




Peer Community In Evolutionary Biology

Differential effect of genes in diverse environments, their role in local adaptation and the interference between genes that are physically linked

Sebastian Ernesto Ramos-Onsins  based on peer reviews by **Tanja Pyhäjärvi** and 1 anonymous reviewer

Katie E. Lotterhos, Sam Yeaman, Jon Degner, Sally Aitken, Kathryn Hodgins (2018) Modularity of genes involved in local adaptation to climate despite physical linkage. bioRxiv, ver. 1, peer-reviewed and recommended by Peer Community in Evolutionary Biology. <https://doi.org/10.1101/202481>

Submitted: 15 October 2017, Recommended: 18 May 2018

Cite this recommendation as:

Ramos-Onsins, S. (2018) Differential effect of genes in diverse environments, their role in local adaptation and the interference between genes that are physically linked. *Peer Community in Evolutionary Biology*, 100050. [10.24072/pci.evolbiol.100050](https://doi.org/10.24072/pci.evolbiol.100050)

Published: 18 May 2018

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

The genome of eukaryotic species is a complex structure that experience many different interactions within itself and with the surrounding environment. The genetic architecture of a phenotype (that is, the set of genetic elements affecting a trait of the organism) plays a fundamental role in understanding the adaptation process of a species to, for example, different climate environments, or to its interaction with other species. Thus, it is fundamental to study the different aspects of the genetic architecture of the species and its relationship with its surrounding environment. Aspects such as modularity (the number of genetic units and the degree to which each unit is affecting a trait of the organism), pleiotropy (the number of different effects that a genetic unit can have on an organism) or linkage (the degree of association between the different genetic units) are essential to understand the genetic architecture and to interpret the effects of selection on the genome. Indeed, the knowledge of the different aspects of the genetic architecture could clarify whether genes are affected by multiple aspects of the environment or, on the contrary, are affected by only specific aspects [1,2]. The work performed by Lotterhos et al. [3] sought to understand the genetic architecture of the adaptation to different environments in lodgepole pine (*Pinus contorta*), considering as candidate SNPs those previously detected as a result of its extreme association patterns to different environmental variables or to extreme population differentiation. This consideration is very important because the study

is only relevant if the studied markers are under the effect of selection. Otherwise, the genetic architecture of the adaptation to different environments would be masked by other (neutral) kind of associations that would be difficult to interpret [4,5]. In order to understand the relationship between genetic architecture and adaptation, it is relevant to detect the association networks of the candidate SNPs with climate variables (a way to measure modularity) and if these SNPs (and loci) are affected by single or multiple environments (a way to measure pleiotropy). The authors used co-association networks, an innovative approach in this field, to analyse the interaction between the environmental information and the genetic polymorphism of each individual. This methodology is more appropriate than other multivariate methods - such as analysis based on principal components - because it is possible to cluster SNPs based on associations with similar environmental variables. In this sense, the co-association networks allowed to both study the genetic and physical linkage between different co-associations modules but also to compare two different models of evolution: a Modular environmental response architecture (specific genes are affected by specific aspects of the environment) or a Universal pleiotropic environmental response architecture (all genes are affected by all aspects of the environment). The representation of different correlations between allelic frequency and environmental factors (named galaxy biplots) are especially informative to understand the effect of the different clusters on specific aspects of the environment (for example, the co-association network 'Aridity' shows strong associations with hot/wet versus cold/dry environments). The analysis performed by Lotterhos et al. [3], although it has some unavoidable limitations (e.g., only extreme candidate SNPs are selected, limiting the results to the stronger effects; the genetic and physical map is incomplete in this species), includes relevant results and also implements new methodologies in the field. To highlight some of them: the preponderance of a Modular environmental response architecture (evolution in separated modules), the detection of physical linkage among SNPs that are co-associated with different aspects of the environment (which was unexpected a priori), the implementation of co-association networks and galaxy biplots to see the effect of modularity and pleiotropy on different aspects of environment. Finally, this work contains remarkable introductory Figures and Tables explaining unambiguously the main concepts [6] included in this study. This work can be treated as a starting point for many other future studies in the field.

References:

- [1] Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F & Bergelson J. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334: 83–86. doi: [10.1126/science.1209244](<https://doi.org/10.1126/science.1209244>)
- [2] Wagner GP & Zhang J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics* 12: 204–213. doi: [10.1038/nrg2949](<https://doi.org/10.1038/nrg2949>)
- [3] Lotterhos KE, Yeaman S, Degner J, Aitken S, Hodgins K. 2018. Modularity of genes involved in local adaptation to climate despite physical linkage. *bioRxiv* 202481, ver. 4 peer-reviewed by Peer Community In Evolutionary Biology. doi: [10.1101/202481](<https://doi.org/10.1101/202481>)
- [4] Lotterhos KE & Whitlock MC. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology* 23: 2178–2192. doi: [10.1111/mec.12725](<https://doi.org/10.1111/mec.12725>)
- [5] Lotterhos KE & Whitlock MC. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* 24: 1031–1046. doi: [10.1111/mec.13100](<https://doi.org/10.1111/mec.13100>)
- [6] Paaby AB & Rockman MV. 2013. The many faces of pleiotropy. *Trends in Genetics* 29: 66–73. doi: [10.1016/j.tig.2012.10.010](<https://doi.org/10.1016/j.tig.2012.10.010>)

