

Jarne et al. PCI EvolBiol – answers to managing board / recommenders / reviewers.

Authors' replies = AR (*in italics*)

Dear Philippe Jarne,

Your article, entitled *Connectivity and selfing drives population genetic structure in a patchy landscape: a comparative approach of four co-occurring freshwater snail species*, has now been reviewed. The referees' comments and the recommender's decision are shown below. As you can see, the recommender found your article very interesting but suggests certain revisions.

We shall, in principle, be happy to recommend your article as soon as it has been revised in response to the points raised by the referees.

AR – This is excellent news. Thank you for your time and effort, and we extend these warm thanks to the recommenders and referees. Note that we provide two versions of the paper files, one with track change to visualize our modifications and one without track change. Data have already been deposited in Zenodo (see Data accessibility section in manuscript) and we also provide statistical (R) codes (archive address in Manuscript).

Recommandation by Trine Bilde and Matteo Fumagalli

Minor revision

Dear Dr Jarne and co-authors

We are pleased that we have now received three reviews of your pre-print. The reviewers and we as co-handlers of your manuscript all enjoyed reading the pre-print, and that we would like to recommend minor revision.

All reviewers agree that your study is interesting and presents a very interesting and comprehensive study. The reviewers have provided several comments and useful suggestions on ways to improve the presentation. These mainly relate to improving clarity of various parts of the manuscript.

One potentially critical point was raised relating to the collinearity between predictors in some of the models, in addition to a question regarding whether an additional variable should be included in the analyses.

It should be fairly straightforward to address the comments in a revised version of the manuscript.

Best wishes

Matteo Fumagalli and Trine Bilde

AR – Thank you very much for your work on this manuscript. We tried to address the two main points you highlighted above, and the more minor points made by the reviewer. See our detailed answers below.

Review #1

The authors present an elegant approach to assess the distribution of genetic variation in Guadeloupe. They analyzed four co-occurring snail species with distinct mating systems and ecological traits. This allowed them to perform a comparative approach, evaluating similarities

and idiosyncrasies in the obtained patterns. There are very few studies of this type, and it is important to note the great sampling effort, during several years, that made their study possible. The main strong points of their manuscript, in my opinion, are the solid methodology, in special the amazing sampling effort and the suite of appropriate analyzes applied; together with an Introduction containing relevant and updated research on the studied questions, and a Discussion very well written, considering both the expected and unexpected aspects of the obtained results. I see some opportunities to improve their manuscript, which I detail below:

Introduction

As I stated above, the introduction contains relevant and updated research on the studied questions and the motivation of the study is very clear. As a suggestion for improvement, I found the structure of this session somewhat confuse. There are several instances where the authors merge methods with the Introduction itself. For example, the start of the second (Lines 73-75), third (Lines 126-128) and fourth (Line 145-157) paragraphs. I believe some of these stretches would better fit in Mat&Met session. Some of this information could be condensed in the last paragraph of the Introduction, in the form of clear objectives and hypotheses (which are well organized in Table 1) that the authors are willing to test, to tackle the problems raised throughout their Introduction. This would make the Introduction much more clear and organized.

AR – What should come in the Introduction vs. M&M (Material & Methods) of a paper is not a subject with simple solutions. Enough should be provided in Introduction to make sense of the issues that are addressed, but not too much. L. 73-75 is just general information that introduces the general structure of the following paragraph, and indicates the factors that are considered in our work (environmental factors, invasion dynamics and mating system). Note that to be consistent with the order in which these factors are analyzed in Discussion, we now present them in the following order: (i) environmental factors, (ii) mating system, and (iii) invasion dynamics in the Introduction and M&M. We therefore fully rewrote the section going from L. 126 to the end of Introduction (presentation of biological system with regard to these three factors and questions addressed), and we somewhat restructured the second section ('The species studied and population sampled') and the last section ('Drivers of genetic diversity') in M&M. We hope that this answers the point made by the reviewer.

Methods

I believe that a figure containing the landscape properties (containing topography or pond location and size, or any other aspect the authors believe is important) could be useful. This could be added to Figure 1 or to the Supplementary Material.

