



## The missing link in biogeographic reconstruction: Accounting for lineage extinction rewrites history

HerreraAlsina, Leonel; Algar, Adam C.; Lancaster, Lesley T.; Ornelas, Juan Francisco; Bocedi, Greta; Papadopoulos, Alexander S. T.; Gubry-Rangin, Cecile; Osborne, Owen; Mynard, Poppy; Sudiana, I. Made; Lupiyaningdyah, Pungki; Juliandi, Berry; Travis, Justin

### Journal of Biogeography

DOI:

[10.1111/jbi.14489](https://doi.org/10.1111/jbi.14489)

E-pub ahead of print: 19/09/2022

Peer reviewed version

[Cyswllt i'r cyhoeddiad / Link to publication](#)

*Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):*

HerreraAlsina, L., Algar, A. C., Lancaster, L. T., Ornelas, J. F., Bocedi, G., Papadopoulos, A. S. T., Gubry-Rangin, C., Osborne, O., Mynard, P., Sudiana, I. M., Lupiyaningdyah, P., Juliandi, B., & Travis, J. (2022). The missing link in biogeographic reconstruction: Accounting for lineage extinction rewrites history. *Journal of Biogeography*. <https://doi.org/10.1111/jbi.14489>

#### Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 Article type: Research Article

2 The Missing Link in Biogeographic Reconstruction: Accounting for Lineage Extinction Rewrites  
3 History

4 Short title: EXTINCTION IN BIOGEOGRAPHIC RECONSTRUCTIONS

5

6 Leonel Herrera-Alsina<sup>1\*</sup>, Adam C. Algar<sup>2</sup>, Lesley T. Lancaster<sup>1</sup>, Juan Francisco Ornelas<sup>3</sup>, Greta  
7 Bocedi<sup>1</sup>, Alexander S. T. Papadopulos<sup>4</sup>, Cecile Gubry-Rangin<sup>1</sup>, Owen G. Osborne<sup>4</sup>, Poppy Mynard<sup>1</sup>, I  
8 Made Sudiana<sup>5</sup>, Pungki Lupiyaningdyah<sup>6</sup>, Berry Juliandi<sup>7</sup> and Justin M.J. Travis<sup>1</sup>.

9

10 <sup>1</sup>School of Biological Sciences, University of Aberdeen, Aberdeen, UK, AB24 2TZ.

11 <sup>2</sup>Department of Biology, Lakehead University, Thunder Bay, Canada, P7B 5E1.

12 <sup>3</sup>Departamento de Biología Evolutiva, Instituto de Ecología, A.C. (INECOL), Xalapa, Veracruz  
13 91073, Mexico.

14 <sup>4</sup>School of Natural Sciences, Bangor University, Bangor, Gwynedd, UK, LL57 2DG.

15 <sup>5</sup>Research Center for Biology, Indonesian Institute of Sciences, Indonesia.

16 <sup>6</sup>Zoology Division, Museum Zoologicum Bogoriense, Research Center for Biology, Indonesian  
17 Institute of Sciences (LIPI), Cibinong, Indonesia.

18 <sup>7</sup>Department of Biology, Faculty of Mathematics and Natural Sciences, IPB University, Bogor,  
19 Indonesia 16680.q.

20

21 \*Correspondence:

22 *School of Biological Sciences, University of Aberdeen,*

23 *Aberdeen, UK, AB24 2TZ*

24 [leonelhalsina@gmail.com](mailto:leonelhalsina@gmail.com)

25

26 ACKNOWLEDGMENTS

27 This work was supported by a grant from the Ministry of Research and Technology, Indonesia to  
28 Berry Juliandi (No. 2020/IT3.L1/PN/2021) and funding from NERC/NEWTON (NE/S006923/1 and  
29 NE/S006893/1) to Adam C. Algar, Greta Bocedi, Cecile Gubry-Rangin, Lesley Lancaster, Alexander  
30 S. T. Papadopulos and Justin M.J. Travis. The manuscript was improved by the constructive feedback  
31 from two anonymous reviewers. No permits were required to conduct this research.

32

33 ABSTRACT

34 Aim

35 In the most widely used family of methods for ancestral range estimation (ARE), dispersal, speciation,  
36 and extirpation events are estimated from information on extant lineages. However, this approach fails  
37 to consider the geographic distribution of extinct species and their position on the phylogenetic tree,  
38 an omission that could compromise reconstruction. Here we present a method that models the  
39 geographic distribution of extinct species and we quantify the potential inaccuracy in ancestral range  
40 estimation when extinction rates are above zero.

41 Location

42 Global applications, with an example from the Americas

43 Taxon

44 All taxa, with an example from hummingbirds (*Amazilia*)

45 Methods

46 Methods capable of explicitly modelling extinct branches along with their reconstructed geographic  
47 information (GeoSSE) have been overlooked in ARE analysis, perhaps due to the inherent complexity  
48 of implementation. We develop a user-friendly platform, which we term LEMAD (Lineage Extinction  
49 Model of Ancestral Distribution) that generalizes the likelihood described in GeoSSE for any number  
50 of areas and under several sets of geographic assumptions. We compare LEMAD and extinction-free  
51 approaches using extensive simulations under different macroevolutionary scenarios. We apply our  
52 method to revisit the historical biogeography of *Amazilia* hummingbirds.

53 Results

54 We find that accounting for the lineages removed from a tree by extinction improves reconstructions  
55 of ancestral distributions, especially when rates of vicariant speciation are higher than rates of *in situ*  
56 speciation, and when rates of extinction and range evolution are high. Rates of *in situ* and vicariant  
57 speciation are accurately estimated by LEMAD in all scenarios. North America as the most likely  
58 region for the common ancestor of hummingbirds.

59 Main conclusions

60 Methods that neglect lineage extinction are less likely to accurately reconstruct true biogeographic  
61 histories of extant clades. Our findings on an empirical dataset reconcile the Eurasian origin of  
62 *Amazilia* with biogeographic reconstructions when lineage extinction is considered.

63

64

65 Key words: ancestral distribution, BioGeoBEARS, centre of origin, diversification events, extinction  
66 rates, hummingbird evolution, in-situ speciation, vicariance.

67

68

70 Identifying the geographic centre-of-origin for diverse clades has long been of interest in  
71 biogeography. This endeavour is made difficult because the presence or absence of a species at a  
72 given location varies over time and, over longer time scales, species continuously appear and  
73 disappear from the Earth (Barraclough & Vogler, 2000; Jablonski & Sepkoski, 1996; Losos & Glor,  
74 2003). The distribution of clades is the result of shifts in the distribution of constituent species via  
75 range shifts and speciation and extinction, but, in many cases, these processes may leave little fossil or  
76 other tangible evidence of their history, meaning that inferences of centres-of-origin must be inferred  
77 from data on extant species and extant ranges. A foundational field in modern biogeographic research  
78 has been the reconstruction of the geographic distributions of ancestral lineages, in order to relate  
79 biogeographic processes to extrinsic events (e.g., geological shifts, onset of ice ages) while  
80 increasingly taking intrinsic, evolutionary processes into account.

81 For Ancestral Range Estimation (ARE), the two popular methods (DIVA, Dispersal  
82 Vicariance Analysis; Ronquist, 1997 and DEC Dispersal-Extinction-Colonization model; Ree &  
83 Smith, 2008) use the term “extinction” to refer to extirpation (i.e., local extinction), while true lineage  
84 extinction is ignored. These approaches (hereafter Extinction Free approaches; EF) consider the  
85 following events: dispersal, extirpation and speciation, and are therefore appropriate when all lineages  
86 and speciation events are represented in the phylogeny, i.e., no branches are missing due to extinction.  
87 However, the vast majority of available phylogenetic trees are reconstructions where extinction has  
88 removed many branches, such that a pair of extant species that appear as sister species (or clades) in a  
89 reconstructed tree might not be true sisters due to missing nodes. Using the geographic distributions  
90 of the apparent pair of sister clades to infer whether in-situ speciation or vicariance occurred at the  
91 node where they diverged (the putative common ancestor) may be unreliable because any extinct,  
92 intermediary lineages are not only absent from the tree but any information on their geographic  
93 distribution is also missing. Thus, attempting to infer in-situ speciation and vicariance events across a  
94 phylogenetic reconstruction without accounting for extinct lineages could compromise the ancestral  
95 range estimation (Fig. 1). Although the problem of extinct lineages in macroevolution and  
96 biogeography has been pointed out by Sanmartín and Meseguer (2016) and more specifically for  
97 ancestral range estimation by Crisp et al. (2011), the consequences of ignoring extinct lineages for  
98 ARE are still unknown and unquantified.