Reviews

Evaluation round #2

DOI or URL of the preprint: [10.1101/202481](https://doi.org/10.1101/202481)

Version of the preprint: 2

Authors' reply, 09 April 2018

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Sebastian Ernesto Ramos-Onsins](#) , posted 09 April 2018

This preprint merits a revision

In this second version of the manuscript the authors have answered the questions addressed by the editor and the reviewers and have made major modifications in relation to the first version of the manuscript. Consequently, this second version has substantially improved. Nevertheless, several points have not been completely answered, as they are indicated by the reviewers. Although the authors made a serious effort in clarifying a number of concepts in this new version of the manuscript, still some work is needed in this direction. One of the main points emphasised by the reviewer 2 is to clarify the concepts of modularity, genetic architecture and the definitions concerning aspects related with the environment. It is also my opinion that in several cases the vocabulary used is confuse for a non-expert reader. I am also interested in question 1 from reviewer 2 (Why do the authors choose 4 clusters in figure 2, rather than 3, or 5, or 6?); from my point of view, the number of clusters may be related in some way to the number of independent fitness components affecting the life of the species, then, the orthogonality of this components would be fundamental for discriminating different models of pleiotropy.

Please do not add more paragraphs in the text, if they are not strictly necessary, because the manuscript will become too long and difficult to read. Please answer carefully all the questions from the reviewers and modify conveniently the manuscript.

Minor comments from Editor:

-Table 1 and figure 1 are really helping to understand the experiment and the concepts used in this manuscript. Nevertheless, Table 1 is still unclear for non-expert readers. For example, the meaning of selectional pleiotropy does not say anything about pleiotropy but on the components of the fitness, it seems at least a partial definition. Also, the meaning indicated in Table 1 for antagonistic pleiotropy may be confused as the norm of reaction. -Section Methods, simulations. Define 1R and 2R in the text.

Dr. Sebastian E. Ramos-Onsins

Reviewed by [Tanja Pyhäjärvi](#), 06 February 2018

The preprint <https://www.biorxiv.org/content/early/2018/01/26/202481> has clearly improved due to revision. Conceptual Figure 1 and the introduction convey the message of environmental pleiotropy. The manuscript is now much more clear on the terminology of modularity.

Minor comments I would call linkage disequilibrium just LD, not "statistical LD". Isn't all LD statistical? Row 918 has an incomplete sentence: "See or more details." Add citation to 55 to Figure 7. Figure text 8 has the various simulated selection strengths listed, please add them also to the main text. Figure 2 is nice because a reader can now connect it with the conceptual Figure 1. However, Figure 2 has a lot of small details, text and numbers (in yellow) that are hard to see without zooming. In Figure 3, there is still a reference to figure 1G, that should probably be 2G.

Reviewed by anonymous reviewer 1, 06 February 2018

[Download the review](#)

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/202481>

Version of the preprint: 1

Authors' reply, 27 January 2018

[Download author's reply](#)

[Download tracked changes file](#)

Decision by **Sebastian Ernesto Ramos-Onsins** , posted 27 January 2018

Revise

This manuscript addresses an important subject regarding the differential effect of genes in diverse environments, their role in local adaptation and the interference between genes that are physical linked but have divergent behaviour in different environments. The authors aim to contrast the hypothesis that “evolution in complex environments should select for modular genetic architectures with limited pleiotropy among modules”. The manuscript uses innovative approaches in this field to analyse the interaction between the environmental information of each individual, the genetic linkage of genes and the genetic polymorphisms. The use of co-association networks and galaxy byplots are especially informative and show the main results of this work. Also, it is of crucial importance the simulation analysis performed in this manuscript to understand the possible expected patterns under different scenarios; this analysis includes three different demographic histories (isolation by distance from (i) equilibrium, expansion from (ii) single-refugia or from (iii) two-refugia) and explore the allele frequencies of loci under neutral and under positive selection, contrasted with information on a number of generated environmental variables.

The authors used as a model species for such analysis the species *Pinus contorta* (lodgepole pine). Using co-association networks identified several non-overlapping modules of genes associated with environmental factors (Aridity, Freezing, Geography and Multi, this last not clearly associated to a single environmental factor). Surprisingly, physical linkage was observed between genes associated with different climate modules but seem not to affect importantly the different modularity of genes to different environments. The results obtained here using novel approaches in this field may help to understand the effects of complex and heterogeneous environments over the genetic variability and over the gene interactions in the genome.

