

UC Merced

Frontiers of Biogeography

Title

Lineages through space and time plots: Visualising spatial and temporal changes in diversity

Permalink

<https://escholarship.org/uc/item/1521z9pb>

Journal

Frontiers of Biogeography, 11(2)

Author

Skeels, Alex

Publication Date

2019

DOI

10.21425/F5FBG42954

Supplemental Material

<https://escholarship.org/uc/item/1521z9pb#supplemental>

Copyright Information

Copyright 2019 by the author(s). This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>



Lineages through space and time plots: Visualising spatial and temporal changes in diversity

Alexander Skeels

Macroevolution and Macroecology
Group, Research School of Biology,
Australian National University,
Canberra 0200, Australia.
alexander.skeels@anu.edu.au;
tel. +61 2 6125 9138

Abstract. During the radiation of a clade, diversification rates can show temporal patterns such as a speedup or slowdown, which might relate to different ecological and evolutionary mechanisms. The temporal dynamics of diversification of whole clades are often visualised as a lineage-through-time (LTT) plot, which traces the number of reconstructed lineages at different time points. However, clades do not radiate evenly across space and may show different temporal dynamics in different regions. As such, a biogeographic approach is required to more completely understand temporal diversification dynamics. Here, I present a tool to extract temporal diversity information across different biogeographic regions from the output of commonly used ancestral range estimation models implemented in the R package BioGeoBEARS. The lineages through space and time (LTST) plot allows for visualisation of diversification dynamics in different regions, formatted in an accessible way which can be used for further quantitative analysis.

Keywords: diversification, BioGeoBEARS, slowdown, biogeography, phylogeny, lineage-through-time plot (LTT)

Introduction

The temporal dynamics of the build-up of diversity in different regions is key to understanding the emergence of biogeographic patterns in diversity. A common method to explore temporal patterns of diversification within whole clades is to plot the logarithm of the reconstructed number of lineages in different time-slices drawn through a time-calibrated molecular phylogeny; a lineages-through-time plot (LTT; Figure 1A; Nee et al. 1992; Harvey et al. 1994). LTTs can be used to visually explore when diversification is deviating from a straight line as expected under a constant rates model, for example as the net diversification rate of a clade slows down during the late stages of an adaptive radiation (Nee et al. 1992; Phillimore and Price 2008). Temporal diversification information can be analysed quantitatively to see when diversification deviates from the expectation of constant rates (e.g., Pybus and Harvey 2000), and the shape of LTT curves have recently been shown to be a useful summary metric for diversification diagnostics (Janzen et al. 2015). LTTs have been used extensively in evolutionary biology, forming a standard visualisation procedure in the toolbox of evolutionary biologists. However, diversification dynamics are not spatially homogenous and within a single clade diversity may accumulate more or less rapidly in different regions. To visualise the spatial

context of diversification through time first requires an estimate of the geographic distribution occupied by each reconstructed lineage in the phylogeny.

A suite of ancestral range estimation (ARE) methods are available to infer when in history lineages have occupied different regions based on models of geographic range evolution (Ronquist 1997; Ree and Smith 2008; Landis et al. 2013; Matzke 2013a). These methods model range dynamics by taking geographic range information, as measured by the present-day occupancy of species in discrete regions (often coded broadly at the scale of ecoregions, biomes, or biogeographic realms), as well as a time-calibrated phylogeny, and fitting models of geographic range evolution to these data. Different implementations of AREs model range shifts via dispersal between regions, speciation amongst or between regions, or extirpation from a region. ARE models can be constrained to account for the effect of geographic or environmental distance on dispersal (Van Dam and Matzke 2016), trait-based dispersal capacity of species (Klaus and Matzke 2019), historical information on geographic ranges from the fossil record (Matzke 2013a), or can be time-stratified to account for the ontogeny of different regions, such as the formation of islands or rise of mountains (Matzke 2014).

