

Round #3

Decision

by Joaquín Hortal and Juan Arroyo, 2019-02-06 16:45

Manuscript: <https://doi.org/10.1101/236646>

Needs a revision; see AE report

First of all, please accept our apologies for the delay with this review. You present a novel and very bold hypothesis about (one of) the causes of latitudinal diversity gradients, and you do so in a quite complete and thorough way. So analysing all aspects of it has required some time, which together with the change of recommender and one of the busiest periods of the last years has done the rest. We are very sorry for that. Briefly, the current version of your study is in great shape, but needs some final tweaking. Besides our own reading, two additional reviewers made an in-depth review of it, and they identify some key points that merit some revisions. Most of these revisions are of relatively minor importance, as they are directed to improve the clarity of the manuscript and, importantly, the presentation of your results and analyses. As stated above, you tackle a novel issue and you do it thoroughly, so it is not surprising that some bits remain obscure even though the main lines of your work are quite convincing (but see my cautionary note below). So please follow the advice made by the reviewers, and try to clarify and/or discuss the points they make in a way that (a) the reasons behind your methodological choices are clear for the readers; and (b) the uncertainty associated to your results and their and implications are communicated better.

We want to warmly thank the editors for all the work they have done on the manuscript. We realize our study is not an easy one to evaluate, as it includes different types of analyses (biogeography, diversification), with or without fossils and for three different groups, which additionally comes (sometimes) with contradictory results (e.g. those based on extant data vs. fossil data). Therefore we are very grateful for all the effort they put into it.

Although this was not required, we have introduced some modifications in the text concerning the way our main hypotheses are presented; the AGE

hypothesis is now presented in the discussion section, as an outcome of our results rather than in the introduction. We think this makes a simplified and clearer message for the readers. We also think with this modification our manuscript will become more similar to the current scientific style. Otherwise, nothing else has been modified except based on the reviewers' suggestions. We hope the editors do not consider this last minute change inappropriate.

With regard to (a) both reviewers provide some clues about where the text needs improvement and some clarifications. Perhaps the most significant of them is Joaquín Calatayud's concern #2. This may be solved by reporting the state node probabilities as supplementary material, and making a brief statement that, given their distribution, the uncertainty associated to their estimations does not affect the overall estimation of the number of times each type of dispersal event (i.e. equatorward/poleward) happened. However, if the actual results do not provide solid evidence when uncertainty in state node estimations is displayed (i.e. if it not self-evident that there is no significant effect of uncertainty after a visual review), then I would recommend that you follow Dr. Calatayud's recommendations for estimating the uncertainty in the dispersal event associated to each branch of the phylogeny.

We have followed this recommendation and we now present a summary of the uncertainty in our biogeographic results for the 10 most basal nodes in Supplementary Tables 12-14, and the uncertainty for all nodes on Appendix 1. This is an important point, and we should have provided this information before. Please, see the answer to reviewer #1 below.

Re (b) we urge you to follow the reviewers's advice, in particular Dr. Calatayud's point #1, which I believe will improve the fairness in the presentation of your results. But importantly, we think that at the very end your manuscript may overstate the importance it has for the current configuration of the Latitudinal Diversity Gradient for the two ancient clades of turtles and crocodiles, and other tropical-adapted clades. It is not that your do not present convincing evidence that the ancient loss of tropical diversity has an imprint on the current configuration of the gradient. But the way you word things in the discussion and conclusions give the impression that current

