

The manuscript entitled “How ancient forest fragmentation and riparian connectivity generate high levels of genetic diversity in a micro-endemic Malagasy tree” by Salmona et al. evaluates the influence of landscape variables in the genetic differentiation of the olive species *Noronhia spinifolia*, micro-endemic to northern Madagascar. The main result shows the forest cover as the most important variable modulating genetic differentiation of this species.

The manuscript is well written, and the analyses are appropriate and thoroughly performed. I particularly appreciate the effort made in analyzing separately the organelle and nuclear genetic signals. However, I have two main concerns. The first one is about the emphasis of an ancient forest fragmentation that, from my point of view, has little support on the results (see the last comment below). The second main concern is about the landscape variables used and how the resistance layers were created from them (see below for more details), but I am sure that with some clarification and modifying a bit the text, it will be easily solved.

In addition, I have some suggestions about the structure of the manuscript, particularly regarding the introduction. The introduction is a bit disconnected from the other sections, particularly when referring to the landscape effects on the genetic differentiation of *N. spinifolia*. During the introduction, some landscape variables are referred as important for other species, such as the forest matrix, or the Manankolana River. In Methods, it is said that little information on the landscape variables that may affect *N. spinifolia* connectivity is available and hence all available landscape variables were assessed (L. 266-269). However, in the discussion I could learn that from fruit and flower morphology, and from knowledge of close related species some hypotheses may be formulated, and consequently, some expectations about the landscape variables as well. Including these hypotheses and expectations in the introduction may help the flow of the text between the different sections. Without these previous hypotheses and expectations a wide range of possibilities is open. For example, the authors include rivers and streams as landscape variables, but they did not consider the directionality of the water flow which could be connecting upstream forest patches with downstream ones. The same happens with wind direction and slope (upslope vs downslope movements).

Following the manuscript flow, see below my detailed comments, most of them minor comments:

L. 72-79. This paragraph changes the focus from the importance of studying species genetic diversity and connectivity across landscapes and landscape features affecting them, to the importance of the antiquity of open-canopy environment. Although I have no doubts on the importance of confirming the

antiquity of this kind of habitat, I recommend to remove this paragraph as it is not the main objective of the study and is disconnected from the other paragraphs.

L. 93-94. Tree species usually have long generation times, so changes in the landscape will need long time to be reflected by these species and hence the effect of recent changes will not be noticeable. I would rather expect scrubs or herbs associated with the forest and with shorter generation times to be better models for studies in fragmented habitats.

L. 182: 220 leaves from 220 trees?

L. 183: if these variables are not used in this study there is no need to mention them.

L. 189-192: I find confusing that 220 leaves were collected (from 220 different trees, based on Methods S1), 137 samples were used to extract DNA, and 72 were selected for genotyping. Why was it not possible to use all collected samples?

L. 202-208: Some information is missing here to fully understand this paragraph. Particularly, what the authors called “loci identified *in silico*”. In Methods S5 is clearly explained but the methods section in the main text should be comprehensible by itself.

L. 268: I recommend to specify the ten landscape variables used. A table may be a good way to show all of them and a summary of the results, or at least specifying which ones were discarded for further analyses (L. 391).

L. 276: It is not clear how the authors created the resistance layers. It says that 14 conductance-resistance values were tested, and in Methods S10 this values are 1:20, 1:15, 1:10, 1:8, 1:5, 1:4, 1:2, 2, 4, 5, 8, 10, 15, 20. From the table S6 I assume that each of these 14 values is assigned to the landscape feature (cost variable) and to the rest of the surface (cost non-variable), and a resistance surface is created from each combination of two values. The x-axis in figures S21-S24 is the logarithm of the ratio between cost variable and cost non-variable. If I am right, discrete features only have one cost level? i.e. discrete forest cover was considered as forest vs non-forest, independently of the density of the forest cover. How the procedure is for continuous variables?

About the categorization of variables in discrete and continuous, slope and wind speed were considered as discrete variables, why?

This part of the methods should be clarified in the manuscript.

I find very interesting considering the wind as landscape variable, as it is not quite usual. However, I wonder why the authors considered only wind speed and not direction, as direction may be even more

important in modulating the connectivity by wind. If the authors find it relevant, more information can be extracted from analyzing wind connectivity between patches and I recommend using rWind (Fernández-López & Schliep, 2018) for this purpose [Fernández-López, J., & Schliep, K. (2018). rWind: download, edit and include wind data in ecological and evolutionary analysis. *Ecography*, 42(4)].

L. 373: it is northwest-southeast, right?

L. 391-392: “showing uniform response at varying costs” is this variable response showed in Fig. S21-S24? If it is the case, slope exhibit a uniform response and based on Table S6 it was not discarded, why?

L. 521-525: this difference between years is not relevant since all of them have a strong relationship with the genetic differentiation of the species, the 2000's layer has better resolution that could bias the result (L. 528), and since the hypothesis of ancient forest fragmentation is been discussed.

L. 539-540: I find this conclusion rather strong. How many generations would be necessary to detect the genetic consequences of a recent fragmentation? I find difficult to detect a recent forest fragmentation event considering the long generation time of *N. spinifolia*. For example, if it has occurred at the beginning of the XX century only between two and five generations have passed, few to detect its effects.