



Simultaneous estimation and selection of species diversification model

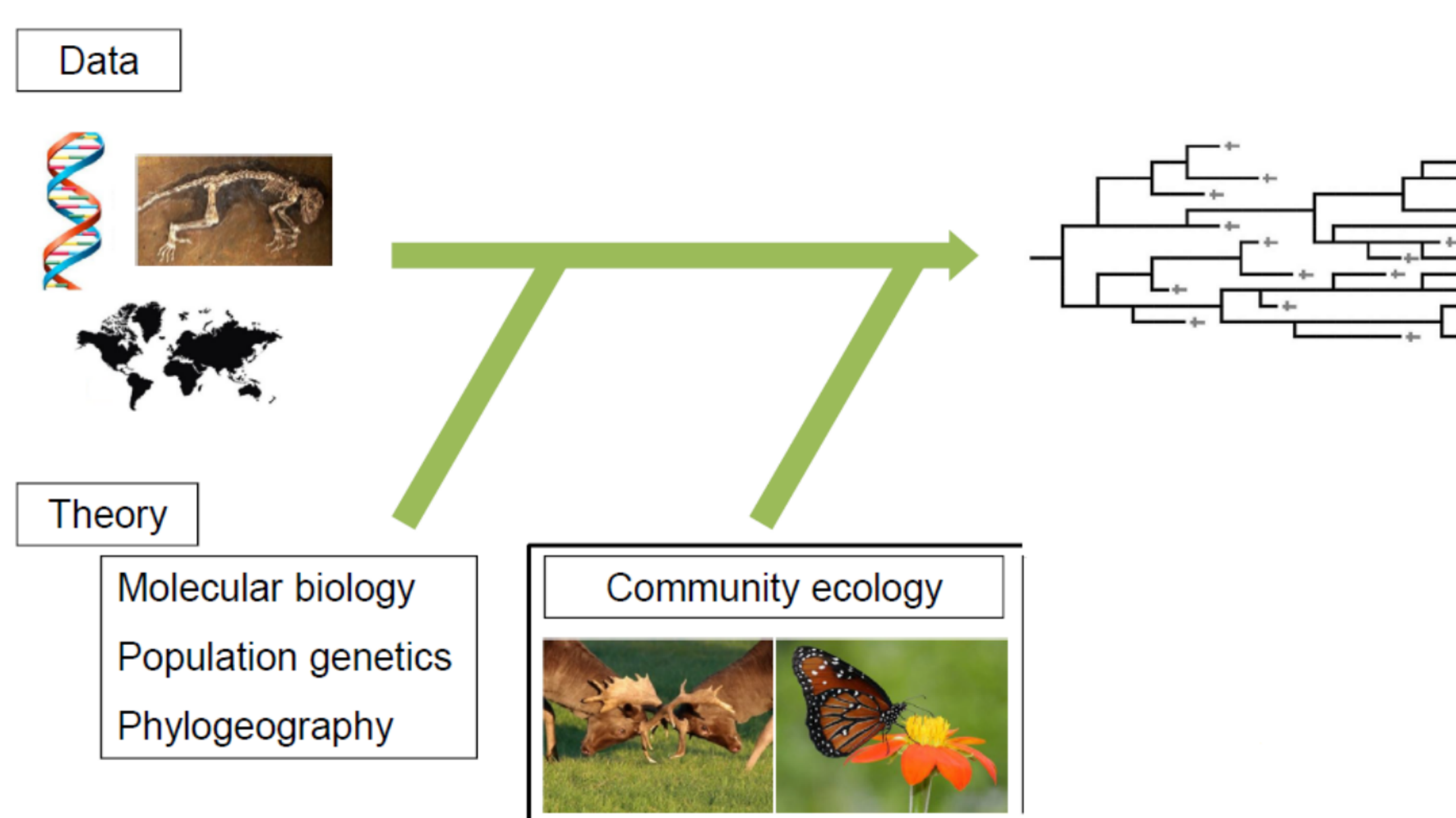
Francisco Richter, Ernst Wit & Rampal Etienne

University of Groningen, Faculty of Mathematics and Natural Sciences

Background

The mechanisms that control the diversification of species are poorly understood. In the last decades sophisticated diversification models have been developed, but they perform on a case-by-case basis. We propose a general speciation model with potentially many covariates in order to consider ecological interactions. This complex stochastic differential equation model can be written equivalently as a combination of two generalized linear models.

Set up



The phylogenetic tree is mathematically determined by

- A set of branching times \mathcal{T}
- The topology Υ .

