

Dear authors,

Thank you for soliciting the Peer Community in Evolutionary Biology to assess your study.

We have now received the feedback of three reviewers for your preprint study (see below). You will see that the three referees bring up very interesting and useful comments as well as suggestions that I am sure will improve the study. Overall, I agree with the reviewers that the study has many merits and that the findings are interesting. I also think the approach proposed here is original and may be useful for further studies. However, the study suffers from some methodological issues. I think the main issues concern the phylogeny and dating analyses, but because these results are the cornerstone of the other analyses, the corresponding results may be inconclusive as it stands. The referees also felt that the manuscript suffers from a lack of clarity in several parts of the text. The manuscript would certainly benefit from a careful reading by a native English speaker.

To summarize, I have identified six major points raised by the reviewers that you would need to carefully address. This includes the following: (1) The writing and structure of the paper (in many places the referees have suggested to rephrase the sentence, clarify the meaning, or be more accurate); (2) Clarification of the hypotheses tested (this is an important part of the paper and it needs support from previous studies and the biology of the group); (3) Revising the phylogenetic and dating analyses based on the comments (e.g. using RAxML / MrBayes for the phylogeny and BEAST for the dating with a fixed topology if necessary); (4) Redoing the downstream analyses depending on the newly generated time-calibrated phylogeny (would be great to compare your results with the newly generated tree, and the results obtained with the Deuve et al.'s tree as suggested); (5) Revising the results and discussion based on that new analyses if any change has to be done and including the uncertainty of your results (confidence intervals or credibility intervals); (6) Emphasizing your results in the context of existing literature regarding Pleistocene climatic oscillations and associated effect on biodiversity.

Based on the referees' comments and my reading, I believe the manuscript will benefit from a revision and a second round of reviews. If you chose to resubmit a revised paper, please make a point-by-point reply to the comments (like for a traditional journal). For the moment, I do not recommend the study in PCi Evol. Biol. but if the revision is thorough (satisfies the reviewers) and the results still support the conclusions, I will be supportive for the paper as being recommended.

Dr. Fabien Condamine, recommender for PCi Evol. Biol.

Dear Dr. Condamine, thank you very much for handle our manuscript and thanks also to the reviewers for their detailed revisions and insightful comments. We feel that reviewers' comments were both sharp and sound, and hence they were quite helpful. We would like also to apologize for the delay, but suggested re-analyses were computationally exigent and took longer than expected. These new analyses corroborated previous results on the impact of Pleistocene glaciations on the phylogenetic structure of *Carabus* faunas, even when using different time calibrations in phylogenetic reconstruction. The remaining major concerns have also been thoroughly addressed and are discussed in detail below.

*Reviewed by anonymous reviewer, 2017-07-06 10:19*

The manuscript entitled "Pleistocene climate change and the formation of regional species pools" seeks to statistically identify distinct biogeographic regions within the Palearctic region, focusing on the genus *Carabus* (Coleoptera), in order to test which variables explain best the species turnover across these regions, and whether they fit with the hypothesis the Pleistocene climatic variation affected the pattern of diversity. The authors focus on four (actually three) explanatory variables: i) the ecological niche (climatic niche + habitat), ii) geographic connectivity among regions and iii) phylogenetic relationships. The study does not bring new data and rely on sequences publicly available and distribution data from previous publications of the same authors. Globally, while the Material and Method section and the Results are well written, clear and concise, I found that the Introduction and Discussion are much more difficult to follow, some ideas would require to be better explained, in particular when tightly related to the hypotheses (see below). I really like the approach chosen to identify the biogeographic regions. It gives a real statistical support for describing the

pattern of diversity. However, I have more questions about the method used for reconstructing the phylogeny and interpretation of the ancestral range estimation (see details below).

Thank you very much for your detailed revision and your accurate and constructive comments. We have reconstructed the *Carabus* phylogeny using more up-to-date methods and reinterpreted ancestral estimation results. Detailed comments are provided below.

Introduction L33. I find “determine” too strong. Also I am not sure about this sentence and the whole idea of this introductory paragraph. The authors cite for example Ricklefs & He (2016) as an example supporting the idea that “regional biota [...] determine ecological and evolutionary processes [...] at finer scale”. But the reference above actually supports the idea that the PROCESSES acting at large scale and structuring the regional biota (speciation, extinction, dispersion) can affect local processes. This may sound like a detail but otherwise I do not see how the “regions biota” per se can affect local processes. I also think this clarification is extremely important since the conclusion of the paper (the last sentence) directly refers to this question of PROCESSES acting at large scale and potentially indirectly affecting the local scale.

