Round #2 Decision

by Sebastian Ernesto Ramos-Onsins, 2018-03-20 15:31

Manuscript: https://doi.org/10.1101/202481 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.

We clarified these concepts at the beginning of the introduction and semantics throughout; see specific responses below. It has been difficult to use these terms in a way that is consistent with the way that they have been historically used in the literature, while making the manuscript accessible to non-expert readers.

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.

The number of clusters we choose for visualization does not change the number of co-association modules in the results, because the modules are based on drawing edges among SNPs based on a threshold in their similarity in associations. So, we could have chosen to plot the networks within 5 or 6 groups instead of 4, but the number of modules and the edges connecting them would not have changed. This is explained further in our responses to reviewer 2 below.

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.

The length of the revised manuscript is the same.

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.

We added some more text to clarify the definition. This definition it was taken directly from Paaby and Rockman 2012 and we prefer not to add or change the term as they defined it, given the substantial confusion already associated with these terms in the literature.

Also, the meaning indicated in Table 1 for antagonistic pleiotropy may be confused as the norm of reaction.

We clarified that the definition was referring to “antagonistic pleiotropy at a single locus” and clarified the definition itself. As referenced, Savolainen et al. (2013) defined the term in this way to describe fitness effects at a single locus across environmental extremes, and it has been widely used in this way in the landscape genomics literature. Reaction norms are used in many different ways in the literature. If a reaction norm is measured on a trait in different environments, it can represent plasticity of that trait. If reaction norms are used to measure fitness, then interactions can represent local adaptation caused by antagonistic pleiotropy.

-Section Methods, simulations. Define 1R and 2R in the text.

Done.

Dr. Sebastian E. Ramos-Onsins

Reviews

Reviewed by Tanja Pyhäjärvi, 2018-02-06 12:20

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.

Thank you!

Minor comments I would call linkage disequilibrium just LD, not “statistical LD”. Isn’t all LD statistical?

Done.

Row 918 has an incomplete sentence: “See or more details.”

The reference was accidently deleted, this sentences was fixed to “See [62, 63] for more details.”

Add citation to 55 to Figure 7.

Done.

Figure text 8 has the various simulated selection strengths listed, please add them also to the main text.

Done.

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.

We improved this plot by using black font that can be more easily seen, and listing number of genes involved in the larger environmental response modules. The small details are irrelevant to the take-home message, but are there in case the reader really wants to zoom in, and more visible in the supplemental plots.

In Figure 3, there is still a reference to figure 1G, that should probably be 2G.

Done.

Reviewed by anonymous reviewer, 2018-02-06 12:27

Summary:

The following is a review of a new version of the manuscript entitled, “Modularity of genes involved in local adaptation to climate despite physical linkage”. In general, this version of the manuscript is a nice improvement. In particular, the new introduction and especially the new Figure 1 are far clearer.

This clarity, and the improved understanding of the methods that it provides, bring up a few new technical questions. These are described below under ‘major comments’.

Additionally, there are some concepts that remain unclear, in particular, the abstract seems like it could only be understood by a very narrow group of biologists that are familiar with the last decade of literature on pleiotropy. Much of this issue is that important concepts, like modularity, are not explained. Further, terms that have particular meanings like, "genetic architecture," appear to used in a non-traditional way. This is extremely confusing. Most of this issue cannot be resolved by defining terms because these concepts are complex; they merit more detailed explanation than a simple definition. These issues are described below in detail under 'major comments'.

We respond to these suggestions in detail below.

The more I understand, the more I feel confident that this type of analysis, called coassociation networks, represents a very cool and useful way to investigate pleiotropy, and further, to define which environmental variables (or combinations of variables) are relevant to natural selection.

Thank you! We think so too, although we acknowledge that it has been difficult to explain and we appreciate your thorough comments on semantics.