99 One way forward is to model the potential past existence of lineages at any point of a tree  
100 branch, which extinction subsequently removed, and to account for all the possible geographic  
101 distributions of those extinct lineages. This is achieved in ClaSSE (Cladogenetic State change  
102 Speciation and Extinction) and GeoSSE (Geographic State Speciation and Extinction) models  
103 (Goldberg et al., 2011; Goldberg & Igić, 2012) where speciation, lineage extinction, dispersal and  
104 extirpation events are part of the biogeographic dynamics. These models have been used for  
105 describing how biodiversity accumulates over time in a dynamic context, and in particular, to explore  
106 spatial differences in diversification rates (e.g., Ding et al., 2020; Meseguer et al., 2020). However,  
107 their potential for estimating ancestral distributions has been almost overlooked (but see Lancaster &  
108 Kay, 2013, Caetano et al. 2018). This is surprising, because the spatial distribution of ancestors is  
109 estimated during -SSE likelihood calculation. Matzke (2014) and Ree & Sanmartín (2018)  
110 acknowledged the utility of -SSE models but found that existing implementations were not easy to  
111 use, especially with more than two regions. Here, we use extensive simulations to quantify the impact  
112 of including lineage extinction in ancestral range estimation by comparing the outcome of EF  
113 approaches to that of our new -SSE implementation, which we have also made available as a user-  
114 friendly R package. We simulated biogeographic scenarios that differed in the relative rates of in-situ

115 and vicariant speciation along with different rates of lineage extinction to document variation in  
116 performance of both approaches. Finally, we apply our approach to estimate the biogeographic history  
117 of *Amazilia* hummingbirds. Evolutionary studies using DIVA and DEC have supported South  
118 America as the most likely location of the first speciation event in hummingbirds (McGuire et al.,  
119 2014), however the fossil record points to Eurasia as the source region of the hummingbird lineage  
120 (Louchart et al., 2008; Mayr, 2004). This leaves a time gap of several million years and a geographic  
121 gap of thousands of kilometres. Our approach shows that this spatiotemporal gap is explained when  
122 lineage extinction is no longer neglected.

123

## 124 METHODS

### 125 *Extinction Free models: differences between DIVA and DEC*

126 Extinction Free methods (EF) require that the distribution of a clade be divided into regions  
127 (letters are used for convention) so that the geographic distribution of a given species is coded by its  
128 presence in those regions, and occupancy of multiple regions is indicated by the combination of  
129 corresponding letters. A time-calibrated tree that includes all extant species is also needed (our  
130 approach requires the same data, see below). Matzke (2013) upgraded DIVA (Ronquist, 1997) from  
131 parsimony to likelihood in his BioGeoBEARS R package while maintaining its properties, whereas  
132 DEC is a model originally developed in a likelihood framework (these methods differ in some  
133 assumptions, see below). In such models, speciation is typically modelled as *in-situ* speciation  
134 (occurring within a region, increasing local diversity) or vicariance (geographically mediated  
135 divergence resulting in allopatry, i.e., complementary ranges). Here, we use the notation  $DIVA_{events}$   
136 and  $DEC_{events}$  to refer to the two sets of biogeographic assumptions and leave DIVA and DEC to refer  
137 to the models of ancestral range estimation implemented in BioGeoBEARS.  $DIVA_{events}$  assumes that  
138 widespread species can split their ranges (vicariance) in any combination regardless of the number of  
139 areas where daughter lineages inhabit (e.g., a species presents in region A, B, C and D can split in  
140 AB-CD or A-BCD; widespread vicariance *sensu* Matzke, 2013) while  $DEC_{events}$  assumes that one of  
141 the daughter lineages will be present at a single region (e.g., ABCD species splits in A-BCD or B-  
142 ACD; narrow vicariance). For *in-situ* speciation and in contrast with  $DIVA_{events}$ ,  $DEC_{events}$  allows  
143 widespread lineages to speciation by having one population (i.e., one of the regions where it is  
144 present) diverging from the rest and coexisting with the parental lineage: for instance, ABCD species  
145 produces one daughter lineage which is present at ABCD and the other daughter which is restricted to  
146 region A (*in-situ* subset hereafter; sympatry subset *sensu* Matzke, 2013).

147

### 148 *Lineage Extinction Model of Ancestral Distribution (LEMAD)*

149 We use the area/trait-dependent diversification framework (State-dependent Speciation and  
150 Extinction, -SSE models; Maddison et al., 2007; Goldberg et al., 2011; Herrera-Alsina et al., 2019) to  
151 model past changes in species' geographic distributions. We generalize the computation of the  
152 likelihood described in GeoSSE (Goldberg et al., 2011) for any number of areas and under several  
153 sets of geographic assumptions that facilitate its use in ancestral range estimation (ARE). Notice that  
154 GeoSSE and ClaSSE (Goldberg & Igić, 2012) models have the same system of equations. During the  
155 R package building process, we calculated the likelihood under GeoSSE (from diversitree package)  
156 and LEMAD for a dataset (model parameters, tree and geographic distribution of species in two areas)  
157 to confirm that the likelihoods are identical (Fitzjohn, 2012). Unlike EF methods, the -SSE framework  
158 considers that, at any point along a tree branch, a lineage could have been present but went extinct,

159 with or without first producing (also extinct) descendants. To this end, the algorithm uses two coupled  
160 differential equations (Appendix S1), where one accounts for the probability of a lineage being at a  
161 given region (or set of regions), and the other reflects the probability of a lineage going extinct for the  
162 same region (or set of regions). These equations are numerically integrated to obtain a likelihood  
163 value for the data given the model with its parameters (dispersal/contraction, *in-situ* and vicariant  
164 speciation). Different parameter combinations are tested to find the best combination (likelihood  
165 optimization). With the parameters that maximize the likelihood, we compute the change in  
166 probability for a lineage to be at each distribution from the present (tree tips) to the past (root) and  
167 extract those probabilities at the nodes. Ancestral range probabilities were estimated by taking the  
168 partial likelihoods from the downpass and rescaling them so that they summed to 1 at each node  
169 (Nguyen, 2011). In summary, the model simultaneously considers the probabilities of dispersal,  
170 extirpation and speciation (via *in-situ* or vicariance) for extant and extinct lineages. The likelihood of  
171 the model is optimized, and the rates of geographic change, *in-situ* speciation and vicariance are  
172 estimated. Lineage extinction can be estimated or fixed to a specific rate by the user. In short, Lineage  
173 Extinction Model of Ancestral Distribution (LEMAD) computes the likelihood of the current  
174 distribution of species (given the parameters of the model) where lineage extinction is a fundamental  
175 part of the calculation. The R package lemad is available at <https://github.com/leonelhalsina/lemad>.

