Round #1

 # Dear Dr. Mooers,
We are very grateful for your comments and suggestions, as well as those of the three reviewers. We have addressed all the reviewers' comments and think the manuscript has benefited from all the suggestions. See our detailed answers below preceded by "#".

Associate Editor's Comments

This is an intriguing study. I commend the authors for all the work they have done, and I echo the reviewers that this could be a qualitative contribution to the LDG literature. I very much liked the review and Table 1 (though I wish it had been annotated just a wee bit more).

#We would like to thank you for your work and relevant comments on our study. We have followed your suggestion and expand the legend on Table 1 to describe the assumptions of each hypothesis (and thus our annotations in the table). Now we think the table is self-explanatory and makes a better review of previous evolutionary hypotheses.

The AGE set-up does indeed seem synthetic, but it is formalizing and extending a lot that we already knew (e.g. that the LDG is not static because equatorial conditions have waxed and waned across latitudes through deep time, and the LDG could be due to tropical forms dying out in a cooler world farther from the poles) does not come till very late in the set-up (lines 133-147). This is confusing, because it is not clear if we are being exposed to a strong or weak test of a theory or to a new theory altogether.

 # We agree with the editor that this point could be confusing. Ideas of extinction in the Northern Hemisphere already existed to explain the LDG. Even since Darwin and Wallace, extinction has been considered an explanation for the lower diversity in the Holarctic in comparison with equatorial regions, but has often been attributed to the effect of Pleistocene glaciations. Recent studies also suggested higher turnover in the Holarctic to explain the pattern (i.e. high speciation and high extinction; e.g. Pyron 2014 – Glob. Ecol. Biogeogr.; Weir & Schluter 2007 – Science). However, diversity losses (i.e. higher extinction than speciation) in the Holarctic have been difficult to prove beyond high regional turnover, yet the later only explains the slow accumulation of lineages but not diversity losses. In addition, recent fossil investigations now show that the LDG has varied through time from a flat to a steep LDG (Mannion et al. 2014 –TREE). This suggests the processes that have generated these different patterns might have also changed through time. Hence, what we attempt to do in this paper is to formalize previous ideas on extinction but also expand them in the context of a time-variable pattern. In this sense, our AGE framework is novel and structured around testable predictions and hypotheses. We have now tried to clarify this in the Introduction.

I am also a bit more cautious than the reviewers regarding the results. I did not see a comparison of standard BiSSE analyses (which mirror those by Rolland et al. PloS Biol. 2014) with an AGE specific BiSSE analyses. Would this not be a relevant test of AGE? If not (perhaps because the authors feel all BiSSE analyses are suspect due to, e.g., the lack of fossil constraints on dispersal), then this needs to be clarified, and the BiSSE results recast as "unhelpful" – at the moment, I am not sure if I am to try to interpret them, or not. (And if not, does this mean I should discount the Rolland et al., results too?).

 # It is true that BiSSE models are conceived to test scenarios of differential diversification and asymmetric transitions as the one suggested here. Nonetheless, a classic BiSSE time-constant model, as is often used on LDG studies, would not be sufficient to test our hypotheses (*i.e.* the LDG varied across latitudes and time). Testing our scenario requires the implementation of time-variable trait-dependent models, such as the recently developed BiSSE.td, in which speciation, extinction, and dispersal rates

are allowed to vary between regions and to change at specific times (FitzJohn 2012 – Meth. Ecol. Evol.; Spriggs et al. 2015 – New Phytol.). In this study we performed both analyses, BiSSE and BiSSE.td. For the later, we introduced two shift times to model different diversification dynamics between greenhouse, coldhouse periods and their transitions.

Unfortunately, we found that none of the SSE models was completely in agreement with fossil findings; the time-variable BiSSE.td analyses correctly identified the main direction of dispersal in greenhouse and coldhouse periods, but did not detect diversity losses in the Holarctic as evidenced by fossils (*i.e.* negative diversification rates in the region during coldhouse periods) – in this sense mirroring Rolland et al. (2014 – PLoS Biol.) and Pyron & Wiens (2013 – Proc. B). We don't know yet how to explain this, but it is possible that there is not enough signal in the phylogenies to recover this pattern, such that under certain levels of differential extinction the model lost its power. In that case, this artifact might have limited other LDG studies as well, such as the one from Rolland et al. (2014 – PLoS Biol.). However, we cannot confidently extend this criticism beyond our study, as we have not investigated the LDG of mammals using fossils and thus we don't know if diversity losses occurred across Holarctic mammals in Rolland et al.'s study.