AR – Pond location is already indicated in Figure 1. A map including topography (given the geographic scale; ca. 10 km per 0.1 min latitude/longitude) would not provide much relevant information, since the Grande-Terre topography varies at very small scales that would not be apparent in Figure. Nonetheless, we provide two maps illustrating variation in connectivity and size across the 75 ponds in which at least one species was sampled. This new Figure is now available in the Supplementary Material (Figure S1), and mentioned in the 'Drivers of genetic diversity' section of the M&M. The Supplementary Figures have been renumbered accordingly.

It is not clear by reading the methods when the genetically analyzed individuals were sampled (it gets somehow more clear in the Discussion, but could be better explained here too). Were they sampled throughout the years or in one specific season? If throughout the years, then the genetic statistics generated are including information of demographic/climatic events from

different years. This is fine, but needs to be better stated. Also, how did you define invasion age given this inter-year sampling scheme?

AR – We mentioned (L. 183) that “Field sampling takes place annually in January-February” and then on L. 232 that “The four species were sampled in Grande-Terre during the yearly field surveys ...”. To be clearer, we now write “The four species were sampled in Grande-Terre during the yearly field surveys (January-February) ...”. Invasion age was defined in L. 225-231 (as “the time in years since the first detection of the invasive P. acuta and sampling for the genetic analysis”). Not sure how we can be more specific.

I like the approach of comparing the results from different species, and the authors performed this with great detail and rigor. Yet, I still feel that some statistical analysis explicitly testing across-species correlations (i.e. including more than one species in the same analysis) with the environmental/demographic variables could be included, in addition to qualitatively comparing the results from different species.

*AR – This is a good idea, and we know from other studies that for example P. acuta and A. marmorata are interacting both in terms of population demography and of evolutionary trajectories (Chapuis *et al.* 2017, Dubart *et al.* 2019), as mentioned several times in the paper. The reason why we did not run the kind of analyses suggested by the reviewer is that the four species are not occurring together in all sites (Figure 1, Table S3), such that species Y cannot serve as an explanatory variable for the genetic diversity of species X, except to the cost of considerably decreasing the number of sites – not a good idea given the number of factors already considered. However, we ran some analyses including pairs of species, for example for comparing genetic variation in sites in which two species co-occurred (again, this decreased the number of data points). This was reported in P. 23 (L. 45-48), and we essentially found no significant results.*

By the way, line numbering was re-initiated (to L. 1) from P. 22 on – we fixed that in order to have continuous numbering over the whole paper in the current version.

Please provide your R scripts, especially for the linear models.

AR – This is requested by PCI EvolBiol, and now provided in [Zenodo ??](#) as mentioned in the Data accessibility section.

Discussion

The authors defend that an asymmetric island or a metapopulation model would be better suited for the studied species. Could this pattern affect the results of their mantel test? Moreover, I think the authors should consider the potential effects of IBD on their analyses of population structure and correlation with environmental factors. An extensive literature is available on such effects (for a review, see Meirmans 2015).

AR – We first note that isolation by distance was detected in two species (P. acuta and D. depressissimum), but the relationships between genetic and geographic distances were weak ($r = 0.301$ and 0.318 respectively). This was confirmed by the spatial clustering analysis in which only three and one clusters were detected in P. acuta (Grande-Terre) and D. depressissimum respectively. Moreover, a possible distance effect is captured in those variables which integrate a spatial dimension, namely connectivity (local spatial component) and density of favorable habitats (regional component). We already largely discussed their effects in the main text, but

now add a sentence in the Discussion section (end of paragraph 'Which other factors affect genetic variation?') to connect them to possible spatial correlation.

References

Meirmans, P.G. (2015), Seven common mistakes in population genetics and how to avoid them. *Mol Ecol*, 24: 3223-3231. <https://doi.org/10.1111/mec.13243>

Review #2

The manuscript "Connectivity and selfing drives population genetic structure in a patchy landscape: a comparative approach of four co-occurring freshwater snail species" proposes a multi-species approach to detect how consistent the effects of environmental and demographic factors, as well as mating strategies, are across closely related species with similar ecologies and overlapping geographic distributions. The main conclusions are that mating systems and connectivity are the main predictors of neutral genetic variation and structure. I found the manuscript quite complete, though there are instances where I think it could use some re-writing so as to make it easier for readers to follow. My comments mainly address the form and I have some questions concerning choices in statistical analyses.

