



Peer Community In Evolutionary Biology

The colonization history of largely isolated habitats

Andrea S. Meseguer based on peer reviews by **Simon Joly**, **Florian Boucher** and 2 anonymous reviewers

Michael D Pirie, Martha Kandziora, Nicolai M Nuerk, Nicholas C Le Maitre, Ana Laura Mugrabi de Kuppler, Berit Gehrke, Edward GH Oliver, Dirk U Bellstedt (2019) Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotropics by *Erica*. bioRxiv, ver. 1, peer-reviewed and recommended by Peer Community in Evolutionary Biology. [10.1101/290791](https://doi.org/10.1101/290791)

Submitted: 09 April 2018, Recommended: 02 January 2019

Cite this recommendation as:

Meseguer, A. (2019) The colonization history of largely isolated habitats. *Peer Community in Evolutionary Biology*, 100065. [10.24072/pci.evolbiol.100065](https://doi.org/10.24072/pci.evolbiol.100065)

Published: 02 January 2019

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

The build-up of biodiversity is the result of in situ speciation and immigration, with the interplay between geographical distance and ecological suitability determining the probability of an organism to establish in a new area. The relative contribution of these factors have long interested biogeographers, in particular to explain the distribution of organisms adapted to habitats that remained largely isolated, such as the colonization of oceanic islands or land waters. The focus of this study is the formation of the afrotropics flora; patches of temperate vegetation separated by thousands of kilometers in Africa, with high levels of endemism described in the Cape region, the Drakensberg range and the high mountains of tropical east Africa [1]. The floristic affinities between these centers of endemism have frequently been explored but the origin of many afrotropics lineages remains enigmatic [2]. To identify the biogeographic history and drivers of biogeographic movements of the large afrotropics genus *Erica*, the study of Pirie and colleagues [3] develops a robust hypothesis-testing approach relying on historical biogeographic models, phylogenetic and species occurrence data. Specifically, the authors test the directionality of migrations through Africa and address the general question on whether geographic proximity or climatic niche similarity constrained the colonization of the Afrotropics by *Erica*. They found that the distribution of *Erica* species in Africa is the result of infrequent colonization events and that both geographic proximity and niche similarity limited geographic movements (with the model that incorporates both factors fitting the data better than null models). Unfortunately, the correlation between geographic and environmental distances found in this study limited the potential evaluation of their roles individually. They also found that species of *Erica* have dispersed from Europe to African regions, with the Drakensberg Mountains representing a colonization sink, rather than acting as a “stepping stone” between

the Cape and Tropical African regions. Advances in historical biogeography have been recently questioned by the difficulty to compare biogeographic models emphasizing long distance dispersal (DEC+J) versus vicariance (DEC) using statistical methods, such as AIC, as well as by questioning the own performance of DEC+J models [4]. Behind Pirie et al. main conclusions prevails the assumption that patterns of concerted long distance dispersal are more realistic than vicariance scenarios, such that a widespread afrotemperate flora that receded with climatic changes never existed. Pirie et al. do not explicitly test for this scenario based on the idea that these habitats remained largely isolated over time and our current knowledge on African paleoclimates and vegetation, emphasizing the value of arguments based on empirical (biological, geographic) considerations in model comparisons. I, however, appreciate from this study that the results of the biogeographic models emphasizing long distance dispersal, vicariance, and the unconstrained models are congruent with each other and presented together. Pirie and colleagues [3] bring a nice study on the importance of long distance dispersal and biome shift in structuring the regional floras of Africa. They evidence outstanding examples of radiations in *Erica* resulting from single dispersal events over long distances and between ecologically dissimilar areas, which highlight the importance of niche evolution and biome shifts in the assembly of diversity. Although we still face important limitations in data availability and model realism, the last decade has witnessed an improvement of our understanding of how historical and environmental triggers are intertwined on shaping biological diversity. I found Pirie et al.'s approach (and analytical framework) very stimulating and hope that will help movement in that direction, providing interesting perspectives for future investigations of other regions.