The comparison of the unconstrained and constrained DEC analyses is useful to show that uninformed reconstructions may be biased, but I am not sure that this was tested (do we have a metric to evaluate bias?), nor am I convinced that reconstructions assuming the fossil record is biogeographically "complete" for these groups over these time scales (the HFC DEC) can be considered data. I would want a terrestrial vertebrate paleontologist to comment on this specifically.

Unfortunately we don't have a specific metric to evaluate the power of fossil-informed biogeographic reconstruction vs. ordinal analyses beyond the self testimony of fossils; the presence of a fossil in a given locality indicates the ancestral presence of the taxa to which the fossil is assigned in this locality. Popular metrics in model comparisons such as the log-Likelihood or AIC values are not useful in this case, because biogeographic reconstructions constrained by fossils tend to have lower (worst) likelihood values than unconstrained models (even if these models are more realistic; Meseguer et al. 2015 – Syst. Biol.; Sanmartín & Meseguer 2016 – Frontiers in Genet.). This is simply because we are forcing the biogeographic character to evolve through "unlikely" character states given the data for extant species.

Nonetheless, we agree with the editor that assuming the fossil record is complete (the HFC model in our study) is probably a strong assumption for the data, while the SFC model implies lesser assumptions. If the editor agrees, however, we would like to keep the HFC model in the main text. We think the SFC model, despite being an improvement to classic models, is in some cases insufficient to incorporate known fossil information. In our experience, when extant taxa are not a good proxy of past diversity (*e.g.* because high levels of extinction erased the signal), the SFC model tends to misrepresent fossil information, that is, to not incorporate in the ancestral reconstruction of nodes all the fossil constraints we implemented. For the SFC approach, fossil data are incorporated into the anagenetic component of the likelihood calculation, and thus the fossil constraints impact the particular branches to which they have been assigned (Moore et al. 2008 - Proc. 52nd Annu. Syst. Symposium). If the area of the fossil is not represented at all among extant descendants, the fossil will still condition the estimated geographic-transition probability matrix for that branch, but

the fossil area tends to disappear along the particular branch without impacting other branches or nodes. In summary, we believe the "true" scenario is something in between the SFC and HFC models. In all cases, we would like to stress that both the HFC and SFC models are overall congruent and support the AGE model.

In this revised version of the manuscript, we have tried to clarify why we used the HFC model (methods section) but, if the editor agrees, we would like to keep the HFC reconstruction of the previous figure 3 (now figure 6) in the main text to clearly illustrate the discrepancies between biogeographic reconstructions based on present distributions only and reconstructions informed by fossils.

I think the tests of the AGE suggested by Reviewer 2 would be very interesting, but these suggestions also highlight what may be a major issue with this manuscript – it seems to at once want to introduce a new framework and also test the predictions from that new framework, and also present a lot of very deep-time and large-bin PyRate and DEC reconstructions across three large datasets. I am not convinced one can do all this at once effectively. The fact that none of the three groups actually present patterns consistent with AGE also gives me pause (am I correct with this interpretation? – figures 6 and 7 seem the most relevant here, but the fact I have to ask should give pause).

 # We might have not been clear enough when explaining our results. All fossil investigations for Testudines and Crocodylia are in agreement with the AGE model: we found strong evidence for diversity declines in the Holarctic (*i.e.* with net diversification rates in the Holarctic being negative during climatic transitions to icehouse; new Figure number 4), range extinctions in the Holarctic (*i.e.* species preferentially lost their high latitude distributions as evidenced by solid blue and red lines overpassing green lines on the new Figure 7a), as well as into the equatorial migrations (solid blue lines overpassing red lines on Figure 7b). Meanwhile, this scenario is not supported by analyses based only extant data (see dotted lines on Figure 7a and b, showing equivalent levels of range extinction and dispersals between Holarctic and Equatorial regions). This conflict between extant-data and fossil-informed analyses suggests extant data might not represent the past diversity, thus being probably insufficient to study the LDG.

For Squamata, however, the scenario is different, and we found the AGE model does not fully explain the current LDG pattern. For example, in Figure 4, we did not find more Holarctic than Equatorial extinctions as predicted by the AGE hypothesis.

I wonder if a paper that focused on, e.g. the Testudines as a test case might have higher impact. One could explore the AGE in detail with a view to presenting how one might properly test the AGE with these sorts of data. I leave this, and the reviewers comments, for the authors to consider as they revise and improve the paper.

 # We appreciate very much this suggestion. We are aware that presenting analyses for three different groups could confuse the main message of our work. However, if the editor agrees, we prefer to keep the three study cases. The truth is that we already submitted this paper to other journal including the turtle analyses only. However, the paper was rejected on the basis that turtles were apparently not sufficient to test our case because their subtropical preferences. This is the main reason why we decided to expand the study to other groups and include more empirical data to confront our AGE model.