Thank you for this clarification. We have modified this sentence as suggested (lines 39-40).

L55-59. I find these two sentences contradictory. Do the authors mean that it is well known that species/population ranges are the result of Pleistocene climate but we do not know how general this is? These sentences should be clarified.

This was indeed the meaning of the sentence. We have reworded it to clarify that Pleistocene effects are well known in species ranges and local diversity patterns, but whether its signature scales up to a regional level is unknown.

L.67-82. This part deserves to be clarified. It gives the framework of the study, and many of the patterns tested are based on this, yet it is confusing. For example l. 67-70, I do not understand the last part of the sentence. I don't know if this may help but a way of presenting some of the processes would be to talk about the balance (or ratio) between speciation and dispersal/adaptation for a region. Variations in this ratio may be mediated by important barriers to dispersal (reducing dispersal) or climatic fluctuations (increasing dispersal) and lead to phylogenetic clustering or overdispersion.

Thanks for your useful suggestion. We have rewritten the paragraph including the balance idea.

L70-71. Throughout the manuscript the authors use “phylogenetically constraint”. To me this is an over-interpretation of the pattern called phylogenetic signal or phylogenetic conservatism (depending on the definition one wants to use). Using phylogenetic conservatism or signal would make things much more clear I think. Also, it is not the “species’ resemblance” that is conserved but the climatic niche, which leads to a resemblance among closely related species.

We definitively agree with these suggestions and have modified both terms throughout the text.

L85. The authors study the “European *Carabus*”. Does that mean there are non-european *Carabus*? In other words, is *Carabus* endemic to the Palearctic? If not where are the other *Carabus* in the phylogeny, and how this may affect the results (ancestral range estimation in particular) and interpretation?

This is a crucial point that was not explained in the previous version of the manuscript. We thank you very much for raising it. Some lineages of *Carabus* have species inhabiting Europe as well as Asia. This might certainly affect ancestral range estimation methods as they are intended to deal with complete-sampled phylogenies. However, the effects of incomplete lineage sampling should be like the ones produced by unnoticed extinctions (i.e. generally almost all) and those have been shown to be weak (Matzke 2014 *Systematic Biology* 63:51–70). Nevertheless, we are conscious that the actual effects of extinctions and incomplete sampling in ancestral estimations are still largely unknown. For this reason, we were cautious when interpreting these results, including also in the supporting information two complementary methods that do not violate any assumption. These methods corroborated the (perhaps more reliable) result from ancestral estimations, regarding the abrupt increase of marginal probabilities coinciding with the beginning of the Pleistocene. The

agreement among different methods provided confident support for the idea about the glaciation's signature on the lineage composition of Carabus faunas, validating as well the ancestral range estimation results. We have explained this in more detail (lines 378-393), moving also one of these complementary methods from the SI to the main text.

L99. "presumed low dispersal capacity": reference needed, especially since hypothesis H1 at least is based on this statement.

Done.

L103. Hypothesis H3. This is a question that came to me several times throughout the paper: do we need the Pleistocene climatic fluctuations to explain that lineages living at higher latitudes are adapted to cold environments? After all, temperatures decreasing toward the poles, one would expect to see lineages adapted to cold environments at higher latitudes compared to the Mediterranean basin. I think the authors should think of a way to distinguish between a "simple" latitudinal gradient effect and the Pleistocene climatic fluctuations.

We agree with this, although the hypothesis was referring to the greater influence of climate shaping species ranges in Northern Europe as compared to the south. We reworded this sentence to make this clear.

Material and methods L168-172. I wonder: the authors do not give the sign of the expected correlation. Here, the only hypothesis is a positive correlation between niche distance and phylogenetic distance (leading to a phylogenetic conservatism). However one may expect a negative correlation, i.e. closely related species have very dissimilar niches, a pattern one may expect for example in the case of adaptive radiation. Was the test designed to detect both negative and positive correlations?

Thank you for raising this interesting point. The test can certainly detect the sign of the correlation. We have included them in Table S4.4 although we did not discuss about this option since these correlations were not significant.

L195-196. The sentence need to be revised.

Done

L215. Should remind what a "node" is here.

Done

L239-242. The sentence need to be revised.

Done

L256. Is there a specific reason for using kernel smoother instead of "polygons"?