Still, there are a few things that remain a bit unclear to me, and I hope my comments are helpful to the authors in improving the clarity of the paper to enable a larger collection of biologists to understand and benefit from their work.
Major points:
1. Why do the authors choose 4 clusters in figure 2, rather than 3, or 5, or 6? The authors hint at an answer to this question by stating that the number itself is not important, rather it is the "underlying clustering of the SNPs revealed by co-asssociation networks that is relevant". Still, could that underlying clustering be affected by the number of clusters chosen? It seems important to show that you get the same conclusions about the modular pleiotropy vs. universal pleiotropy debate regardless of how many clusters are chosen for study. Perhaps repeat with different numbers of clusters and show the results do not change? The discussion on lines 538 through 550 also hints at this issue, maybe some of this should be moved up, or there should be more explicit references to the decision about the number of clusters.

The number of clusters we choose for visualization does not change the number of co-association modules in the results, because the modules are based on drawing edges among SNPs based on a threshold in their similarity in associations. This threshold was determined by simulation as a number that enriched for edges among unlinked SNPs selected by the same environment. Our results were not very sensitive to increasing this threshold by 0.05 (e.g., the number of modules and number of SNPs per module were similar), but our results were sensitive to decreasing the threshold 0.05 (a stricter threshold resulted in modules of SNPs with extremely similar associations, and a large number of "modules" comprised of a single SNP unconnected to other SNPs, even SNPs in the same gene). Given that we showed with simulations how the threshold was chosen, we chose not to show these results in the main paper. In the discussion we simply encourage others to make sure their conclusions are not sensitive to this threshold and added some text.

So, we could have chosen to plot the networks within 5 or 6 groups instead of 4, but the number of networks and the edges connecting them would not have changed because they are based on the threshold. We re-arranged the section "Results: Environmental response modules" to explain this more clearly. The groupings only help to visualize the hierarchical nature of the modules.

2. Definition of modularity. First off, the definition given in table one is circular, i.e. the definition of modularity contains the word "modules". This is not right. The modularity concept is a complex one, and perhaps does not belong in the definitions table but rather in a few sentences, or even a paragraph of the introduction. As written, for a long time I assumed the authors were referring to some specific definition of modules, e.g. genes linked together on a chromosome. But what I think the authors are describing is a more general definition of modules, e.g. genes that all contribute to the same function or process for any number of underlying reasons. In this more general concept of modularity, the underlying reason does not matter, and the modules are defined by the processes to which they contribute, not by some molecular mechanism such as participation in the same gene regulatory network, or reliance on the same transcription factor. This should be described much more thoroughly and earlier.

Your general description of modularity is indeed the one we were aiming for, and you are correct in that readers who may not be familiar with the various ways "modularity" is used in the literature may be confused. As pointed out by (Wagner et al. 2007), modularity is "an abstract concept that seeks to capture the various levels and kinds of heterogeneity found in organisms." We expanded the definition of modularity Table 1, and revised two paragraphs in the introduction where the term is introduced.

3. Use of the phrase "genetic architecture". To me, this phrase implies something different from what the authors use it to describe. To me, genetic architecture refers to the location of genes on chromosomes. Therefore a modular genetic architecture is referring to the grouping of genes with similar properties nearby one another on a chromosome. This does not seem to be the only thing the authors mean when they talk

about modular genetic architecture. Even in table one, modular genetic architecture says nothing about the locations of genes on chromosomes but says something very generic. I would advise removing the phrase "genetic architecture" from the manuscript. The authors seem to use this phrase to imply something very broad. Perhaps a better word for what the authors mean is "pleiotropic structure", for example, Wagner uses the phrase "pleiotropic structure" in his paper, "the pleiotropic structure of the genotypephenotype map". The authors of the current manuscript seem to be asking whether the pleiotropic structure is modular, or not.

We use "genetic architecture" from the definition given by Hansen in his seminal paper "Evolution of genetic architecture", where he refers to the pattern of genetic effects that build and control a component of the organism (character, trait, or fitness). A description of genetic architecture includes statements about gene and allele number, the distribution of allelic and mutation effects, patterns of pleiotropy and epistasis, and recombination rates among causal loci on chromosomes.

While we do ask why the pleiotropic structure is modular, we also characterize the strength and direction of association of alleles with environmental variables, and the linkage among loci in different modules. Thus, Hansen's definition of genetic architecture seems more fitting. We added this definition to Table 1 and added some text to the second paragraph of the introduction to clarify our use of the term.

