Running head: The Andes and the evolution of Liolaemidae lizards

How important is it to consider lineage diversification heterogeneity in in macroevolutionary studies: lessons from the lizard family Liolaemidae

Olave Melisa[a], Avila Luciano J. [b], Jack W. Sites, Jr.[c] and Morando Mariana[b,d]

[a]*Department of Biology, University of Konstanz, Konstanz, Germany.*

[b]*Instituto Patagónico para el Estudio de los Ecosistemas Continentales, Consejo Nacional de Investigaciones Científicas y Técnicas (IPEEC-CONICET), Boulevard Almirante Brown 2915, U9120ACD, Puerto Madryn, Chubut, Argentina.*

[c]*Department of Biology and M.L. Bean Life Science Museum, Brigham Young University (BYU), Provo, UT 84602, USA; current address: Department of Biology, Austin Peay State University, Clarksville, Tennessee, 37044.*

[d]*Universidad Nacional de la Patagonia San Juan Bosco, Sede Puerto Madryn, Boulevard Almirante Brown 3700, U9120ACD, Puerto Madryn, Chubut, Argentina.*

Abstract

Macroevolutionary studies commonly apply multiple models to test state-dependent diversification. These models track the association between states of interest along a phylogeny, but they do not consider whether independent shifts in character states are associated with shifts in diversification rates. This potentially problematic issue has received little theoretical attention, while macroevolutionary studies implementing such models in increasing larger scale studies continue growing. A recent macroevolutionary study has found that Andean orogeny has acted as a species pump driving diversification of the family Liolaemidae, a highly species-rich lizard family native to temperate southern South America.

This study approaches a distribution-dependent hypothesis using the Geographic State Speciation and Extinction model (GeoSSE). However, more recent analyses have shown that there is a clear heterogeneous diversification pattern in the Liolaemidae, which likely biased the GeoSSE analysis. Specifically, we show here that there are two shifts to accelered speciation rates involving species groups that were classified as "Andean" in their distributions. We demonstrate that this GeoSSE result is meaningless when heterogeneous diversification rates are included. We use the lizard family Liolaemidae to demonstrate potential risks of ignoring clade-specific differences in diversification rates in macroevolutionary studies.

Key words: GeoSSE, diversification, speciation, extinction, macroevolution, biogeography, *Liolaemus*, *Phymaturus*, *Ctenoblepharys,* Andes

Introduction

Macroevolutionary modeling of diversification plays important roles in inferring large-scale biodiversity patterns (Schluter 2016). Several studies have focused on quantifying differences in macroevolutionary patterns linked to geographic, ecological, life-history and other traits, based on the variation in speciation and extinction rates (Jablonski 2008; Rabosky and McCune 2010; Ng and Smith 2014). Given that the mechanisms underlying the correlations between characters and diversification are generally poorly understood (Rabosky and Goldberg 2015), models have been developed to test the role of a range of different states promoting diversification, including binary traits (Maddison 2006), quantitative traits (FitzJohn 2010), geographic character states (Goldberg et al. 2011), multiple characters (FitzJohn 2012), punctuated trait changes (Goldberg and Igic 2012; Magnuson-Ford and Otto 2012), and time-dependent macroevolutionary rates (Rabosky and Glor 2010). These models

have been shown to perform very well on simulated datasets when using reasonably large trees (FitzJohn et al. 2009; FitzJohn 2010; Rabosky and Glor 2010; Goldberg et al. 2011; FitzJohn 2012; Goldberg and Igic 2012; Magnuson-Ford and Otto 2012; Stadler and Bonhoeffer 2013; Davis et al. 2013), and they have been implemented in hundreds of empirical studies (Rabosky and Goldberg 2015).

These models track associations between the states of interest and speciation and extinction rates along a phylogenetic tree, but they do not consider whether independent shifts in trait state are associated with shifts in diversification (Maddison and FitzJohn 2014; Rabosky and Goldberg 2015). Therefore, even if the shift is unrelated to the state targeted, a strong correlation with the diversification can be inferred from a rate shift (Maddison et al. 2007; FitzJohn 2010; Maddison and FitzJohn 2014; Rabosky and Goldberg 2015). Thus, all heterogeneity in diversification rates could potentially be linked purely to the states included in the analysis. Consequently, while larger trees are preferred due to the presumable increase of power, this also increases the risk of including clades with differences in states that can affect diversification along a tree (factors such as ecological requirements, dispersal abilities and life history [Li et al. 2018]). These potential issues have received little theoretical attention, while macroevolutionary studies implementing such models at increasingly larger scales continue to rise (Rabosky and Goldberg 2015).