configuration is the direct result of this process, thus giving a secondary importance to all the ulterior processes niche-conservatism, post-glacial dispersal, adaptations to dryer environments, etc. This is a matter of wording, of course, but the devil is in the details. We believe that the text will be more fair to the complexity of processes and factors that give rise to the current diversity gradient if you would simply include your evidence within the framework provided by current state-of-the-art in latitudinal diversity gradients. Such current state-of-the-art may not be clear, but it comprises a lot of different pieces of evidences and perspectives. Perhaps starting with the reviews of Jansson & Dynesius (Ann Rev Ecol Evol Syst 2003) and Willig et al. (Ann Rev Ecol Evol Syst 2003), and including the outcome of, e.g., Hawkins et al. (Ecology 2003); Wiens & Donoghue (TREE 2004); Ricklefs (Ecology 2006; Am Nat 2008; Ecol Lett 2011); Mittelbach et al. (Ecol Lett 2007); Svenning & Skov (Ecol Lett 2004); Allen & Gilooly (Ecol Lett 2006); Arita & Vázquez-Domínguez (Ecol Lett 2008); or (of course) Condamine et al. (Ecol Lett 2012); and ending with the (arguably) mixed evidence provided by Hanly et al (Am Nat 2017); Rabosky et al (Nature 2018) or Rangel et al. (Science 2018), among others. Here I'm not claiming that you have to be exhaustive, but rather to reflect that there are several causes of the gradient that all are operating over the pattern we see today. This is something that can perhaps be solved tweaking a few sentences here and there and/or adding one or two more in the conclusions, perhaps without even adding any citation. But without it we believe that the text does not provide a fair account of its true importance. Note that this does not reduce its importance.

We thank the editor for all this information and agree with this comment. The AGE hypothesis proposes a potential mechanism to explain the transition from a flat to a steep gradient, but (1) this might not be the only mechanism at play, and (2) the AGE hypothesis is not an entirely new idea, but mostly the contextualization of previous ideas in the frame of a time-variable scenario. Accordingly, we have edited the conclusions of the study. Many of the references he suggests are already cited in our study, but we have also included others to better represent the state of the art on the LDG.

Besides these (relatively) minor concerns, we believe that your manuscripts presents an overall solid work, both conceptually and analytically, that has the potential to have a significant effect on current debate on the origin(s) of the latitudinal diversity gradient. In fact, it is not every year (or even decade) that we see a truly novel addition to the latitudinal diversity gradient debate. We are looking forward to recommend a final version of this work.

Again, we are more than grateful for all your work on our manuscript and your positive input.

Reviews

Reviewed by Joaquin Calatayud, 2018-11-23 05:38

This is a very interesting and conceptually and methodologically impressive study. The authors present a new hypothesis that integrates and reconciles previous ones through a time-variable framework. In agreement with comments in previous revisions, I find the contextualization of the AGE hypothesis to be excellent, including also testable predictions from other hypotheses (such as those ones presented in table 1). In my opinion, this synthesis in its own deserves credits enough as to make an important contribution to the LDG field. In this line of merits, the authors used a throughout methodological approach to test predictions from the AGE hypothesis, including three taxa and phylogenetic (SSE models and DEC), fossil informed phylogenetic (DEC + fossils) and fossil analyses (PyRates). This provides inciting results (such as the disagreement between fossil and phylogenetic analyses), which again in its own may have an important impact on biogeographical and comparative studies. Overall, I find the study very interesting and I appreciate the outstanding synthetic and analytical effort behind it.

We would like to thank the reviewer for the positive feedback and his comments, which have helped us to improve our study.

While I really like the manuscript as it is now, in my opinion there are few minor aspects that could be further improved. Nevertheless, it is worth noticing that I'm just a short-experience practitioner on some of the analyses used. Thus, I can not deeply evaluate the adequacy of some analyses and especially those based on fossils.

1) Perhaps the most important thing is that results in general could be presented in a more informative way. That is, while I like the figures and appreciate the effort to synthesize results from all taxa and type of analyses, I feel that important piece of information are lacking or at least difficult to access. For instance, credibility intervals in figs. 4 and 3 are difficult to visualize, especially for diversification rates. In my opinion the manuscript would benefit from presenting credibility intervals where they were computed (i.e. SSE and PyRates analyses) numerically (e.g. in Table 2). I think this is quite important to clearly demonstrate that the results support predictions from the

AGE hypothesis. For instance, in fig. 4, it seems to me that the credibility intervals of diversification rates in the Holarctic and Equator datasets overlap (at least for turtles). If so, the crucial statement of lines 322-325 (i.e. “the slowing of diversification was much stronger in the Holarctic than at the equator”) would be weakly supported and perhaps some rewording would be needed.

