Response to “Comparing SNPs and microsatellites for phylogeography”

Below, we detail our responses (in **bold**) to each comment made by the Associate Editor and the reviewers.

**ANSWER: New title, “Comparative assessment of range-wide patterns of genetic diversity and structure with SNPs and microsatellites: a case study with Iberian amphibians”.**

April 29 2020

ID: EVA-2020-098-OA  
Title: Comparing SNPs and microsatellites for phylogeography

Dear Dr. Camacho-Sanchez:

Your manuscript has now been peer-reviewed and assessed by me as the Associate Editor. Comments follow below. On the basis of this, I must unfortunately inform you that this paper cannot be accepted for publication in Evolutionary Applications. You will see that both reviewers are very congruent  in their evaluation. I also found that it is unlikely that the present manuscript could be improved to the point of acceptability. I know you will be disappointed with this negative response, but I hope that the comments below will be useful to you in planning resubmission elsewhere.

**ANSWER: thank you for taking the time to review the manuscript and for your constructive response. The manuscript has been deeply reviewed and reformatted for Ecology and Evolution.**

I thank you for considering Evolutionary Applications for the publication of your research and hope that the outcome of this specific paper will not discourage you from the submission of future manuscripts.

Sincerely,  
Dr. Louis Bernatchez  
Editor in Chief, Evolutionary Applications  
[Louis.Bernatchez@bio.ulaval.ca](mailto:Louis.Bernatchez@bio.ulaval.ca)

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author  
The manuscript proposes to compare the use of SNP and microsatellite analyses for phylogeographic analysis through the study of two amphibian species on the Iberian Peninsula.  Approximately 80-90 individuals for each species are used, for 14 or 18 microsatellite loci, from 25 or 43 locations.  A ddRAD SNP data set is applied on both species, and over 15,000 loci are produced for one species, and over 30,000 for the other.  In short, the authors find multiple reasons for emphasizing the advantages of SNPs over microsatellites.  While there have been a number of recent studies discussing the pros and cons of these two sets of markers, and there is continued interest in understanding the detailed expectations of what one can draw from results based on SNPs or msats, this manuscript in my view does not really address anything new of interest. Furthermore or most importantly, the study begins with the implicit (explicit in the title) premise that this comparison is relevant for phylogeographic analysis.  But virtually no one uses microsatellite loci explicitly for phylogeographic analysis.  Thus, the study as presented provides a kind-of strawman (msats for phylogeographiy), which would be easy to shoot down, even without a comparison to SNPs. There are a number of other methodological details in the ms that seriously bias the results in this direction because the methods used would typically not be applied to the analysis of microsatellite data set, and in some cases even to SNP data.  I discuss these issues in more detail below.

**ANSWER: thanks for the critical comments. Regarding the comment “this manuscript in my view does not really address anything new of interest”, our manuscript provides for the first time a comparable framework in terms of samples and analytical approach to assess the patterns in genetic structure and diversity between microsatellites and SNPs. Previous studies comparing these two markers continuously failed in not providing comparable sampling, metrics to evaluate the nor comparable clustering approaches. Our approach allows to discuss in detail the pros and cons of using both marker types, which can be relevant for setting up a research project or aid with the interpretation of results using any of these marker types.**

**We address in comments below the concern regarding the use of microsatellites for “phylogeography”.**

Comparing SNPs and microsatellites for phylogeography

Title: The title is very general and give the impression that microsatellites are a marker of interest for phylogeographic analysis. Perhaps there is a semantic problem, but I have never had the impression that microsatellites themselves are very useful for phylogeographic analysis in the traditional sense.

**ANSWER: we have replaced “Phylogeography” with “structure”, and narrowed the title: “Comparative assessment of range-wide patterns of genetic diversity and structure with SNPs and microsatellites: a case study with Iberian amphibians”**

Almost without exception, phylogeographic analysis is traditionally based on gene genealogies correlated to geographic structure. When microsatellites are applied in phylogeographic studies, they are most often used as a compliment to mtDNA or other gene genealogies, whereby msats provide the within lineage structure and diversity component, and the genes genealogies provide the lineage component that is most readily correlated to geography.