General comments:

Methods:

Line 357: You have chosen to run linear models to test the effects of the environmental factors on the observed genetic structure of populations/species. From Table S9, there exist correlations between these factors, but it is not clear to me whether this was accounted for in the linear models (I'm under the impression that it was not). I can understand the choice to look at each factor separately, but at the same time I think it pertinent to account for interactions between predictors of a linear model. Or have I missed something? If interactions/correlations were accounted for, could the code be made available somewhere?

*AR – The reviewer is right when pointing out the correlation between some variables (Table S9), but note that they were computed over all sites (N = 74) considered in the study while the linear models were run per species. We also computed correlations per species, meaning (i) on fewer data points (21 to 43), and therefore less power, and (ii) an inflation in the number of tests (120 extra tests), calling for caution in interpretation. There were less significant correlations per species (1 to 4 vs. 6 for the whole dataset), giving less weight to the reviewer's concern. Note also that the within-species correlations went generally in the same direction than those calculated over all sites, but not in all cases (for example, we found a negative correlation between vegetation cover and stability in *D. depressissimum*).*

*It is also true that the linear models presented in the manuscript do not include interaction factors, because the number of factors tested was large compared to the number of data points (sites per species), as explained in M&M. However, we systematically considered interactions in the simplest model per species when two factors, or more, were retained (i.e. in five situations, Table S11; e.g., connectivity and invasion age for R_A in *A. marmorata*), and never detected any significant interactions. This was mentioned at the bottom of Table S11, but not explained in the M&M. We therefore added a section in the M&M to make that clear.*

Though it is not an environmental factor, as I read through the manuscript, it was not clear to me why the mating system was not included in the linear models. It would be interesting to show whether this trait overshadows any of the other predictors when it comes to the genetic

structure, and to point out in which case it is not (if ever) the main predictor. Checking for correlations between it and the observed environmental factors may also provide some insight.

AR – The mating system does not exhibit variation within species – the selfing rate was not different from zero in the outcrossing species, and very high in the selfing species (when it was possible to get an estimate). This is a classical result in freshwater snails, despite our long-term effort to detect such variation (review in Escobar et al. 2011), and the very reason not to consider the selfing rate in the linear models.

Results:

The results are presented in a detailed way, with main findings appropriately highlighted. However, the sheer density of information is overwhelming. I greatly appreciated the tables, that I found to summarise the results quite well, and was wondering whether the text could be organised differently so as to rely more on the readers' referral to the tables and make the reading lighter. There are also several instances when describing the results where the authors go back and forth between species. This left me utterly confused and I needed to re-read the text several times to untangle everything. Could there be a way to avoid this? Maybe describing each species separately then making comparisons (i.e. between sister species, selfing vs outcrossing, invasion ... but this could also just be done in the discussion as is already the case), or starting with the specific comparisons without going into the details of each species? Here the different layers of comparison are done simultaneously, and I personally had a tough time keeping up.

AR – We made the choice chose to present the results per parameter / factor across the four species, generally presenting the results in the same order (P. acuta, A. marmorata, D. depressissimum and D. surinamense). However, this order was not always followed to lighten the writing. We agree that this might not always be easy to follow. Presenting the results per species, then comparing, as suggested by the reviewer, would unduly enlarge the Results section, which is already long enough. We therefore propose a new version in which we stick more strictly to our original plan, hoping that reading will be easier.

P23 line 36-37: This sentence is confusing since the results of the correlations are not mentioned further in the main text: "It is less expected to find a negative correlation between size and connectivity and a positive one between stability and density of favorable habitats." Maybe change it to something like: "There were however unexpected correlations, notably, a negative correlation between size and connectivity and a positive one between stability and the density of favorable habitats." On that last point, is stability expected to be negatively correlated with the density of favourable habitats? I would have thought that on the contrary, these two would be positively correlated. Would you mind explaining this?

AR – Thanks for the suggestion. We did not expect a significant correlation between stability and the density of favorable habitats (the latter is based on the number of ponds in a circle of 4 km around the focal site). We rewrote this sentence to be clearer about the 'expected absence' of correlation in this situation.

Discussion:

P29 - line 161: What do you mean by "selective interference" in this context?

AR – Arguably a technical term, widely used in the literature on the evolution of sex and of selfing (see e.g. Otto 2021), meaning that the efficiency of selection is weakened by limited recombination. For example, good mutations could be trapped in a background of bad

mutations in the absence (or near-absence in selfers) of recombination. We added the reference to Otto (2021).

