\tau_{f^{\phi}})\right)$$

$$f^{\pi} \sim \text{MVN}\left(\begin{matrix} U\gamma^{\pi}_{j=1} \\ \vdots \\ U\gamma^{\pi}_{j=J} \end{matrix}, \text{diag}(\tau_{f^{\pi}})\Omega_{f^{\pi}}\text{diag}(\tau_{f^{\pi}})\right)$$

U matrix group-level covariates; γ^{ϕ}, γ^{π} vectors group-level reg coefs.

Ω_{ϕ}, Ω_{π} corr matrix of FG by time; τ_{ϕ}, τ^{π} scale of FG by time.

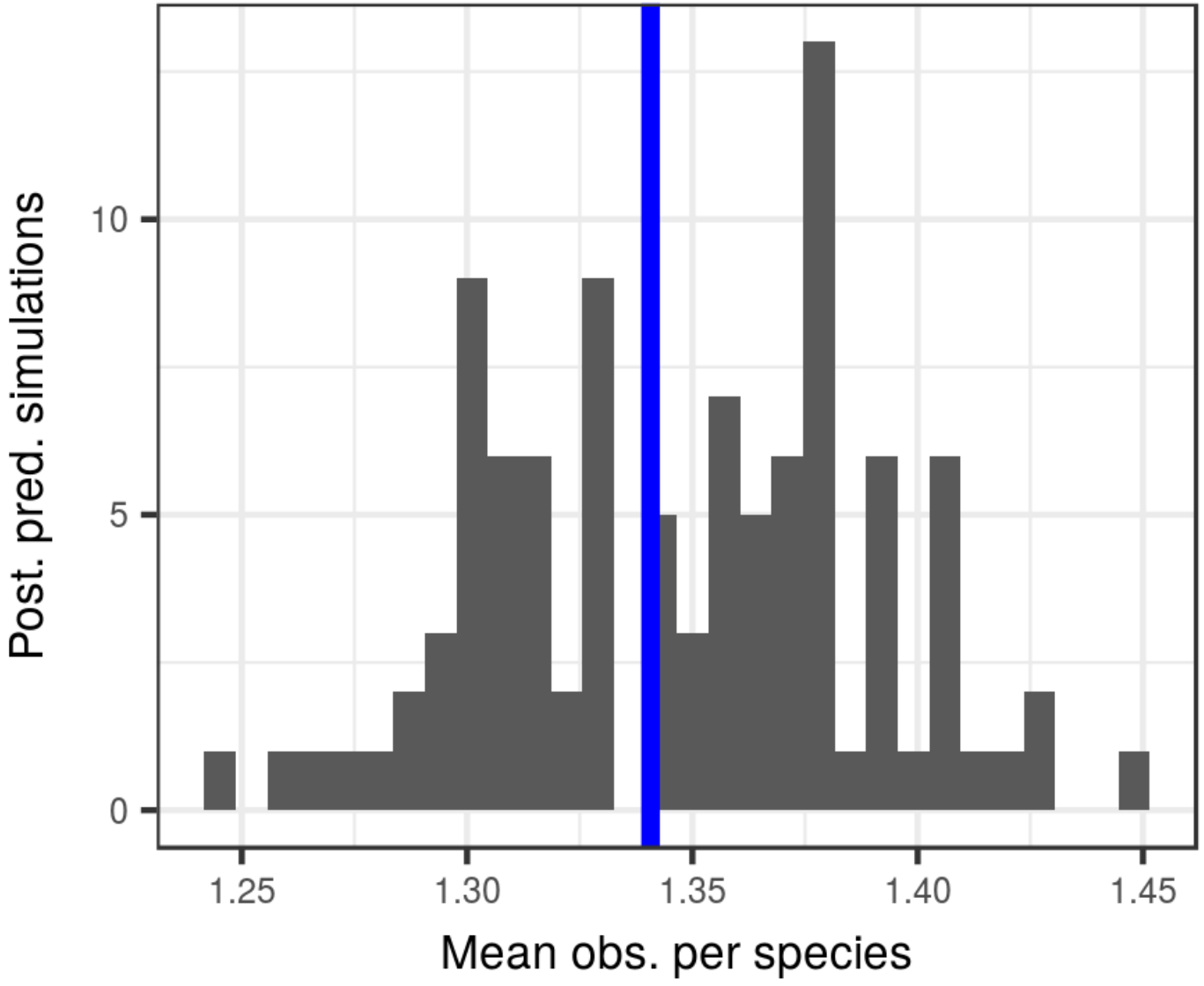
Final details (priors, implementation)

- Random-walk priors for time-varying intercepts to control.
- Strong priors against effects of mass on observation, origination, survival (e.g. $\mathcal{N}(0, 0.5)$).
- Strong priors against effects of group-level covariates on observation, origination, survival (e.g. $\mathcal{N}(0, 0.5)$).
- Strong prior against correlations between functional groups over time (e.g. LKJ(4)).
- Programmed as marginalization problem b/c gradient based estimation.