And its likelihood function is defined as

$$L(Y|\Theta) = \prod_{i=1}^p \sigma_i e^{-\sigma_i t_i} \frac{\rho_i}{\sigma_i} \quad (1)$$

Where $\sigma_i = \sum_{j=1}^{n_i} \lambda_{i,j} + \mu_{i,j}$; $\lambda_{i,j}$ is defined as the **speciation rate** of the species j at time t_i and $\mu_{i,j}$ corresponds to the **extinction rate** of the same species. $\rho_i \in \Upsilon$ is the topology variable corresponding to the speciation or extinction rate taking place on waiting time t_i . Both λ and μ are linear functions of many potential explanatory variables.

MCEM & Case studies (work in progress)

The fact that we typically only have data on currently existing species is described as a missing data problem. Thus, we perform an MCEM algorithm

$$Q(\theta|\theta_{(i)}) = \int \log L(\theta|\mathbb{T}) d\mathbb{T} \rightarrow \theta_{(i+1)} = \arg \max_{\theta} Q(\theta|\theta_{(i)})$$

Diversity-dependence

For the first case studies we plug in the diversity-dependence model [1] under the framework:

$$\begin{cases} \lambda_{i,j} = \lambda_0 + (\lambda_0 - \mu_0) \frac{N_i}{K} \\ \mu_{i,j} = \mu_0 \end{cases}$$

Case study 1: Dendroica

EM iter	λ_0	K	μ_0
1	4.00	30	1.00
2	1.51	36	0.77
3	0.97	116	0.61
4	0.88	811	0.56
5	0.85	3165	0.53
6	0.85	∞	0.52
7	0.85	∞	0.52

Table: MCEM iterations for Dendroica phylogeny.

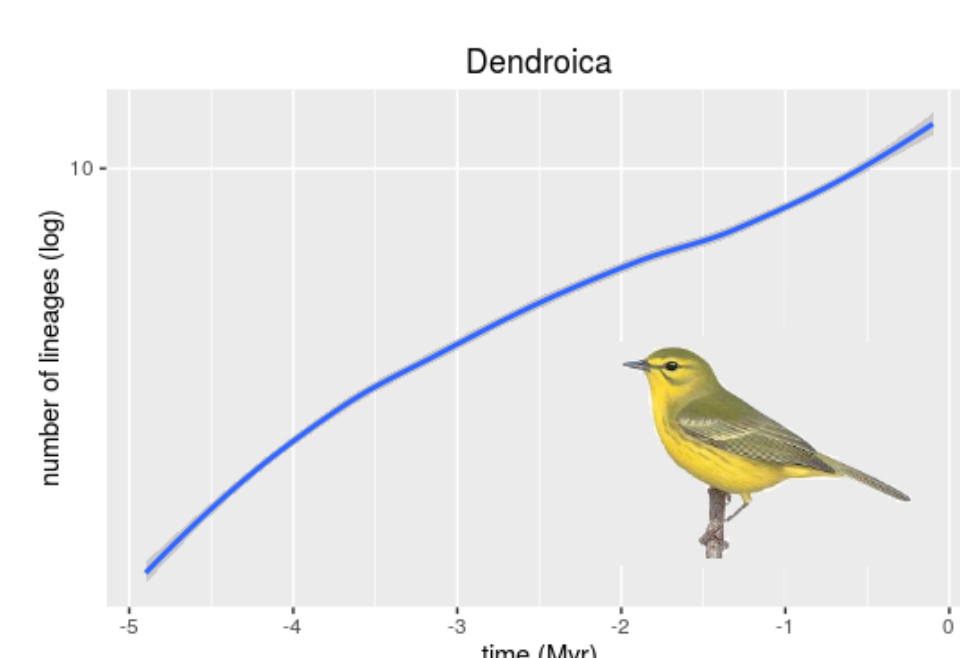


Figure: Expectation of lineages through time for the obtained parameters

Case study 2: Foraminifera

EM iter	λ_0	K	μ_0
1	4.00	30	1.00
2	0.79	35	0.30
3	0.37	42	0.13
4	0.25	47	0.07
5	0.19	46	0.05
6	0.19	46	0.04
7	0.16	45	0.03
8	0.16	45	0.03

Table: MCEM iterations for Foraminifera phylogeny.