#### 176 *General Assumptions in LEMAD*

177 Although the LEMAD model can account for differences in diversification rates across  
178 regions (like in GeoSSE/ClaSSE original application), in LEMAD the rates of speciation and  
179 extinction are constant across regions. This is achieved by assigning the same rate of speciation and  
180 extinction to each area or combination of areas during the parameter setup. This simplification is  
181 necessary to reduce the otherwise immense complexity of parameter space when the analysis is  
182 performed for many regions; note that this assumption is the same in DIVA and DEC models. In  
183 LEMAD we assume that shifts in the geographic distribution of species are the product of expansion  
184 and contraction. For example, a species present in region A cannot instantaneously change to region  
185 B. It has first to expand to region B (to be present in AB) followed by an extirpation event in A. These  
186 assumptions are the same as in EF methods. Lineage extinction can be modelled in two ways:  
187 extinction by extirpation and instantaneous extinction. In the former case, a lineage can undergo  
188 extirpation events in different regions of its distribution (range contraction) and eventually go extinct  
189 when it is extirpated from its last remaining region. This is similar to the idea of the empty range ( $\emptyset$ )  
190 in Ree and Smith (2008). In the case of instantaneous extinction, a species can go extinct regardless of  
191 the number of regions where it is present. Although extinction by extirpation is appropriate when  
192 regions are small and each of them represents a single population (the extinction of a species takes  
193 place once the last population disappears), the scale at which ARE is normally conducted renders this  
194 type of extinction inappropriate (Polly & Sarwar, 2014). Furthermore, by using instantaneous  
195 extinction we account for those events that involve a sudden decline in total population size that are  
196 not related to standard dynamics of region colonization/extirpation, so we can measure the  
197 contribution of each process independently. We therefore used instantaneous extinction in LEMAD,  
198 but extinction by extirpation could also be enabled. Our model assumes that lineages, including  
199 extinct lineages and ancestors, can be present in multiple regions, even if extant species are not. For  
200 instance, with three regions (A, B, and C), LEMAD calculates the probability of the ancestors being  
201 present in A, B, C, AB, AC, BC, or ABC (all possible combinations). By allowing this, we do not  
202 constrain the model to only consider region-endemic lineages, which could lead to underestimation of  
203 the importance of widespread historical lineages in shaping more narrow modern distributions.  
204 However, the model is flexible enough to set any restriction in the number of permitted regions per

205 ancestral species. Note that LEMAD can handle any number of regions, however computation time  
206 will exponentially increase with the number of regions. For instance, a phylogenetic tree with 66  
207 species and 3 areas (yielding 7 possible ancestral areas) can take around 10 minutes of computing  
208 time. With 4 areas (and 15 possible ancestral distributions) the calculation can take around 75  
209 minutes. With 6 areas (and 63 possible states) the computing time can be as long as 100 hours.

210 LEMAD enables two different sets of biogeographic assumptions (i.e., LEMAD<sub>diva\_events</sub> and  
211 LEMAD<sub>dec\_events</sub>; we refer to both models under the term LEMAD) that match DIVA<sub>events</sub> and  
212 DEC<sub>events</sub>. As they are different parameterizations of the same model, the comparison of their  
213 likelihoods is valid and straightforward.

#### 214 *Accuracy Assessment*

215 In order to compare the accuracy of LEMAD and EF approaches under different extents of  
216 extinction, we modelled a number of scenarios in which we: (i) simulated the evolutionary history of a  
217 clade along with the geographic evolution of its species, (ii) fit both models and (iii) compared their  
218 ancestral range estimations.

#### 219 *Simulation procedure*

220 The simulation started with one lineage in a random region (A, B and C), or combination of  
221 regions (AB, AC, BC, or ABC); lineages undergo the following events: dispersal, extirpation,  
222 speciation, and extinction. The simulation runs in continuous time where the waiting time between  
223 events is drawn from an exponential distribution (Gillespie algorithm; Doob, 1945; Gillespie, 1977).  
224 The duration of the simulation is chosen to ensure a final clade size of 150 species given the  
225 speciation rates (scenarios with high extinction were allowed to run longer, see below).

226 We kept track of the geographic distribution of lineages over time and of ancestor-descendant  
227 relationships and used this as a record to build a phylogenetic tree of the clade. As a result, the  
228 simulation produces a phylogenetic tree (without extinct branches, similar to standard reconstructed  
229 trees) and the geographic distribution of extant species. Notice that species (ancestors and extant  
230 lineages) could be in any of the seven states of the system (A, B, C, AB, AC, BC, or ABC).

#### 231 *Model fitting*

232 We simulated two datasets that differed in modes of vicariance and in-situ speciation, following the  
233 assumptions in DIVA<sub>events</sub> and DEC<sub>events</sub>. For the simulations under DIVA<sub>events</sub>, we fitted DIVA (from  
234 BioGeoBEARS) and LEMAD<sub>diva\_events</sub>. Similarly, the simulations under DEC<sub>events</sub> were fit with DEC  
235 (from BioGeoBEARS) and LEMAD<sub>dec\_events</sub>. Next, we extracted the most likely ancestral distribution  
236 estimated by LEMAD and EF for every node in the phylogenetic reconstruction and compared to the  
237 record of ancestors directly from simulated datasets. This is, for a given ancestor/node, we took the  
238 distribution with the highest probability and compared to the distribution that was logged during the  
239 simulation. We defined a node successfully inferred when both distributions matched completely (if A  
240 is the simulated truth, only A would be a successful reconstruction. Neither AB nor ABC would be  
241 correct). We counted the number of nodes that were successfully recovered by both models in two  
242 sections of time during the history of the clade: recent and ancient time windows. We repeated the  
243 simulation-inference procedure under 18 different parameter combinations: rates of in-situ speciation  
244 = 0.02, 0.03, 0.04 and vicariance = 0.02, 0.03, 0.04 to combine into 3 scenarios with overall  
245 speciation of 0.06; extinction = 0, 0.003, 0.03; dispersal/extirpation = 0.06, 0.6 (30 runs for each  
246 combination). To measure the accuracy in parameter estimation, we used the rates (geographic  
247 change, in-situ speciation and vicariance) that are estimated during the analysis and compared them to

248 the simulation generating rates. Lineage extinction was not estimated but was fixed to the generating  
249 rate as we were interested in the performance of the other (more informative) parameters.

250 We were also interested in measuring whether phylogenetic reconstructions and geographic  
251 data are informative about the modes of in-situ and vicariant speciation, which constitute the main  
252 difference between  $DIVA_{events}$  and  $DEC_{events}$ . Specifically, we measured the power of LEMAD to  
253 detect different sets of biogeographic assumptions. To this end, we simulated datasets under  
254  $DIVA_{events}$  and fitted  $LEMAD_{diva_{events}}$  and  $LEMAD_{dec_{events}}$  models and compared their likelihoods. It  
255 is expected that  $LEMAD_{diva_{events}}$  model should have higher likelihood than  $LEMAD_{dec_{events}}$  because  
256 the generating model was indeed, a  $DIVA_{events}$  process. We counted the number of simulated datasets  
257 where this was the case. We also conducted the complementary analysis: we simulated datasets under  
258  $DEC_{events}$  to fit and compare  $LEMAD_{diva_{events}}$  and  $LEMAD_{dec_{events}}$  models (30 runs for each case).

### 259 *An Empirical Example*

260 The geographic origin of the American avian family Trochilidae (Hummingbirds) is still  
261 debated (McGuire et al., 2014). Previous ARE analyses have supported South America as the most  
262 likely area where the common ancestor of hummingbirds lived (22 million years ago; McGuire et al.,  
263 2007, 2014). Interestingly, the fossil record points to Eurasia as the source (Louchart et al., 2008;  
264 Mayr, 2004) from which the first hummingbird lineage spread via the Bering Strait 34–28; therefore,  
265 early diverging hummingbird lineages are expected to be found in North America. However, this is  
266 not the case, which leaves a time gap of several million years. To determine whether LEMAD could  
267 provide insights on this, we reconstructed the geographic distribution of a widespread and  
268 representative hummingbird clade (*Amazilia sensu lato* and closely related species) using both  
269 LEMAD and EF models. The phylogenetic tree was taken from McGuire et al. (2014) in combination  
270 with geographic information from Ornelas et al. (2014). Extant species and extinct lineages could be  
271 present in three regions: A) South America, B) Mesoamerica and C) North America (West from the  
272 Isthmus of Tehuantepec) or a combination of them. We did not include Eurasia as a possible region as  
273 1) no living species are present, and 2) the artificial inclusion of a Eurasian branch into the  
274 phylogenetic reconstruction would bias the analysis and model the distribution of recent ancestors in  
275 Eurasia which disagrees with the fossil record. As no information exists on how *in-situ* and vicariant  
276 speciation occur in *Amazilia* (see first paragraph of Methods), we could not assume either  $DIVA_{events}$   
277 or  $DEC_{events}$  so we ran  $LEMAD_{dec_{events}}$  and  $LEMAD_{diva_{events}}$  and compared the fit using AIC weights.  
278 Additionally, the models were combined with three different assumptions for rates of lineage  
279 extinction: one in which extinction is the same as the estimate for speciation rate (using a standard  
280 birth-death model: 0.15), one in which extinction is 10 times less frequent than speciation (0.015) and  
281 one in which extinction is 10 times more frequent (1.5). Notice that by fixing extinction to a certain  
282 rate, the rates of *in-situ* and vicariant speciation will adjust accordingly during the likelihood  
283 optimization. Phylogenetic reconstructions often do not include all species in a group (due to a lack of  
284 DNA samples for instance); LEMAD features functionality where the number of missing extant  
285 species is taken into account during the calculation (the so-called sampling fraction in diversification  
286 models; (Fitzjohn et al., 2009). We included this completeness information for the *Amazilia* dataset.

