Peer Community In
Evolutionary Biology

Pros and Cons of local adaptation scans

Guillaume Achaz *based on peer reviews by* ***Lucas Gonçalves da Silva*** *and 1 anonymous reviewer*

Pratlong M, Haguenauer A, Brener K, Mitta G, Toulza E, Garrabou J, Bensoussan N, Pontarotti P, Aurelle D (2018) Separate the wheat from the chaff: genomic analysis of local adaptation in the red coral Corallium rubrum. Missing preprint_server, ver. Missing article_version, peer-reviewed and recommended by Peer Community in Evolutionary Biology. `https://doi.org/10.1101/306456`

Submitted: 24 April 2018, Recommended: 13 December 2018

Cite this recommendation as:
Achaz, G. (2018) Pros and Cons of local adaptation scans. *Peer Community in Evolutionary Biology*, 100061.
10.24072/pci.evolbiol.100061

Published: 13 December 2018

The preprint by Pratlong et al. [1] is a well thought quest for genomic regions involved in local adaptation to depth in a species a red coral living the Mediterranean Sea. It first describes a pattern of structuration and then attempts to find candidate genes involved in local adaptation by contrasting deep with shallow populations. Although the pattern of structuration is clear and meaningful, the candidate genomic regions involved in local adaptation remain to be confirmed. Two external reviewers and myself found this preprint particularly interesting regarding the right-mindedness of the authors in front of the difficulties they encounter during their experiments. The discussions on the pros and cons of the approach are very sound and can be easily exported to a large number of studies that hunt for local adaptation. In this sense, the lessons one can learn by reading this well documented manuscript are certainly valuable for a wide range of evolutionary biologists. More precisely, the authors RAD-sequenced 6 pairs of 'shallow vs deep' samples located in 3 geographical sea areas (Banyuls, Corsica and Marseilles). They were hoping to detect genes involved in the adaptation to depth, if there were any. They start by assessing the patterns of structuration of the 6 samples using PCA and AMOVA [2] and also applied the STRUCTURE [3] assignment software. They show clearly that the samples were mostly differentiated between geographical areas and that only 1 out the 3 areas shows a pattern of isolation by depth (i.e. Marseille). They nevertheless went on and scanned for variants that are highly differentiated in the deep samples when compared to the shallow paired samples in Marseilles, using an Fst outliers approach [4] implemented in the BayeScEnv software [5]. No clear functional signal was in the end detected among the highly differentiated SNPs, leaving a list of candidates begging for complementary data. The scan for local adaptation using signatures of highly divergent regions is a classical problem of population genetics. It has been applied on many species with various degrees of success. This study is a beautiful example of a well-designed study that did not give full satisfactory answers. Readers will especially appreciate the honesty and the in-depth discussions of the authors while exposing their results and their conclusions step by step.

References:

[1] Pratlong, M., Haguenauer, A., Brener, K., Mitta, G., Toulza, E., Garrabou, J., Bensoussan, N., Pontarotti P., & Aurelle, D. (2018). Separate the wheat from the chaff: genomic scan for local adaptation in the red coral Corallium rubrum. bioRxiv, 306456, ver. 3 peer-reviewed and recommended by PCI Evol Biol. doi: [10.1101/306456](`https://dx.doi.org/` 10.1101/306456)

[2] Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics, 131(2), 479-491.

[3] Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

[4] Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics, 74(1), 175-195.

[5] de Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST-based method to uncover local adaptation using environmental variables. Methods in Ecology and Evolution, 6(11), 1248-1258. doi: [10.1111/2041-210X.12418](`https://dx.doi.org/` 10.1111/2041-210X.12418)

Reviews

Evaluation round #2

DOI or URL of the preprint: **10.1101/306456**
Version of the preprint: 2

Authors' reply, 05 December 2018

Our reply is in attached file. Please note that the BioRXiv file includes all corrections, except formatting for PCI (we are waiting for the formal recommendation), the formatted version can be uploaded here: `https://amubox.univ-amu.fr/s/rRjj7YXiC4JZRAL`

Download author's reply

Decision by Guillaume Achaz, posted 23 November 2018

Recommendation, pending very minor revisions

Dear authors,

Please first apologize for the delay in responding to your submission. My agenda has been sadly very compact. I think your revised version is even better than the first one. I will be happy to recommend your preprint for PCI Evol Biol, provided that you can fix few minor details listed below. I believe this would not you very take long. I shall then quickly post my recommendation.

- Gene diversity, although now defined, is used in your ms. Please replace all its occurrences by expected heterozygosity that is a much more explicit wording.

- Table 3: MAF<1% actually means MA Counts<4 for most sites, unless there are too many missing data. In this last case, could you provide some extra information on missing genotypes ? (maybe mean and 95% of the distribution)
- Table 2, 4 : could you add one letter for the region in parenthesis after the name. This will considerably ease the reading.
- Table 4: remove Tajima's D from the legend. Fourth column head is "number of private alleles" and non integers are reported. Is this a mean. If yes, please use "average number of private alleles". Otherwise please explain.
- Table 5 : disturbing to read that Fst of 0.01 is "highly significant". I guess this is only due to the size of the dataset in terms of loci, that are assumed to be independent.

May I suggest (but feel free to refuse) to change the title from "Separate the wheat from the chaff: genomic analysis of local adaptation in the red coral Corallium rubrum" to "Separate the wheat from the chaff: genomic scan for local adaptation in the red coral Corallium rubrum"

as the ms is not really an "analysis" of local adaptation but rather a quest for it.