By using a kernel smoother we were trying to avoid sampling biases. However, given the nature of our data (range maps) its use was not justified. We have repeated the analyses without kernel smoothing, which, by the way, allowed us to include all species. Results remained mostly unaffected with the same interpretation.

L287-296. The section about the molecular phylogeny is extremely short and would benefit from adding more information here, instead of simply referring the supporting information. The authors unlink the substitution models in the RAXML across regions (genes I presume) – supposing that each gene follows a different substitution model. However they consider only one clock, estimated only based on the branch length of the RAXML tree and time constraints. I do not understand the choice of using the chonos function for time-calibrating the tree, i.e. using a method that does not even use the molecular matrix to estimate one clock, while this information is available. To me seems like this is under-exploiting the information available. Applying a molecular clock is not simply

rescaling a topology based on some constraints. I do not pretend that the results will drastically change but it seems to me like the method used here is quite weak, especially in the context of inferring historical biogeography, where the timing of events is particularly important.

Thank you very much for this comment. The method we were using was suboptimal as you and another reviewer stated (see below). We have repeated the phylogenetic reconstruction using BEAST and a Random Local Clock. Furthermore, by using a Bayesian approach we were able to consider phylogenetic uncertainties related to the molecular phylogenetic reconstruction. Finally, we also included two different time calibrations to account for uncertainties in *Carabus* origin (see below). After reanalyses using these new phylogenetic hypotheses (n=20,000), we obtained very congruent results to previous ones. That is, the same lack of correlation between phylogenetic distances and (sub)regional cooccurrence and a breakpoint in the relationship between ancestral estimation of marginal probabilities and node age coinciding with the beginning of the Pleistocene. This confirms the important role of Pleistocene glaciations, showing also that our results are robust against methodological uncertainties and inaccuracies. For the sake of simplicity, we removed results based on RAXML trees.

L287. Phylogenetic distances were computed from the time-calibrated tree?

Yes. We have clarified this in the text.

L323. Ancestral range estimation: Have the authors modified the adjacency matrix or multipliers? Is there a reason for not reducing the adjacent areas?

The main reason for not modifying the multipliers is that the history of Europe is too complicated as to infer past dispersal probabilities among regions. For instance, during the Messinian the Mediterranean Sea went desiccated, potentially connecting southern Mediterranean Peninsulas. Yet, after desiccation a salty desert remained which might have precluded dispersal events among these regions. On the other hand, the repeated glaciations during the Pleistocene produced decreases in sea and ocean levels. This might have also opened land bridges between European islands and peninsulas. Although their duration and specially their suitability to allow the dispersion of fauna is largely unknown. All these events and their unknown consequences on dispersal processes persuaded us to not modify the adjacency matrix (even in a time-stratified way).

Results L414. Revise the supporting information numbering.

Done.

L416. I do not understand which information allows the authors to conclude that colonization occurred "recently". Because there is no phylogenetic signal?

Yes. Theoretically, the lack of both geographical (i.e. connectivity) and phylogenetic signals should be expected under recent colonisations, where species should be mostly sorted by their climatic tolerances. That is, if colonisations are old enough as to allow speciation processes some phylogenetic signal might be expected. We have clarified this in the text.

L420. In 59% of cases BAYAREA+j has the lowest AIC. But I am curious, what is the mean AIC difference with the second best model?

Thank you for noticing this. We have included a new column on Table 1 showing averaged AICc weights (which are more directly comparable among different models). The number of cases in which BAYAREA+j model had the lowest AICc is very similar to average AICc weights. This indicates that BAYAREA+j was by far the most probable model when it had the lowest AICc.

L422-425. The authors discuss the parameter j. However they do not providing any discussion about the fact that BAYAREA is chosen over other models such as DEC. BAYAREA models speciation event in which both descendant lineages inherit of the entire ancestral range, leading to speciation events that occurred over the entire distribution of the ancestral lineages. How does such model fit when interpreting the results?

This is a very good question that might be potentially interpreted in the light of Pleistocene glaciations, especially attending to the higher contraction (or extinction) parameter in BAYAREA than DEC. That is, the mixing of lineages apparently produced by glaciations might promote that probabilistic ancestral estimation finds as more likely this “widespread sympatry” and subsequent range contractions. Nevertheless, we believe that this interpretation is quite speculative given that some model assumptions were not fulfilled and especially due to the low predictive power of ancestral estimations. Hence, we preferred to not include it in the text and focused in results that were corroborated by other methods as explained above.