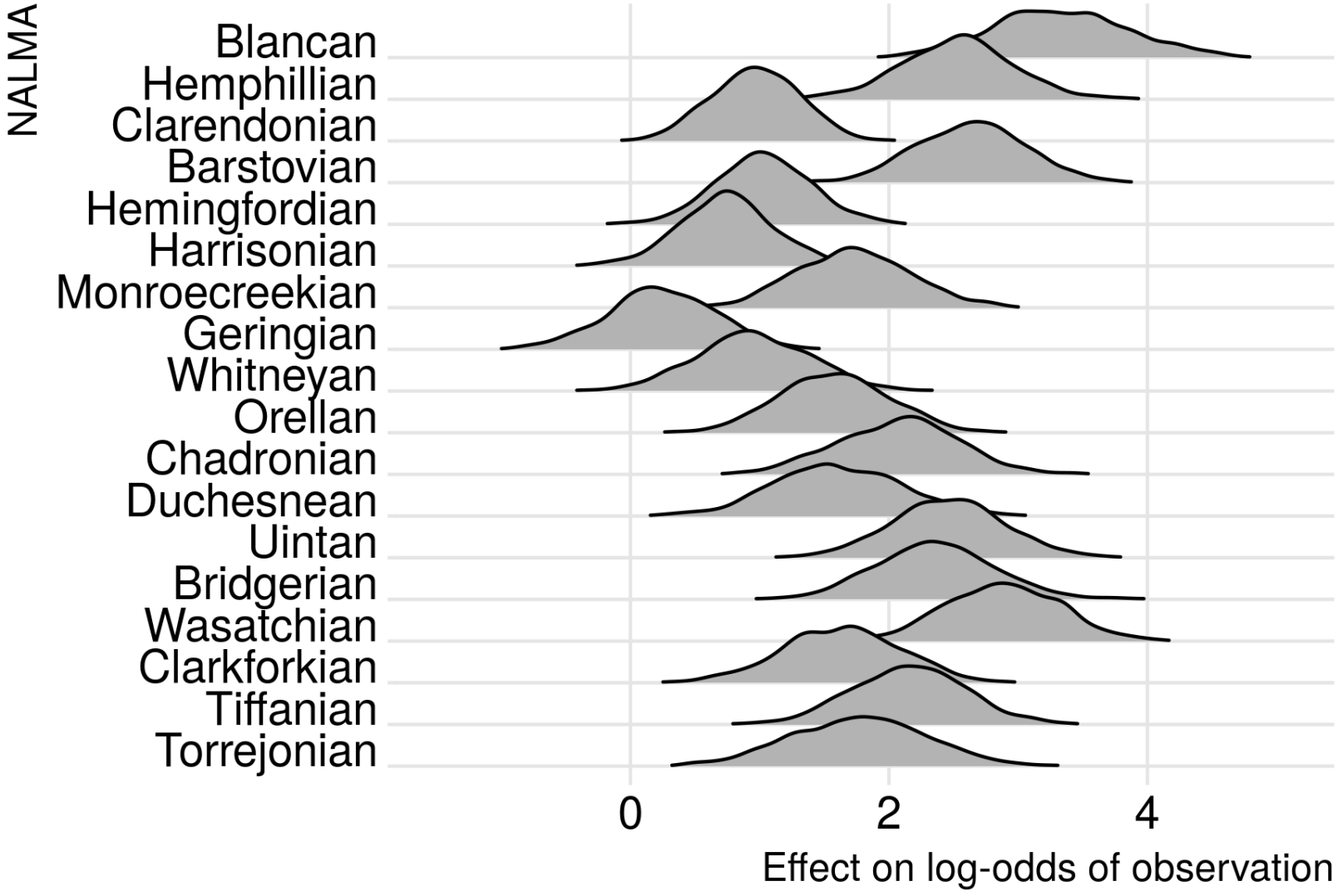
Parameter estimation and inference

- full HMC/MCMC slow
- **Automatic Differentiation Variational Inference (ADVI)**
  - approximate Bayesian inference
  - assumes posterior is Gaussian, no correlation between parameters
  - true Bayesian posterior