287

## 288 RESULTS

### 289 *Increase in accuracy by modelling extinct branches*

290 Our simulations indicate that the reconstruction of the biogeographic history of a clade is  
291 notably improved when the set of branches that potentially existed and went extinct is incorporated



292 into the analysis. The extent of the improvement depends on the relative rates of *in-situ* and vicariance  
293 speciation, extinction and dispersal/extirpation (range evolution). For instance, LEMAD is more  
294 accurate than extinction free approaches (EF) when vicariance is higher than *in-situ* speciation and  
295 there are high rates of range evolution. We find no parameter combination where EF outperforms  
296 LEMAD.

297 Although we find that low rates of range evolution led to few differences between EF and  
298 LEMAD, data simulated under the biogeographic assumptions of DEC<sub>events</sub> shows that ancient nodes  
299 are better estimated by LEMAD when lineage extinction is higher than zero. Under DIVA<sub>events</sub> and  
300 low rates of range evolution, neither ancient nor recent nodes are better estimated with LEMAD (Figs.  
301 2 and 3).

302 The scenarios with high rates of range evolution show increased accuracy in ancestral range  
303 estimation (ARE) when using LEMAD than when using EF. Under DIVA<sub>events</sub>, the improvement is  
304 limited to recent nodes but also ancient ones when rates of vicariance are higher than in-situ  
305 speciation. Datasets with DEC<sub>events</sub> show that LEMAD outperforms EF in recent nodes in all  
306 scenarios; ancient nodes are also better estimated except when *in-situ* speciation is dominant.

307 Even though the differences between LEMAD and EF are more important as extinction rate  
308 increases, simulations with zero extinction also suggest a better performance of LEMAD over EF  
309 approaches in most cases. However, recent ancestors are correctly recovered by both approaches at  
310 similar numbers when simulations featured low rates of range evolution. Finally, we find that the  
311 LEMAD estimates for dispersal/extirpation, *in-situ* and vicariant speciation are accurate across all  
312 parameter combinations and, importantly, the model can correctly detect statistical differences in their  
313 relative contributions (Figs. S1 – S3). In summary we recommend using LEMAD when rates of  
314 vicariant speciation are equal or higher than rates of in-situ speciation, and when range expansion and  
315 contraction are highly dynamic (Table 1).

316

### 317 *Ability to distinguish the signal of DIVA<sub>events</sub> and DEC<sub>events</sub>*

318 For the simulations where *in-situ* subset was not assumed (DIVA<sub>events</sub>; see methods), we fitted  
319 LEMAD model in two versions: LEMAD<sub>diva\_events</sub> and LEMAD<sub>dec\_events</sub>. We find that in 86% of the  
320 simulated datasets, LEMAD<sub>diva\_events</sub> has the highest statistical support and for the remaining 14% of  
321 the simulations, LEMAD<sub>dec\_events</sub> was wrongly selected as the best model. It is important to note that in  
322 the datasets where LEMAD chose the right [generating] model (i.e., DIVA<sub>events</sub>) the average  
323 difference in AIC weights is 0.43. In contrast, in simulations where the wrong model was selected, the  
324 difference was minimal (mean of AIC weights = 0.02). When DEC<sub>events</sub> was the generating model  
325 (i.e., *in-situ* subset, see methods), LEMAD<sub>dec\_events</sub> is correctly selected 76% of the times over  
326 LEMAD<sub>diva\_events</sub>. A potential explanation on why LEMAD failed to select the correct model in some  
327 datasets is that in a three-area system like the one we are simulating, events of widespread vicariance  
328 are not possible which makes DIVA<sub>events</sub> and DEC<sub>events</sub> less different from one another. Therefore, this  
329 analysis mainly explored the traces of the *in-situ* subset assumption left in phylogenetic trees.

### 330 *Reconstruction of *Amazilia* biogeography*

331 We find higher likelihood for LEMAD models with DIVA<sub>events</sub> than for LEMAD with  
332 DEC<sub>events</sub> (difference in AIC weights = 0.95) which suggests that widespread species speciate by  
333 vicariance and not by in-situ speciation. Within LEMAD<sub>diva\_events</sub> we find models with smaller  
334 extinction rate more likely, however, this result is not surprising as the estimated rate of extinction

335 from a birth-death model was close to zero (Table 2). Instead of comparing across extinction rates and  
336 choosing the DIVA<sub>events</sub> model with the best AIC, we explore the parameter estimates and the  
337 reconstructed ancestral distributions for each model. Regardless of the assumed extinction rate, all  
338 reconstructions point to North America as the most likely region for the common ancestor of  
339 hummingbirds (Fig. 4). In such a scenario, our simulation analysis finds that LEMAD is 50-100%  
340 more effective than EF approaches in inferring the clade's common ancestor.

341

## 342 DISCUSSION

343 We showed that ancestral range estimation can benefit from the -SSE framework by  
344 modelling lineage extinction, and that methods that neglect lineage extinction are less likely to  
345 accurately reconstruct true biogeographic histories of extant clades in a wide variety of scenarios. The  
346 parametrization of the model allows competing hypotheses for centres-of-origin and *in-situ* versus  
347 vicariant speciation to be distinguished. With it, we found that North America is the most likely place  
348 of origin of *Amazilia* hummingbirds, which resolves a previous spatiotemporal disconnect between  
349 the hypothesized source region and the origin of first species divergence.

350 Empirical studies in island biogeography provide insights on how vicariance/*in-situ* rates  
351 contribute to biodiversity patterns. Speciation after dispersal largely contributes to building species  
352 richness in small-sized islands and is responsible for the uniqueness of their assemblages (Losos &  
353 Schluter, 2000; Stuart et al., 2012). Archipelagos with small islands are expected to have high rates of  
354 vicariance and therefore LEMAD might be more appropriate for ancestral range estimation (ARE)  
355 than EF approaches. Nonetheless *in-situ* speciation becomes more frequent than vicariance as the size  
356 of the island increases which amounts to higher chances of geographic isolation and diversity of  
357 habitats (Kisel & Timothy, 2010); in fact, islands over a threshold size show evidence of rapid  
358 diversification (Algar & Losos, 2011; Losos & Schluter, 2000). Because the large geographic scale at  
359 which ARE is normally conducted (continents or large-sized islands), *in-situ* speciation can be as  
360 frequent as vicariance. In this scenario and when DEC<sub>events</sub> are assumed, the improvement provided by  
361 LEMAD is expected in recent and basal nodes. With DIVA<sub>events</sub>, recent nodes are better estimated  
362 than EF methods whereas basal nodes show a non-significant improvement.

363 Similar to EF models, LEMAD assumes constant rates (extinction, vicariance and *in-situ*  
364 speciation), which might not match empirical datasets in some cases. For instance, McGuire et al.  
365 (2014) reports an important variation in richness across hummingbird subclades. This can be due to  
366 differential speciation (or extinction) rates among lineages (e.g., via diversity-dependent  
367 diversification; Etienne and Haegeman, 2012). McGuire et al. (2014) found that the difference in  
368 speciation rate between two subclades can be as large as 15-fold according to BAMM analysis.  
369 Heterogeneity in diversification rates which is independent from trait states or geographic  
370 distributions, is likely to be ubiquitous across taxonomic groups besides hummingbirds and it is  
371 necessary to develop an ARE method that can handle this complexity. We argue that this should be  
372 the next methodological step forward. If the variation in speciation rates across lineages is  
373 consistently a result of regional differences (i.e., lineages inhabiting a given area experience higher  
374 rates than in other regions), GeoSSE and GeoHiSSE (Caetano et al., 2018) are the proper tools to use.  
375 However, if more than three regions (or trait states) are to be analysed, SecSSE (Herrera-Alsina et al.,  
376 2019) can be used and with the right setup, it allows for character changes at cladogenetic events and  
377 not just along the branches extends (like in ClaSSE model; for an example with habitat preference see  
378 Aduse-Poku et al., 2022). Any tool that would consider variable speciation rates across lineages  
379 should also incorporate variation in rates of expansion as the chances of vicariant events intrinsically

380 depend on lineage dispersal (i.e., only multi-region lineages can undergo vicariance). Furthermore,  
381 opportunities for dispersal across regions can vary over time and assuming a single rate of range  
382 expansion/contraction might not be realistic in some cases (e.g., Buerki et al., 2011). Unlike DEC, the  
383 current implementation of LEMAD cannot handle this variation but the framework could be adapted  
384 include it.