The lizard family Liolaemidae is the most species-rich lizard clade the southern half of South America (307 species; Reptile Database 11 February 2019). The clade includes three genera: *Ctenoblepharys, Liolaemus* and *Phymaturus* (Fig. S1; Table 1). *Ctenoblepharys* is a monotypic genus with a distribution restricted to the coastal desert of Peru (Table 1), whereas *Liolaemus* is the world's richest temperate zone genus of extant amniotes (Olave et al. 2018), with 262 described species (Reptile Database 2 February 2019). *Liolaemus* includes a highly diverse group of species inhabiting a wide range of different environments (Table 1). The

76 sister genus of *Liolaemus*, *Phymaturus* (44 species; Reptile Database 2 February 2019) is
77 distributed along both the eastern and western Andean slopes in Argentina and Chile
78 (*palluma* clade), and through Patagonia (*patagonicus* clade). *Phymaturus* are strictly
79 saxicolous and largely restricted to volcanic plateaus and peaks (Cei 1986).
80 The three genera have clear differences in species richness, ecological requirements,
81 behaviors, and life histories (Table 1). A recent macroevolutionary study currently has found
82 disparate patterns of diversification among the three genera (Olave et al. *in review*), while
83 another recent study has focused on the entire clade, unknowingly the shifts in the
84 diversification rates along the tree (Esquerré et al. 2019). This study (Esquerré et al. 2019)
85 represents a major contribution to evolutionary biology and herpetology in that it: (i) presents
86 the largest Liolaemidae time-calibrated phylogeny to date (258 taxa), (ii) the most extensive
87 compilation of habitats, altitudes, and temperature data for all taxa, (iii) it hypotheses
88 ancestral range reconstructions, and (iv) opposite to previous findings, it demonstrates that
89 multiple origins of viviparity are not intrinsic properties in speciation rates.
90 However, Esquerré et al. approach the distribution-dependent hypothesis using the
91 Geographic State Speciation and Extinction model (GeoSSE; Goldberg et al. 2011) to test for
92 differences in speciation rates in Andean vs non-Andean (low elevation) species. The
93 GeoSSE model detected higher speciation rates in the Andean areas, and authors infer that
94 the Andean orogeny has acted as a "species pump" driving diversification in the Liolaemidae.
95 These authors performed further analyses to support this hypothesis (e.g. ancestral
96 distribution reconstructions and a time variable diversification model). Nonetheless, the
97 GeoSSE test was clearly key to identifying the role of the Andean orogeny in driving the
98 diversification of this clade. However, here we show that a clearly heterogeneous
99 diversification history of Liolaemidae is not considered by the GeoSSE analysis. Specifically,
100 we detect two shifts to accelerated speciation rates involving clades that were identified as

101 “Andean”. We show that the less diverse genus *Phymaturus* is characterized by the highest
102 speciation rates, and that there is a second shift within *Liolaemus*, specifically in the *L.*
103 *elongatus* clade. Consequently, the differences on speciation rates detected between Andean
104 vs. non-Andean species for the distribution-dependent diversification test is meaningless. We
105 demonstrate that the “Andean orogeny” hypothesis is not supported when the heterogeneous
106 diversification rates among these lizards is considered. The speciation history of the clade
107 Liolaemidae clearly demonstrates potential risks of the implementation of GeoSSE (and
108 likely other models of the family) when ignoring clade-specific differences in diversification
109 rates in macroevolutionary studies.
110
111 **Materials and methods**
112 *Phylogenetic tree*
113 We incorporated here the time-calibrated phylogenetic tree of Esquerré et al. (2019), which
114 includes the monotypic *Ctenoblepharys*, 188 described + 11 undescribed species of
115 *Liolaemus*, and 35 *Phymaturus* species (73% species coverage of all recognized
116 Liolaemidae). A consensus tree was obtained using TreeAnnotator 2.4 (Bouckaert et al.
117 2014).
118
119 *Speciation and extinction rates*
120 We estimated net diversification, speciation and extinction rates using BAMM 2.5 (Rabosky
121 et al. 2014). BAMM is a Bayesian approach that uses a rjMCMC to estimate lineage-specific
122 speciation and extinction rates, and rates of phenotypic change. Because the method
123 estimates rates per branch, it allows us to compare changes of these rates among clades and
124 species (i.e., tips) of interest. As in similar models, BAMM assumes the given topology is the
125 true phylogenetic tree, so to account for the topological uncertainty, we ran the analysis using