Given this flexibility and suitability of ARE models to a range of diverse systems, the application of

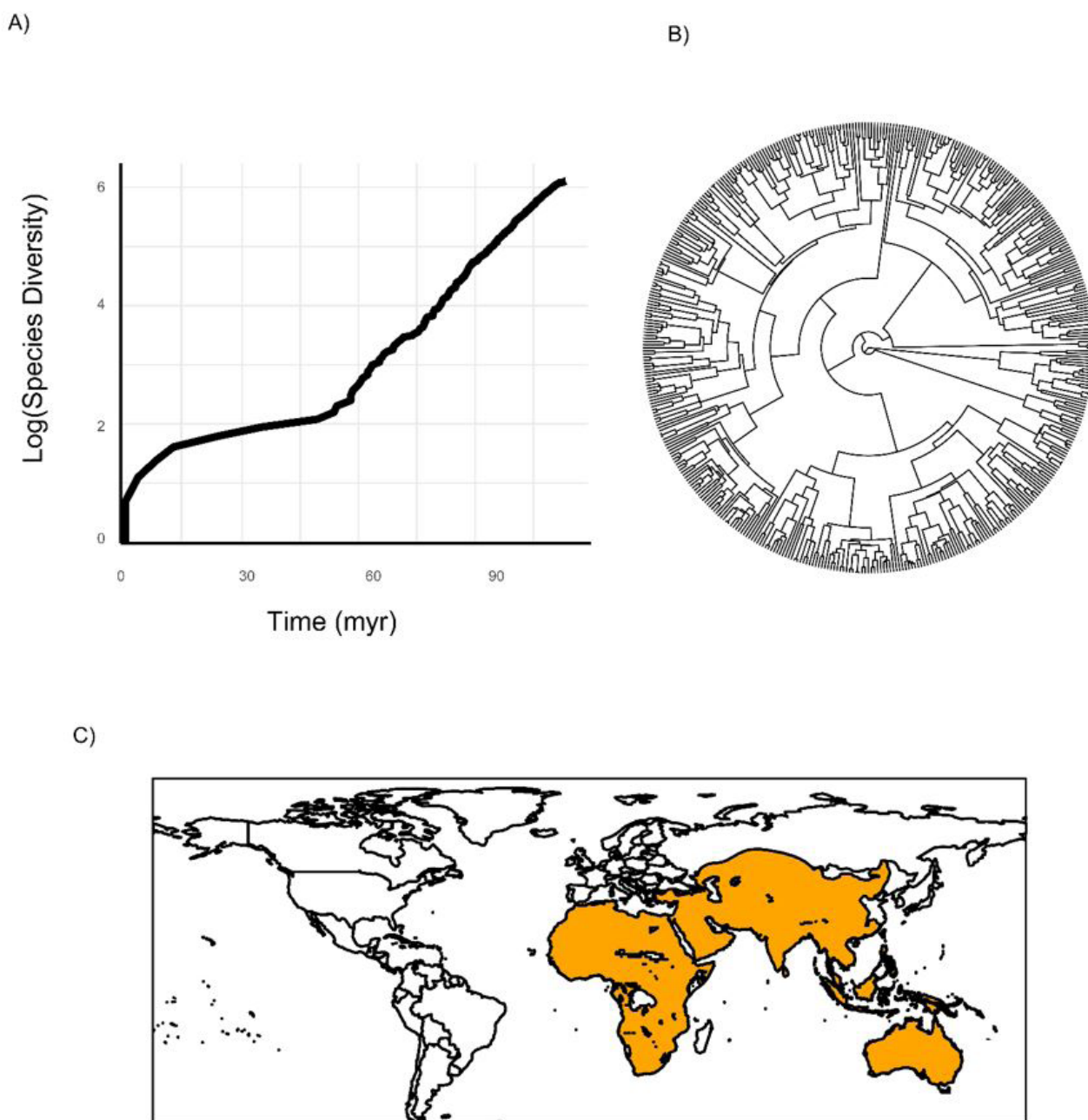


Figure 1. Lineages-through-time plot (A), time-calibrated phylogeny (Tonini et al. 2016) (B), and distribution (Roll et al. 2017) (C) of agamid dragons (family Agamidae).

ARE methods is prolific in the field of biogeography and many different models, including variants of the dispersal-extinction-cladogenesis model (DEC; Ree and Smith 2008), dispersal-vicariance-analysis (DIVA; Ronquist 1997), and BayArea model (Landis et al. 2013), are commonly implemented in the R package BioGeoBEARS (Matzke 2013b, 2014). However, despite the advancement of these analytical methods for ARE, methods to visualise the temporal dynamics of range evolution and diversification tends to be limited to painting the branches of a phylogeny based on the

inferred ancestral condition according to a model of range evolution (e.g., Figure 2 in Matzke 2014). This limitation generally makes it difficult to extract the temporal information contained in the output of these models. In this paper, I present a method for post-hoc data extraction and visualisation of temporal species diversity information contained in the output of a BioGeoBEARS analysis. These lineage-through-space-and-time (LTST) plots form a bridge between clade-level LTTs and infra-clade regional diversity dynamics.