Reviewers' comments

Reviewed #1 (Dr. Joaquín Hortal)

This is a very nice piece, trying to provide more realistic models of the macroevolutionary processes that drive latitudinal gradients of diversity on Earth. As such, models provide a bold approach to model extinctions and dispersions between latitudinal bands, accounting for long-term climatic oscillations. The modelling part seems sound to me, although I am just a well-informed reader rather than an expert in that part. But the main hypotheses behind the models provide an incomplete coverage of Latitudinal Diversity Gradients (LDGs), and the data itself has some limitations that have not been accounted for, or at least discussed.

You state explicitly along the whole text that the Latitudinal Diversity Gradient (LDG) is universally a decrease from richer tropics to less diverse temperate and, progressively, polar regions. This is far from being an absolute truth. I can concur that such LDG stands out for many biological groups (reptiles are certainly one of these groups), but a large number of taxa and functionally-consistent groups show other LDG patterns. Besides the typical textbook examples of, say, penguins (that are just a small clade and therefore not suitable for your coarse-evolutionary extent model), there are many groups that present their maximum diversity peaks at the temperate zones, including grasses, conifers, bryophytes, many groups of freshwater invertebrates (e.g. dive beetles, trichopterans, plecopterans, ephemeropterans, etc.; Vinson & Hawkins Ecography 2003; Boyero & Pearson J N Am Benthol Soc 2009; Morinière et al. Sci Rep 2016), parasitoid wasps, aphids (Kouli et al. Ann Zool Fennici 1994), certain families of dung beetles (but not others), among many other examples. These groups encompass a large part of terrestrial biodiversity (there are more known dung beetle species than birds and mammals altogether, for example). Patterns in the marine environment are also heterogeneous, and for many groups correlate more with the distribution of upwelling sources than with latitude. So there are many inverse LDGs due to the constraints imposed by their bauplans (see Kindlmann et al. in Scaling Biodiversity, ed. Storch et al., 2007). These basal bauplans (i.e. structural plans) are a direct consequence of the origin of each one of these major clades, that does not correspond to warm tropical periods and/or equatorial origins (e.g., the evolution of mosses in the Ordovician Ice Age, the clades originated in Laurasia during the Carboniferous, the C3/C4 transition in the Miocene and the adaptation of the major grass clades to arid and seasonal environments, etc.). Turtles show some deviation to the classic tropical-peak LDG pattern, but I miss any reference on how this affects the observed LDG for them in the context of your model.

Does this affect the validity of your alternative model? I do not think so, but questions its generality. I do think that it is has the potential to provide a nice representation (but see below) of the patterns and processes followed by many groups that currently show apparent "tropical niche conservatism" and processes followed by many groups that currently show apparent "tropical niche conservatism" patterns. In fact, that physiological limits rather than dispersal limitations shape the patterns of diversity for many groups in the interglacials seems more than reasonable, and many groups may follow that pattern, but sometimes these patterns are mediated by other requirements, promoting peaks of diversity at higher latitudes. Compare, for example, Hortal et al. Ecol Lett 2011 with Calatayud et al. J Biogeogr 2017; I quote these because I know them well, but they show how the same processes with different basal adaptations to climate lead to contrasting latitudinal patters (within Europe in this case). In what respects to recovering the LDG, your model only applies to taxa evolved in warm-tropical environments; and it works for reptiles (which are certainly a warm-adapted taxon). But I wonder if it could be generalized to origins in other latitudinal bands.

 # The reviewer is right, and we are grateful for this careful explanation; the accumulation of species in the equatorial latitudes is probably one of the most notable geographic pattern of diversity, but not the only one, and many groups present an inverse LDG. In addition, as the reviewer points, diversity patterns of many understudied groups are yet to be described prior to get a complete picture of the latitudinal diversity distribution.

The fact that other patterns exist, however, do not exclude our affirmations stating that the LDG is one of the most conspicuous patterns in ecology and evolution. In fact, the recent of study Kinlock et al. (2018 – Glob. Ecol. Biogeogr.) found the LDG pattern to be weaker than previously described (Hillebrand 2004 – Am. Nat.), although its generality was still confirmed. Many studies claim the LDG as the Earth's first-order biodiversity pattern (Krug et al. 2009 – Astrobiol.). A simple search in Google Scholar with the key word 'latitudinal diversity gradient' provided 82,000 entries (https://scholar.google.fr/scholar?hl=fr&as_sdt=0%2C5&q=latitudinal+diversity+gradient&btnG=&oq=latitudinal+divers), showing the extent of interest such biodiversity pattern has created.