We have followed the reviewer suggestion and improved these figures (now presented under the numbers Figures 2, 3, 4) to make credibility intervals “CIs” more readable. We acknowledge this is still not perfect, but as we want to show the uncertainty in parameter estimation we hope the main messages are understandable. However, we think that providing 95% CIs numerically in the form of a table is problematic here. Although we agree with the reviewer that such a table is often useful, given the number of parameter values (and thus credibility intervals) estimated in our study, this potential table would be huge and not very informative (unreadable) for the audience; please consider that in BiSSE we have estimated 3 parameters (speciation, extinction and dispersal; thus 6 CIs) for two groups (turtles and squamates), using two different models (time constant and time variable), and in three different periods (greenhouse, transition and colhouse) per model and group. The number of parameters estimated increases for PyRates, with 3 parameters (speciation, extinction and dispersal) for 3 different groups (turtles, squamates and crocodiles), in 2 different datasets (geography and ecology), and for 3 to up to 6 periods (depending on how the main time intervals were defined; by climate or geological periods). We would also like to stress that 95% CI are otherwise reported in the figures so this table could be redundant.

Concerning the overlap in credibility intervals of diversification rates in the Holarctic and Equator datasets, we would like to clarify that in our study and other similar studies (e.g. Silvestro et al. 2015 – New Phytol.), the rates of two adjacent intervals are considered significantly different if the mean of one lays outside the 95% credibility interval (CI) of the other, and *vice versa*. In our study, and for crocodiles, mean diversification rates in the Holarctic during the Paleogene climatic transition do not overlap with 95% CI of the previous interval, indicating that we detected a significant decrease of diversification rates in this

period. They also do not overlap with the 95% CIs of the equator, and *vice versa*, indicating that diversification rates differ significantly between regions on this period of interest (new Figure 3). For turtles, mean diversification rates in the Holarctic during the Paleogene climatic transition do not overlap with 95% CIs of the previous interval, suggesting a significant decrease of diversification rates in this period. Similarly, the mean diversification rate of the equator does not overlap with the CI of the Holarctic. The mean diversification rate of the Holarctic does overlap with the credibility interval of the equator. Nonetheless, we consider this to be an artifact of the huge 95% CI for the equator, which is probably overestimated because of the poor data quality of this dataset. As we discuss in the text; *“diversity dynamics for species distributed in the equator are difficult to trust in our study, because the equatorial dataset is the poorest of our study due to the small number of fossil lineages and the small number of records per lineage (Supplementary Table 15). This is clearly reflected by uncertainties on the estimated rates for equatorial clades, with wide credibility intervals probably due to geographic biases in the fossil record”*.

2) If I understand properly, during ancestral range estimation analyses the number of dispersal and range extinction events are calculated based on the most probable node state. While this seems to be a commonly used procedure, I think it ignores that the most probable state of a given node can still present a low probability. In others words, the uncertainty inherently associated to the model is not considered as the node state is taken as certain. As the node state probability is not provided it is difficult to evaluate whether this may influence subsequent estimations on the number of “into the equator” and “out of the equator” dispersal events. At first glance and for instance, this may be overcome by calculating the probability that such events occur for each phylogenetic branch. Then, these probabilities could be used along with any resampling technique to estimate the number of dispersal and range contraction events as well as its confidence. Please, forgive if I’m wrong and take this with caution since I’m not an expert in this sort of approaches. Alternatively, I think the state node probabilities should be provided and caution should be taken if these probabilities are low.