4. Definitions of complex environments, multivariate environments, intercorrelated environmental variables, aspects of the environment. A key contribution of this manuscript is a better way of defining, understanding and describing the aspects of the environment that influence evolutionary outcomes. **As humans, we can measure whatever parameters we can dream up, e.g. temperature, aridity, latitude, but the variables we define by measuring things may not correspond to the primary variables that matter for an organism's fitness. This distinction between environmental variables, and fitness 'aspects', is a really tough one to make. And it is a key part of what this manuscript investigates.** These ideas need to be described carefully earlier in the introduction and in plain English, not in fancy statistical terms. In particular, the abstract was really lacking clarity in this area. All it says is that, "intercorrelated environmental variables prevent testing of the predictions…". This statement by itself is so vague! The authors need to think carefully about what issues to tackle in the abstract in more detail, and which issues to omit to save space.

Thank you for these useful suggestions on how to explain these concept concepts in more understandable terms. We added/edited some text in the second paragraph of the introduction to clarify this, and much of the abstract has been re-written.

5. Abstract. As mentioned, the abstract was very unclear. I think that in the beginning of the abstract, the authors are trying to describe a problem: linkage among genes under different selective pressures can make natural selection less efficient. This problem should be stated more clearly.

The second sentence of the abstract is also unclear
because the authors posit a theoretical solution to another, separate problem that they never described. Why would complex environments select for modular genetic architectures? Why would complex environments select for high recombination rates? Why is pleiotropy costly? **It is because when genetic changes affect many traits, it becomes more likely that they will have a negative consequence on at least one.** This simple fact underlying the "cost of complexity" is never really stated. It should probably be stated in the abstract, so that readers can understand why a modular structure would be expected given complex environments. Also, what is a complex environment? How about something like, "**in order to test whether complex environments select for modularity, we must first define what a complex environment is. This is complicated because environmental variables, such as temperature, humidity, and soil composition, are often correlated such that seemingly large numbers of these variables can often be reduced to a single parameter.**" Please see previous major comment #4 for more thoughts on the abstract.
The abstract should stand alone as a clear description of the paper. In the results section of the abstract, authors use the words, "modules of genes". Does this refer to a set of genes that are physically linked? Modules often refer to genes that perform similar functions but are not necessarily linked on chromosomes. This was very confusing and it took me all the way through to the results section to really understood what was meant by "modules". I am still only 90% sure I understand.
What could this mean:" We observed limited evidence for environmental pleiotropic effects on distinct aspects of climate?" What is an environmental pleiotropic effect? What is a distinct aspect of climate? Even if this is defined later in a table, it is not OK to use that word here in the abstract without definition.
The abstract should be re-written using simpler concepts and terms.

We liked many of these suggestions and re-wrote the abstract in simpler concepts and terms.

6. Selection as an all-powerful force. Often times the text implies that natural selection is very powerful and that if something is favored by selection, it should indeed exist in nature. This is not true. For example, in small populations, genetic drift dominates selection. Here is an example of a case where natural selection appears to be described as an all powerful force, lines 553-555: "This was somewhat unexpected from a theoretical perspective, as selection would be expected to disfavour linkage and increase recombination between genes adapting to selection pressures with different spatial patterns of variation [34–36]." In this case, it is not exactly clear what should be expected from a "theoretical perspective". It is entirely possible that selection is weak in this population of trees and does not have the power to rearrange genomes such that genes experiencing similar selection pressures are physically linked. Typically, this kind of strong selection is only expected in microbes with huge populations sizes, and not large organisms with smaller population sizes. In general, the language should be changed to be more careful.

We re-worded this sentence to say: “This was somewhat unexpected from a theoretical perspective: while selection pressures due to genome organization may be weak, if anything, selection would be expected to disfavour linkage and increase recombination between genes adapting to selection pressures with different spatial patterns of variation.”

In general, we did not mean to imply that selection is an all-powerful force, only that theory sets up some predictions that we can test with data. If the loci we study are true causal loci, then selection was strong enough to reduce LD among linked loci in the genome. (See added sentenced in 1st paragraph of the discussion).