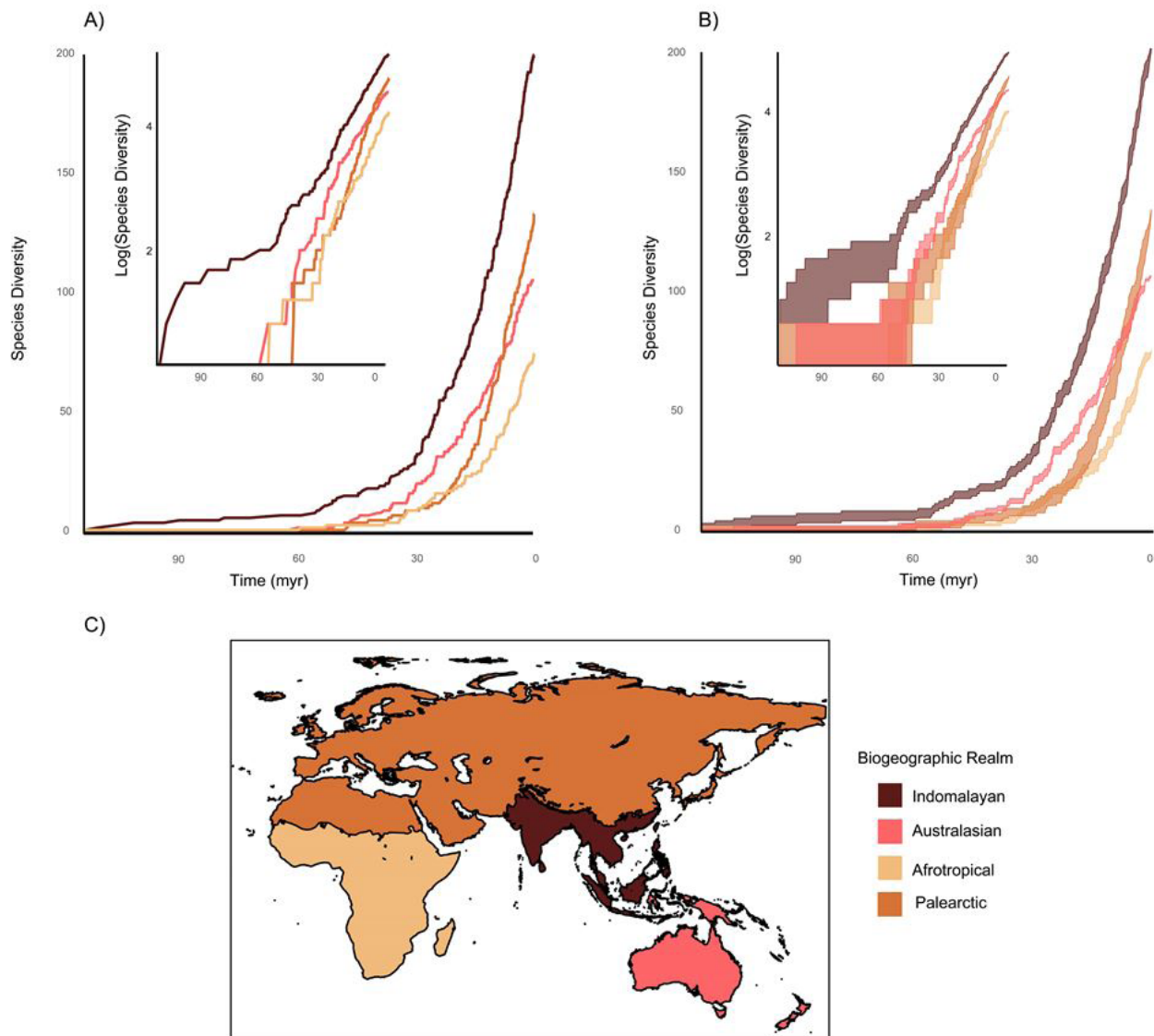


Figure 2. Lineages through space and time plots (LTSTs) for agamid dragons. This figure shows a LTST from a single stochastic map (A), a LTST displaying the distribution of diversity through time from 50 stochastic maps by showing the 95% quantile diversity recorded at each time step (B), and a colour-matched map of the biogeographical realms used in the biogeographical analysis (C).