However, we have now revised our text and included the fact that our study focuses on clades that were ancestrally adapted to tropical conditions and had difficulties to evolve cold tolerances under climate change. We think this might be the case for turtles, crocodiles and squamates, as evidenced by their current distribution and the inferred diversity declines in the Holarctic during major Cenozoic climate cooling (new Figure numbers 4 and 7).

As the reviewer points, Testudines do not have a typical LDG today, but instead show a hump-shaped diversity gradient centered around subtropical latitudes (10°S–30°N; Angielczyk et al. 2015). The paleolatitudinal distribution of turtles, however, was concentrated in the northern Holarctic, between 30–60°N during the Cretaceous (Nicholson et al. 2015 – Nat. Comm.; Nicholson et al. 2016 – Roy. Soc. B), suggesting that climate-driven extinction might have also affected this group. Our findings for turtles confirm this hypothesis, with tropical extinctions in the Holarctic and equator-ward dispersals increasing during global cooling of the Cenozoic. This result suggests that we can probably expand the AGE framework to groups adapted to subtropical conditions.

The processes are certainly there, it may be only a matter of constructing alternative hypotheses for groups generated in cooler moments/regions, where taxa retain "memory" of the conditions where they appeared. I honestly think this would make a great paper. Should you decide not to incorporate this into the models, the partial character of the richer-in-the-tropics LDG should be acknowledged throughout the text, making clear that your models seek to represent LDGs centered around the equator.

 # Our initial idea was to emphasize the role played by ancient Eocene extinctions over tropical-adapted taxa, since most accepted views generally focus on more recent Pleistocene extinctions affecting temperate taxa. However, your suggestion is interesting and could help to broaden the scope of our AGE framework. We have now tried to clarify in the introduction that our model mostly applies for taxa that were ancestrally adapted to tropical conditions, but that this model could be extended for taxa with different physiological tolerances. High levels of extinction are not expected under climate cooling for temperate adapted taxa, but on the contrary we can expect diversity increases as their environmental preference expands. In fact, we found that

diversification rates of turtles, crocodiles and squamates that originated under temperate climatic conditions were significantly higher than rates of tropical-adapted taxa right after the late-Eocene onset of temperate biome (Figure 5). The new temperate niches could have constituted an opportunity for diversification because of increased geographic ranges and ecological opportunities, drove an inverse LDG for other groups (book chapter of Kindlmann et al. 2008; Leslie et al. 2012 - PNAS).

Please read: *"the AGE model focuses on the fate of tropical-adapted clades under climate cooling. Nonetheless, for groups that originated or adapted to temperate conditions (evolved the relevant adaptations to cope with climate change), we expect that the AGE model could be extended to predict that temperate lineages more likely diversified in the Neogene habitats and contributed to an inverse LDG."*

Apart from that, I miss an assessment of whether your results are flawed by eventual biases in the fossil record. It is well known that we lack enough fossil data and adequate models to represent the tree of life, what has been termed the Darwinian shortfall (Diniz-Filho et al. TREE 2013; Hortal et al. Ann Rev Ecol Evol Syst 2015). So it is really hard to buy the assumption that the fossil record is complete; should it be unbiased, I could live with it, assuming random error and undersampling of the tree. But there is a high chance that biases in the fossil record affect the results of your paper.

In this particular case, most neotropical palaeontologists advise about the strong taphonomic bias towards the Holarctic; fossils are scant in tropical regions, in part because lack of paleontological effort, and in part due to the lower adequacy for the formation of deposits in many of the old soils of tropical South America, which is the continent I know better in this aspect. I'm not so sure about other tropical regions, but my guess is that only Chinese territories hold a comparable sampling effort in tropical Asia. The problem with the comparative oversampling of the Holarctic is that it spuriously increases the probability of finding dispersions into the tropics, compared to out of the tropics (for the Holarctic branches of the tree are more populated). There may be other more sophisticated ways I'm not aware of (perhaps adding probability of fossilization as a prior), but an straightforward way to account for this would be to resample n times the fossils to obtain samples with the same number of fossils in each latitudinal band (or at least, in tropics and Holarctic), and redo the trees n times. This sensitivity analysis (or a Bayesian alternative based on an adequate prior) will provide more robust results. In any case, a fair assessment and/or statement of how the limitations of the fossil record may affect your results is needed.