Evaluation round #1

DOI or URL of the preprint: `https://doi.org/10.1101/306456`

Version of the preprint: 1

Authors' reply, 17 October 2018

Download author's reply

Decision by **Guillaume Achaz**, posted 17 October 2018

Revise

Dear authors,

Two external reviewers and myself have carrefully read your ms entitled "Separate the wheat from the chaff: genomic analysis of local adaptation in the red coral Corallium rubrum" and we all agree that this ms is sound, interesting and certainly deserves to be recommded by PCI Evol Biol. As pointed out by the reviewers, the main strength of this ms is the indepth discussion of the pros and cons of the methods and results you provide. I agree with their opinion.

We nonetheless have few suggestions that could potentially improve the overall quality of this ms. I believe the suggested corrections will be easily addressed and that I will be able to recommend the revised version without any further opinion from the two external reviewers. Please revise and respond to all comments (see below) before I can recommend this article.

Best regards,

Guillaume Achaz

:: Suggestions from G Achaz ::

Major: (A) can the authors provide basic analysis of within samples diversity (sample size, #sites, paiwise difference, Tajima's D, etc). (B) did the authors try to infer model parameter (using a mix of demography and structure) using appropriate methods like the ones implemented in dadi (or any other equivalent approach). (C) please shorten and/or lower the importance of the detection of selection as there isn't obvious positive conclusions from it.

Minor: l43 "between" is repeated in tandem l70 replace "high" by "strong"? l189, l270, table4 what is gene diversity? pairwise differences (i.e. heterozygosity)? l194 I don't understand what is the meaning here of

’centered’ l282-283, l294, l345 this seems surprising. Why would the sex contribute to genetic differentiation? Am I missing something obvious? l309 you mean 20 instead of 10? l357 which axis are you refering to? pc1? l384 I would be more cautious as this pattern is not visible in the Fst analysis.

Reviewed by anonymous reviewer 1, 16 May 2018

This manuscript describes the population genetic analysis of the red coral Corallium rubrum, using RAD-Seq data from three populations in the mediterranean see, and samples at two depth. In order to describe the genetic structure of the species, the authors used a combination of principal components analysis (PCA), bayesian inference (STRUCTURE), analysis of molecular variance (AMOVA), and measures of fixation idenx (Fst). The authors find that geographical location is the main factor explaining the genetic variance, depth showing an effect only in one population. The authors conduct a careful discussion about potential confounding factors, in particular regarding outlier genes with high Fst values.

This is a well-conducted study with an interesting discussion, both with methodological and biological implications, notably with respect to adaptation in the context of climate change.

I only have minor comments:

- l276: ”principal components reflected the geographical and depth origin”... As discussed later in this paragraph, the depth only appear to have an effect in one population. It would be more accurate to write something like ”principal components reflected the geographical and, to some extant, depth origin”
- l58: as written, the authors seem to oppose natural selection and gene flow, while the two phenomena are not at the same level... Just like mutations, gene flow can be neutral or selected. Maybe rephrase?
- l235-236: a threshold of 1% was used for the chi2 test and then corrected with a FDR or 5%? This is unusual... Either a threshold on the FDR was used, or p-values were adjusted for multiple testing, but one does not ”correct” p-values using a threshold.
- l271, Wilcoxon test: not clear what was compared to what: Marseille vc. pooled Corsica + Banyuls? A Kruskal-Wallis test + posthoc would be more appropriate here (see ”agricolae” package in R for non-parametric posthoc tests)
- l283: Reference is a bit unclear. Do the author mean something like ”The 5th axis of the PCA separated all individuals according to their sex, independently of their geographical origin, as observed in a previous study (Pratlong et al. 2017)” ?
- l298-299: two times ”the results”
- l309, ”even for populations separated by 10m”: unclear which populations are refered to
- l324: it is unclear which test was performed there, as there are two Fst values per population and it seems like the authors made one comparison per population?
- l376, at a genomic SCALE ?
- Figure 1: while the depth of the sampling site is coded in the sample name, it would be ore clear to have it explicitly indicated (for instance using two distinct plotting symbols)

Reviewed by Lucas Gonçalves da Silva, 23 May 2018

Dear editors and authors,

I would like to congratulate the authors, because the study is extremely interesting, relevant, with well designed, with consistent methods and great results.

I think a very positive point of the work is the discussion around the limitations of the study. The authors assume and describe these limitations, and this is always very important in genomic scale studies.

The only point left to be desired from my point of view was the discussion and characterization of the study in the context of implications for species-focus conservation. This species is considered endangered by IUCN and included within the Mediterranean Red List of Endangered Species. I believe that the isertion of informations about conservation status (introduction and discussion) will help the applicability of work in decision making, and can enrich the work in general.

Figures and tables are satisfactory.

L36 - ”... we also discuss ...” L42 - ”adaptation by species ...” L43 - remove ”between” L87 - missing space

I would love to see a niche modeling with this data, applying maxent, maxlike or some other likelihood algorithm for SDM (or more than only one). It is not something to be included in this paper, but I recommend authors to do so, since they have enough data for this. If you are interested, I leave the communication channel open with me via the editor of PCI Ecol Evo. Editors are authorized to send my email contact, if the authors want it. I strongly recommend the acceptance of this paper.

Best Regards.

Dr. Lucas Gonçalves lucas_gonc@yahoo.com.br