Methods

Data collection and ancestral range estimation background

The method for generating a LTST is presented in the R package *ltstR*¹ and described below using a global radiation of agamid dragons (family Agamidae) as an example. The agamid dragons are a large clade of over 400 species, which are broadly distributed through much of the Australasian, Indomalayan, Afrotropical, and Palearctic biogeographical realms (Figure 1C). This radiation is over 90 million years old and appears to be showing a slowdown in diversification towards the present (Figure 1A), which is confirmed by estimating

the γ statistic on this group ($\gamma = -2.624$; Pybus and Harvey 2000). The following analysis was performed on a randomly selected phylogeny from the post-burn-in posterior distribution of phylogenies from Tonini et al. (2016; Figure 1B) and spatial information obtained from Roll et al. (2017; Figure 1C).

The first step towards generating a LTST is to perform an ARE analysis using BioGeoBEARS. BioGeoBEARS implements user-specified ARE models from which the probability that each node and internal branch in a phylogeny occupies a given state (geographic range) can be estimated. Biogeographic stochastic mapping (Dupin et al. 2016) uses stochastic simulations based on the phylogeny and the specified ARE model and parameters to assign a state to internal branches

¹ Available at github.com/alexskeels/ltstR

and nodes. Because they are probabilistic, stochastic maps will tend to differ from one another and account for alternative biogeographic histories based on the ARE model and parameters. As such, LTSTs require at least a single stochastic map and, to account for biogeographic uncertainty LTSTs, can be produced using a distribution of stochastic maps. Phylogenetic uncertainty is another source of variation for LTSTs as ARE is typically performed on a representative phylogeny such as a maximum clade credibility phylogeny from a Bayesian posterior. Phylogenetic uncertainty can be accounted for by replicating ARE estimation across, for example, alternative phylogenetic trees from a Bayesian posterior. However, replicating AREs is likely to be computationally prohibitive for many large data sets. Therefore, it is advised that users acknowledge this source of uncertainty when interpreting LTSTs.

For the agamid dragons, I scored the biogeography of species by categorising them as present or absent from the major biogeographical realms of the world. The agamids were present in four of these, the Afrotropical, Indomalayan, Australasian, and Palearctic realms, which we used to fit a DEC model to illustrate the application of LTST. From the maximum likelihood estimates of the DEC model parameters, we obtained 50 biogeographic stochastic maps from which we generated LTST plots (Figure 2).

LTSTs are generated with the R package *ltstR*¹ or with R scripts found in the Supplementary Appendix S1 and requires four objects: 1) a time-calibrated phylogeny, 2) a BioGeoBEARS geography file, and two objects from the output of biogeographic stochastic mapping under an ARE model with BioGeoBEARS; the 3) anagenetic events data table, and 4) cladogenetic events data table. The first two objects are required input for a BioGeoBEARS analysis, and the last two are the standard output from biogeographic stochastic mapping with BioGeoBEARS. Users having performed biogeographic stochastic mapping with BioGeoBEARS will have all the necessary components.

Step-by-step guide to producing LTSTs

The following outlines a step-by-step guide to producing LTST plots. A vignette and associated data to replicate this analysis for the agamid as well as further details on the R code can be found within the Supplementary Appendix S1 or on GitHub¹.

1) The method works by first going through the cladogenetic and anagenetic event tables and extracting all the times in which a state change occurs on the phylogeny for a given stochastic map (*getEventTiming* function). The output from this function is the events timing table, a data frame which gives the time of each event, the node at which the event is located on the phylogeny, whether the event was cladogenetic (occurring at a speciation event, at a node) or anagenetic (occurring in between speciation events, along a branch), the number of species extant at that time, and a running total of the number of species present in each unique state combination. From the events timing table we can infer whether the

diversity of a region increases through immigration or speciation or decreases through emigration or extirpation, to keep a record of diversity changes through time.