We have followed this recommendation and we now present a summary of the uncertainty in our biogeographic results for the 10 most basal nodes in Supplementary Tables 12–14, together with the uncertainty for all nodes in Appendix 1. This is an important point, and we should have provided this information before. In general, uncertainty is higher for the fossil-informed biogeographic models than for the unconstrained model (Supplementary Tables 12–14). Biogeographic reconstructions constrained by fossils tend to have lower likelihood values and relative probabilities for each node than unconstrained models (even if these models are more realistic; Meseguer et al. 2015 – Syst. Biol.; Sanmartín & Meseguer 2016 – Front. Genet.). This could be explained because in the fossil models, we are “forcing” the biogeographic character to evolve through “unlikely” character states given the data of extant species. And so, unfortunately, we don’t have a specific metric to evaluate the power of fossil-informed biogeographic reconstruction vs. traditional analyses beyond that the presence of a fossil in a given locality is a hard evidence for the ancestral presence of the taxa to which the fossil is assigned in this locality.

Comparing relative likelihoods for each different ancestral scenario based on fossils, we find that fossil-informed models generally identify the most likely ancestral scenario for each node with a high relative probability (> 0.8), or present probability values considerable higher (i.e. double) than the values of the next most supported scenario for particular nodes (Supplementary Tables 12–14, Appendix 1). When relative probabilities are similar between two different ancestral scenarios, we estimated the same ancestral range but a different mode of range inheritance (i.e. for each node DEC provides the ancestral range and how this range splits in the descendant lineages), and in all cases reconstructions include the Holarctic in the second most likely ancestral range for the concerned nodes (and in contrast with the unconstrained model). This suggests that our results are not very sensitive to the uncertainty in biogeographic estimation. We include a comment on this, but since the manuscript is already long it is not extended.

We also agree that estimating the number of dispersal events and range contraction events together with their confidence would be very interesting. Biogeographic Stochastic Mapping might allow doing so. Unfortunately, as far as

we know, it is not possible yet to include fossil information in the form of constraints on ancestral nodes using this method (fossils could only be considered as “fossil tips”, i.e. branches with tips significantly below the top of the tree; see <http://phylo.wikidot.com/biogeographical-stochastic-mapping-example-script>).

3) I do not understand very well why the number of dispersal events are normalized by the number of lineages. To my understanding by doing so results cannot be longer related to diversity patterns (as diversity is being controlled for).

When the number of dispersal events is controlled by the number of species, we are transforming the absolute number into a proportion, for example, instead of showing that X lineages of crocodiles dispersed into the tropics, we say that X% of crocodiles dispersed into the tropics. We think doing this is important to make the results comparable with other groups.

4) From the complete ignorance, is it possible that some incongruences between phylogenetic and fossil evidence come from the different taxonomic level used (i.e. species vs genus, respectively)?

This is a very interesting point. The difference in taxonomic levels between fossils (compiled at the genus level) and phylogenetic data (at the species level) could certainly be a problem if fossils are not assigned to their correct phylogenetic node. Here, we include fossil information in the form of ‘fossil constraints’ at certain nodes, according to the range of distribution of fossil occurrences assigned to a particular taxon during the corresponding time frame. For example the ancestor of turtles in our tree dates back to the Late Triassic (210 Ma, node 1 Fig. 5), and we set a constraint on this node with the distribution of all the Late Triassic fossils assigned to turtles. We acknowledge there is uncertainty on the placement of fossils to phylogenetic nodes and this is a probable source of biases in our study. However, we consider this bias much less problematic than the bigger bias of not considering fossils at all; it is alarming that the ancestral reconstructions based only on present data do not found a

single ancestor of turtles (for example) in the Holarctic despite the hundred of fossils that indicate they were in this region before the Cretaceous (Table S7).

We therefore consider that most of the incongruences between these two datasets come from the effect of extinction erasing evidence in phylogenetic reconstructions. Moreover, phylogenetic comparative methods have already been proven to be highly sensitive to high levels of extinction (Meseguer et al. 2015 – Syst. Biol.; Sanmartín & Meseguer 2016 – Front. Genet.).