 # The reviewer is right; the fossil record is incomplete and biased towards the Holarctic where sampling effort has been considerably higher. In our study, we used the fossil record in two different ways: *(i)* **speciation and extinction dynamics were estimated entirely from the fossil record (new Figure numbers 4 and 5), including all available Holarctic and equatorial occurrences;** *(ii)* **dispersal and range extinction dynamics (Figure 7) were derived from phylogenetic-based biogeographic reconstructions constrained with fossils (8 fossils for crocodiles, 23 for turtles, and 30 for squamates; Tables S7–9). However, we think that in both cases our main results are not affected by the Holarctic bias of the fossil record:**

(i) **Concerning the inferences of diversification (speciation and extinction) rates (Figures 3–5): the PyRate method used in this study 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. 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., 2011, 2015, 2018; Antonelli et al., 2015; Favre et al., 215). We have thus followed this trend. However, this does not warrant that the fossil record is complete and we have acknowledged that point in the manuscript. As the reviewer points out, our inferences are much more solid 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 12). 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 significant amount of fossils (Table S12) and calculated independently from the equatorial dataset. Therefore these inferences are robust showing 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.

(ii) <u>Concerning dispersal estimates (Figure 7):</u> The reviewer's concerned is that "*the comparative oversampling of the Holarctic is that it spuriously increases the probability of finding dispersions into the tropics, compared to out of the tropics*". It is true that we have a comparative oversampling of ancestral Holarctic lineages in the fossil record, but at the same time, we have a comparative oversampling of equatorial lineages in the present (due to extinction of Holarctic species in the past). Inferences of ancestral biogeographic ranges and dispersal rates in our study are derived from present distributions and phylogenetic relationships, with the fossils used only to "constrain" or "correct" the ancestral reconstructions. The fossil record tells us that the ancestors of turtles, crocodiles and squamates were widely distributed in the Holarctic and the equator in the past. However, our biogeographic reconstructions based on extant data only recovered equatorial ancestors. This is because in maximum-likelihood biogeographic analyses, ancestral states are inferred conditional on the distribution of extant data (Ree et al. 2005 – Evolution). Meaning that if an ancestral geographic area is no longer represented among extant descendants (e.g. because of the effect of extinction), this area would never be reconstructed as an ancestral state (Sanmartín & Meseguer 2016 – Frontiers in Genet.). In our dataset, most turtle, crocodile and squamate species are currently distributed in the equator, and therefore the model only recovered an equatorial distribution for most basal nodes in the tree (e.g. Figure 6a), as well as higher dispersal rates out of the tropics than into the tropics (green dotted lines on Figure 7). However, the reconstructions based on extant data only were not able to recover ancestral Holarctic distributions for basal nodes, despite extensive fossil evidence attesting an ancestral Holarctic distribution for these lineages. So the

problem we found here was actually the opposite to what the reviewer suggests: we had a _comparative oversampling of equatorial species in the present (due to extinction of Holarctic species in the past), what lead to spuriously increases of the probability of finding dispersions out of the tropics, compared to into the tropics in the past._ Therefore, to overcome this sampling bias, we implemented fossil constrains to correct the model, but please note that we implemented both Holarctic and equatorial fossil constrains for different nodes.

Besides these two major concerns, the manuscript is nice, well-written and clear. As a minor comment, I'd move the sentences in lines 232-236 to the paragraph starting in line 153.

 # Thank you for your positive comment. Regarding the suggested move of the sentences, the problem is that different analyses were based on different data. Phylogeny-based diversification analyses (BiSSE and BiSSE.td) were based on phylogenetic data only, while fossil-based diversification analyses (PyRate) used fossil data only. Since on lines 232-236 we describe the fossil dataset, we cannot move this paragraph to the phylogeny-based diversification analyses (on line 153). We realize, however, that the way we structured the paragraphs could be confusing for a potential reader. We have thus reorganized the methods/and results sections to clarify this.

There are perhaps too many figures for an standard paper, but that is a matter of the journal you aim to send it to.

 # Yes we agree. But if the reviewer agrees we would like to keep them since we feel all of the figures are relevant to follow the text.

My own biases are obvious in this review, particularly in what respects to being picky with data quality and completeness, but generalizing current LDGs as holding a general peak of the tropics clashes with my experience, and I've seen LDGs for a fair number of groups as author, reviewer and editor, plus teaching LDGs for over ten years. So no matter whether you generalize your model or not, please avoid stating that LDG is universally centered in the equator. It undermines a really superb piece of work.

I hope this helps, Abrazos, Joaquín

 # We are very grateful for your work on our manuscript, and also for your comments and all the references suggested improving our work. We think your suggestions have made our paper more interesting and appealing for a broader audience.