- 2) Lineages can occupy multiple states (be present in multiple regions simultaneously), and BioGeoBEARS gives each state combination a unique identifier. For example, in the agamid lizards, a lineage that occupies Australasia is given state 2 while a lineage that occupies Australasia and Indomalaya simultaneously is given state 8. An intermediate step is required which matches the state combination identifier to the actual geographic ranges. The function *getRangeStates* generates a lookup table (Table 1) to perform this matching.
- 3) Next, to obtain the LTST data for each geographic region separately we employ the functions *getLTSTDataTable*, for a single stochastic map, or *getLTSTFromStochasticMaps*, for multiple stochastic maps. These functions take the events timing table from *getEventTiming* and the lookup table from *getRangeStates* to produce the LTST data table, a data frame (or list of data frames for multiple stochastic maps) which contain the number of species present in each geographic range at each event time (Table 2). At this stage the LTST data table for a single stochastic map can be plotted (see vignette in the Supplementary Appendix S1 or on GitHub¹; Figure 2A). However, comparing LTSTs across multiple stochastic maps requires further processing because the timing and number of events will be different between iterations of the stochastic mapping and therefore not align perfectly.
- 4) To account for the discrepancy in event timing between stochastic maps, the function *slidingWindowForStochasticMaps* performs a sliding window analysis to get the number of species

Table 1. Lookup table for geographic ranges from state combination identifiers for agamid dragons. Agamid species are found in ten different combinations of the four geographic regions; W = Australasia, I = Indomalaya, P = Palearctic, and A = Afrotropics.

Geographic Ranges	State Combination Identifier
W	2
I	5
WI	8
PI	11
P	4
AP	9
A	3
AI	10
WPI	14
API	15

Table 2. First four rows of the LTST data table for agamid dragons showing the number of species present in the four geographic regions at each time point. W = Australasia, I = Indomalaya, P = Palearctic, and A = Afrotropics.

Time	W	I	P	A
113.3	0	1	0	1
110.1	0	2	0	1
105.3	0	3	0	1
101.0	0	4	0	1

present in each region within regular time intervals based on the window size. The default is 0.1 time units, but users should select values that make sense for their particular datasets. For example, if the phylogeny is measured in units of million years, then 0.1 represents 100,000 years. If the depth of the phylogeny is, for example, 30 million years, then this may be a fine enough resolution to assess meaningful patterns, but if the depth of the phylogeny is only 1 million years, then it may make sense to narrow the window size (e.g., 0.01). The output is a time-standardised LTST data table.

- 5) To visualise the LTSTs from multiple stochastic maps we then summarise the variability of species diversity values for each region at each time point in the list of time-standardised LTST data tables. The function *timeRangeAcrossStochasticMaps* finds the upper (97.5%) and lower (2.5%) quantiles for each region at each time step and returns this in the LTST quantile data table which can be used to plot the distribution of diversity values for a given LTST (see vignette in the Supplementary Appendix S1 or on GitHub¹; Figure 2B).

Importantly, not only does this method provide a tool for plotting this biogeographic, temporal-diversification information but it also returns it in a format that is user friendly, making quantitative analysis of these data more accessible to many users. For example, logistic growth models can be fit to the temporal diversity data contained within the LTST data tables to estimate the diversity asymptote, which might be interpreted as the regional diversity carrying capacities (Skeels and Cardillo 2019).

Benefits of a biogeographical approach to temporal lineage diversification dynamics

LTSTs and a biogeographical approach to investigating the temporal dynamics of diversification are important to understand the evolution of biodiversity. Diversity is not evenly distributed, with some regions being more diverse than others, for example, the Indomalayan realm contains a greater amount of agamid diversity (Figure 2). This uneven distribution of diversity may be because clades have been, 1) diversifying more rapidly, 2) diversifying for longer, or 3) have a greater carrying capacity in particular regions. LTSTs allow us to more easily visualise these kinds of spatiotemporal dynamics of diversification within clades. In the agamids we can

see that diversification appears to have been occurring much earlier in the Indomalayan region, followed by the Australasian, Afrotropical, and Palearctic regions. We can also see that diversity began to build in the Palearctic, Afrotropics and Australasian regions at a similar time, but different temporal dynamics have led to stark differences in diversity between these regions. Notably, we can see that the accumulation of diversity was more rapid in Australasia during the period of between roughly 60-15myr, before slowing towards the present. Palearctic diversity, on the other hand, has been increasing more steadily in the past 20myr and at some point around 10myr overtook Australasia in species diversity. Comparatively, the Afrotropics have had a lower rate of species accumulation.