126 each of the last 500 trees inferred during the MCMC of BEAST. We informed the proportion
127 of missing taxa using *globalSamplingFraction* = 0.73, thus the program accounts for the
128 missing tips (i.e. 73% coverage). Priors were generated using *setBAMMpriors* in
129 BAMMtools (Rabosky et al. 2014), and we used all 500 obtained means for target groups
130 (genus, subgenus, clades and tips) to construct the final distributions used for all downstream
131 comparisons. All BAMM analyses were run for 5 x 10^6 generations, sampling every 1,000
132 generations, and with 25% burnin. We constructed parameter distributions per genus that
133 captured topological uncertainty. We calculated summary statistics using R (mean, standard
134 deviation and quartiles), and compared statistical differences among specific target clades
135 with ANOVA tests using the R function aov().
136
137 *Hypothesis testing: role of the Andean orogeny in diversification of the Liolaemidae*
138 To quantify the association between speciation and extinction rates to the Andes range, we
139 extracted species-specific speciation and extinction rates for different target clades, including
140 the whole family, genus (*Phymaturus* and *Liolaemus*), subgenus (*Eulaemus* and *Liolaemus*
141 sensu stricto), clades within *Phymaturus* (*P. palluma* and *P. patagonicus*) and several smaller
142 clades within *Liolaemus* (Table S1). We performed linear regressions using the R function
143 lm(), between the speciation (and extinction) rates and the maximum altitudes for all species.
144 The maximum altitude data were taken from the Esquerré et al. (2019) recompilation (their
145 Table S3). We also calculated linear models using the R function aov(), with the formula: rate
146 ~ "target clade" * "maximum altitude".
147 We implemented the GeoSSE (Geographic State Speciation and Extinction) models (Goldberg
148 et al. 2011) to test the hypothesis of higher speciation rates associated with the Andean species,
149 and used the same classification and tested the same set of models as in Esquerré et al. (2019;
150 Table 2). However, we do not fully agree with the original classification; as an example, the

151 “Patagonia” group is distributed across a huge area that was assumed to be “Andean”, which
152 we consider a poor classification for many species. For example, both the *P. patagonicus* and
153 the *L. lineomaculatus* clades are restricted mainly to the lowland Patagonian steppe. However,
154 here we respect the authors’ original classification and address the issue of heterogeneous
155 diversification rates in our analyses and discussion. We ran all analyses for the Liolaemidae as
156 a single clade, and then also for different nested clades. We used ML to estimate the parameters
157 as a starting point for an MCMC chain of 30,000 generations with a 20% burnin. All analyses
158 were performed in the R package *diversitree* (FitzJohn et al. 2009).
159
160 **Results**
161 *Heterogeneous diversification within the family Liolaemidae*
162 BAMM estimation of speciation and extinction rates on the Liolaemidae phylogeny (Figure
163 1A-B), displays two prominent shifts (PP = 0.4; Table S3), including the origin of the genus
164 *Phymaturus* (red), and the *Liolaemus elongatus* clade (light blue). There are significant
165 differences in speciation and extinction rates among genera ($p < 0.001$), as clearly shown by
166 the distributions of parameter estimations (Fig. 2). Specifically, the genus *Phymaturus* has the
167 highest speciation rate, is also associated with a high extinction rate. This result is concordant
168 with another study currently under review, using an different phylogenetic tree (Olave et al.
169 *in review*).
170
171 *Hypothesis test: the role of the Andes mountains in diversification of the Liolaemidae*
172 We constructed linear models between the maximum altitude (MA) of species occurrence
173 records, and the species-specific speciation and extinction rates. When considering all species
174 of Liolaemidae (258 tips), we find highly significant differences among genera ($p < 2^{-16}$), and
175 no significant effect of MA ($p = 0.808$), or their interaction ($p = 0.207$; Table S3A; Fig. S2).