L419. Concerning the ancestral range estimation: I still think that a short description of the pattern is necessary. I understand that the authors are interested in assessing the role of recent climatic event, but the group is more than 20 my old (according to the phylogeny). I think the authors cannot really ignore the 20 my of diversification before the Pleistocene events.

We agree that 20 My of diversification should not be ignored. Indeed, interpreting older diversifications is both interesting and tempting. However, given the low marginal probabilities of nodes before the Pleistocene it is quite difficult to make solid statements on the potential distributions of ancestral species. We believe, however, that this was a very interesting result by itself, indicating that the spatial signature of older diversifications is eroded, rather than simply ignoring 20 My of evolutionary history.

And I would like to see the distribution of species at the tips of the phylogeny, otherwise this information is not available and it will also help better understanding the regional structure. Should we also observe a directional pattern of dispersal events if Pleistocene climatic fluctuations affected speciation and dispersal? Shouldn't we observe recent northward colonizations? Do we?

It is important to note (as another reviewer pointed out, see below) that the phylogeny presented in figure 4 is one out of 20,000 (based on new analysis). Hence, by plotting and interpreting tip and ancestral estimated distributions of just one hypothesis some mistakes might be expected. To be conservative and avoid this potential source of error we removed the distribution of internal nodes from figure 4, which now focuses only in marginal probabilities. In any case, as far as we know ancestral estimation methods could only show northward colonisations in the case that some species diversified in the north. This should be nonetheless quite unlikely since northern regions were glaciated until very recently (less than 20,000 years ago).

Figure 4. Based on this time-calibrated tree, there seems to be parts of the tree that were clearly not affected (at least in terms of speciation) by Pleistocene climate changes. For example the clade including Tomocarabus, Diocarabus, Orinocarabus, Eurycarabus, Aulonocarabus, Pachystus (which is inferred to have occurred in the northern part for a long time) or the clade Archicarabus. Are there any hypotheses that may explain why some lineages have been more affected by Pleistocene climatic events than others?

This is a very interesting question. Unfortunately, although there is a fair knowledge of the natural history of the group, knowledge on the specific details of the different subgenera is still insufficient so as to allow developing plausible hypotheses on how glaciations may have affected them differently. Perhaps different dispersal abilities, climatic tolerances or simply historical contingencies might be behind different responses to extreme climatic changes, but besides the coarse-grain habitat preferences we have included with the help of Achille Casale, there is nothing apparent in the morphological differences between lineages that may provide a clue on how they were selected during glaciations, at least to our knowledge. This surely deserves future investigations. We have included this in the discussion (lines 540-541).

Discussion L451. Sentence needs to be revised.

Done

L451-452. Could be interesting to give an example of how local diversity is affecting to clearly contrast with the result of this paper

Sorry, we were not able to completely understand this suggestion.

L460-462. Or simply a lesser geographic complexity?

We included it.

L480-482. Revise the sentence, which in itself cannot be understood.

The sentence was perhaps too speculative, so we decided to remove it.

L490-491. Could this also be a consequence of the lower number of species in northern regions?

Theoretically, the number of observations should not affect the strength of a correlation so we do not think so.

L502. Then vicariant events should be recent; otherwise local diversification would have led to phylogenetic clustering.

Agree. Indeed, we were trying to explain that if speciation is mostly driven by vicariant processes a mixing of lineages should be expected. We reworded the sentence.

L516-519. Then where are these founder speciation events in the ancestral state estimation? This statement could be supported by some examples of such events (or event show it on the figure).

As explained above given the low marginal probabilities a detailed interpretation of ancestral estimation results is risky.

L522-254. It seems like the authors are still discussing about the +j, but I do not see how this sentence related to the +j parameter. Some rewording is needed.

Done.

*Reviewed by anonymous reviewer, 2017-07-12 10:29*

The paper aims at depicting the processes that contribute to the formation of regional species pools. Using Carabus as a biological model, the authors used a network approach to characterise regional and subregional faunas across Europe. Similarities in species co-occurrence within these faunas were then modelled as a function of climatic niche and habitat similarities, spatial connectivity and phylogenetic distances. I'm not familiar with the analyses used in the paper, but I found the issue very interesting. I think that such studies should enable a better understanding of the origins and dynamics of regional biotas. My main concern was for the structure of the paper. The authors made the effort to define specific hypotheses (H1-H6) in the introduction, which they did not refer hereinafter, especially in the discussion.