References:

- [1] Linder, H.P. 1990. On the relationship between the vegetation and floras of the Afrotropical and the Cape regions of Africa. *Mitteilungen aus dem Institut für Allgemeine Botanik Hamburg* 23b:777–790.
- [2] Galley, C., Bytebier, B., Bellstedt, D. U., & Peter Linder, H. (2006). The Cape element in the Afrotropical flora: from Cape to Cairo?. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609), 535-543. doi: [10.1098/rspb.2006.0046](<https://dx.doi.org/10.1098/rspb.2006.0046>)
- [3] Pirie, M. D., Kandziora, M., Nuerk, N. M., Le Maitre, N. C., de Kuppler, A. L. M., Gehrke, B., Oliver, E. G. H., & Bellstedt, D. U. (2018). Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotropical by *Erica*. *bioRxiv*, 290791. ver. 5 peer-reviewed and recommended by *PCI Evol Biol*. doi: [10.1101/290791](<https://dx.doi.org/10.1101/290791>)
- [4] Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography*, 45(4), 741-749. doi: [10.1111/jbi.13173](<https://dx.doi.org/10.1111/jbi.13173>)

Reviews

Evaluation round #3

DOI or URL of the preprint: [10.1101/290791](https://doi.org/10.1101/290791)

Version of the preprint: 3

Authors' reply, 14 December 2018

Dear *PCI Evol Biol*,

Thanks again for your time spent on our preprint. The new version is now live, and I am again uploading responses and a tracked-changes version to indicate how we have used the last set of very helpful comments.

On behalf of the authors, with best wishes, Mike Pirie

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Andrea S. Meseguer](#), posted 12 December 2018

Needs a revision

I have decided not sending the paper again to reviewers in order to speed up the process, but a few important modifications are still necessary before the paper could be recommended.

Regarding the biogeographic models: I found satisfying that the results of the unconstrained model are presented together with the results in which jump dispersal was forced to occur between Europe and the Cape region. Although I am still sceptical with this assumption, I agree this approach could be valid if we assume that widespread ancestral reconstructions are biologically unlikely. I regret, however, the authors don't provide a more detailed justification of this assumption; if they argue in the introduction that ancestral continuous (widespread) distributions fragmented by vicariance probably explain disjoint patterns of (arid-adapted) African plant groups (e.g., Sanmartín et al., 2010; Pokorný et al., 2015; Bellstedt et al., 2012), why in this particular case this biogeographic scenario (an ancestral widespread distribution) does not make biological sense? I imagine it has something to do with the temperate adaptations of the genus and the fact that temperate conditions were never continuous in Africa, but this needs to be clarified (maybe in Appendix 3 or in the main text?).

Reviewer 3 was concerned for the potential correlation of the studied variables (i.e. geographic distances and niche similarity). In this new version, the authors evaluated their correlation and found what is generally considered a moderate to strong value of correlation (Kendall's R = -0.64), but was this result taken into account at all? It has strong implications, as one of the main questions of the study (i.e. the relative importance of niche vs. distance in dispersal patterns) cannot be answered. I am convinced this limitation does not affect the main conclusions of the study since the preferred model was the combined "niche/distance" model. In addition, this is an interesting result "per se". However, in light of this, I have the impression that comparisons of the "geographic distance" and the "niche similarity" models do not make sense anymore because these variables are not independent. Only comparisons of the null and the combined "niche/distance" model might remain informative. I leave to the authors the decision on whether they want to exclude these models from the model comparison table or not, but this limitation needs to be taken into account and the results/discussion sections modified accordingly.

Again, I hope you will find these last comments helpful and look forward to read the final version of the manuscript!

Best Andrea [Download recommender's annotations](#)

Evaluation round #2

DOI or URL of the preprint: <https://doi.org/10.1101/290791>

Version of the preprint: 2

Authors' reply, 04 December 2018

Dear PCI Evol Biol,

Thanks again for the valuable input. I am including a cover letter responding to each of the comments and an annotated tracked-changes version of the new preprint text.