176 Analyses of *Liolaemus* alone (194 tips) show a highly significant subgenus effect ($p = 3.10^{-8}$),
177 but non-significant MA effect ($p = 0.365$) or interaction effects ($p = 0.57$; Table S3E).
178 Equivalent results (i.e. no effect of the MA, but significant clade effect) were found for the
179 subgenus *Liolaemus* sensu stricto (s.s.) when including (97 tips), or excluding the *L.*
180 *elongatus* clade (71 tips), as well for the *Eulaemus* subgenus (97 tips; see Table S3F-G).
181 We also found a significant negative linear correlation between MA and speciation ($p = 1x10^{-}$
182 ^{4}) and extinction (p-value = 0.0019) rates in the subgenus *Eulaemus*, but a poor fit of the
183 model (R-squared < 0.2; Fig. 3). Analyses of *Phymaturus* alone (58 tips) show a positive
184 linear correlation between speciation rate and MA (Fig. 3), but there is also a clear clustering
185 of the *P. patagonicus* and *P. palluma* clades, both detected by the linear model (clades $p < 2^{-}$
186 ^{16}) with a non-significant contribution of the MA ($p = 0.158$), or their interaction ($p = 0.769$;
187 Table S3B). Finally, we found a significant correlation between speciation and extinction
188 rates for the *Phymaturus palluma* clade alone (28 tips; Fig S3 and Table S3C).
189 We performed distribution-dependent diversification tests using the GeoSSE program, first
190 testing the entire clade Liolaemidae, and found highly significant results (p-value =
191 0.0001735; Fig. 4) for the constrained model of equal speciation in Andean and sub-Andean
192 regions (Table S4). Thus, the GeoSSE model returns significantly higher speciation rates in
193 the Andean clade (= 0.27) relative to "lowland" species (= 0.11). This result is consistent
194 with previous findings by Esquerré et al. (2019); i.e., high-elevation Andean environments
195 are significantly associated with high speciation rates in the Liolaemidae. This analysis
196 included all *Phymaturus* species as Andean (Table 2), which also displayed a speciation rate
197 three times higher than *Liolaemus* (Fig. 2). We re-ran the analyses for the *Liolaemus* species
198 only, which returned only a slightly significant p-value = 0.04114 (Fig. 4) for a higher
199 speciation rate in Andean species (0.1883 vs. 0.1225; Table S4). This signal disappears
200 completely with the removal of the *L. elongatus* clade (which was classified entirely as

Andean; p-value = 0.16863), or when running the test with the *Eulaemus* subgenus alone (p-value = 0.182782; Fig. 4).

Discussion

Incorporating large trees for macroevolutionary studies has the advantage of providing larger datasets, and presumably more power. However, it is important to keep in mind the assumptions that go into such analysis: it treats all clades as evolving according to the same model, with the same values for the rate parameters. Here, we used the lizard family Liolaemidae to test for errors associated the use of large trees where clade-specific differences could bias results and lead to wrong conclusions. Our results clearly indicate that the signals of accelerated speciation rates associated with the Andean uplift in the distribution-dependent diversification test implemented in GeoSEE are biased (Fig. 4), due to the two diversification rate shifts along the tree (Fig. 1). Specifically, we demonstrated that the genera *Phymaturus* and *Liolaemus* display clear disparate patterns of diversification and that, when incorporated into this study, show that there is no apparent signal of Andean orogeny increasing speciation rates in the Liolaemidae. Earlier studies have confounded clade-specific rate accelerations with the distribution-dependent diversification results.

We do not argue against the implementation of GeoSSE (or any other model) in macroevolutionary studies, and do not doubt about the utility of state-dependent diversification models in general. However, our study calls attention to identify diversification rate heterogeneity for subsequent partitioning for the GeoSSE model (or other models within the family; see also Rabosky and Goldberg 2015), and we show that the BAMM program is a good option to identify such changes.

Acknowledgments

We thank D. Esquerré for providing clarification of how they performed their analyses. We thank all members of the Grupo de Herpetología Patagónica (IPEEC-CONICET) for continuing support. Financial support was provided by ANPCYT-FONCYT 1252/2015 (MM), and a postdoctoral fellowship (MO) from the Alexander von Humboldt Foundation at Meyer Lab, Konstanz, Germany.

Author contributions

MO and MM designed the study. MO carried out the analyses. MO, MM and JS wrote and edited the manuscript. LJA and MM provided recommendations based on the biology of the focal organism. All authors read and approved the final manuscript.

References

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology*, *10*(4), e1003537.

Cei, J. M. (1986). Reptiles del centro, centro-oeste y sur de la Argentina: Herpetofauna de las zonas áridas y semiáridas.

Davis, M. P., Midford, P. E., & Maddison, W. (2013). Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evolutionary Biology*, *13*(1), 38.

Esquerré, D., Brennan, I. G., Catullo, R. A., Torres-Pérez, F., & Keogh, J. S. (2019). How mountains shape biodiversity: The role of the Andes in biogeography, diversification, and reproductive biology in South America's most species-rich lizard radiation (Squamata: Liolaemidae). *Evolution* (in press).

FitzJohn, R. G. (2010). Quantitative traits and diversification. *Systematic biology*, *59*(6), 619-633.

FitzJohn, R. G. (2012). Diversitree: comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, *3*(6), 1084-1092.

FitzJohn, R. G., Maddison, W. P., & Otto, S. P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic biology*, *58*(6), 595-611.