Nevertheless, the manuscript needs to be more clearly explained. Although there are many concepts extensively described in the introduction, it is often difficult to understand the meaning of sentences including words such as modularity, architecture and pleiotropy, which can have different meanings in different context. The two reviewers also coincide in the difficulty to understand some sections of the manuscript. I am especially curious about the simulation methods. I believe the methodology used for simulation, as well as the specific parameters used, should be included with more detail in order to facilitate the replication of the analysis. I am wondering about the possibility of including the effect of recombination in such simulations, which may also help to the interpretation of the data.

Finally, the two reviewers give an essential number of comments that the authors must follow before having the recommendation of PCI in Evolutionary Biology. Please answer all the comments of the reviewers separately in a separated text and detail all the modifications included in the manuscript. I encourage the authors to revise this manuscript following comments from this round of review and resubmit to PCI in Evol. Biol.

Reviewed by **Tanja Pyhäjärvi**, 05 December 2017

Review of Modular environmental pleiotropy of genes involved in local adaptation to climate despite physical linkage □Katie E. Lotterhos, Kathryn Hodgins, Sam Yeaman, Jon Degner, Sally Aitken □<https://doi.org/10.1101/202481>

The paper presents (to my knowledge) a novel approach to analyse environmental adaptations using genetic polymorphism and environmental data. The main idea is to inspect the modularity of genetic architecture by joint co-association network analysis of environment and genetic polymorphism. The authors use experimental data from *Pinus contorta* exome sequencing with climate data. In addition, they simulate multiple demographic, selective and neutral scenarios to test the behaviour of the method. The concept and evolutionary implications of multivariate nature of climatic variation were nicely explained. Also, discussion on the caveats of using the principal components (or equal approach) to summarize environmental variation's effect on biological organisms was important to bring up. The new approach is attractive and applicable to wide set of systems. Clarifying some aspects of the analysis would improve getting the main message through.

Major comments:

The preprint presents what I presume to be a reanalysis of exome resequencing genetic polymorphism data across 281 *Pinus contorta* populations from Yeaman et al. 2016 (Science). However, the authors should state clearly whether this is the same data or something that has not been published yet.

Because the modularity is taken into account for both gene groups as well as for environment, it was a bit confusing at some points which modularity was referred to. First I assumed that the paper only presents modularity in environment (as suggested by the title). However, the abstract presents modularity mostly for the genetic architecture. The basic idea of modularity of genetic architecture could be presented more clearly from the beginning (or in the title?) via an example or maybe even a figure presenting the both to make it easier to grasp from the beginning. Please explain clearly what is meant and is there a difference between: "evolutionary modules of loci", "selectional modularity of architecture", "modularity if selectional pleiotropy", "developmental/functional modules" and "environmental module" or are they essentially the same thing.

Why were SNPs that were already top candidates based on preliminary analysis only used in the co-association analysis? If the univariate model is likely to miss some signals of adaptation, why limit only to SNPs based on correlation with single environmental variable?

The authors state that the results are not sensitive to the distance threshold used in the clustering of the networks. What other thresholds were tested and why 0.1 was chosen as a threshold? Further, as the co-association analysis is in the core of the paper, it would be interesting to know, why the authors ended up using the division into 4 clusters. Are the conclusions dependent on the number of clusters? For example, the distance between "Freezing" and "Geography" is not very large and the clusters are actually quite similar based on visual inspection. Also, will the clustering be identical when structure corrected associations are used?

Minor comments:

Simulations were a great addition. Please give more details about the demographic simulations. For example, what was the strength and duration of selection in relation to demographic history in the simulations?

Exactly how many individuals were sampled per population? Is each of the 281 populations represented by single individual? If only one per population was sampled, how were the allele frequencies for Bayenv2 obtained? Do seedlots refer to a set of seeds from multiple trees, e.g. can it be assumed to be a random sample from the population?

How does your approach relate to the idea of analysing gene networks jointly when identifying the genetic basis of local adaptation that was presented e.g. by Daub et al. 2013?

Add citation to Hill and Robertson 1966

Structure correction is mentioned briefly, but what is the overall structure pattern in *P. contorta*?

"Across" repeated in Linkage disequilibrium part.

What do you mean by “SNPs from most genes associated with only a single climate module”? Based on figure 1, each SNP can only be associated with one network.

Could you present what kind of results would reject the Hypothesis of Modular Pleiotropy?

Replace “Thus, it..”, with “Thus, it is...” in Discussion.

How sensitive the method is to the choice of environmental variables or SNPs? Are the networks and modules highly dependent on single variables? Also, how many nuisance variables (environmental variation that are not selectively relevant) the analysis tolerates?

Figure 1a, text in the gray background is tiny and hard to read.

Figure 3, the shading of the quadrant does not reproduce in the printouts.

Figure 2, The among-group LD patterns are almost invisible in the screen and completely invisible in the print-out.

Tanja Pyhäjärvi, University of Oulu, Finland

Reviewed by anonymous reviewer 1, 04 December 2017

[Download the review](#)