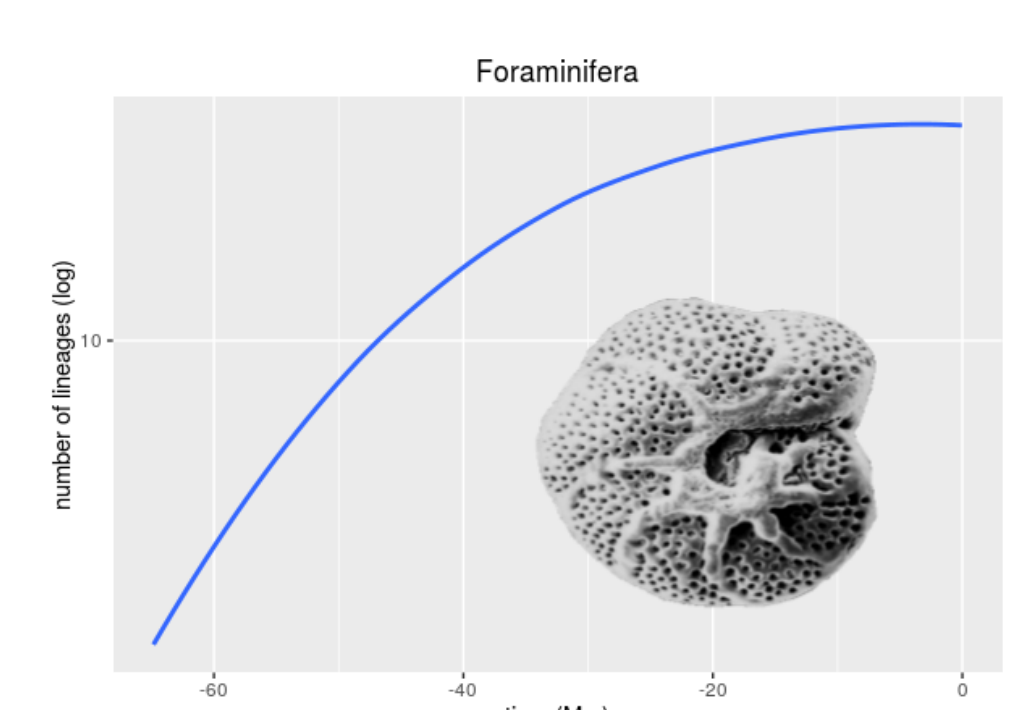


Figure: Expectation of lineages through time for the obtained parameters

Challenges

- Decay and fossilization degrade crucial evidence useful for phylogenetic analyses that could infer underlying mechanisms. In fact, we normally observe extant species only.