Goldberg, E. E., & Igić, B. (2012). Tempo and mode in plant breeding system evolution. *Evolution: International Journal of Organic Evolution*, *66*(12), 3701-3709.

Goldberg, E. E., Lancaster, L. T., & Ree, R. H. (2011). Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, *60*(4), 451-465.

Jablonski, D. (2008). Species selection: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, *39*, 501-524.

Li, J., Huang, J. P., Sukumaran, J., & Knowles, L. L. (2018). Microevolutionary processes impact macroevolutionary patterns. *BMC evolutionary biology*, *18*(1), 123.

Maddison, W. P. (2006). Confounding asymmetries in evolutionary diversification and character change. *Evolution*, *60*(8), 1743-1746.

Maddison, W. P., & FitzJohn, R. G. (2014). The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic biology*, *64*(1), 127-136.

Maddison, W. P., Midford, P. E., & Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic biology*, *56*(5), 701-710.

Magnuson-Ford, K., & Otto, S. P. (2012). Linking the investigations of character evolution and species diversification. *The American Naturalist*, *180*(2), 225-245.

Ng, J., & Smith, S. D. (2014). How traits shape trees: new approaches for detecting character state-dependent lineage diversification. *Journal of evolutionary biology*, *27*(10), 2035-2045.

Olave, M., Avila, L. J., Sites Jr, J. W., & Morando, M. (2018). Hybridization could be a common phenomenon within the highly diverse lizard genus Liolaemus. *Journal of evolutionary biology*, *31*(6), 893-903.

Olave, M., Gonzalez Marin, A., Avila, L.J., Sites, J.W. Jr. & Morando, M. Disparate patterns of diversification within Liolaemini lizards. In *Neotropical diversification*, edited by Rull & Carnaval. Springer Nature (*in review*).

Rabosky, D. L., & Glor, R. E. (2010). Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proceedings of the National Academy of Sciences*, *107*(51), 22178-22183.

Rabosky, D. L., & Goldberg, E. E. (2015). Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic biology*, *64*(2), 340-355.

Rabosky, D. L., Grundler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., ... & Larson, J. G. (2014). BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, *5*(7), 701-707.

Schluter, D. (2016). Speciation, Ecological Opportunity, and Latitude: (American Society of Naturalists Address). *The American Naturalist*, *187*(1), 1-18.

Stadler, T., & Bonhoeffer, S. (2013). Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1614), 20120198.

297 **Tables and Figures**

298

	Ctenoblepharys	*Phymaturus*	*Liolaemus*
Described species	1	44	262
Distribution	Perú	Argentina Chile	Argentina Chile Perú Bolivia Southern Brazil Uruguay
Habitat	coastal desert	saxicolous	terrestrial arboreous arenicolous saxicolous
Diet	insectivores	herbivores	herbivores omnivores insectivores
Time for sexual maturity	unknown	7-8 years	2 years
Reproductive mode	oviparous	viviparous	viviparous oviparous parthenogenesis