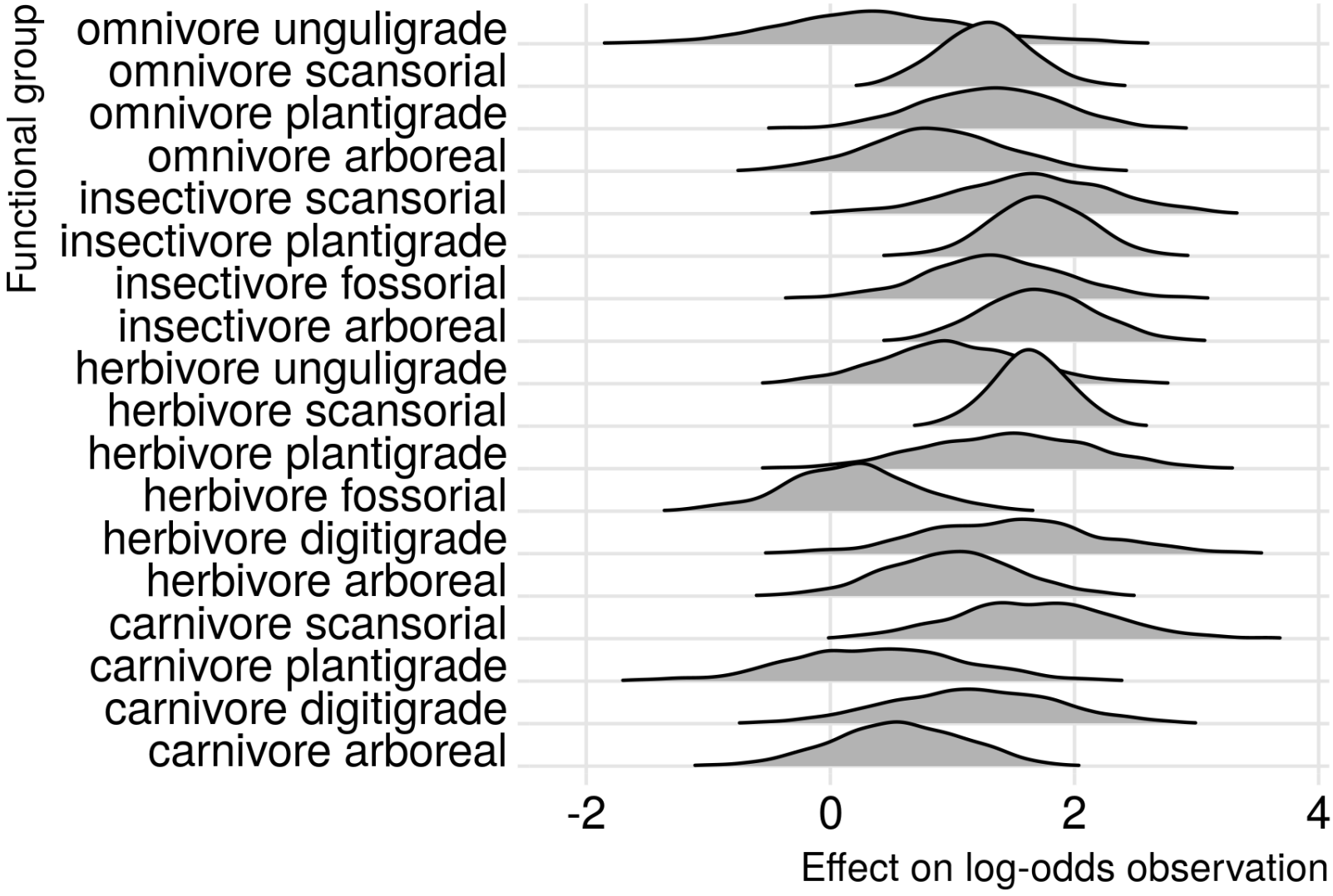
Model adequate? Posterior predictive check