7. Relationships between the text and the figures. The figures are really beautiful and help explain the methods and the results more clearly than the text alone. Often I did not perfectly understand the text, but did understand the figure. In the minor points below, I highlight a number of cases where the overall clarity and impact of the paper could improve substantially if the text takes more time to describe the figures. For example, when describing the opposite associations in the multi vs. the aridity groups in figure 5, point the reader specifically to the green and red dots.

We worked on the main text describing Figure 5 substantially, and text relating to the minor comments below.

Also, the figure legends often seem to describe a specific panel of the figure (e.g. either the left or right side of Figure 1) and yet do not state that this is the case. This leads to erroneous statements about the figures. This should be checked; specific cases are highlighted in the minor points below.

We edited the legend for Figure 1 to address this comment.

Finally, the figures are often miscalled, for example, Figure 1G is repeatedly mentioned but does not exist.

We found and edited erroneous references to Figure 1G.

8. Discussion. Often the discussion makes statements that seem bold given the observations. I recommend spending more time connecting the dots: for example, if modularity exists, why does it imply that selection has acted? Lines 612 – 613 describe “drivers of selection”, without having made a clear connection between pleiotropy, modularity, and selection pressures.

Throughout the manuscript we tried to be careful in our interpretation, especially to remind the reader that our conclusions are based on associations.

We edited the sentence to read “Our results provide a framework for developing hypotheses that will help to disentangle the selective environments and provide genotypes for assisted gene flow in reforestation”

In the section of the discussion entitled: , we also state “These results are in line with theoretical predictions that modular architectures should be favored *when there are many sources of selection* in complex environments [26].”

Minor points:

1. The concept of pleiotropy, as well as the second sentence of the introduction, both need more explanation. As previously mentioned, the concept of the “cost of complexity” is a tough one. It is not well summarized by this statement: “Genes that have effects on seemingly unrelated traits may influence the rate of adaptation”. What the author means is: “Pleiotropy is when a gene has effects on multiple traits. Pleiotropy may hinder the rate of adaptation by increasing the likelihood that genetic changes have a deleterious effect on at least one trait.”

We replaced the sentence with the suggested wording.

2. The following sentence starts out clear, but becomes rather unclear. The word “in” does not seem appropriate here. Is this an error? “Because of the conceptual issues described above, certain aspects of genetic architecture have not been well characterized in adaptation to multivariate environments”

This clause has been highly revised to say “certain aspects of the genetic architecture of adaptation on landscapes have not been well characterized”

3. In the first part of the following sentence, it seems like the authors define modularity as a functional module, but in the second part, it seems like they define modularity as a physical module. Which is it? This issue should be resolved by responding to major comment #2 about defining modularity. Here is the confusing sentence: “particularly modularity (i.e., when mutations affect traits that are part of the same functional complex, Table 1) and recombination rates among genes in different modules (i.e., physical proximity in the genome).”

We reworded this sentence in response to major comment #2.

4. The phrase on lines 73 and 74 is a bit awkward “aspects of fitness”. Is it typical to refer to fitness as being comprised of multiple aspects? This will be resolved when the authors respond to major comment #4, above.

The wording here was not critical to the sentence, so we deleted it.

5. The figure legend for figure 1 does not explain what 1a and 1b represent. Are there 4 genes (i.e. 1 through 4) with two SNPs each (i.e. a and b)? If I have this right, then the following sentence in the figure legend is confusing because on the right side the aridity module consists of 4 genes, "For example, the aridity module is a group of SNPs within 2 genes adapting to aridity, and shows associations with both temperature and climate moisture deficit." Perhaps specify that this refers to the left column?
6. Also, the following sentence in the figure 1 legend is confusing for two reasons: (1) There is a typo (i.e. "are have"), and (2) it is unclear is this is true of both the left and the right examples. "In this example all SNPs within a module are have the same associations with multiple environmental variables."

The first sentence of Figure 1 explains: "each gene (identified by numbers) contains two causal SNPs (identified by letters)". We further clarified the figure legend and the sentence in question was modified.