299

300 Table 1: Summary of distribution, habitat use, diet and reproductive mode among the three

301 Liolaemidae genera.

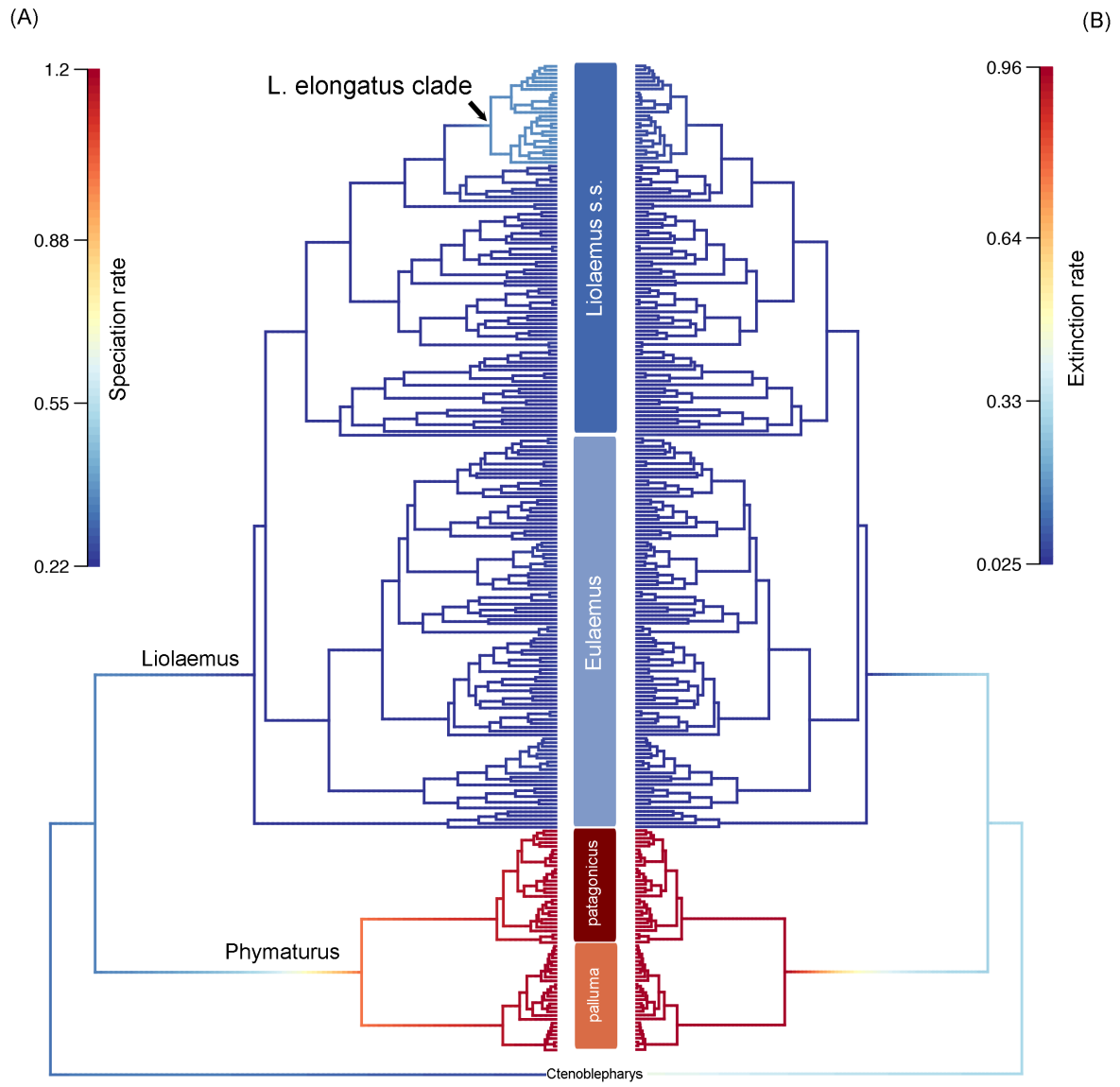
302

	Considered "Andean species"			Considered "Non-Andean species"		
	Patagonia	Central Andes	Altiplanic Andes	Central Chile	Atacama Desert	Eastern lowlands
Liolaemus	63	48	56	17	14	32
Phymaturus	33	23	5	0	0	1
Ctenoblepharys	0	0	0	0	1	0

303

304 Table 2: Species count for the geographic classification from Esquerré et al. (2019). Taken

305 from their supplementary material.



306

307 Figure 1: Color-coded phylogenetic trees for the speciation (A) and extinction (B) rates

308 through time for the Liolaemidae.