**ANSWER: we have re-oriented the manuscript to assess current patterns of genetic structure and diversity, without any mention to genealogies (i.e. phylogeography), since microsatellites are not appropriate for the reconstruction of phylogenetic relationships and we are mainly evaluating demographic processes.**

The authors correctly state and cite the relevant literature on high rates of homoplasy in microsatellites, and that this is increasingly problematic the more divergent populations (or lineages) are.  But this a very well known and well described phenomena.  A good study to cite demonstrating how extreme this can become in a phylogeographic context is Queney at al 2001, on rabbit populations on the Iberian Peninsula. In that study, two mtDNA lineages with ca. 2 million years of divergence show almost no divergence across a set of microsatellites. The author’s are also investigating taxa with a similar phylogeographic structure as this rabbit study (there are, as the others mention, a number of animals that have a roughly congruent structure on the IP).  Although the demographic patterns of amphibians are expected to be different then for rabbits and the divergence is cited as approximately 1 million years instead of two, there are numerous parallels in the theory and practice between that study, and many statements by the authors in the present study.  Additionally, there is a very large literature supporting not only homoplasy in microsatellites but also the failure of particular distances to reflect higher divergence times (or even the difficulty in having any lineage correlation at all between divergence times and msat divergences).

**ANSWER: we have cited Queney et al. (2001) in lines 54 and 58 to refer to the homoplasy problem in microsatellites.**

Another issue in this ms that is somewhat troubling is the use of very low numbers of individuals per site (< 5, sometimes as little as 1) for microsatellites. I know of few studies applying such numbers to an analysis, as the strength of msats is in population analysis, and due to the high level of polymorphism, samples sizes of 30 or more are targeted in an attempt to accurately estimate allele frequencies per population.  To compare these two markers the authors use a simple distance measure (Manhattan distances) to build a NJ joining tree for each marker type.  I understand the desire to use the same method for both markers, but this approach would ideally never be used for either marker set alone, and does not take make the best use of the information for either type of marker. Furthermore, microsatellites are almost never used to build a phylogenetic tree, and it seems that this, and not phylogeographic inference is the major thing that the authors are actually comparing.

**ANSWER: we acknowledge that when characterizing populations, it is important to have representative sample samples, often in the order of 20 individuals for microsatellites. However, in our study, we are describing broad patterns in the Iberian Peninsula. Therefore, we are not interested in identifying sampled localities to populations, but rather having representative numbers on the main biogeographical regions in Iberia. For this latter purpose we are able to adequately represent the different genetic clusters identified in previous works with microsatellites (north/south regions) with sufficient sample sizes (>20 samples) to get a reliable perspective on the broad genetic patterns. We have added a sentence in lines 112-114: “*They were evenly distributed across the main genetic clusters determined in previous works with microsatellites (Gutiérrez-Rodríguez et al., 2017; Sánchez‐Montes et al., 2019), securing the representation of more than 20 samples per north/south clusters.*”**

**We agree that phylogenetic trees are not appropriate for representing intra-species genetic structure. We have carefully removed any mention to “phylo-” in the manuscript and stated clearly that our interest genetic structure and not phylogenies. To prevent misinterpretations, we now have given priority in the manuscript to STRUCTURE analysis, and we have moved genetic structure from Neighbor Joining trees to supplementary material.**

I think there is still great need for understanding the differences or similarities in costs, as well as inference power for SNPs and msats in various population genetic contexts, but in my view phylogeography or phylogenetics are not the areas where we need this type of comparison.   
**ANSWER: we make clear now that the comparison restricts to population genetics and not to phylogenetics.**

References:

Queney G, Ferrand N, Weiss S, Mougel F, Monnerot M (2001) Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (Oryctolagus cuniculus) Molecular Biology and Evolution 18(12), 2169-2178.

Reviewer: 2

Comments to the Author

In this study the authors compare the results obtained for phylogeographic studies of two Iberian amphibians using a small number of microsatellites or a large panel of SNP markers. Previous studies had already addressed the population structure of those amphibians in the Iberian Peninsula using microsatellites and for this study the authors have selected a few of the individuals that had been studied for one of the species and have typed SNP markers on them. For the other species, the authors could only use a few of the previously typed samples and have complemented those with a few more samples from the same locations. This implies that for this second species the results obtained with the two sets of markers are not completely comparable. The authors use 1-4 individuals per population to characterize patterns of genetic diversity, what is uncommon with microsatellites. In general, the authors confirm that a larger number of markers provide higher resolution and show that a large panel of SNPs allow more robust results. The fact that more robust results are obtained when using more markers is not surprising. However, a few aspects of the analyses make me doubt about some important aspects of the study.