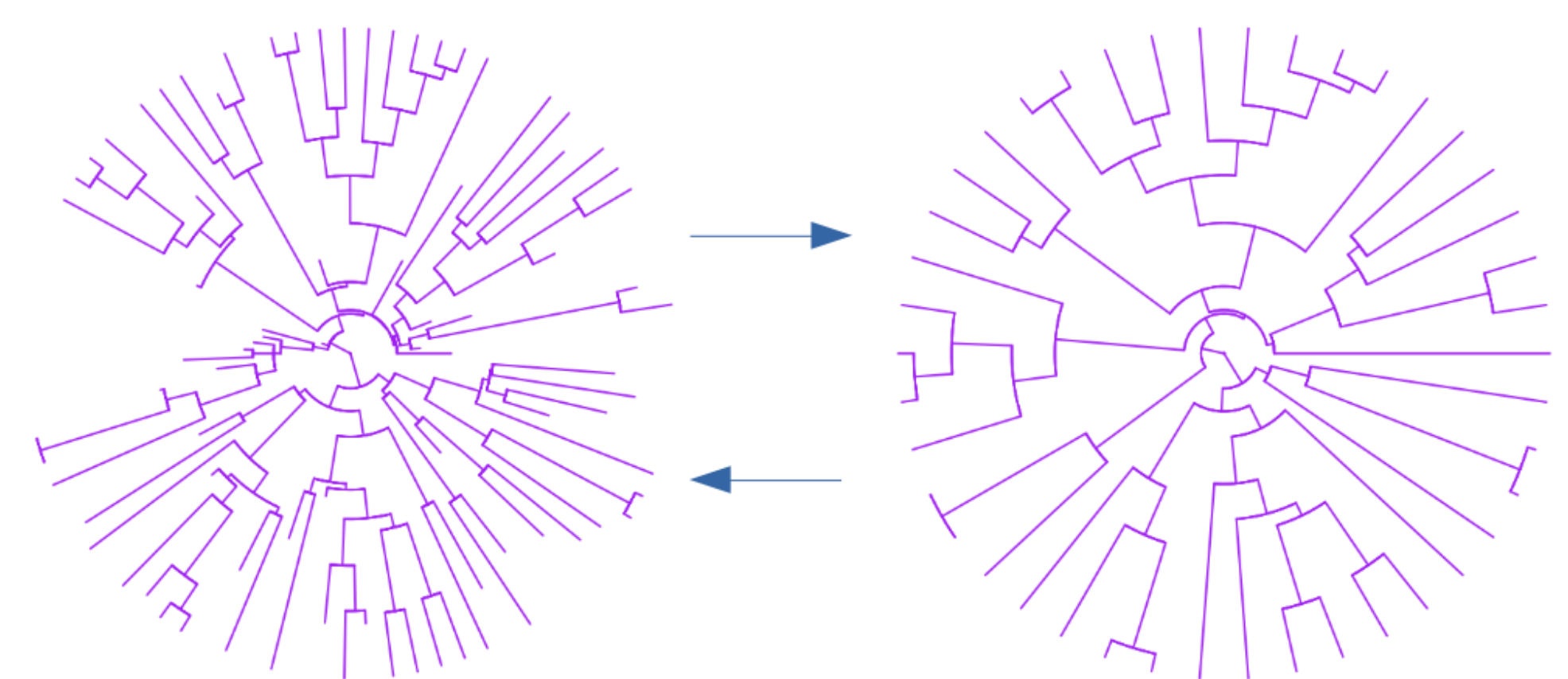


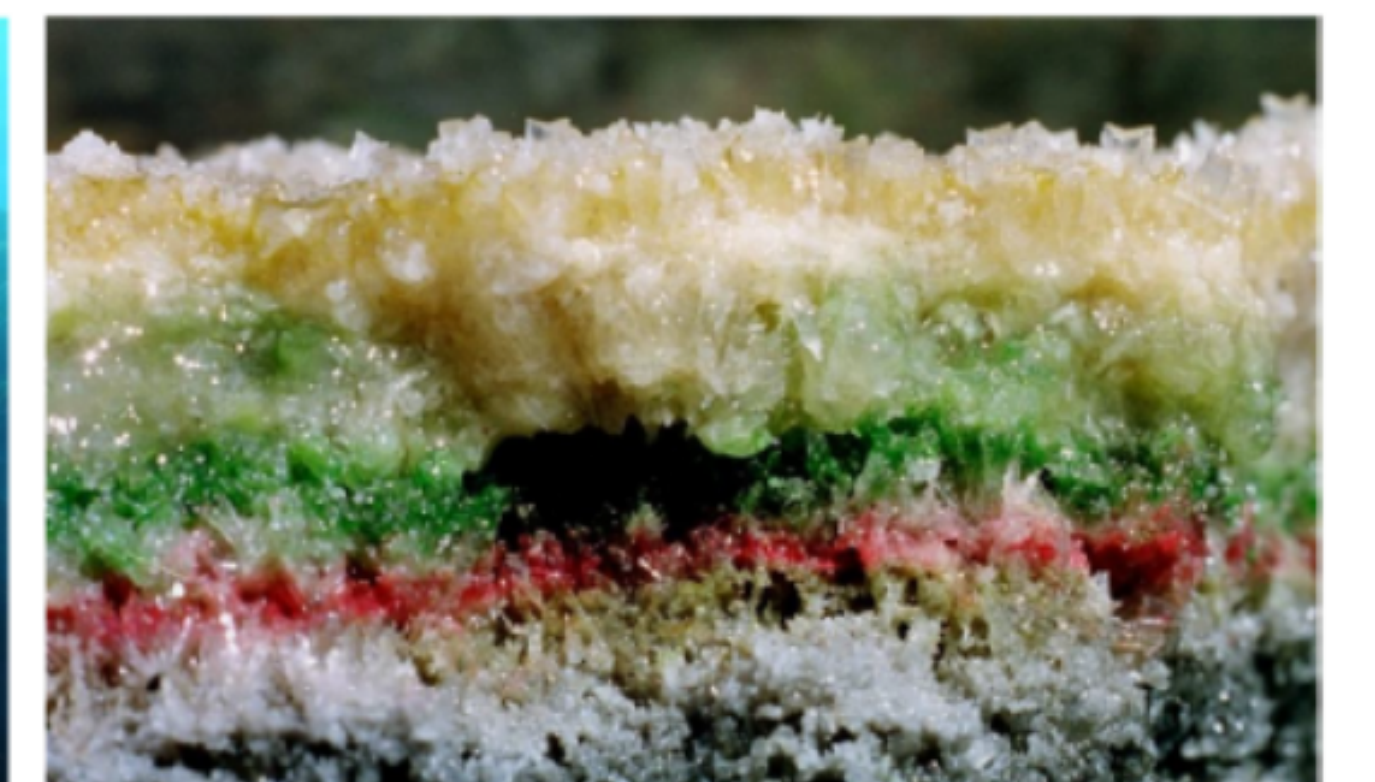
Figure: Phylogenetic trees where we can visualize the loss of information on reconstructed trees. At the left we have a tree with all extinct species whereas the right plot shows the same tree with only observable species.

- Diversification processes have many potential explanatory variables, which increases the dimensionality and the complexity of current models enormously.

Coral and fish communities



Microbial communities



Tropical forest communities



Savanna communities

Contact Information

- f.richter@rug.nl
- <https://nl.linkedin.com/in/francisorichter>

References

- [1] R. S. Etienne, B. Haegeman, T. Stadler, T. Aze, P. N. Pearson, A. Purvis, and A. B. Phillimore. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20111439, 2011.