5) It seems that supplementary figures are mislabelled (e.g. those referred in lines 265-268)

We have double-checked and the figure numbers are correct in this sentence.

6) In line 530 I think it should be “23/34”.

This has been corrected.

7) In my opinion lines 604-606 should be in the results or discussion sections (right now it is only discussed for the group apparently not following the first AGE postulate, which seems unfair).

We think we are not seeing the same line numbers. In our biorXiv version (BIORXIV-2017-236646v3-Meseguer), the lines 604-606 correspond with the methods of PyRate and the paragraph is relevant where it is placed. “*With these datasets, we reproduced the same PyRate analyses as for the whole dataset (see above). In general, the fossil datasets included mostly Holarctic fossils, with a smaller number of occurrences for the equator. Caution is therefore required when drawing conclusions from the equatorial datasets*”.

Reviewed by anonymous reviewer, 2018-11-23 05:42

The present manuscript by Meseguer and Condamine address and interesting and complex question using different approaches. The present version looks to be quite mature, well written and structured. Authors discussed general patterns explaining the Latitudinal Diversity Gradient in a scenario of Asymmetric Gradient of Extinction. They discussed all different results from each analysis instead choosing “the best” option to explain their hypotheses. This keep in some cases open conclusions, which is in my opinion, a positive thing considering the methodological complexity of the analyses in the manuscript and the assumptions that are implied on each one, which makes difficult to choose one alternative over the others. They also present ancestral area reconstructions with, and without fossil records in the analyses, comparing both results, again, without choosing one over the other.

Nevertheless, I found some methodological aspects that must be, at least, discussed and better justified:

We would like to thank the reviewer for his/her work on our manuscript.

In Methods, lines 594 to 594, authors assume that “taxa inhabiting the warm Holarctic were adapted to tropical-like conditions”. This assumption is supported by a reference for turtles (Testudines) but not for Squamata nor Crocodylia. You must include at least some references supporting this assumption for these two groups as in Testudines, to justify it before performing the analyses.

The reviewer raised a good point. We have now added references for the other groups studied here. It now reads as follows: “*For turtles, crocodiles and squamates this assumption is supported by Cenozoic climatic niche modelling (Waterson et al., 2016; Pie et al., 2017), stable isotope analyses and other climate proxies (Markwick 1998; Tütken & Absolon 2015)*”.

Additionally, authors used a DEC model for their biogeographical analyses without any specific motivation (at least it is not specified in the manuscript). Why are you using this model and no alternatives like a DIVA or BayArea-like? I would like to see some justification for the use of this model in your analyses.

Over the past decade, several promising biogeographic methods have been proposed, but the use of DEC (Ree et al. 2005 - Evolution) has become a standard practice due to its flexibility and robustness (Beeravolu & Condamine 2016 - bioRxiv), the possibility to account for branch lengths measured in units of time, to account for different area configuration scenarios or changing dispersal opportunities, as well as for extinction and different types of dispersal (see review of methods in Ronquist & Sanmartin 2011 - Annu. Rev. Ecol. Evol. Syst.; Matzke 2014 - Syst. Biol.; Sanmartín & Meseguer 2016 - Front. Genet.). More importantly, DEC allows including fossil information in the reconstruction of ancestral nodes, which is crucial in the context of our study. BayArea is basically an extension of DEC to hundreds or thousands of areas (Landis et al. 2013 - Syst. Biol.). Our study only includes 2 discrete areas, and thus the “data augmentation” approach of BayArea is not necessary. RASP and BioGeoBEARS are graphical interfaces of DEC models, and BioGeoBEARS includes the DEC+J model that assumes dispersal to occur at cladogenetic events (i.e. jump dispersal; Matzke 2014 - Syst. Biol.). Event-based parsimony biogeographic methods (e.g. DIVA, TreeFitter) do not account for branch lengths measured in units of time, use parsimony as an optimal criteria and have been shown to severely underestimate dispersal and extinction events (Ronquist & Sanmartín 2011 - Annu. Rev. Ecol. Evol. Syst.; Sanmartín 2012 - Evol. Edu. Outreach). These methods aim at finding congruent distribution patterns among organisms as evidence of shared biogeographic history (*i.e.* vicariance), and treat extinction and dispersal as a source of homoplasy (noise) in biogeographic reconstructions (Sanmartín 2012 - Evol. Edu. Outreach). Therefore they are less appropriate to estimate dispersal histories as we intend to do in our study. Finally, and perhaps more importantly, all these alternative models are not comparable in a statistical framework (Ree & Sanmartín 2018 - J. Biogeo.). Accordingly, we prefer to not include this discussion in the text, since the use of DEC is standard but also because our manuscript is very long, including multiple alternative methods in biogeography and diversification, with and without fossils.