All the best, Mike Pirie

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Andrea S. Meseguer](#), posted 25 November 2018

Needs a revision

This paper has improved from the last version and the authors have taken into consideration and answered many of the concerns raised in the initial review. Reviewer 1 believes it is in good shape for recommendation after a few minor revisions. Reviewer 2 did not see the original, but only the revised version and agrees with Rev 1 that the paper is of general interest. However, Rev 2 has some substantive comments that need to be dealt with before the paper can be accepted for recommendation. The most critical comments centre on the niche modelling approach and how this affects your results. Rev 2 has a number of other critical but supportive and constructive comments that will make the paper more useful for the biogeographic community.

The authors' response to my critique of their use of the DEC+J models is satisfying. I agree that this ms is not the place to go into the detail of this argumentation, although I appreciate they acknowledge the controversy and provide the DEC results. Just as a note of caution, the recent paper of Ree & Sanmartin (2018) shows that DEC and DEC+J models are not directly comparable using statistical methods such as AIC that assume probabilistic equivalency of events. Therefore, I recommend privileging arguments based on empirical (biological, geographic) considerations rather than AIC scores. AIC comparisons might otherwise work to compare DEC (and DEC+J) models among them.

I regret however, the authors have not fully taken into account my comments about the adjacency matrix. I apologize if this is because I was not clear enough. The "Cape to Cairo" hypothesis (dispersal from Europe to the Cape, followed by migrations to the Drakensberg; Figure 1), could not be properly tested if the adjacency matrix does not allow ancestral connections between Europe and the Cape regions (0 values on this connection). By setting 0 values you are specifying that these two areas were not connected in the past (an ancestral distribution in Europe and the Cape is disallowed). Hence, although in theory dispersal is still allowed, in practice this constraint forces species to follow another route to disperse between these areas or forces jump dispersal. In my opinion this is a too strong assumption when the "Cape to Cairo" is the hypothesis to test, specially given the possibility that before the Miocene aridification of northern Africa, Erica was distributed in regions where it does not occur today. This limitation seems to apply only to DEC+J and not to DEC analyses where the adjacency matrix was not implemented. I suggest testing again the hypotheses without this constriction, which will make the results more solid.

I also have a few minor points: - The biogeographic reconstructions on Appendix 13 don't have any legend. You need to explain the symbols on the trees, the colours, the abbreviations or the analyses that have produced these figures. In addition, the size of the pie charts on the trees is way too big and needs to be reduced. - The unconstrained biogeographic DEC and DEC+J reconstructions need to be presented the Appendix. - Please, clarify if the "Cape to Cairo" model fits the data better than "Southerly stepping stone" model. Discussion lines 285-287 state: "Cape to Cairo" and "Drakensberg melting-pot" mostly fit the data better than "Southerly stepping stone". Meanwhile, in the results: "Under DEC+J given the best tree, the Drakensberg melting pot, geographic distance, and southerly stepping stone models revealed the lowest AIC; under DEC the Drakensberg melting pot model alone scored best, but with higher AIC ". Moreover, in Table 1, the CtoC model does not even appear among the best models when considering the best tree.

I hope that the authors will find these comments helpful to improve their work.

Yours sincerely,

Andrea S. Meseguer

Reviewed by [Florian Boucher](#), 05 October 2018

[Download the review](#)

Reviewed by anonymous reviewer 2, 05 October 2018

General comments

(1) English sounds good. However, I am not a native English speaker and then I am unable to judge the quality of English language.

(2) I am not a specialist of biogeographic models and then I am unable to give a strong opinion on biogeographic inferences and model selection.

(3) As my specialty is niche comparison among species, I will overall focus my comments on this particular aspect of the MS.

Major comments

-L282: Authors depicted the biogeographic regions by using an arbitrarily set buffer of c. 1° lat/lon in radius around presence records.

I am afraid that a 1° radius buffer will introduce erroneous/unsuitable climate conditions for *Erica* species in analyses. It is worth to calculate niche dimensions using only real presence records ...My biggest concern is about the arbitrary selected 1°C buffer. It could have negative impact on inferences. This is particularly true for species occurring in tropical mountains where 1°C radius could mean an abrupt change in climate conditions. I invite the authors to run again models with true presence records only.