We are very thankful for the positive comments and the suggested changes. We have incorporated direct references to the proposed hypotheses in both results and discussion sections.

Moreover, the material and method section is very long. The Data origin section should be embedded in the other subsections for easy readability. If the structure of the paper is revised, I think that it deserves the label of PCI Evol Biol.

Thanks for this comment. We agree that the material and method section was quite long. We have reduced it as much as possible, imbedding also the information on the data within other subsections.

Minor comments l.69-82: the link between the text and Fig. 1a is not obvious, because the terms used were not the same. Please clarify, or delete reference to Fig. 1a. l.

Done.

128: I have not found the map at <http://dataservice.eea.europa.eu> I.

The site seems to have discontinued some information, in this perhaps because it moved to the INSPIRE website for spatial information in Europe, that we now include in the text. The LAEA grid we used as a base for the information is the grid recommended by the relevant INSPIRE working groups. More detailed specifications can be found at <https://inspire.ec.europa.eu/id/document/tg/gg>.

134-142: the relevance of the environmental and geographical data concerning Carabus could be explained in this paragraph

Done

I. 165: "lineages" I.

Done

175: "Further, the independent effects...": why "independent"?

This was certainly not clear. We meant the effects of climate and geography alone (i.e without phylogenetic effects). The sentence is reworded now.

I. 177: " to geographic isolation (Fig. 1a.iii), whereas..."

Done

I. 261-269 were vegetation categories defined at the scale of each grid cell, or more finely?

No, they were defined more finely (at the resolution provided in the data, 5 minute). This is now explained in the text (line 248). Thanks for noticing that.

I. 273: "pixel" = grid cell?

Thanks for noticing this as well. The cost surface was calculated at 1 km<sup>2</sup>. This is also explained in the text.

I. 369-386: what about the geographical consistency of submodules?

The p values supporting submodules are provided in line 405.

Please choose between the terms "region/subregion" and "module/submodule" throughout the text to facilitate reading.

Done.

I. 398-400: spatial connectivity was not significantly related to species co-occurrence in Modules 5 and 6.

In this sentence we referred to the relationship between connectivity and submodule co-occurrence across the entire study area. We modified the sentence to clarify this.

I. 408-409: "The effects of connectivity were stronger in southern regions (i.e., modules 1 and 3)": in module 3 it was not stronger than in modules 4 and 7.

True. We corrected this.

I. 412-414, "Indeed, niche similarity and spatial connectivity...": this sentence is confusing and may be deleted.

Done.

I. 414: “Only niche similarities (mainly climate)”: were all the significant effects of niche similarities among regions and subregions related to climate?

In general, both climatic and forest similarities were significant (but in module 6 where only climatic similarities showed significance). This information is now provided in table S4.4.

I. 415-417: “ as expected if...”: for instance here, the authors should refer more clearly to hypothesis H3.

Done.

I. 424: “which could be seen as a signal of Pleistocene glaciations...”: or of more ancient events?

True, although it is difficult to identify other more ancient events that may have generated island-like areas in Europe.

I. 426 “the range contraction parameter”: named e (extinction) in the Table legend?

Yes. We included this explanation.

I. 489: replace “stronger related” by “more strongly related”?

Done.

*Reviewed by anonymous reviewer, 2017-07-27 19:24*

In this study, Catalayud et al. look into how species pools of the genus *Carabus* are formed in response to Pleistocene climate change. I find the study interesting but I have some issues with parts of the methodology.

Thank you for your time revising our manuscript and insightful comments.

I am assuming the methods used to detect species pools and assess the determinants of species pools are sound. I have no expertise in these methods and the section describing them is a little tedious to read and understand.

We have reduced this subsection, clarifying some points. We believe that now it is easier to read and understand.

The matrix design seems a little odd to me. I understand how the matrices are designed but I find it hard for example to understand why the connectivity matrix is designed based on solely topography and presence of water bodies. The slope and the presence of water bodies are not necessarily the most crucial factors determining *Carabus* beetle dispersal. A more inclusive view would be to take into account ecosystem/niche change/breaks. At the moment, this is decoupled in the different matrices. It is not clear to me if this makes sense in a biological way. Knowing the biology and distribution of *Carabus* in Western Europe, it is not clear to me why the authors chose the slope and presence of water bodies as proxies for barriers limiting dispersal.