385 In previous studies, selecting DEC over DIVA was mostly based on the superior statistical  
386 properties (likelihood based) of DEC when compared to the parsimony method used in DIVA. After  
387 BioGeoBEARS was made available, researchers could confidently fit both methods to datasets and  
388 compare likelihoods, but surprisingly analysis is generally conducted with DEC, rather than DIVA,  
389 without justification. We recommend fitting both  $LEMAD_{dec\_events}$  and  $LEMAD_{diva\_events}$  to data and  
390 comparing likelihoods, instead of discarding either biogeographic model a priori. Our simulations  
391 show that LEMAD is capable of telling the two models apart, even though  $DEC_{events}$  are slightly less  
392 likely to be correctly detected than  $DIVA_{events}$ . We find that not only the relative contributions of *in-*  
393 *situ* and vicariant speciation, lineage extinction and range evolution directly influence the precision of  
394 the ancestral reconstruction, but the set of biogeographic assumptions is also of paramount  
395 importance. For instance, unlike  $DIVA_{events}$ ,  $DEC_{events}$  attribute some speciation events as *in-situ*  
396 subset instead of vicariance followed by dispersal (Ree et al., 2005). This is reflected in the estimates  
397 for both processes in our analysis: even if the contributions of *in-situ* and vicariant speciation are the  
398 same, we found high variability in vicariance estimates (higher than *in-situ* events) when  $DEC_{events}$   
399 underlies simulations. Similarly, when using  $DIVA_{events}$ , the estimates for *in-situ* speciation are likely  
400 to be more variable than those for vicariance. In both cases, high rates of lineage extinction increase  
401 the variability of rate estimates.

402 High rates of dispersal/extirpation have two main consequences on these biogeographic  
403 analyses. First, the impact of ignoring extinct branches in accurate ARE is higher than in the presence  
404 of low rates of range evolution. LEMAD is more likely to correctly track the change in geographic  
405 distribution of ancestors along the branches of a phylogenetic tree than EF methods, even with zero  
406 extinction. This could be due to how the likelihood at the root is handled by both approaches. In  
407 LEMAD, the probabilities of all the areas are multiplied by speciation rates whereas EF approaches  
408 do not consider speciation (Ree & Smith, 2008). This multiplication at the root (also called  
409 “conditioning on extinction” because we are looking at a tree Nee et al., 1994) is used in all SSE  
410 diversification models. This may be responsible for its overall higher precision, which is magnified in  
411 systems with many range shifts. Second, with elevated rates of dispersal/extinction, the uncertainty  
412 around speciation estimates is high. This is likely to occur because dispersal taking place right after  
413 *in-situ* speciation (something expected with high rate of dispersal) looks similar to a vicariance event.  
414 In a similar way, an extirpation event following vicariance could be mistaken for *in-situ* speciation.  
415 Importantly, although the estimates show important variation, the model can correctly detect  
416 statistical differences between rates of *in-situ* and vicariant speciation.

417 LEMAD allows for the evaluation of contrasting models that make explicit assumptions  
418 regarding the rates of evolutionary events. Nonetheless, more meaningful hypotheses can be  
419 contrasted with fossils or other extinction estimates, which in turn would render a more accurate  
420 reconstruction of ancestral distributions (Mao et al., 2012). Alternatively, LEMAD can be modified to  
421 include extinct tree branches along with their last known distribution (Zhang et al., 2022; for an  
422 example of body size and extinct branches in a SSE implementation see Porto, 2022). The  
423 incorporation of known distributions of ancestors (i.e., constraining an internal node to have a certain  
424 distribution; see Meseguer et al., 2015) in LEMAD would be treated in a similar manner as the total  
425 likelihood is computed at the tree root, when giving different weights to the various regions. This

426 feature, however, is not implemented yet. Dispersal could also be fixed to a specific rate, however,  
427 empirical evidence for rates of dispersal is challenging to collect. Unsurprisingly, the large  
428 geographical scale in ARE implies that regions are likely to be different from one another in both  
429 biotic and abiotic factors. Lineage dispersal in this context does not only entail the mobility to new  
430 localities but the successful arrival and further adaptation to potentially new conditions. It is likely  
431 that dispersal estimates from mark-release-recapture techniques (e.g., Hill et al., 1996), or other field-  
432 based measures would not be appropriate for ARE. One promising concept for testing with LEMAD  
433 is the taxon cycle, which posits that phases of range expansion and contraction occur along with  
434 habitat shifts (Ricklefs & Bermingham, 2002). The duration of these phases might offer a sensible  
435 starting point for developing hypotheses on rates of dispersal/extirpation. Finally, LEMAD enables  
436 the comparison of different assumptions on the distribution of the very first common ancestor to the  
437 entire clade, i.e., the location of the centre-of-origin.

438         The biogeographic history for *Amazilia* hummingbirds reconstructed by LEMAD model  
439 showed clear differences with its EF counterpart. LEMAD found North America as the most likely  
440 region for the common ancestor of hummingbirds (Fig. 4). This finding contrasts with previous  
441 studies where South America was found as the ancestral distribution. McGuire et al. (2014) proposed  
442 a northern arrival of hummingbirds to America with further expansion into South America  
443 immediately followed by extinction events that wiped out all hummingbird species from North  
444 America (recolonization of North America came at a later stage). However, their EF analysis could  
445 not prove this hypothesis. By considering extinction explicitly, our LEMAD analysis provides the  
446 missing piece of this puzzle, reconciling the South American distribution of the common ancestor of  
447 extant hummingbird species when ignoring extinction with North American distribution of the  
448 ancestor when extinction is considered.

449

## 450 CONCLUSION

451         Lineage extinction seems less tangible than lineage formation; yet, we have shown that  
452 incorporating it into biogeographic models is crucial for a better reconstruction of ancestral areas.  
453 When using extinction-free methods, taxonomic groups can be inferred to have different centres of  
454 origin, however, this could be the result of dissimilar extinction rates across clades rather than actual  
455 differences in biogeographic histories. As a corollary, many taxa might have originated at the same  
456 place, we think that there are broad patterns which are yet to be discovered.

## 457 DATA AVAILABILITY STATEMENT

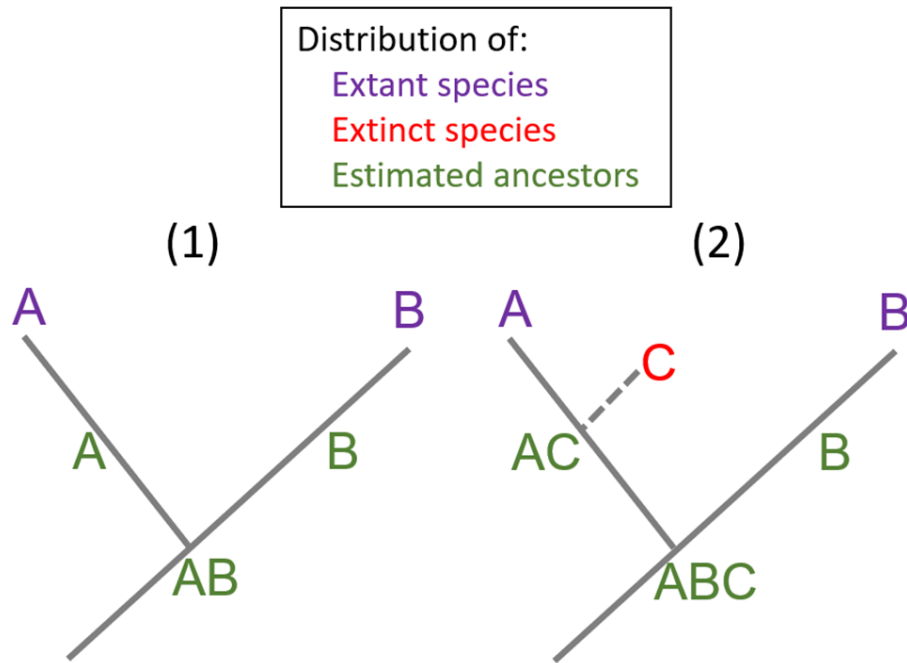
458         The R package `lemad` is available at <https://github.com/leonelhalsina/lemad>. And our code to  
459 reproduce simulations will be available in a public repository.