-Are niche similarity/dissimilarity proxies (i.e. Schoener's D values) correlated to geographic distances? If there is correlation, it could impact inferences and model selection procedure. Collinearity is always a problem in modelling and the absence/presence of collinearity between both descriptors should be addressed in the MS. The presence of colinearity could make tricky results interpretation.

- Authors could be a bit more careful when interpreting realized niche differences. Actually I would be happy to see at some point in the discussion a statement of differences between the realized and fundamental niche and implications of these concepts on the significance of their results. As authors use the term 'niche' without discriminating the fundamental and the realized niche, the MS is sometimes confusing.

Minor comments

L51-53: Authors say 'The distribution pattern of the more than 800 *Erica* species across Europe and Africa provides an opportunity to disentangle the effects of geographical and ecological distance on biogeographic history'

I am not sure that this statement is pertinent. I would say that combination of genetics, distribution data and fossil information could help to disentangle....

L82-83: Authors say : 'Nonetheless, similar distribution patterns across Europe and Africa are observed in different plant groups.'

References for that statement ?

L84-85: Authors say :Organisms adapted to different habitats respond differently to changing environmental conditions (Mairal, Sanmartín & Pellissier, 2017; Chala et al., 2017).

In my opinion, this statement is obvious and thus useless. Could be removed from the MS.

L604: Authors say : 'In this study, we modelled shifts between biomes and dispersals over larger distances in the evolution of *Erica*, in order to test six hypotheses for the origins of Afrotemperate plant groups (Fig. 1).'

Can we use the term biome here? I am not sure.

L643: Authors say : The dispersals to Tropical Africa and to Madagascar both involved large shifts in niche (Schoener's D of 0.298 and 0.274 respectively) .

Too speculative in my opinion. Authors use D Schoener index as a proxy for niche similarity. I would like authors to moderate the reliability of this index. In my opinion, strong differences in D schoener values among clades could just reflect very slight fundamental niche differences (or perhaps no differences at all) since all *Erica* species are adapted to temperate climates. I am pretty sure that niche similarity tests according Broenimann (2012) would indicate that all *Erica* lineages have climatic niche more similar than expected at

random. Perhaps these realized niche differences could not reflect real differences in their fundamental niche. Again, the term 'niche' is used without accounting for potential differences between realized and fundamental niches.

Evaluation round #1

DOI or URL of the preprint: [10.1101/290791](https://doi.org/10.1101/290791)

Version of the preprint: 1

Authors' reply, 19 September 2018

Thanks for your patience - please find responses and a tracked changes version of the text attached; bioRxiv of the new version 2 appears to be live. We look forward to your assessment! Mike Pirie

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Andrea S. Meseguer](#), posted 19 September 2018

Revise

All the reviewers now chimed in with their opinions and agreed on the interests of the manuscript. I commend the authors for all the work they have done, and I echo the reviewers. I think this study represents a nice piece of work investigating the factors and processes mediating dispersal of plant clades from Europe to Africa. I very much appreciated the thoroughness of the analyses and the clarity of the text.

I recommend the authors undertake a thorough revision based on the constructive comments of the reviewers, taking particular care to address the reviewer's methodological concerns. I found particularly interesting the criticism of one reviewer regarding the conceptual problem of proposing hypotheses that are not mutually exclusive (or framed at different levels), as well as all methodological suggestions proposed by the three reviewers.

I am, however, a bit more cautious than the reviewers regarding the results of this study. In agreement with one reviewer, I regret the choice of biogeographic models. The DEC+J model has been shown at best to not be directly comparable to the DEC model, and at worst to present some statistical problems (Ree & Sanmartin 2018). I would recommend excluding this model from the comparisons and if the authors still wanting to present it, to do it on supplementary material. I am also suspicious with the DEC+* model of Massana & al. (2015). This model has been published as a preprint with no reviewer assessment on their quality/performance. I thus suggest excluding this model from the comparisons as well.