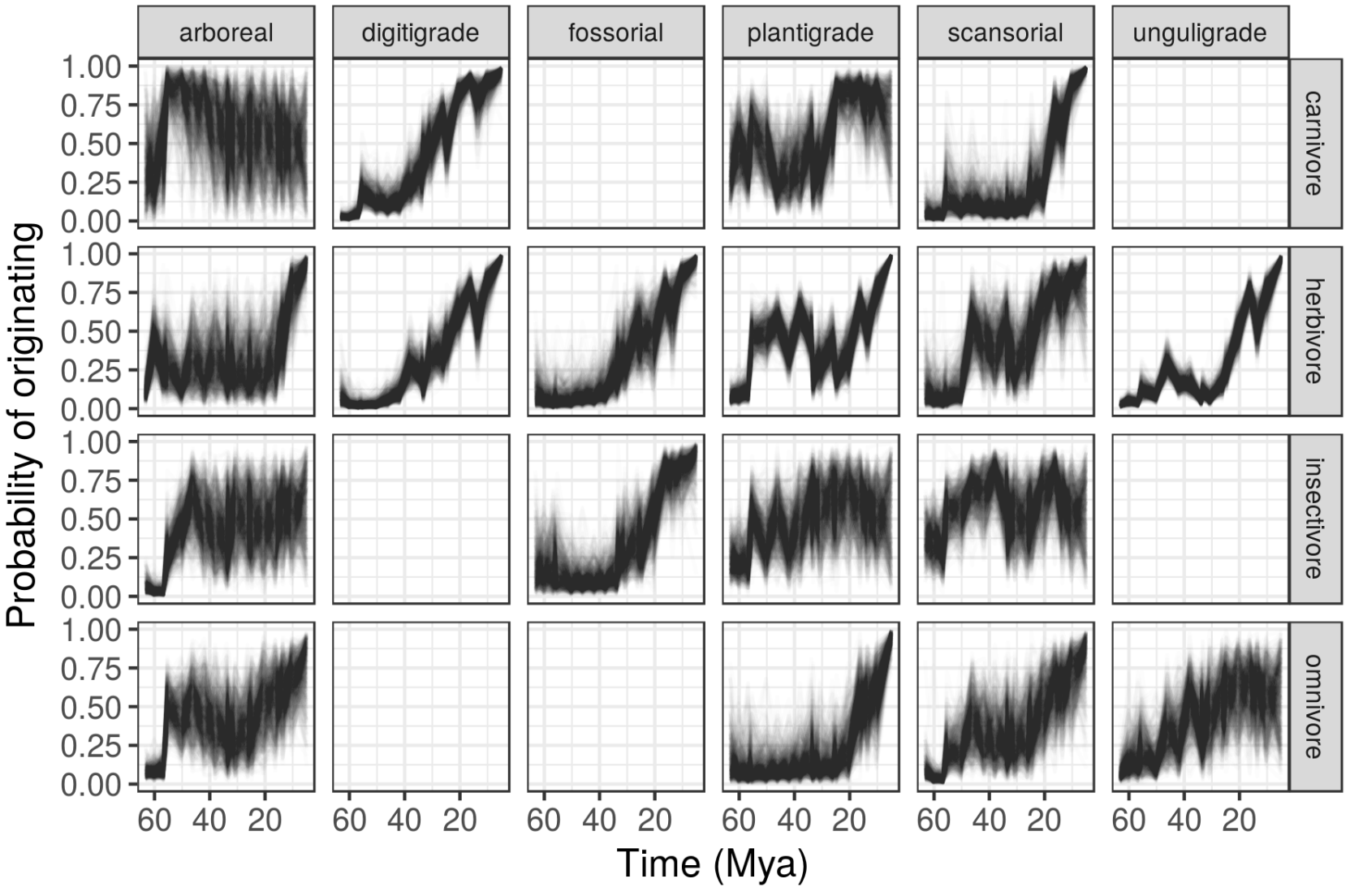
Observation probability; NALMA



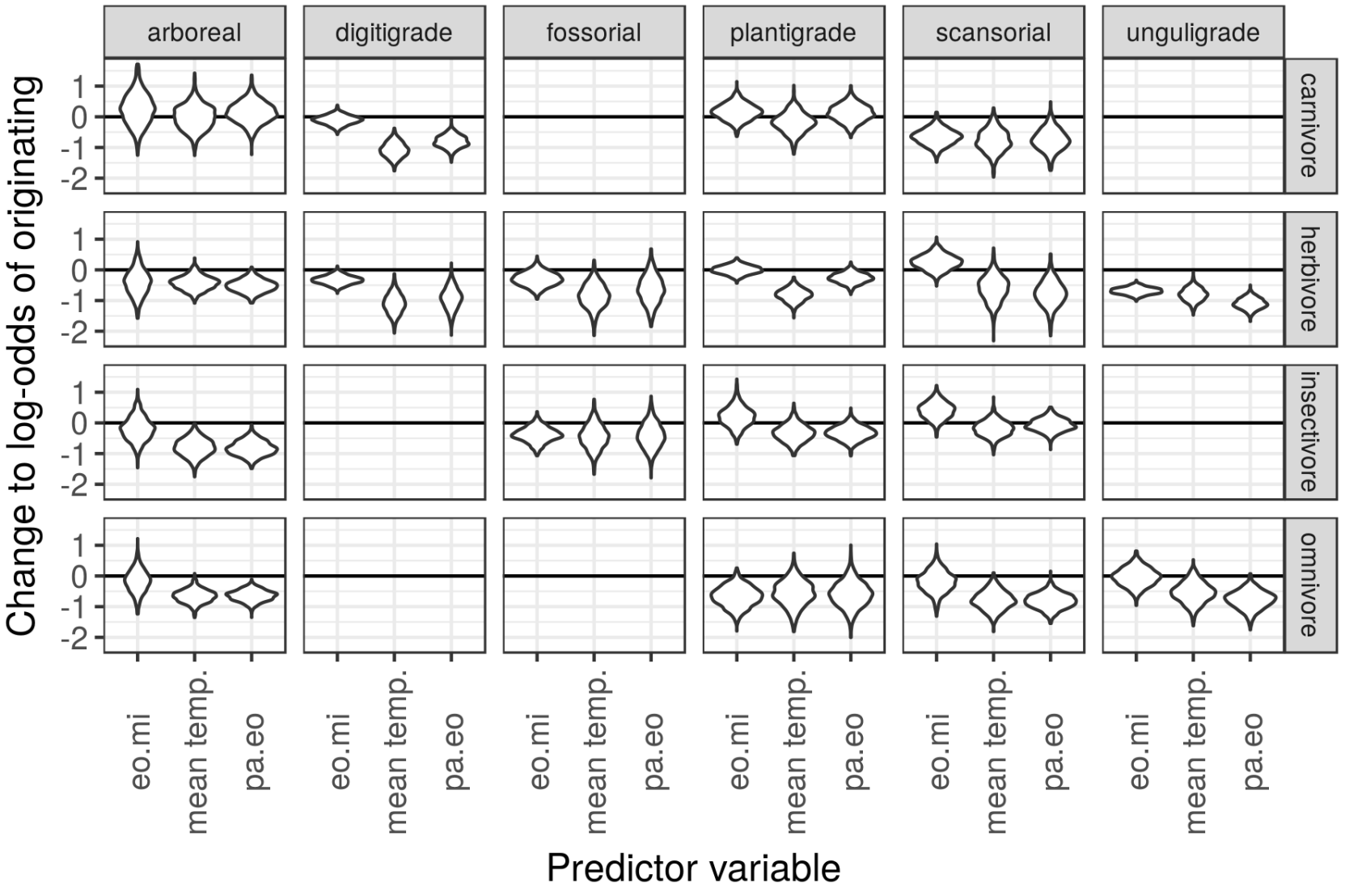