460

461 FIGURE

462

463 Figure 1. The reconstruction of the ancestral distribution for a two-species clade inhabiting a region  
464 with subregions A, B and C. We show the main difference between 1) Extinction Free (EF) method  
465 and 2) Lineage Extinction Model of Ancestral Distribution (LEMAD). In contrast to EF, LEMAD  
466 considers the missing lineages due to extinction and their geographic distribution in the analysis.

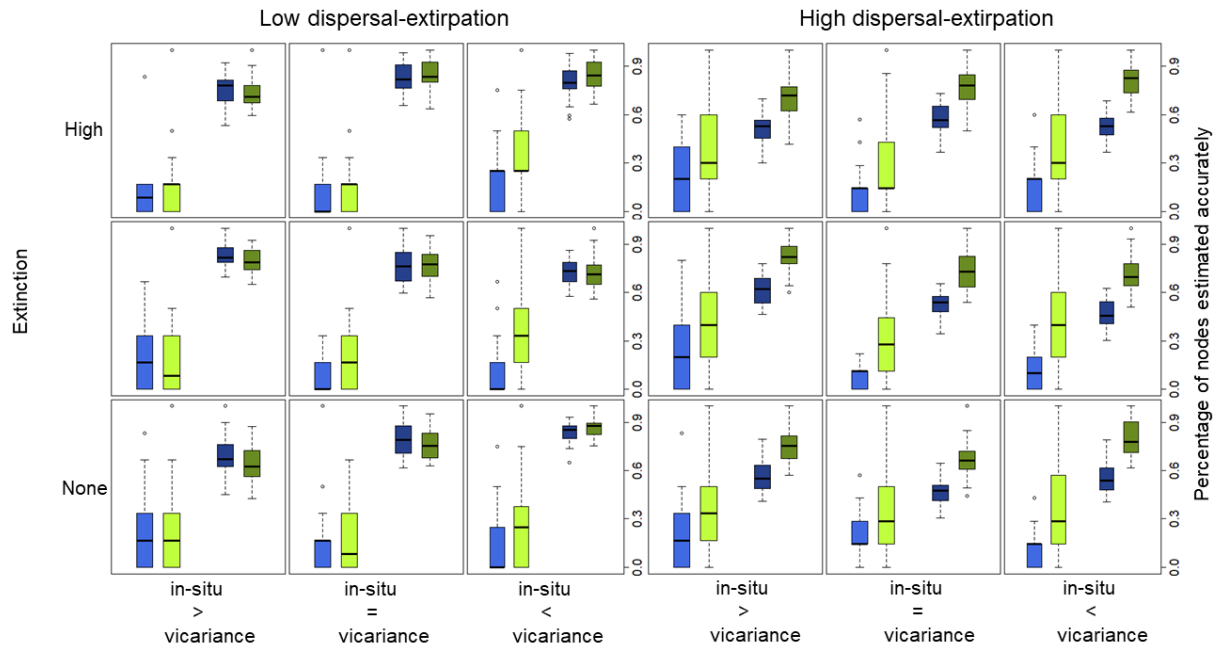


467

468

469

470 Figure 2. Accuracy in Ancestral Range Estimation under DEC (in blue) and LEMAD (in green)  
 471 models at recent (from half simulated time to present; dark shades) and ancient nodes (light shades).  
 472 Eighteen scenarios were simulated with different rates of lineage extinction, range evolution  
 473 (dispersal/extirpation) and relative contributions of *in-situ* speciation and vicariance. For each panel,  
 474 the y-axis shows the standardized number of ancestors whose distribution was correctly estimated by  
 475 the models.