P29 - line 180: "As mentioned above, the four species differ in their colonization rate (not in extinction rates), but the difference is between the two families, not between selfers and outcrossers (Pantel et al., submitted). Since the difference is not between the two taxonomic families, it is likely that demographic extinction and colonization do not affect genetic variation strongly, as already shown in *D. depressissimum* (Lamy et al., 2012), but rather modulate the influence of the mating system." - You mention that the difference is between the two families then go on to say that it is not. Above it is that the difference is colonisation is between the two taxonomic families, so are the authors referring to demographic size in the second sentence? Please clarify.

AR – Oops! We meant “Since the difference is not between the mating systems”. This is now corrected.

Some minor comments:

P8 - line 184: 24 species were sampled?

AR – Yes for the annual long-term survey (demographic data). Only four were considered in the genetic analyses.

P16 - line 331: the same effect as the first three

P29 - line 216: This dynamic

AR – These two typos have been corrected

Review #3

The authors present a comparative analysis of population structure across four species of freshwater snails. The underlying questions concern the extent to which structure is determined by environmental features common to multiple species as opposed to species-specific factors that are species-specific. This question has broad interest but has rarely been studied using direct comparison, as the authors demonstrate with a survey of examples. The study system used here is ideal because of the defined and isolated region and the clearly delineated habitat patches. As the authors acknowledge, four species will not lead to very general conclusions but some interesting comparisons are possible (particularly between mating systems and between an invasive, expanding species and three native, more stable species). This study can also lead the way to further comparative approaches.

The Introduction does a very good job of defining the problem. The authors divide factors influencing neutral population structure into three categories: environmental factors, invasion dynamics (an example of an historical factor) and organismal traits. Of course, all of these factors will have effects. It would be helpful to be explicit about the quantitative nature of the question, i.e. to ask about the relative contributions of these factors. It would also be good to acknowledge immediately that they interact. For example, organismal traits can influence how each species experiences the same environmental factors.

AR – A very good remark. We would indeed be extremely happy to be more quantitative in our predictions of the relative contributions of these factors (environmental factors, invasion

dynamics and organismal traits). We know from the literature and previous work on freshwater snails (including in Guadeloupe) that they all shape genetic diversity. The mating system consistently has a strong effect, and the effect of the other ones depends on how they affect the effective size of populations. As this is not known (quantitatively), it is difficult to produce quantitative predictions on the respective effects. We add a sentence in the M&M section (at the beginning of the section “We built general predictions ...” to recognize this issue, and also the possible interactions between factors.

The Methods section is also clear. I am not sure of the value of including the few samples from outside Grande Terre since we clearly expect different patterns between, as opposed to within islands.

AR – They do not bring much to the whole picture, but some results are not uninteresting (e.g., the very low variability in P. acuta populations from Marie-Galante, presumably resulting from a marked genetic bottleneck). We prefer to keep them, say, as a window to future work.

A modest number of microsat loci was used, but sufficient for the description of population structure. Since the loci used differed among species, some comparisons may be difficult (particularly for absolute diversity levels as opposed to patterns). Predictions are summarised in Table 1 and explained in the text. These are rather general and not directly connected to the measures of population structure derived here. This means that there is no explicit rationale for the choice of measures. A closer connection between the predictions and the variables used to test them would be helpful.

AR – We derived expectations for genetic diversity within populations and genetic structure among populations, both within species. We then used classical estimators for both genetic diversity (number of alleles and of genotypes, gene diversity) and genetic structure (F_{ST}) – the predictions hold for these estimators. We now explain why these estimators were retained and make a closer connection between the predictions (Table 1) and these estimators in the section entitled “Genetic diversity and differentiation” in M&M.

On ll.340-341, the difference in predicted effect direction for *P. acuta* is not clear. On l.349, I would expect other things, not just site size, to influence long-term population size.

AR – We clarified why we expect this effect on L. 340-341. Vegetation cover affects negatively the colonization rate which depresses the local effective population size, even if populations are more stable. We rewrote this sentence (also to correct the incidence of vegetation cover which was wrongly worded). On L. 349, we just mentioned that LT population size and site size should have the same influence. So not clear what are the “other things”.