Reviewer #2 (anonymous reviewer)

The authors of Ancient tropical extinctions contributed to the latitudinal diversity gradient present an intriguing paper that takes a historical approach to examine one of the most prominent macroecological pattern on Earth – the latitudinal diversity gradient (LDG). They briefly review the various hypotheses proposed to explain the general decline in species richness from equatorial to polar regions, and introduce a new frame-work, the asymmetric gradient of extinction, that ties together previous LDG theories and attempts to rectify theory with the observation that the steepness of the LDG varies considerably over geological time. The authors should be commended for bringing together diverse data sources and analyses to

support their theory, and this paper represents a major advancement in the study of the LDG.

Thank you very much for this positive input and your review.

I have only one major suggestion on how this manuscript could be improved. The crux of their argument is that during 'greenhouse' periods, tropical-adapted lineages have the opportunity to expand and establish into Holarctic regions – resulting in a shallow LDG as diversity accumulates through dispersal and in situ speciation. During the transition from 'greenhouse' to 'coldhouse' global climate eras, there should be a disproportionate loss of tropical-adapted lineages from the Holarctic, creating a steep LDG. This is an intuitive and intriguing idea, however I feel that the key test of this hypothesis has not been shown – that lineages originating in the tropics, but present in the Holarctic should have much higher extinction rates than 1) lineages with Holarctic origins and 2) tropical/Holarctic lineages in the equator.

While they do demonstrate what appear to be broad-scale differences in extinction and origination for Holarctic vs. Equatorial species across this transitional period – what would really solidify their hypothesis would be to test the following predictions using this same dataset:

1. Species from ancestrally tropical lineages have higher extinction (and lower or equal origination) in the Holarctic than species from Holarctic-derived lineages during coldhouse/transitional periods

2. Species from ancestrally tropical or Holarctic lineages have equivalent origination and extinction rates in the equator during this period

It may be difficult to assign 'Holarctic' vs 'Tropical' origin to all species, but presumably the DEC analysis could potentially be used to categorize lineages as Equatorial or Holarctic origin. Alternatively one might expect that species from lineages that have longer evolutionary histories in the Holarctic would have lower extinction rates during the transitional and coldhouse periods. There are a few methods that may work to assign origin in this case. The authors could then repeat their analyses (Fig. 6) with rates estimated for these four groupings (Tropical in Holarctic; Temperate in Holarctic; Tropical in Equator; Temperate in Equator), this would significantly strengthen the narrative of the manuscript and support their main hypothesis.

#We are grateful for this interesting discussion. Including the area of origin of clades in the analyses could bring interesting information, allowing for example, to better estimate evolutionary rates. However, implementing such information in the estimation of speciation and extinction rates would be methodologically difficult: Phylogenetic-based methods such as BiSSE allow estimating speciation and extinction rates while simultaneously inferring the ancestral characters and character transition rates. However, as far as we know, these methods do not allow including fossils. Biogeographic models, such as DEC, allow estimating the origin of areas, but not speciation and extinction rates. In addition even with DEC, it is very difficult to infer the area(s) of origin of ancient clades. As you can see for example with the example of turtles, the ancestors are inferred as being widespread (Figure 6b). Speciation and extinction rates of Figures 4 and 5 are therefore entirely derived from the fossil record without associating a phylogenetic context. At the moment it is not possible to include phylogenetic relationships in the fossil-based analyses of PyRate, and thus we cannot estimate clade-specific rates based on their area of origin.

Nonetheless, we think that including the area of origin in our analyses is not necessary to support our main hypothesis (*i.e.* the current LDG resulted from diversity

losses in the Holarctic following the contraction of the tropical biome). This hypothesis has two components: *(i)* we need to uncover the pattern, in this case prove that there has been more extinction in the Holarctic than in the equator during climate cooling (note that for this part, knowing the ancestral area of origin of the clade is irrelevant); and *(ii)* we need to uncover the mechanisms and prove that extinction affected tropical-adapted taxa. For this second point, knowing the area of origin could be relevant if tropical-adapted taxa appeared only in the equator and temperate taxa in the Holarctic. However, this was not the case and tropical lineages were also distributed across the Holarctic before the Oligocene (Greenwood & Wing, 1995). Therefore, we consider that more interesting than assessing the ancestral area of origin is to assess the ancestral ecology of the taxa. And this is something we could actually do, and have approximated in our study according to the general macroclimatic conditions of the fossil locality at that particular time (*i.e.* overall tropical or temperate).