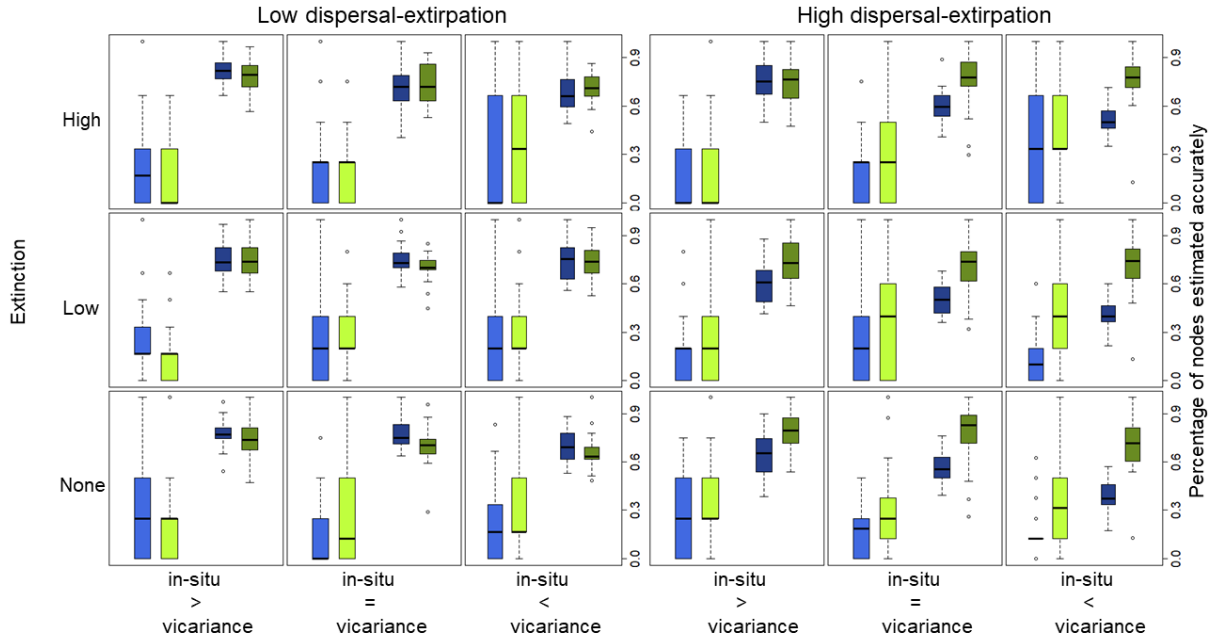


476

477

478

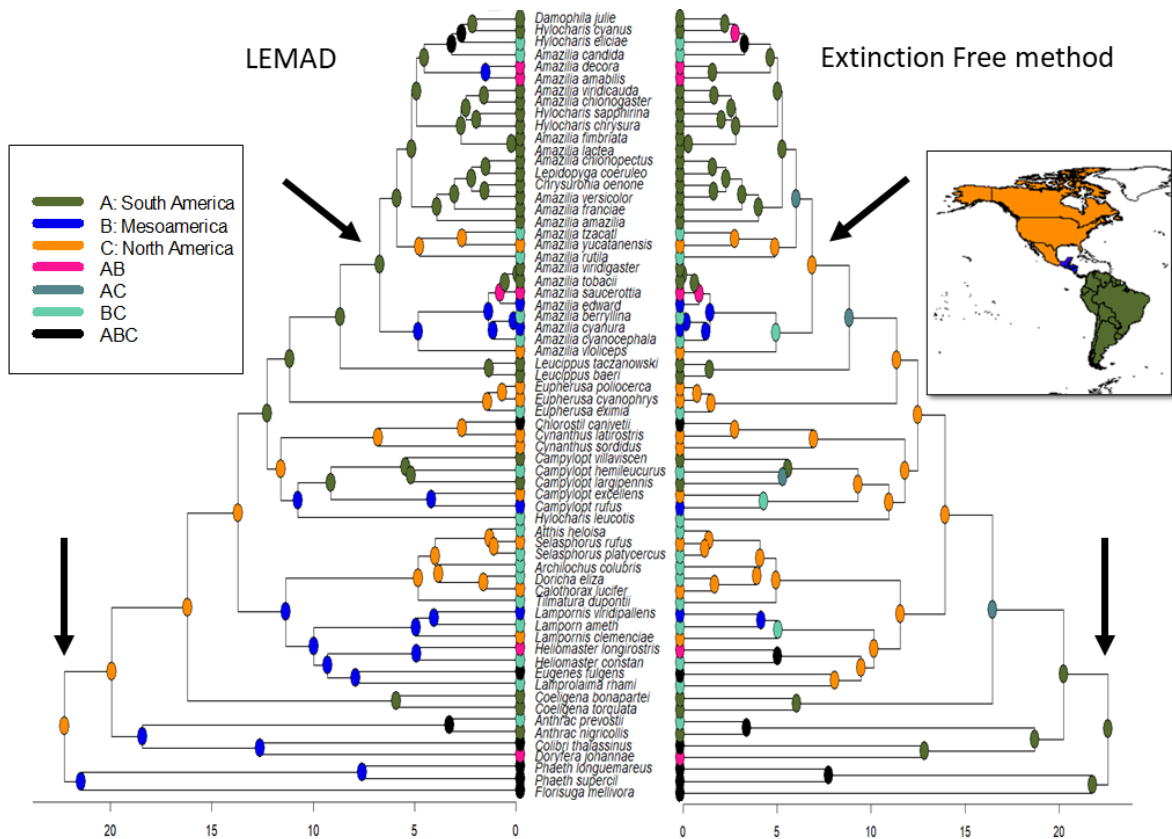
479 Figure 3. Accuracy in Ancestral Range Estimation under DIVA (in blue) and LEMAD (in green)  
 480 models at recent (from half simulated time to present; dark shades) and ancient nodes (light shades).  
 481 Eighteen scenarios were simulated with different rates of lineage extinction, range evolution and  
 482 relative contributions of *in-situ* speciation and vicariance. For each panel, the y-axis shows the  
 483 standardized number of ancestors whose distribution was correctly estimated by the models.



484

485

486 Figure 4. Estimated geographic distribution of “Amazilia” hummingbirds’ ancestors under two  
 487 different approaches of state reconstruction. Extant and ancestral species (and extinct species in the  
 488 case of LEMAD) could be present in A) South America, B) Mesoamerica and C) North America  
 489 (West from Tehuantepec) or a combination of them. Coloured circles show the most-likely  
 490 distribution. Arrows show some discrepancies between LEMAD and its Extinction Free method  
 491 counterpart on the ancestral range estimation of the entire hummingbird clade and “Amazilia” group.



492



493 Table 1. Recommended scenarios to use LEMAD. Necessary condition is indicated with \*

	Range evolution rate	Main speciation mode	Lineage extinction
When reconstructing nodes:			
Ancient	Any	Vicariance; Equal contribution of both modes	Intermediate, High
Recent	High*	Any	Any

494

495 Table 2. Summary of LEMAD models fitted to “Amazilia” hummingbird dataset under different  
496 assumptions on rates of extinction and modes of in-situ and vicariant speciation.

Biogeographic model	Assumed extinction	Log likelihood	Free parameters	AIC weights
DIVA <sub>events</sub>	0.015	-289.49	3	0.81
DIVA <sub>events</sub>	0.15	-291.10	3	0.16
DEC <sub>events</sub>	0.015	-293.08	3	0.02
DEC <sub>events</sub>	0.15	-295.54	3	<0.01
DIVA <sub>events</sub>	1.5	-333.82	3	<0.01
DEC <sub>events</sub>	1.5	-347.36	3	<0.01

497

498

499

500

501

## 502 REFERENCES

- 503 Aduse-Poku, K., van Bergen, E., Sáfián, S., Collins, S. C., Etienne, R. S., Herrera-Alsina, L.,  
504 Brakefield, P. M., Brattström, O., Lohman, D. J., & Wahlberg, N. (2022). Miocene Climate and  
505 Habitat Change Drove Diversification in *Bicyclus*, Africa's Largest Radiation of Satyrine  
506 Butterflies. *Systematic Biology*, *71*(3), 570–588. <https://doi.org/10.1093/sysbio/syab066>
- 507 Algar, A. C., & Losos, J. B. (2011). Evolutionary assembly of island faunas reverses the classic  
508 island-mainland richness difference in *Anolis* lizards. *Journal of Biogeography*, *38*(6), 1125–  
509 1137. <https://doi.org/10.1111/j.1365-2699.2010.02466.x>
- 510 Barraclough, T. G., & Vogler, A. P. (2000). Detecting the geographical pattern of speciation from  
511 species-level phylogenies. *American Naturalist* *155*: 419–434.
- 512 Buerki, S., Forest, F., Alvarez, N., Nylander, J. A., Arrigo, N., & Sanmartín, I. (2011). An evaluation  
513 of new parsimony-based versus parametric inference methods in biogeography: A case study  
514 using the globally distributed plant family Sapindaceae. *Journal of Biogeography*, *38*(3), 531–  
515 550. <https://doi.org/10.1111/j.1365-2699.2010.02432.x>
- 516 Caetano, D. S., O'Meara, B. C., & Beaulieu, J. M. (2018). Hidden state models improve state-  
517 dependent diversification approaches, including biogeographical models. *Evolution*, *72*(11),  
518 2308–2324. <https://doi.org/10.1111/evo.13602>
- 519 Crisp, M. D., Trewick, S. A., & Cook, L. G. (2011). Hypothesis testing in biogeography. *Trends in*  
520 *Ecology and Evolution*, *26*(2), 66–72. <https://doi.org/10.1016/j.tree.2010.11.005>
- 521 Ding, W.-N., Ree, R. H., Spicer, R. A., & Xing, Y.-W. (2020). Ancient orogenic and monsoon-driven  
522 assembly of the world's richest temperate alpine flora. *Science*, *369*, 578–581.  
523 <https://www.science.org>
- 524 Doob, J. L. (1945). Markoff Chains--Denumerable Case. *Transactions of the American Mathematical*  
525 *Society*, *58*(3), 455. <https://doi.org/10.2307/1990339>
- 526 Etienne, R. S., & Haegeman, B. (2012). A conceptual and statistical framework for adaptive  
527 radiations with a key role for diversity dependence. *American Naturalist*, *180*(4).  
528 <https://doi.org/10.1086/667574>
- 529 Fitzjohn, R. G. (2012). Diversitree: Comparative phylogenetic analyses of diversification in R.  
530 *Methods in Ecology and Evolution*, *3*(6), 1084–1092. <https://doi.org/10.1111/j.2041-210X.2012.00234.x>
- 532 Fitzjohn, R. G., Maddison, W. P., & Otto, S. P. (2009). Estimating trait-dependent speciation and  
533 extinction rates from incompletely resolved phylogenies. *Systematic Biology*, *58*(6), 595–611.  
534 <https://doi.org/10.1093/sysbio/syp067>
- 535 Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of*  
536 *Physical Chemistry*, *81*(25), 2340–2361. <https://doi.org/10.1021/j100540a008>
- 537 Goldberg, E. E., & Igić, B. (2012). Tempo and mode in plant breeding system evolution. *Evolution*,  
538 *66*(12), 3701–3709.
- 539 Goldberg, E. E., Lancaster, L. T., & Ree, R. H. (2011). Phylogenetic inference of reciprocal effects  
540 between geographic range evolution and diversification. *Systematic Biology*, *60*(4), 451–465.  
541 <https://doi.org/10.1093/sysbio/syr046>