Observation probability; functional group



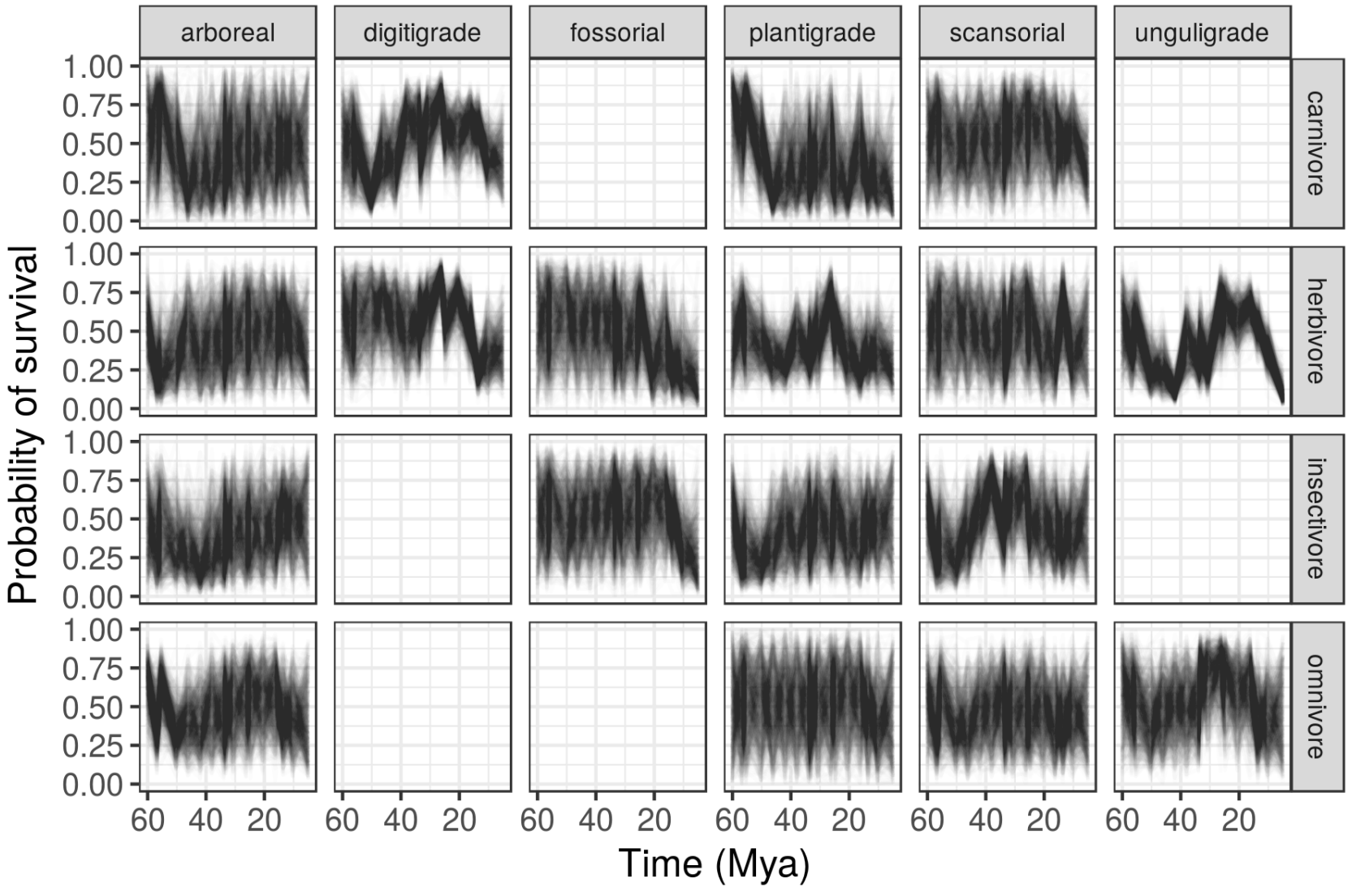
Origination probability; individual-level