To answer our question we have therefore divided the fossil dataset in Holarctic vs. equatorial fossils, and in fossils living under tropical vs. temperate conditions. Please note, that these datasets are not the same, since tropical-adapted taxa were also distributed in the Holarctic before Neogene climate cooling. With our analyses, we have been able to *(i)* show that diversity declines in the Holarctic occurred during climatic transitions to cold (*i.e.* net diversification rates of turtles and crocodiles in the Holarctic being negative during this period (new Figure number 4). In addition, we have been able to approximate *(ii)* the mechanism showing that extinction affected tropical-adapted taxa, and suggesting that physiological limits and climate-driven extinction could be behind the pattern (negative net diversification rates in Figure 5 are only observed in the tropical dataset while temperate clades actually have high rates of diversification coincident with climate cooling).

In summary, we don't think that our results might be affected by the fact that the Holarctic lineages originated in the Holarctic or in the equator. Indeed, we assume high rates of dispersal across these areas when the climate was warm and uniform during the Paleocene. We think the key here is that many lineages living under tropical macro-conditions disappeared from the Holarctic during the Neogene (or lost their Holarctic distribution), independently of whether they were ancestrally originated in the Holarctic or in the equator.

My only minor comment is that there are minor grammatical errors throughout the manuscript, so I'd recommend the authors ask a native English speaker go through and do minor edits to the language.

 # Thank you again. In this new revised version we have made the text be read and corrected by a professional English editor.

<u>Reviewer #3 (Dr. Juan Arroyo)</u>

This manuscript deals with a hot topic of long standing interest among biogeographers, macroecologists, systematists, among others, that of Latitudinal Diversity Gradient (LDG). Despite it being described long ago, it has been reluctant to exclusive hypotheses ready to be tested beyond correlation. However, this pattern has generated a renewed interest within the biogeographic and macroecological community, as modern tools for reconstructing the past of lineages and the areas where they lived, as well as information on fossils record and distribution, are increasing available.

The merits of the manuscript lie on (1) the consideration of competing hypothesis to those canonical assuming a constant through time higher diversity in the tropics; (2) the application of recent tools to explicitly disentangle the role speciation and extinction, and dispersal (although I would prefer the use of migration, just to avoid confusion with the ecological meaning of the term dispersal); (3) the use of massive data of fossil records for the former aim; and finally (4) integrate all this information in a explicit temporal hypothesis o variation according the tropical-temperate type shift.

Despite I am not an expert on the most recent analytical methods on historical biogeography, the manuscript is thought provoking and deserves to be known by the community. I feel that the hypothesis the authors are launching is falsifiable and thus a number of studies will address it, in the extent there are data set suitable for it.

 # We are very grateful for your kind words and constructive suggestions. We have addressed all your suggestions and found them interesting to improve our manuscript.

My comments to specific parts of the study, some of them quite wide across the arguments, follow:

Abstract .
No mention is provided to specific temporal setting of the study, despite the wide citation of specific dates throughout the main text.

 # That is an error. We have now corrected this and mention the time frame of our study in the abstract.

It is convenient that when mentioning "We studied" (line 35) that in fact the study is widely based on existing published phylogenies, which are refined with additional data (e.g. fossils). This is also applicable to lines 147-149, at the end of Introduction.

My view about the challenge the study represents (line 43) is that in fact the authors propose a refinement of the current theories by incorporating meaningful data on temporal context with independent data (fossils), which might change in the future, when more studies on other groups are available, our views about the LDG.

 # Thank you. In this revised version of the manuscript we have modified the abstract to clarify our contribution to the field, as well as the temporal framework of the study.

Introduction
Your approach is to use different groups as to test your hypothesis (three in your study). This brings me two thoughts. One is the selection of groups. Your selection is of three groups which are related, this might limit the power of your analyses. In fact, it would be needed that you mention in the Introduction the rationale for selecting these groups, even thought you justify in detail the Methods section the reader needs to know somewhat about this as soon as possible.

 # We have now moved to the end of the introduction the rational to select these groups. The reviewer is right that these groups are related, however they still represent a significant amount of the vertebrate diversity today. We selected them because there are phylogenetic reconstructions available for these groups, they are ancient and likely experienced all climatic changes that occurred through the Cenozoic, and most

importantly, they have a rich fossil record. Phylogenies of other groups such as birds (who are the closer relatives to crocodiles) or mammals also exist. However, the fossil record of birds is scarce, while the fossils record of mammals is so rich that its study on a global scale is complicated. Please note that before using the fossil datasets we first performed a careful verification (checking the current taxonomy and removing synonym names); all fossil occurrences were collated to exclude ambiguous citations in our fossil datasets.

The second is that, independently of the relationships among groups, it recalls about the correlative value of biodiversity patterns in groups. I mean that it should posed the question of the use the patterns in some groups as proxies for general patterns, which in fact is the ultimate aim of the study.