- 542 Herrera-Alsina, L., van Els, P., & Etienne, R. S. (2019). Detecting the dependence of diversification  
543 on multiple traits from phylogenetic trees and trait data. *Systematic Biology*, 68(0), 317–328.  
544 <https://doi.org/10.1093/sysbio/syy057>
- 545 Hill, J. K., Thomas, C. D., & Lewis, O. T. (1996). Effects of habitat patch size and isolation on  
546 dispersal by *Hesperia comma* butterflies: implications for metapopulation structure. *The Journal*  
547 *of Animal Ecology*, 65(6), 725. <https://doi.org/10.2307/5671>
- 548 Jablonski, D., & Sepkoski, J. J. (1996). Paleobiology, community ecology, and scales of ecological  
549 pattern. *Ecology*, 77(5), 367–1378.
- 550 Kisel, Y., & Timothy, T. G. (2010). Speciation has a spatial scale that depends on levels of gene flow.  
551 *American Naturalist*, 175(3), 316–334. <https://doi.org/10.1086/650369>
- 552 Lancaster, L. T., & Kay, K. M. (2013). Origin and diversification of the California flora: Re-  
553 examining classic hypotheses with molecular phylogenies. *Evolution*, 67(4), 1041–1054.  
554 <https://doi.org/10.1111/evo.12016>
- 555 Losos, J. B., & Glor, R. E. (2003). Phylogenetic comparative methods and the geography of  
556 speciation. *Trends in Ecology and Evolution*, 18(5), 220–227. [https://doi.org/10.1016/S0169-5347\(03\)00037-5](https://doi.org/10.1016/S0169-5347(03)00037-5)
- 558 Losos, J. B., & Schluter, D. (2000). Analysis of an evolutionary species-area relationship. *Nature*,  
559 408(6814), 847–850. <https://doi.org/10.1038/35048558>
- 560 Louchart, A., Tourment, N., Carrier, J., Roux, T., & Mourer-Chauviré, C. (2008). Hummingbird with  
561 modern feathering: An exceptionally well-preserved Oligocene fossil from southern France.  
562 *Naturwissenschaften*, 95(2), 171–175. <https://doi.org/10.1007/s00114-007-0309-0>
- 563 Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character’s effect on  
564 speciation and extinction. *Systematic Biology*, 56(5), 701–710.  
565 <https://doi.org/10.1080/10635150701607033>
- 566 Mao, K., Milne, R. I., Zhang, L., Peng, Y., Liu, J., Thomas, P., Mill, R. R., & Renner, S. S. (2012).  
567 Distribution of living Cupressaceae reflects the breakup of Pangea. *Proceedings of the National*  
568 *Academy of Sciences of the United States of America*, 109(20), 7793–7798.  
569 <https://doi.org/10.1073/pnas.1114319109>
- 570 Matzke, N. J. (2013). Probabilistic historical biogeography: new models for founder-event speciation,  
571 imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of*  
572 *Biogeography*, 5.4. <http://phylo.wikidot.com/biogeobears>.
- 573 Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation  
574 is a crucial process in island clades. *Systematic Biology*, 63(6), 951–970.  
575 <https://doi.org/10.1093/sysbio/syu056>
- 576 Mayr, G. (2004). Old World Fossil Record of Modern-Type Hummingbirds. *Science*, 304(5672),  
577 861–864.
- 578 McGuire, J. A., Witt, C. C., Altshuler, D. L., & Remsen, J. v. (2007). Phylogenetic systematics and  
579 biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data  
580 and selection of an appropriate partitioning strategy. *Systematic Biology*, 56(5), 837–856.  
581 <https://doi.org/10.1080/10635150701656360>

- 582 McGuire, J., Witt, C., Remsen, J., Corl, a, Rabosky, D., Altshuler, D., & Dudley, R. (2014).  
583 Molecular phylogenetics and the diversification of hummingbirds. *Current Biology*, 24, 910–  
584 196.
- 585 Meseguer, A. S., Antoine, P. O., Fouquet, A., Delsuc, F., & Condamine, F. L. (2020). The role of the  
586 Neotropics as a source of world tetrapod biodiversity. *Global Ecology and Biogeography*, 29(9),  
587 1565–1578. <https://doi.org/10.1111/geb.13141>
- 588 Meseguer, A. S., Lobo, J. M., Ree, R., Beerling, D. J., & Sanmartín, I. (2015). Integrating fossils,  
589 phylogenies, and niche models into biogeography to reveal ancient evolutionary history: The  
590 case of *Hypericum* (Hypericaceae). *Systematic Biology*, 64(2), 215–232.  
591 <https://doi.org/10.1093/sysbio/syu088>
- 592 Nee, S., Holmes, E. C., May, R. M., Harvey, P. H., Harvey, P. H., Nee, S., Holmes, E. C., & May, R.  
593 M. (1994). Extinction rates can be estimated from molecular phylogenies. *Philosophical*  
594 *Transactions: Biological Sciences*, 344(1307), 77–82.
- 595 Nguyen, L.-T. (2011). *An Efficient Algorithm for Phylogeny Reconstruction by Maximum Likelihood*.  
596 Technische Universität Wien.
- 597 Ornelas, J. F., González, C., de los Monteros, A. E., Rodríguez-Gómez, F., & García-Feria, L. M.  
598 (2014). In and out of Mesoamerica: Temporal divergence of Amazilia hummingbirds pre-dates  
599 the orthodox account of the completion of the Isthmus of Panama. *Journal of Biogeography*,  
600 41(1), 168–181. <https://doi.org/10.1111/jbi.12184>
- 601 Polly, P. D., & Sarwar, S. (2014). Extinction, extirpation, and exotics: Effects on the correlation  
602 between traits and environment at the continental level. *Annales Zoologici Fennici*, 51(1–2),  
603 209–226. <https://doi.org/10.5735/086.051.0221>
- 604 Porto, L. M. v. (2022). *Patterns of diversification and geographic distribution of Canidae over time*.  
605 <https://doi.org/https://doi.org/10.33612/diss.198160305>
- 606 Ree, R. H., Moore, B. R., Webb, C. O., & Donoghue, M. J. (2005). A likelihood framework for  
607 inferring the evolution of geographic range on phylogenetic trees. *Evolution; International*  
608 *Journal of Organic Evolution*, 59(11), 2299–2311. <https://doi.org/10.1554/05-172.1>
- 609 Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+J model of  
610 founder-event speciation and its comparison with DEC via model selection. *Journal of*  
611 *Biogeography*, 45(4), 741–749. <https://doi.org/10.1111/jbi.13173>
- 612 Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by  
613 dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57(1), 4–14.  
614 <https://doi.org/10.1080/10635150701883881>
- 615 Ricklefs, R. E., & Bermingham, E. (2002). The concept of the taxon cycle in biogeography. *Global*  
616 *Ecology and Biogeography*, 11(5), 353–361. <https://doi.org/10.1046/j.1466-822x.2002.00300.x>
- 617 Ronquist, F. (1997). Dispersal-Vicariance Analysis: A New Approach to the Quantification of  
618 Historical Biogeography. *Systematic Biology*, 46(1), 195–203.  
619 <https://doi.org/10.1093/sysbio/46.1.195>
- 620 Sanmartín, I., & Meseguer, A. S. (2016). Extinction in phylogenetics and biogeography: From  
621 timetrees to patterns of biotic assemblage. In *Frontiers in Genetics* (Vol. 7, Issue 35). Frontiers  
622 Media S.A. <https://doi.org/10.3389/fgene.2016.00035>

623 Stuart, Y. E., Losos, J. B., & Algar, A. C. (2012). The island-mainland species turnover relationship.  
624 *Proceedings of the Royal Society B: Biological Sciences*, 279(1744), 4071–4077.  
625 <https://doi.org/10.1098/rspb.2012.0816>

626 Zhang, Q., Ree, R. H., Salamin, N., Xing, Y., & Silvestro, D. (2022). Fossil-Informed Models Reveal  
627 a Boreotropical Origin and Divergent Evolutionary Trajectories in the Walnut Family  
628 (Juglandaceae). *Systematic Biology*, 71(1), 242–258. <https://doi.org/10.1093/sysbio/syab030>

629

## 630 BIOSKETCH

631 Leonel Herrera-Alsina focuses on how species diversity is spread across space and time. Rates of  
632 diversification and species coexistence are regulated by the distribution of standing diversity and  
633 geographic constraints which vary over time. To understand the interaction of these factors, he  
634 develops dynamic models of diversification which provide theoretical predictions or are applied to  
635 empirical datasets.

636 Editor: Erin Saupe