We agree that *Carabus* dispersion is also constrained by environmental factors. However, as you also noticed, this was already accounted for by the niche similarity matrix. That is, dispersal constraints are in fact included in both topography and climate-based matrices, which arguably correspond to –necessarily simple– representations of the two factors described in our introduction, and depicted in our Figure 1. In this figure, processes i and iii would be climate-driven, and processes ii and iv would be geophysically-driven. This decoupling is fundamental to separate between purely physical and niche constraints, which was one of the main motivations of the study.

That is, rather than trying to capture the effects all dispersal barriers in conjunction, we were trying to explore the contribution of different processes in the configuration of *Carabus* faunas.

Are the 16S alignments with Kalign and MUSCLE strictly identical? If not I don't see why the authors mention the use of a criterion to select their alignments under different programs. They seem to select a better alignment in MUSCLE but decide to take a suboptimal one based on the criterion that they decided to use in the first place.

That is true. We have now used the best alignment for each sequence.

The phylogenetic and dating analyses could be improved. The dataset was not partitioned and was analyzed as a whole. This is most likely suboptimal. Furthermore, the backbone of the topology was constrained to reflect the one of Deuve et al. (2012). I don't see a good rationale to do so, especially considering the rather low nodal support across Deuve et al.'s topology in the first place. The dating is equally odd, with no less than 21 secondary calibrations used to date the resulting RAxML topology. I am wondering why the authors did not directly use the chronogram from Deuve et al. and pruned the tips they were not interested in.

Thanks again for raising this concern. It should be noticed that the main goal of the study was not to provide a new phylogenetic hypothesis for the group, but to generate a reliable set of working phylogenetic hypotheses to test other hypotheses. That was the main reason to follow Deuve et al.'s phylogeny (which is the most recent and complete). That said, the phylogenetic reconstruction and calibration we used could clearly be improved, as another reviewer also noticed (see above). We have repeated this analysis using BEAST, with no *a priori* constraints and a Random Local Clock for calibration. This, in addition, allowed us to consider topological and calibration uncertainties. After using a total of 20,000 trees in subsequent analyses, our previous results and interpretations remained the same, confirming the important signature of Pleistocene glaciation (see also below).

It should be added that the dating of *Carabus* origin is debated even though the literature shows at the moment a unique hypothesis (e.g. several studies by Andujar and colleagues). The latest dating of the beetle-tree-of-life (McKenna et al. 2015; Toussaint et al. 2017) clearly show that Carabinae split from Trachypachidae >150 million years ago. Based on this extremely old divergence, it seems unreasonable to assume an age for the crown of Carabinae at <50 Ma. Most dating studies focusing on Cychrini or Carabini have been based on substitution rates of mitochondrial genes or on biogeographical constraints which are well-known to potentially result in dramatically underestimated ages. This should be kept in mind when analyzing divergence times in *Carabus* and other related lineages. In their results and discussion, the authors mention the perfect fit of their dating and BioGeoBEARS analyses with the Pleistocene glaciations, emphasizing confidence intervals of tens of thousands of years. This is likely not conservative and could be highly biased considering how shaky the divergence times estimates within *Carabus* are likely to be.

This is a crucial point that was ignored in previous analyses. Thank you very much for raising it. To account for uncertainties in the origin of the group we have calibrated the tree using two different crown ages for *Carabus*. The first one is based on Deuve et al.'s (Mol Phylogenet Evol. 2012, 65:259-75) estimation, who found the origin of the group dating at 17.3 Mya (95% HDP: 12.8–23.3). The second one is much older, and it is based on Andujar et al.'s (BMC Evol Biol. 2012 28:12:40) estimation (25.16 Mya; 95% HDP: 18.41-33.04). Notice that using this second calibration, we found that the split between Cychrini or Carabini dates at 36.8 Mya (95% HDP: 30.2-43.2), which is more congruent with the dates you commented. Finally, by using a Bayesian approach we were able to consider calibration uncertainties (i.e. by using several trees sampled from the posterior distribution rather than just a consensus one, see also Appendix S1). Despite using two contrasting calibration times and incorporating uncertainties in the subsequent analyses, we found strongly congruent results supporting the role of Pleistocene climatic fluctuations. Indeed, estimated marginal probabilities show a steeper increase at 2.16 Mya. (45<sup>th</sup> and 55<sup>th</sup> percentiles at 1.68 and 2.94 Mya) using Deuve et al.'s crown age and at 2.28 Mya (45<sup>th</sup> and 55<sup>th</sup> percentiles at 1.78 and 3.19) using Andujar et al.'s one. This provided strong support for the effects of glaciations (which start at 2.56 Mya) on the phylogenetic configuration of *Carabus* faunas, being robust against different divergence times and calibration uncertainties.