In addition, I have concerns on the validity of the results. In addition to the different dispersal matrices, you implemented an adjacent area matrix to constraint the maximum number of areas allowed as ancestral states. I agree on this procedure to decrease model uncertainty, for example as you did by using the "maximum number of areas" command. However, I think the constrictions on the adjacent matrix you implemented are problematic and could have affected the results: 1) my apologies if I'm wrong, but I think that by using this adjacency matrix you force stepping stone dispersal to occur, while this is one of the main hypotheses you want to test. For example, in the matrix on Appendix 3 you are impeding dispersal between Europe and the Cape region, the Drakensberg and Madagascar, while only allowing dispersal through tropical Africa. It is thus not surprising that you found support for the "Drakensberg melting pot" stepping stone scenario in comparison with other long distance dispersal models. 2) In addition, by implementing this matrix you automatically disallow disjoint distributions at ancestral nodes, decreasing the likelihood of extinction to occur.

An important part of the manuscript focuses on whether colonization was mediated by niche changes or occurred across similar habitats. On this regard one reviewer had concerns about the areas used for

comparison. I agree with him that differences between study areas could disappear when compared with other regions in Africa or Europe where *Erica* does not occur. I additionally regret you didn't differentiate the Northern Hemisphere Mediterranean region from other Northern Hemisphere regions. My guess is that the low climatic similarity between southern African and Europe might most likely apply to the Eurosiberian region, but not to the Mediterranean one. I think this differentiation is important to test if long distance dispersals involved niche shifts.

Concerning model comparisons, in addition to AIC scores (e.g. on Appendix 8), I would like to see the differences in deltaAIC values, akaike weights or any other metric that allows to evaluate model improvements and perform model choice. Generally, it is the differences between the likelihoods or AICs that matter, not their absolute values. That is, the larger difference in AIC indicates stronger evidence for one model over the other (Burnham and Anderson, 2002). Delta (AIC differences) within 0-2 has a substantial support for a suboptimal model; delta within 4-7 considerably less support and delta greater than 10 essentially no support.

To conclude, apart from the nexus file I would suggest the author to include a figure with the most likely biogeographic reconstruction, maybe on SI.

I leave this, and the reviewer's comments, for the authors to consider as they revise and improve the paper. I hope the authors will find that many of these will be helpful in improving the manuscript

Reviewed by anonymous reviewer 1, 24 April 2018

This is an interesting manuscript exploring the interactions between geographical distance and ecological niche using the genus *Erica* as a model. The manuscript uses species occurrences and model testing to explore different biogeographic hypotheses. Although the manuscript does not introduce a novel idea, in general it provides new evidences regarding the colonization of new areas with subsequent niche change, and brings new evidences in terms of historical African biogeography and the *Erica* genus. However, in my opinion, the manuscript would benefit by making some clarifications, highlighting better the hypotheses and the general argument. Next time I advise to introduce line numbers so that it is more comfortable to carry out the revision.

The study has a strong background in testing models. Although there are certain aspects of the models that I can not judge since they are outside of my expertise (I recommend that another reviewer or editor assess the robustness of the models), in their current state the models are not sufficiently clear to be reproducible. In the material and methods section there are several points where it is not clear which tools you are using to build the models (e.g. pg.4, paragraph "To incorporate in a solely distance-based biogeographic model"). It is necessary to clarify whether you have used a statistical program or you have programmed bioinformatic scripts. On the other hand, if any script have been programmed, it would be necessary to reference them in the text in order to have access to them and clarify the reproducibility of the study, since in its current state it is not sufficiently clear.

pg. 3 - "which might apply to arid adapted plant groups for which past distributions have been more contiguous (Bellstedt et al., 2012)." Here I would recommend to give credits to recent studies of African groups, both arid and subtropical, which have provided new evidence regarding continuous past distributions in Africa. pg. 3 - "such as the more mesic temperate or alpine-like habitats of the "sky islands" of East Africa (Gehrke & Linder, 2009; Gizaw et al., 2013, 2016)." You are missing the relationships of the African continent with Macaronesia and I would recommend to introduce this concept in the text.

