

Review of: "Testing concordance and conflict in spatial replication of landscape genetics inferences"

Martina Weiss¹

¹ University of Duisburg-Essen

Potential competing interests: The author(s) declared that no potential competing interests exist.

The aim of the paper „Testing concordance and conflict in spatial replication of landscape genetic inferences” was to study landscape genetic processes in the Western Fence lizard *Sceloporus occidentalis* in different regions in the Sierra Nevada mountain range to obtain spatial replication and identify general and site-specific facilitators and barriers to dispersal. In general, the results of the study are of great interest and the paper is very well written. However, there are in my opinion some parts which could be shortened a bit or better summarized and when it comes to the description of PCA loadings I would need some clarification for the variables described or which were named as associated with high or low values. As these analyses are not my main area of expertise, I think it is probably all correctly interpreted, but as it in some cases is not logically for me, some clarification/better explanation is needed in my opinion. I have specified my questions under the specific comments. As there are no line numbers, I copied the sentence to which the comments refer in the review.

Specific comments:

Abstract: “We conducted landscape genetics analyses in five areas within the Sierra Nevada mountain range, using thousands of ddRAD genetic markers distributed across the genome, implemented in the landscape genetics program ResistanceGA.” □ when reading this, I was not sure what is implemented in the program and if the dataset was part of the program. I would rephrase it, to make clear that you generated the ddRAD data yourself.

Introduction:

Page 4:

- “The geologic history of the Sierra Nevada is complex, but is the entirety of the Sierra Nevada mountain range is generally...” □ delete first “is”
- I would use non-breaking spaces before units (e.g. in 1500 m – 2400 m) to prevent the line break between number and unit.
- “Each of the study areas are, to the extent that was possible, composed of similar habitats, elevation gradients, and are of identical size (50 km x 50 km extents) (Figure 1).” □ in my opinion this is not really visible in Figure 1, because there are only the location of the regions shown but not that they are

composed of similar habitats etc.. And as in general not really a lot information are displayed in Fig. 1, it could be smaller.

Methods:

Lizard collection

- “We sampled lizards from June through August in 2016 and 2018 in California, USA over a latitudinal range of >900 km and an elevation range of 100 – 2800 m (Figure 1).” □ were at all sites samples taken in both years or were some sampled in 2016 and 2018? Also the elevation range is not visible in Figure 1. In general, I think a Supplement Table would be needed here including coordinates of specific locations, sampling year, number of specimens per site and information used for the sites. Later tissue samples from a museum collection are mentioned □ when were they taken and from which locations were they? This information should also be included in a supplementary table
- How many specimens were collected in total at each site and area?

ddRAD library prep:

- Information for each specimen uses should also be given in a supplementary table
- Give more information on the enzymes used, add company for restriction enzymes and name the tag you used. For all chemicals and devices company and country of company should be given. Mention in which adapter the indexes were present as you used single indexing? Or did you use double indexing? Give reference to the adapter sequences used. For further studies, you should think of including DBRs into the adapters to be able to get rid of PCR duplicates (see e.g. Schweyen et al 2014: Detection and Removal of PCR Duplicates in Population Genomic ddRAD Studies by Addition of a Degenerate Base Region (DBR) in Sequencing Adapters).
- HiSeq and NovaSeq will give a different amount of output read. Specify why you used two different platforms and how the samples were divided on those platforms? Were they randomized or were different locations sequenced on different lanes? How many specimens were pooled on each lane?

Bioinformatics protocol for SNP data

- why did you use STACKS in population based form and ipyrad individual based? STACKS can also be used individual based. Give the parameters for the STACKS analysis and for what were the data produced with stacks used? Did you compare stacks and ipyrad results? Were there differences?

Spatial data layers

- provide the data you used in supplementary tables (i.e. temperature, precipitation ... data)

Spatial data processing

- reference missing for R and used packages
- “These principal component spatial layers represent different aspects of the environment in each spatial area as detailed in the results and supplemental materials” □ state to which supplemental materials you refer
- “km-sq extents for a total size of 2,500 sq-km, each with 25,159 cells, and a cell size of ~1 km-sq.” □

why not km²; otherwise sq-km or km-sq?

Description of the study areas

- this can be shortened a lot in my opinion. Important features could be summarized in a table for all regions together. That would make it much easier to follow the description and to see similarities and differences
- Why was the number of sampling sites different between the regions? Did you test if the differing sample size had an influence on the spatial analyses?

Results

Sequoia

- Please adjust the numbering of supplementary Tables and figures in line with the order of its mention in the text (comment for all area descriptions)
- For PC1 you mention that higher values are associated with higher elevation and low values with low elevation, but the variable in the table is only associated with high values? For warm and cold temperature there are e.g. two (or more) different variables in the analysis, but for elevation, slope and ruggedness only one each, which you interpret often for low and high values. Could you please explain that to me? Also important for the other areas
- For PC2 you do not name the variables associated with low values

Sierra

- one “,” too much in line 3
- Here, I do not understand why you state high values of PC1 are associated with cooler temperatures, because in Table S7 both cold and warm are associated with low values and cold even lower than warm? Yet, you mention warm for low values. Also here again I am not sure about the low elevation.

Yosemite

- Also here you write that high values of PC1 are associated with cooler temperatures, but in the table again coldest temperature has the lowest value, while there are similar values for warmer temperatures? Please explain

Tahoe

- For PC3 you state that high values have little relief and low slope, while low values are associated with rugged areas and high slope. Here, again I only see these variables for low values and why do you not mention e.g. seasonality for high values?

Plumas

- in the table captions of S20-22 Tahoe region is written instead of Plumas
- Here you state, that high values are associated with warmer areas, but the highest value in the table is for mean_coldest, followed by mean warmest? And lower values are stated to be associated with wetter areas, while the variable in the table is precip_driest? Please explain here the choice of variables you mention.

- PC3 □ why do you mention slope and relief for high values, but not seasonality etc. which is visible in the table?

Comparison between regions

- It would be good to mention here that the roads greatly vary between areas so it is not really surprising that they have different effects

Table 1

- why not order the table similar to the description in the text, starting with Sequoia and ending with Plumas?

Discussion

Differences in results between study areas

- please discuss here also if the different sample sizes in the regions could have influenced the results
- Please explain better why you think that roads do not have a long term effect on the genetic structure