The placement of taxa based on "taxonomic knowledge" is somewhat dubious and I am having a hard time understanding it. I am not really convinced by the method developed in Rangel et al. (2015). The phylogeny of *Carabus* is complex and the morphological affinity among species of

certain groups has been largely challenged by molecular data in the past decade. Here, the authors place >30% of the species they study in clades of the (moderately to poorly supported) tree based on “taxonomic knowledge”. This is not really appropriate in the absence of proper morphological justification, and all analyses based on the resulting topology are likely to be severely biased.

We agree that uncertainty methods related to the lack of molecular markers are not ideal. However, no method is totally ideal, (even purely molecular reconstructions). Rangel et al.’s uncertainty approach is the best we have found so far, for it allows including a full account of species, thus using information that otherwise would be wasted. We are aware of the complexity of *Carabus* morphological taxonomy and the difficulties to place species based on taxonomical knowledge. Indeed, we asked for help to a reputed expert on the taxonomy of the group (Achille Casale), who reviewed an early version of the tree with taxonomical insertions. This gave us some confidence on the insertions, which were nevertheless very conservative. Here, being conservative means that species were associated to their most likely ancestors. That is, in case of doubt on the taxonomic affiliation of a species (for instance, at the subgenus level) we placed it in higher taxonomic levels so that no doubts remained. It should be noticed that after a species is safely associated to a phylogenetic node, it is then randomly placed within descendant nodes. This process is iterated several times (in our case 100 times for each tree) giving a large space of plausible topologies. These different trees are used in subsequent analyses, using averaged results among them. It could be argued that by being conservative we might incur in Type II errors (i.e. results of the “true” phylogeny might be masked by averaging its results with those of a large number of “false” phylogenies). However, our results were very consistent across all the phylogenetic hypotheses used. For instance, none of the trees (out of 20,000, based on new analyses) showed significance in the relationship between (sub)regional co-occurrence and phylogenetic distances. This shows that, rather than severely biased, our results are quite robust, even when taking into account phylogenetic uncertainties of any type.

Overall, I feel that the topic of this study is important. Some results are very interesting, such as the grouping of 7 zoogeographical regions and the importance of niche similarity and spatial connectivity in shaping the co-occurrence of species. The different impact of factors in Northern and Southern regions is also very interesting. But I also find that while the authors use quite complicated methods, they pay less attention, to some extent, to the crucial steps required to build either their models (matrices) or datasets (phylogenetic tree and chronogram). They also barely mention any bias or a more balanced view that could originate from their results. For instance, the result that phylogenetic distances were not correlated with regional co-occurrence is likely biased by the suboptimal methodology used to build a chronogram in this study. Similarly, the authors are quite speculative in their discussion but seem to systematically not report any lack of statistical support or confidence intervals. For instance, they disregard the overall complete lack of nodal support in their phylogenetic tree (this is not even mentioned in the text, and not indicated in Fig. 4). The dating analyses are not discussed and there is not confidence interval provided. The likelihood of the BioGeoBEARS analyses is not given. The pattern presented in Fig 4 is one among 100 and has almost no resolution, this might be worth discussing. A final note is the absence of discussion in the context of existing literature regarding Pleistocene range contractions/shift, lineage diversification and faunal re-assembly. There is a large amount of studies looking into this topic using different methodologies, mostly derived from population genomics or phylogeography. This input would be interesting to underscore the importance of these new results in the field.

We thank the reviewer for this critical view of our previous manuscript, which may have been too succinct to provide a fair account of the results. We believe that the current version solves all these concerns. The separation of the matrices into geographical costs and environmental distances was completely necessary to answer our questions. The phylogenetic reconstruction and calibration is now much more solid, providing same results. Even though phylogenetic uncertainties related with molecular reconstruction were considered, new resultant trees showed considerably higher node support (see Figures S3.2 and S3.3). The likelihood of BioGeoBEARS analyses is now provided in Table 1. We clarified also that the phylogeny showed in figure 4 is just an example. Finally, we enlarged the discussion in some parts including evidence from phylogeographic studies, as well as aspects related to potential biases.