pg. 3 "One such scenario, inferred from Cape clades with distributions very similar to that of *Erica* involves dispersal north from the Cape to the East African mountains via the Drakensberg ("Cape to Cairo"; Galley & al., 2007). McGuire & Kron (2005) proposed a different scenario for *Erica*: southerly stepping stone dispersal." This is not written clearly enough for a reader not specialized in African biogeography. I understand what you mean, but it should be explained more clearly.

pg. 3 - "clades of different ages (Pokorný et al., 2015) and/or origins, but with similar ecological tolerances, might show convergence to similar distribution patterns (Gizaw et al., 2016)." Here you reference to the main idea of the manuscript. However, previous work on this idea is not clearly introduced or disregarded. This was

already explored in the manuscript of Mairal et al. 2017 in Journal of Biogeography, although you give credit to this manuscript elsewhere in the text, I miss that you introduce this idea with more details and you clearly establish a hypothesis.

Pg. 3 - "we test five biogeographic hypotheses" - this is clear in the figure 1, however in the text you only refer to 4 hypotheses, please clarify.

It seems worrisome that within your hypotheses (figure 1) you have not included clearly the mountains of Eastern Africa (e.g. Harar plateau, Abyssinian plateau, Gregorian Rift ...). In these areas the genus *Erica* is highly diversified within each sky-island and these areas have served as stepping-stones for the colonization of eastern Africa from Europe and west Asia (e.g. *Lychnis* in Popp et al. 2008; *Cardueae* in Barres et al. 2013; *Hypericum* in Meseguer et al. 2013; *Canarina* in Mairal et al. 2015). Please clarify if you have had this area in consideration, and consequently, modify the figure or comment this bias clearly in the text.

Reviewed by [Simon Joly](#), 24 April 2018

I enjoyed reading the manuscript of Pirie et al. entitled "Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotropics by *Erica*". It tests different hypotheses regarding the biogeography of the genus *Erica* present in Europe and in the South of Africa. Specifically, they compare previous hypotheses regarding plant dispersion with hypotheses based on distances alone or on bioclimatic niche similarity. The approach is interesting and the overall manuscript is clear and well written.

I see very little flaws with the manuscript, perhaps with one exception. The distance model that the author test assumes a negative linear relationship between geographical distance and dispersal probability. This seems quite inappropriate. Indeed, most studies on plant dispersion show that the relationship between distance and dispersal probability is not linear (see, for instance, Nathan 2006, Science; doi:10.1126/science.1124975). It is perhaps closer to an exponential function (or lognormal), where seeds have a larger probability to fall close to the plant and the probability to disperse far decreasing exponentially with distance. It seems to me that this is something that the authors should have considered to be thorough with their model testing. They could incorporate such non-linear relationships by using an exponential function with different alpha parameters to derive their dispersal probability, and check which one gives the best probability in their biogeographic model fitting. The same thing could probably be done with the niche model, although we probably know much less regarding the relationship between niche distance and dispersal probability.

I also have a few minor comments:

1. In the methods, it would be nice if the authors list the genes used.
2. The authors associate imprecise geographic coordinates to coordinates with less or equal to three decimals. But a coordinate could also be imprecise even with many decimal, such as when it is placed from general locality information instead of a GPS device (as when using Geolocate, for instance). Did the authors consider this type of uncertainty as well?
3. *"Prior to comparing the different biogeographic hypotheses we tested whether a model without constraint to dispersal or ancestral ranges fitted the data better than setting the adjacent area matrix and maximum areas at nodes to two (as would be implied by the present day distribution of the species, which never exceed two areas)."*
It is a strange idea to test two parameters at once. Then you don't know if the difference is due to one or both.
4. I feel that parsimony optimization is not necessary. But at the same time it is not problematic, so I leave the decision to keep it or not to the authors.

Reviewed by [Florian Boucher](#), 30 April 2018

[Download the review](#)