On the other hand, I am concerned about using fossils in a comparative study for the Holarctic and tropical regions. Fossil preservation is different in both cases due mainly to different climatic conditions. This led to a different number of occurrences (fossilization is easier to occur in Holarctic regions), which can bias the analysis in PyRate due to lack of information in tropical regions.

The reviewer makes a good point. The fossil record is incomplete and biased towards the Holarctic where sampling effort has been considerably higher. This is clearly reflected in the number of available equatorial and Holarctic occurrences for our PyRate analyses (for example 429 vs. 3568, respectively, for turtles; see Supplementary Table 15). However, PyRate (which estimates diversification rates based on fossils only) has been shown to correctly estimate the dynamics of speciation and extinction rates under low levels of preservation or severely incomplete taxon sampling (Silvestro et al. 2014 – Syst. Biol.). That is because in contrast to molecular phylogenies, the incompleteness of the fossil record has a less problematic effect on the estimation of speciation and extinction rates because removing a random set of taxa does not affect the observed occurrences of other lineages. However, this does not warrant that the fossil record is complete and we have acknowledged that point in the manuscript. Our inferences are more robust for the Holarctic than for the Equatorial dataset. As we discuss in the text: *“diversity dynamics for species distributed in the equator are difficult to trust in our study, because the equatorial dataset is the poorest of our study due to the small number of fossil lineages and the small number of records per lineage (Supplementary Table 15). This is clearly reflected by uncertainties on the estimated rates for equatorial clades, with wide credibility intervals probably due to geographic biases in the fossil record”*.

Nonetheless, we would like to stress that the diversity dynamics inferred for the Holarctic are based on a large amount of fossils (Table S12) and calculated independently from the equatorial dataset. Therefore these inferences are robust and show a diversity loss (*i.e.* negative net diversification rates) during transitions to cooling periods, which is the main hypothesis we would like to test in our study. Facing similar problems in previous studies, researchers followed to conduct subsampling procedures (akin to rarefaction tests, Alroy 2014 -

Paleobiology). However, when the fossil record is analyzed with PyRate, applying a subsampling procedure is not relevant because PyRate needs all the fossil data to estimate times of speciation (t_s) and extinction (t_e) for each taxon included in the dataset: t_s and t_e are the 'real' ages of origin and extinction of a given taxon (t_e applies only if the taxon is extinct). If subsampling is applied then the first and last occurrences can be removed in the process, which has a direct impact on the estimates of t_s and t_e . Doing so, we will underestimate t_s (origins of taxa) and overestimate t_e (extinctions of taxa). In turn, this will impact the estimates of diversification rates through time. Currently, studies using PyRate do not perform subsampling procedure for the analyses relying on PyRate estimates (e.g. Silvestro et al. 2014 - Syst. Biol., Silvestro et al. 2015 - PNAS, Silvestro et al. 2015 - New Phytol.; Pires et al. 2017).

Finally, in figure 7 caption, I would specify that black line corresponds to the temperature for an easier interpretation of this figure.

Thank you for this point. We have modified the figure caption accordingly (now figure 6).