Correlations among explanatory variables (Table S9) might influence the linear models, or their interpretation, and this possibility should be considered. The linear model description does not make it clear whether interaction effects were considered. Biologically, they might be expected, but their inclusion might make the models too complex.

AR – Right. This concern was addressed in an answer to Reviewer#2.

The Results first present a summary of the patterns of variation and then consider the impacts of the explanatory variables, but without structuring by the three proposed factors. Not only were there correlations among explanatory variables but also there were strong correlations

among measures of diversity and divergence. This suggests two things. First, that it may be possible to simplify the MS by focusing on the most informative variables rather than on multiple variables that measure more or less the same thing. I hesitate to suggest moving more information to SI, since this is already extensive, but that might be an option. Secondly, it makes one wonder about the potential to test multiple effects, especially when predictions are not clearly distinct.

AR – This is an interesting remark. We indeed have two (or three in selfing species) classical estimators of genetic diversity within populations that do not react exactly in the same way to the factors studied here (we made it clearer, as explained just above) and a single estimator of differentiation. It would be hard to simplify here. Some variables are correlated, but none is correlated to all others, such that we do not believe that they “measure more or less the same thing”. We therefore prefer to keep them all. Moreover, we have only two variables for which the predictions in Table 1 are going in the same direction (size and density of favourable habitats) ... and they are neither correlated, nor measuring the same environmental characteristics. We therefore prefer to keep all variables.

Table 3 summarises effects in terms of significance, with connectivity being the major driver of both diversity and divergence. This is, perhaps, not very surprising. However, this comes back to the underlying question which, at least to my mind, should not be about whether these effects exist but rather about their relative contributions. The Yes/No answers in Table 3 seem poorly adapted to the latter question. Estimates of effect sizes would be preferable, if possible, perhaps with the redundant explanatory variables removed in order to focus on the three major factors.

AR – In Table 3, we report results from the linear models, and details (including effect size) are reported in Table S11 (subdivided in three tables). Table 3 is therefore just a compact way to present these details. When more than a single variable is significant (e.g., Table S11A in A. marmorata), the effect sizes can be compared (here, they are of the same order). This certainly allows to be a bit more quantitative, and we now use more explicitly this information in the last paragraph of the Results section.

The connectivity effect has been demonstrated previously for one of the 4 species. This species has also been shown to have a structure fitting an asymmetric island model. One issue is the extent to which this MS goes further. Model fitting does not seem to be possible for the other species, given the data available (and the likelihood of non-equilibrium for *P. acuta*).

AR – We indeed suggested before that the population structure in D. depressissimum fits an asymmetric island model, rather than a symmetric island or a metapopulation model, in a previous paper (Lamy et al. 2012). This was allowed by the fact that we had temporal genetic data, which are not available in the three other species, as mentioned in the manuscript.

Nevertheless, the comparative approach that the authors advocate surely requires effect size estimates rather than only significance tests and it would be good to see this MS heading in that direction. The Discussion ends up being rather descriptive. That said, the logic is clear and the descriptions are interesting.

AR – We hope that the modifications in the current version (see above) is a step in this direction.

Use of English is generally good. I have commented only where the sense is unclear. However, some smoothing of the language might aid readability.

AR – The whole paper has been thoroughly reread and edited when required.

There is rather high reliance on the unpublished (parallel) paper by Pantel et al. (cited as ‘submitted’).

AR – This paper has now been accepted ... voir comment on le cite ! Et lien HAL ou autre.

References cited in the answers to the reviewers

- Chapuis, E., Lamy, T., Pointier, J.-P., Segard, A., Jarne, P., & David, P. (2017) Bioinvasion triggers rapid evolution of life histories in freshwater snails. *American Naturalist*, 190, 694-706. DOI: 10.1086/693854.
- Dubart, M., Pantel, J. H., Pointier, J.-P., Jarne, P., & David, P. (2019) Reciprocal competition, niche differentiation and coexistence between invasive and resident species analysed through two-species metapopulation models. *Ecology* 100, e02700. DOI:10.1002/ecy.2700.
- Lamy, T., Pointier, J.-P., Jarne, P., & David, P. (2012) Testing metapopulation dynamics using genetic, demographic and ecological data. *Molecular Ecology*, 21, 1394-1410.
- Otto S.P. 2021. Selective interference and the evolution of sex. *Journal of Heredity* 112:9-18. doi: 10.1093/jhered/esaa026.