**ANSWER: we acknowledge that when characterizing populations, it is important to have representative sample samples, often in the order of 20 individuals for microsatellites. However, in our study, we are describing broad patterns in the Iberian Peninsula. Therefore, we are not interested in identifying sampled localities to populations, but rather having representative numbers on the main biogeographical regions in Iberia. For this latter purpose we are able to adequately represent the different genetic clusters identified in previous works with microsatellites (north/south regions) with sufficient sample sizes (>20 samples) to get a reliable perspective on the broad genetic patterns. We have added a sentence in lines 112-114: “*They were evenly distributed across the main genetic clusters determined in previous works with microsatellites (Gutiérrez-Rodríguez et al., 2017; Sánchez‐Montes et al., 2019), securing the representation of more than 20 samples per north/south clusters.*”**

The authors choose to use NJ phylogenetic trees. However, these are bifurcating and this approach should not be expected to reflect evolutionary relationships (see line 146) but just indicate overall similarity. In addition, the authors claim to use a “model-free approach”. However, this is not a correct interpretation of what they do. They are assuming that all nucleotides have the same probability of change and, for microsatellites, this is the same as assuming that all allelic states have the same chance of originating from other allelic state; this is equivalent to the infinite allele model, which does not seem to correspond to the evolution of microsatellites which seems to result from Microsatellite evolution inferred from replication slippage. This seems better represented by step-wise mutation models. I am not sure that choosing one evolutionary model or another would make a difference, but the authors should not claim that they are using model-free approaches (line 149).

**ANSWER: we agree that phylogenetic trees are not appropriate for representing intra-species genetic structure. We have carefully removed any mention to “phylo-” in the manuscript and stated clearly that our interest genetic structure and not phylogenies. To prevent misinterpretations, we now have given priority in the manuscript to STRUCTURE analysis, and we have moved genetic structure from Neighbor Joining trees to supplementary material.**

Since the same individuals were not typed for microsatellite and SNP loci, the analysis of the correlations between sMLH for microsatellite and SNP data for H. molleri is done at the level of locations. This implies assuming that individual genetic diversity are a good proxy for population values. Each population is represented by just 1- 4 individuals. If individual genetic diversity is variable within populations compared to between populations, the selection of one individual or another to characterize populations may have an important impact in the correlation. This could be the reason why the correlation for the estimates done with the two marker systems is not significative (p= 0.08) between localities for H. molleri, but it is significative (p< 0.001) between individuals for P. cultripes. Although we do not know if other factors may be affecting these results, they imply that the analyses for H molleri may not be appropriate for the goals of this study.

**ANSWER: we agree that for *H. molleri t*he correlation of heterozygosity from SNPs and microsatellites is subjected to the inherent bias of not using the exact same samples, and the corresponding sections have been removed from the main text.**

I can not understand the calculation of the Coefficient of Admixture. If this is based in another paper, that should be cited (in line 221 the text reads “a newly developed index” and this makes me think that someone else has developed this Coefficient). If it is described here for the first time, the formulas and logic and not sufficiently explained. It is not clear what K and k represent. I understand that k (with lower case) is an indicator of the cluster referred by a q value. However, in this sense, CAmin would make no sense as defined (the minimum possible value for CA would depend on the cluster that we have arbitrarily defined). The formula indicates that:  
CAmin= k x (1/k)^2   
This implies that CAmin= 1/k. Thus CAmin will have different values for each cluster k when considering a given total number of clusters K. I can not understand the reasoning.

**ANSWER: we provided simplified and intuitive explanation in the main text (lines 199-203) and a full explanation with a worked example in Supplementary File S2.**

In Results, the authors extensively discuss the results obtained with the NJ trees and with STRUCTURE and remark how similar the results are for SNPs. However, this is what should be expected. NJ trees reflect overall similarity and not necessarily evolutionary processes. Similarly, STRUCTURE makes groups of individuals considering overall similarity. The results should be the same, and this is what the authors observe. Usage of the two approaches is redundant.

On the other hand, for microsatellites the results are less clear because the approach used is very rough and does not consider lower distances between individuals that have more similar alleles at a locus. Similarly, STRUCTURE only considers alleles that are shared across individuals and not the difference between them. So, in this case, the comparison of individuals just considering if they carry the same alleles or not is not likely to show identical results using NJ and STRUCTURE and ignores a lot of the information provided by microsatellites. Probably using a step-wise mutation model would provide more reliable results for the NJ tree.

**ANSWER: we have moved NJ results to supplementary. The manuscript focuses now on the genetic structure from STRUCTURE.**

The results on “genetic diversity” are difficult to follow. Geographic patters are expressed without doing statistical tests. Altitude, latitude and admixture are commented but not properly evaluated. In addition, admixture may be associated with the K value considered and could be arbitrary.

**ANSWER: we have simplified the results of “genetic diversity” and we have tested the effects of latitude and longitude (lines 271-288).**

The assessment of changes in the support of the nodes of trees using bootstrap does not contribute much to the paper. Having more markers is always expected to produce higher bootrstrap support. Similarly, bootstrap support is always expected to be higher in longer branches. Thus, I do not really see the relevance of the entire last section in results. In addition, what is the meaning of bootstrap supports in a tree of individuals and populations? More markers can lead to 100% support and will allow separating any pair of individuals. However, this could be the case even with extensive gene flow between populations. Perfect separation in the individual tree does not imply perfect separation of the populations.

**ANSWER: we have removed this section.**