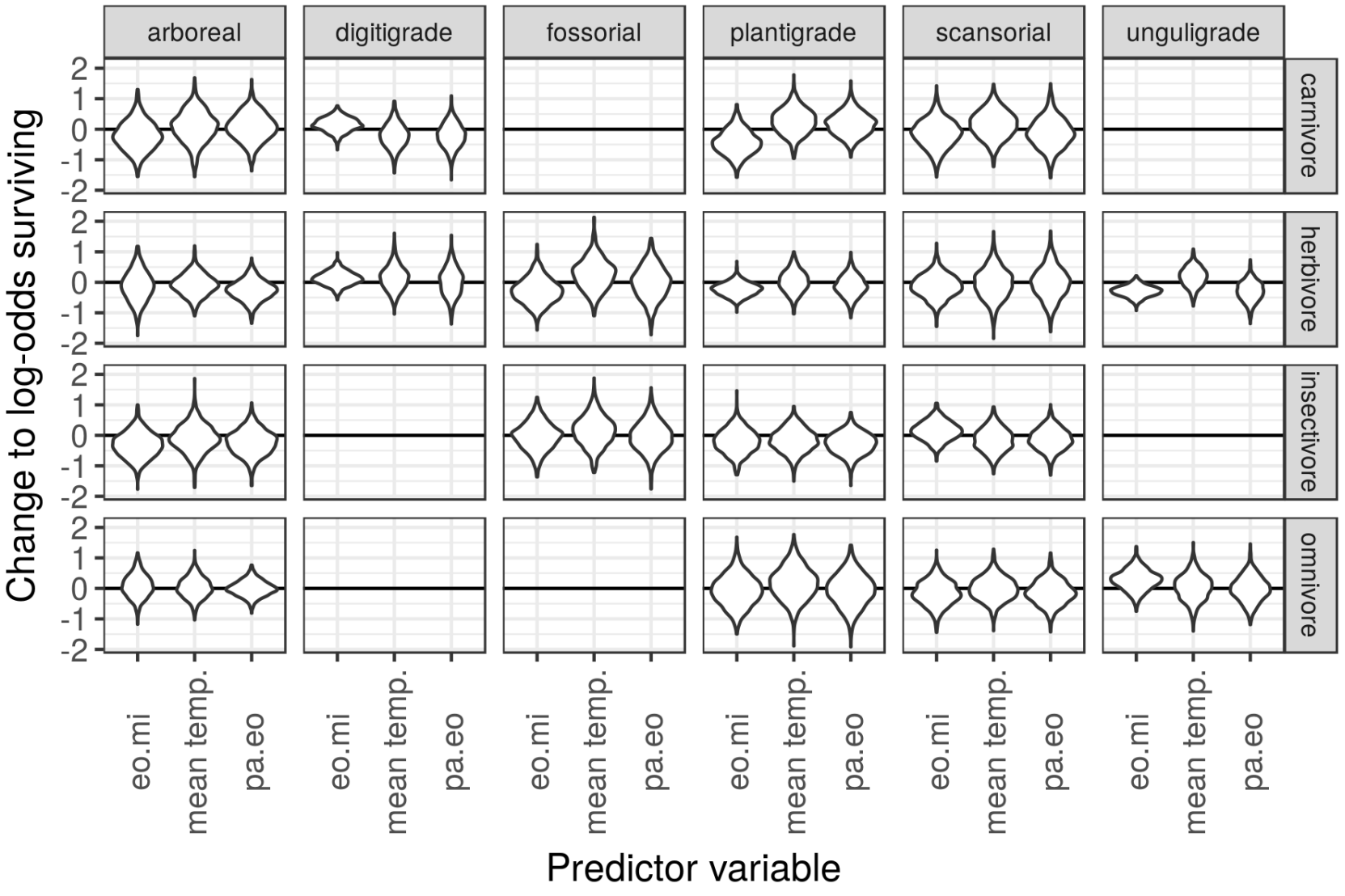
Origination probability; group-level



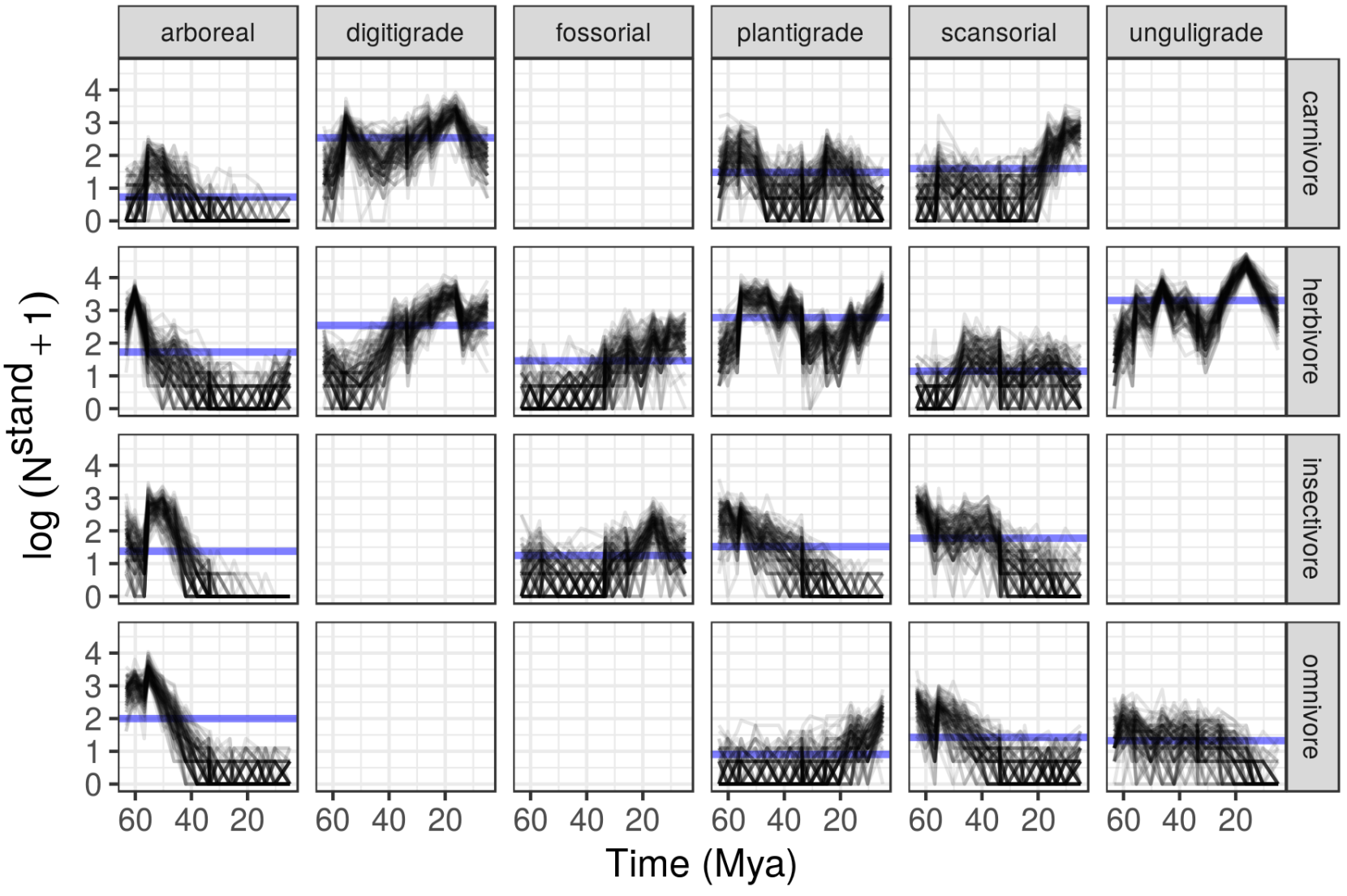
Survival probability; individual-level



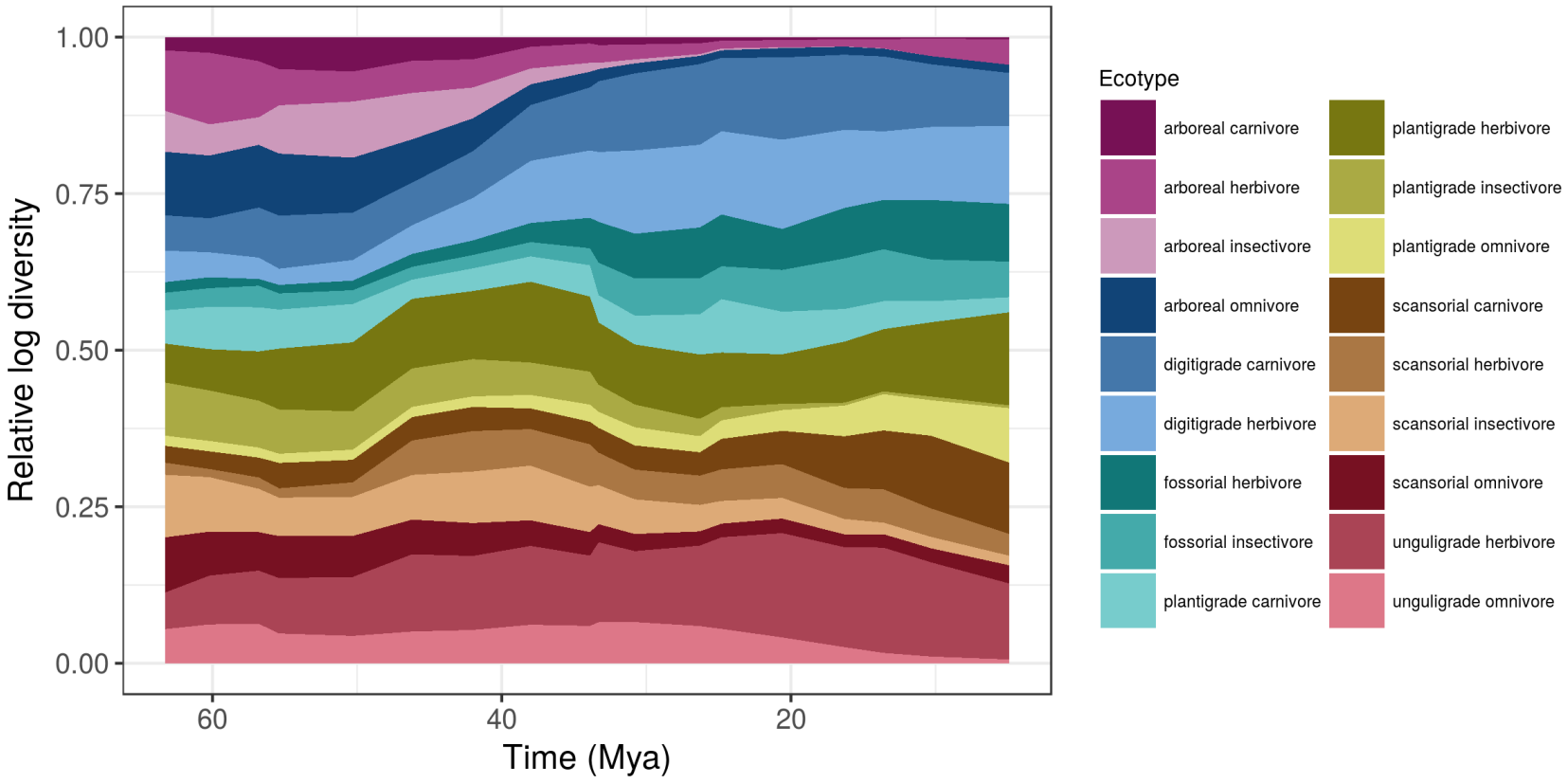
Survival probability; group-level



Standing diversity of functional groups through time



Relative diversity of functional groups through time



Summary and conclusions

-

Acknowledgements

- UC Berkeley
  - **Seth Finnegan**, Adiel Klompmaker, Emily Orzechowski, Larry Taylor, Sara Kahanamoku, Josh Zimmt
- UChicago
  - **Kenneth D. Angielczyk**, **Michael J. Foote**, P. David Polly, Richard H. Ree, Graham Slater

@PeterDSmits

The Paleobiology Database

revealing the history of life