We believe the AGE model might explain the LDG pattern of other tropical-adapted clades that had difficulties to adapt to new temperate environments. Accordingly, we think our hypothesis is generalizable. Nonetheless, as we mention in the introduction: We acknowledge the AGE hypothesis might not be applicable to all organisms showing a LDG today. As exemplified here, tropical extinction is not the most influential cause behind the LDG of squamates. But the AGE hypothesis will be still useful to explain the LDG of groups that once were diverse at high latitudes.

A serious issue throughout the manuscript is the strong bias towards the Northern Hemisphere (Holarctic is included among keywords, but see line 122). Probably this is an effect of the particular distribution of the groups selected. This is not bad at all, but a different selection of groups would have increased the prospective value of the hypothesis, particularly due to the fact that in Southern Hemisphere the climate changes were very probably smoothed due to the sea/land configuration and extension of plates. If the authors are correct, the LDG should vary not only through time, but also through hemispheres.

This is a good comment. It is true that our hypotheses are mainly focused on the Northern Hemisphere, but we don't exclude that similar processes occurred in the southern new temperate regions as well (i.e. diversity losses and into the equator migrations explaining the low diversity on southern regions today). Indeed, we found high levels of range extinctions in these regions (red lines on new Figure number 7). Unfortunately, we were not able to do the PyRate speciation and extinction analyses for the Southern Hemisphere, since there were not enough fossil occurrences in these regions. We have now tried to clarify in the text that our hypotheses are mainly focus on the Northern Hemisphere because of limitations of the data, but they could be extended to southern regions as well.

In table 1 and others there is not indication of the meaning of "r" in the legend. For familiar readers it is clear that this is diversification rate, but should be made explicit.

We have now included the meaning of "r" in the table. We have also expanded the legend on Table 1, following the suggestions of the editor, to describe the assumptions of each hypothesis (and thus our annotations in the table). Now we think the table is self-explanatory and presents a review of previous evolutionary hypotheses.

Results
Line 165. What is the meaning of "region" here? Biogeographical? Any other?

We have clarified this. We refer to biogeographical regions: Holarctic or equatorial

I wonder in what extent results may rely on different rates and probabilities of having reliable fossil records in particular regions (e.g., the tropics). This is central to the whole study. Traits, Ecology...

We think testing the AGE hypothesis is dependent on the quality of the fossil record to some extent, and the reviewer #1 has raised similar concerns. In particular for the equatorial fossil data, we acknowledge the limitation that: "*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 12). 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 significant amount of fossils (Table S12) and calculated independently from the equatorial dataset. Therefore these inferences are robust and show diversity declines (*i.e.* negative net diversification rates) during transitions to cooling periods, which is the main hypothesis we attempt to prove in our study. In addition, the method use here, PyRate, 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.). In any case, the bias of the fossil record might have only affected the equatorial dataset, but likely not the Holarctic dataset, nor the biogeographic reconstructions inferred with fossils.

Although it is true that we have a comparative oversampling of ancestral Holarctic lineages in the fossil record, at the same time we have a comparative oversampling of equatorial lineages in the present (due to extinction of Holarctic species in the past), what lead to spuriously increases of the probability of finding dispersions out of the tropics, compared to into the tropics in phylogenetic-based biogeographic analyses. In our study, inferences of ancestral distributions and dispersal rates were not derived from the fossil record but from distributions and phylogenetic relationships of extant species. For this analysis, fossils were only used to calibrate or "correct" the ancestral reconstruction. The fossil record tells us that the ancestors of turtles, crocodiles and squamates were widely distributed in the Holarctic and the equator in the past. However, our biogeographic reconstructions based on extant data only recovered equatorial ancestors. This is because in maximum-likelihood biogeographic analyses, ancestral states are inferred conditional on the distribution of extant data. Meaning that if an ancestral geographic area is no longer represented among extant descendants (e.g. because of the effect of extinction), this area would never be reconstructed as an ancestral state. In our dataset, most turtle, crocodile and squamate species are currently distributed in the equator, and therefore the model was able to recover an equatorial distribution for most basal nodes in the tree (e.g. Figure 6a), as well as higher dispersal rates out of the tropics than into the tropics (green dotted lines on Figure 7). However, these reconstructions based on extant data only were not able to recover ancestral Holarctic distributions for basal nodes, despite extensive fossil evidence testifying a Holarctic distribution for these lineages in the past. So the problem we found here was that we had a comparative oversampling of equatorial species in the present (due to extinction of Holarctic species in the past), what lead to spuriously increases of the probability of finding dispersions out of the tropics, compared to into

the tropics. Therefore, to overcome this sampling bias, we implemented fossil constrains to correct the model, but please note that we implemented both Holarctic and equatorial fossil constrains for different nodes.