309

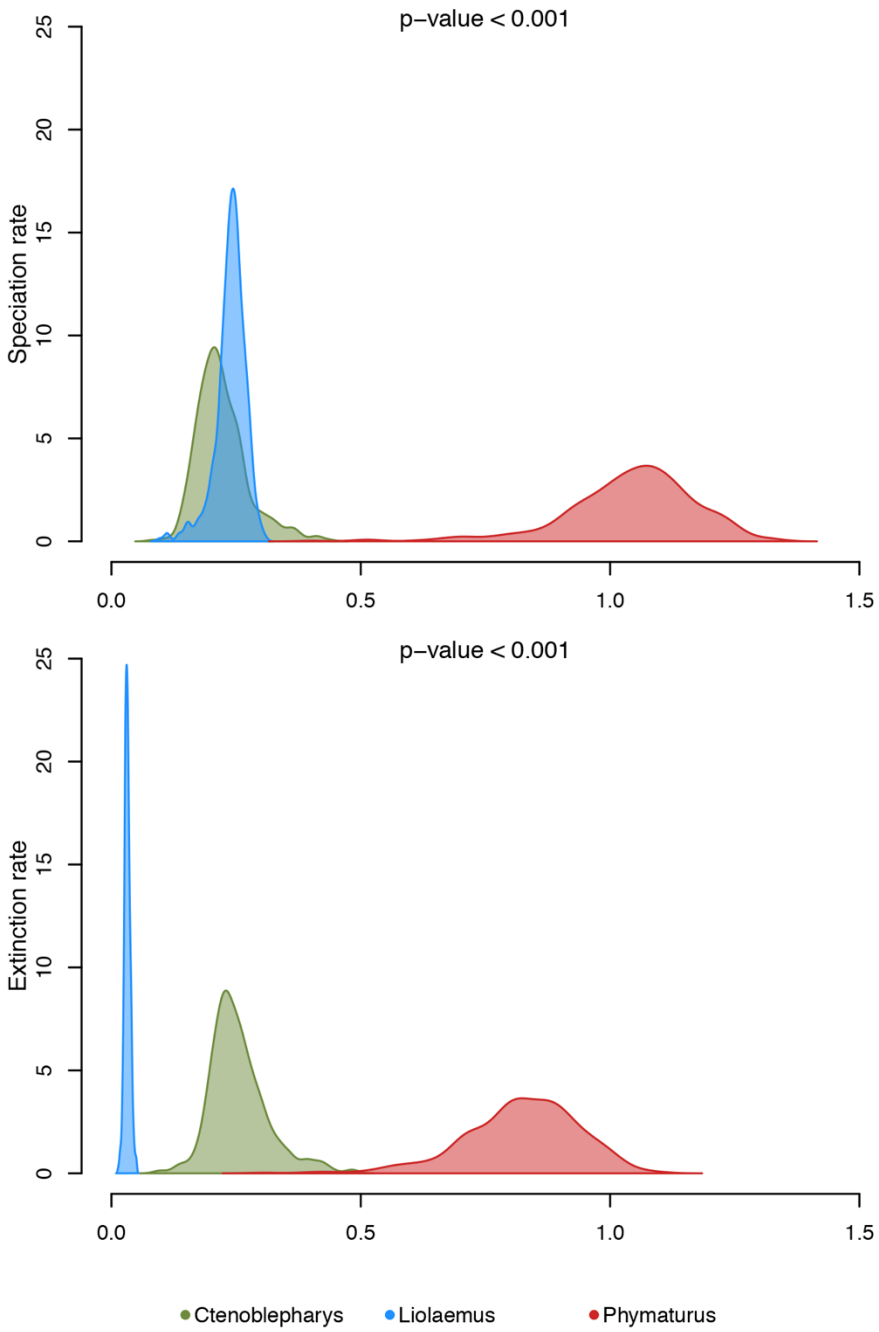


Figure 2: Speciation and extinction rates obtained for *Ctenoblepharys* (green), *Liolaemus* genus (blue) and *Phymaturus* genus (red). The density plots are constructed considering the mean obtained from each of the last 500 trees of the MCMC run for phylogenetic estimation. The p-value corresponds to an ANOVA test comparing distributions.

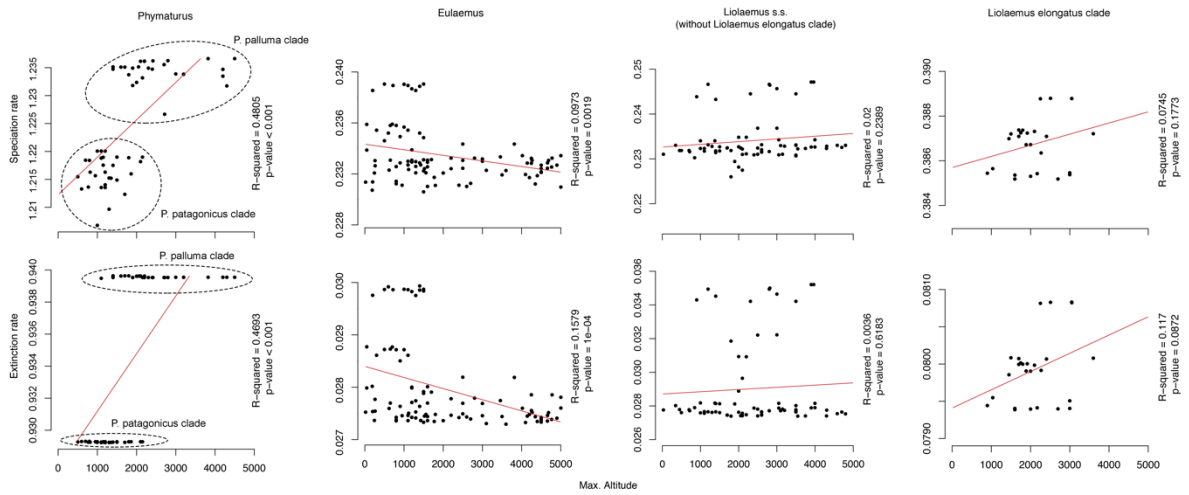


Figure 3: Linear regressions of the speciation/extinction rates as a function of the maximum altitude (meters) of the species occurrence for different target clades: *Phymaturus* genus, *Eulaemus* subgenus, *Liolaemus* sensu strict (s.s.) subgenus when excluding the *L. elongatus* clade and with the *L. elongatus* clade. See also the Figure S2-4 for more regressions, and Table S2 for full results of the linear model.