Investigating regional patterns in diversification may also help clarify processes that are lost when looking at clade-level temporal diversification dynamics. For example, in the Agamidae as a whole we see a diversification slowdown towards the present (Figure 1A). This pattern appears to be repeated in differing degrees within the different biogeographic realms. In the Palearctic, however, we can see that diversification appears to be still increasing steadily or only showing the earliest beginnings of a slowdown (Figure 2B), suggesting ongoing steady diversification from a relatively late-start in this region. This interesting aspect of agamid diversification is overlooked without an explicitly biogeographical approach.

Conclusion

Lineage-through-space-and-time plots contain information that is useful for interpreting how diversity has arisen in different biogeographic regions. Clade-level temporal diversification dynamics such as understood using the classic lineages through time plot may miss key dynamics that occur within and between different regions. Differences in the tempo of lineage accumulation in different regions is particularly obvious in the case of agamid lizards but is likely to be true in many systems across different spatial and phylogenetic scales. LTSTs offer a visualisation tool for post-hoc exploration of these dynamics, and the method presented here also allows users to extract this temporal diversity data which can be used for novel applications (e.g., Skeels and Cardillo 2019).

Acknowledgements

I would like to thank Damien Esquerré, Ian Brennan, and Marcel Cardillo, the editors, Marcus Cianciaruso and Joaquín Hortal, as well as three anonymous reviewers whose feedback greatly helped to improve the final manuscript.

Supplementary Materials

The following materials are available as part of the online article from <https://escholarship.org/uc/fb>
Appendix S1. R scripts to generate lineage through space and time plots, as well as data and vignette to replicate the analysis on agamid lizards.

References

- Dupin, J., Matzke, N. J., Särkinen, T., et al. (2016) Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography*, 44, 887–899.
- Harvey, P. H., May, M. R. & Nee, S. (1994) Phylogenies without fossils. *Evolution*, 48, 523–529.
- Janzen, T., Höhna, S. & Etienne, R. S. (2015) Approximate Bayesian Computation of diversification rates from molecular phylogenies: Introducing a new efficient summary statistic, the nLTT. *Methods in Ecology and Evolution*, 6, 566–575.
- Klaus, K. & Matzke, N. J. (2019) Statistical Comparison of Trait-dependent Biogeographical Models indicates that Podocarpaceae Dispersal is influenced by both Seed Cone Traits and Geographical Distance. *Systematic Biology*, DOI:10.1093/sysbio/syz034.
- Landis, M. J., Matzke, N. J., Moore, B. R., & Huelsenbeck, J. P. (2013) Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, 62, 789–804.
- Matzke, N. J. (2013a) Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, 5, 242–248.
- Matzke, N. J. (2013b) BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R package, version 0.2.
- Matzke, N. J. (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, 63, 951–970.
- Nee, S., Mooers, A. O. & Harvey, P. H. (1992) Tempo and mode of evolution revealed from molecular phylogenies. *Proceedings of the National Academy of Sciences*, 89, 8322–8326.
- Phillimore, A. B. & Price, T. D. (2008) Density-dependent cladogenesis in birds. *PLoS Biology*, 6, 0483–0489.
- Pybus, O. G. & Harvey, P. H. (2000) Testing macroevolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society B: Biological Sciences*, 267, 2267–2272.
- Ree, R. H. & Smith, S. A. (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57, 4–14.
- Roll, U., Feldman, A., Novosolov, M., et al. (2017) The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology & Evolution*, 1, 1677–1682.
- Ronquist, F. (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, 46, 195–203.
- Skeels, A. & Cardillo, M. (2019) Equilibrium and non-equilibrium phases in the radiation of *Hakea* and the drivers of diversity in Mediterranean-Type Ecosystems. *Evolution*, <https://doi.org/10.1111/evo.13769>.
- Tonini, J. F. R., Beard, K. H., Ferreira, R. B., et al. (2016) Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation*, 204, 23–31.
- Van Dam, M. H. & Matzke, N. J. (2016) Evaluating the influence of connectivity and distance on biogeographical patterns in the southwestern deserts of North America. *Journal of Biogeography*, 43, 1514–1532.

Submitted: 3 March 2019

First decision: 24 May 2019

Accepted: 30 May 2019

Edited by Marcus V. Cianciaruso