

Dear Andrea,

Thank you very much for taking the time to go through our revised ms. again – and so promptly. Delighted to address these last points. As ever, our responses are interspersed with your comments below and incorporated into the text in the form of tracked changes.

Best wishes,

Mike Pirie

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; Pokorny 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?).

Yes, this is exactly so. Clarifications added to the introduction to emphasise this distinction:

“Organisms adapted to different habitats respond differently to changing environmental conditions (Mairal, Sanmartín & Pellissier, 2017; Chala et al., 2017). **For example, plant groups with greater tolerances of aridity than *Erica* may have had more contiguous past distributions across Africa** (Bellstedt et al., 2012). Similar distribution patterns of such groups might thus be best described by biogeographic scenarios emphasising vicariance processes, such as for example the “Rand Flora”, representing plant lineages that show similar disjunct distributions around the continental margins of Africa (Sanmartín et al., 2010; Pokorny et al., 2015), or the “African arid corridor” hypothesis that seeks to explain disjunct distributions between the Horn of Africa and arid south-western Africa (Verdcourt, 1969; White, 1983). **By contrast, similar distribution patterns observed across plants such as *Erica* that are adapted, or otherwise restricted, to habitats that remained largely isolated over time** might instead be explained by concerted patterns of LDD (Knox & Palmer, 1998; Galbany-Casals et al., 2014; Nürk et al., 2015; Míguez et al., 2017). Examples include the shared arid adapted elements of Macronesia and adjacent North-West Africa and Mediterranean (Kim et al., 2008; Fernández-Palacios et al., 2011; García-Aloy et al., 2017), and the more mesic temperate or tropical alpine habitats of the “sky islands” of East Africa, in which, for example, multiple lineages originated from northern temperate environments (Gehrke & Linder, 2009; Gizaw et al., 2013, 2016).

I've also added the following (bold) to Appendix 3:

“Prior to comparing the different biogeographic hypotheses we tested whether an unconstrained model fitted the data better than a) restricting the maximum number of areas at nodes to two; and/or b) implementing an adjacent area matrix. Both a) and b) would be consistent with the present day distributions of the species, which never exceed two, adjacent, areas. **Even distributions widespread across just two adjacent areas are rare, and given the geographic structure in the *Erica* phylogeny we would assume that this was the case throughout the time of the radiation (Pirie et al., 2016).** The adjacent area matrix in particular explicitly disallows reconstructions of widespread yet interrupted distributions at ancestral nodes. We would argue that such distributions are unlikely in principle and that dispersal between such areas – if infrequent – would be more appropriately modelled by a process equivalent to jump dispersal than by assuming ongoing gene flow.”

Otherwise, I think this description makes it fairly clear that it is just the interrupted two-area distributions (those with other of the areas in between) that are disallowed, and these make little sense to me even if particular areas had been contiguous at some point.

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.

I'd argue that it is not that relative importance of niche versus distance could not be answered – in principle they could – it is that having analysed the data, our results show not only that this factors are indeed correlated but that **both** are important: The model that incorporates both factors fits the data better than models incorporating either individually; the one is not a direct proxy for the other. This is an important conclusion that would be lost without the comparisons as presented.

Clarified in the Discussion:

“Of the distance models, the combination of geographical and ecological distance fit the data well. **Our results showed that these factors are correlated across the *Erica***

distribution, but nevertheless given the phylogenetic uncertainty it was the combination of both that often fitted the data better than either of factor individually (or indeed the stepping stone models). “

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

Best Andrea