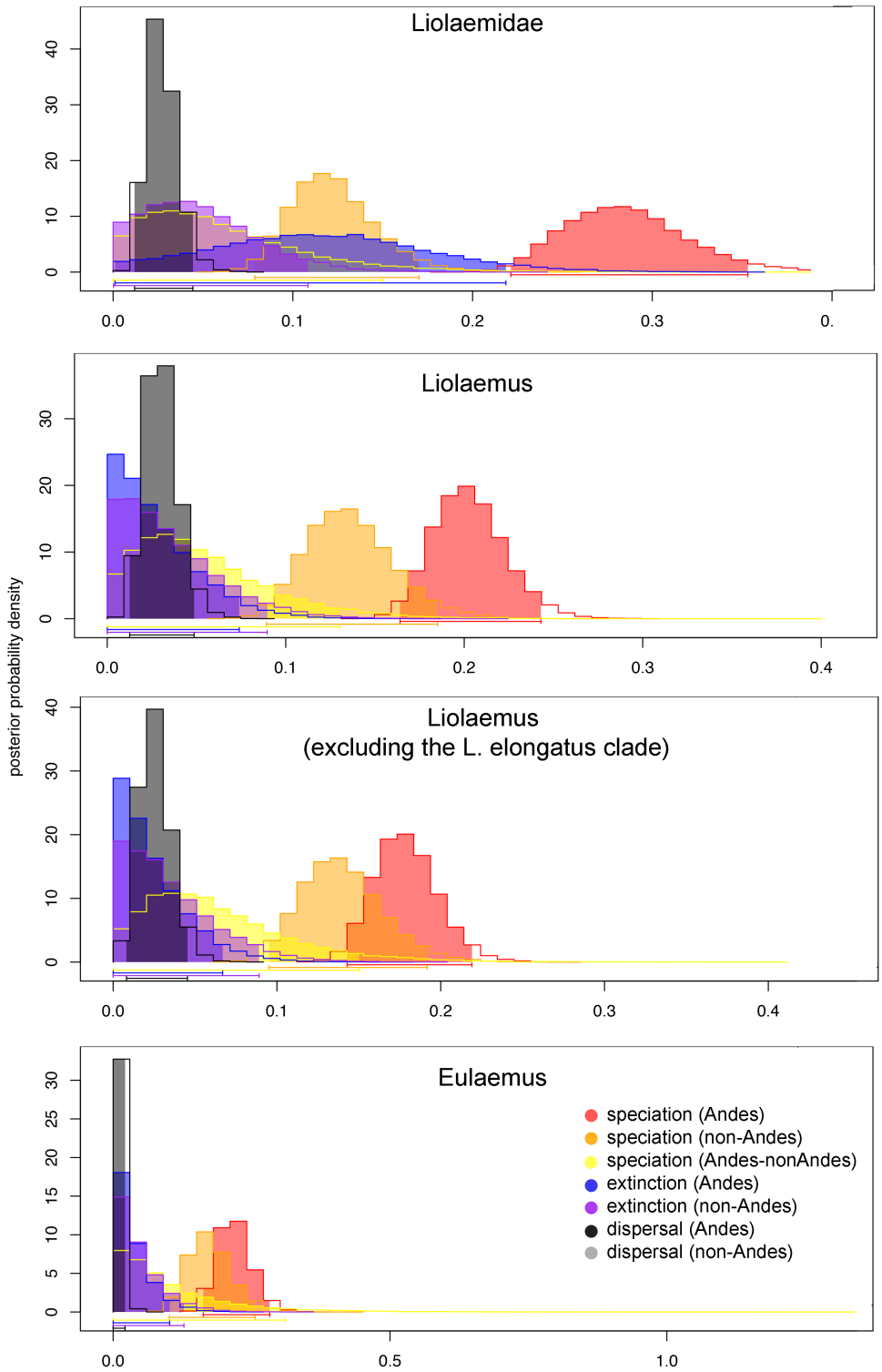


Figure 4: GeoSSE results for the different target clades: Liolaemidae, *Liolaemus*, *Liolaemus* (excluding *L. elongatus* clade) and *Eulaemus*. See also Table S3 for more details.