7. This sentence on lines 210 - 212 is not clear: "Because SNPs in the 4 groups are more dissimilar to SNPs in other groups than the criteria we use to calculate modules, they would not be connected by edges in a co-association network." I do not understand what this sentence means. Does this support the modular pleiotropy hypothesis or the universal pleiotropy hypothesis?

The sentence is in reference to the fact that the number of modules would not have changed if we had presented the data in 4, 5, 6, or any arbitrary number of groups. This entire paragraph has been edited and revised to state this more clearly.

8. The figures referred to in line 235 and 239 are wrong. Figure 1G does not exist.
9. Figure 1G is also referred to in the figure legends, though it does not exist.

We found several erroneos references to Figures 1, which we fixed in the current version.

10. The statement on lines 272 - 273 needs unpacking. "-association networks capture the same information as simple LD-based clustering with the important additional benefit of linking LD clusters to likely environmental drivers of selection." Is this saying that all of the information you gained could have been revealed by another method? Or does this refer to only specific observations?

The first phrase of this sentence was edited to “The clustering in the networks correspond to what we would expect based on LD among genes, with the additional benefit...” If one clustered loci based on LD only, they would find similar clusters in the data, although they would not be able to link them to environmental variables.

11. Figure 3 should be better integrated with the text. Point to specific sections of figure 3 while describing the results in text to improve clarity. As of right now I am confused. Figure 3 shows that the red contigs seem to be in high LD. But the other contigs (yellow and blue) do not look at all like this. Given Figure 3, I do not understand the line on 291-293. Can you tell me what part of figure 3 demonstrates the conclusion on these lines? Also, these lines actually refer to figure 2, not figure 3, and I think this is a typo.

This was a typo and should read: “a large block of loci in low recombination”

12. The lines 304 to 307 merit more explanation. What does it mean these SNPS would have been missed? Please clarify the meaning of this sentence. “Overall, 80% of the geography SNPs, 75% of the Freezing SNPs, 20% of the Aridity SNPs, and 10% of the Multi SNPs were not outliers along the first 10 PC axes and would have been missed by a study based on PC axes.”

We added the clause: “A larger number of top candidate SNPs in our study would not have been found if the investigator did a PCA on the environments and looked for outliers along PC axes”

The section on lines 328 - 336 is very helpful in explaining why co-association is different than PCA.

Thanks!

13. The paragraph on line 346 uses the words “expected genome-wide covariance”, and I have interpreted this to mean that the expectation is based on whether environmental variables are themselves correlated. Maybe add more information to clarify, perhaps something like, “We can visualize (i) the expected genome-wide covariance (given correlations between environmental variables Fig 1A left panel) using shading of quadrants and (ii) the observed genome-wide covariance using a 95% prediction ellipse (Figure 4).

Great suggestion, we added the clause.

14. Maybe help the reader see the following pattern by pointing to it very clearly in text. Something like, “Galaxy biplots revealed that SNPs in the Aridity group showed

associations with hot/dry versus cold/wet environments, while SNPs in the Multi and Freezing groups showed the opposite patterns of associations with hot/wet versus cold/dry environments (Figure 5A; red vs green points demonstrate the opposite correlations). In general, more connections between the text and the figures would be very helpful.

We added details to this paragraph and more generally throughout the methods.

15. The results section should specify how the correcting the associations for population structure were performed in Figure 5B.

We added a clause to this paragraph "structure-corrected allele frequencies calculated with Bayenv2, see Methods"

16. In line 430, it seems premature to call these "adaptation clusters". The analysis on lines 370 through 428 is indeed very intriguing and suggests that these clusters of SNPs adapted to different selection pressures. But calling them adaptation clusters seems like a jump.

Agreed. We think "association clusters" is a more correct term.

17. In lines 578-588, the phrasing is awkward. For example, in the following sentence, the word plastic seems unfair to use: "However, differential expression analyses only identify genes with transcription (i.e., plastic) responses to climate". Plastic and plasticity have specific meanings in the literature, and it seems here those words are adopted for a narrow usage. This can cause a lot of confusion. In general, there is a lot of reliance on definitions in this manuscript, and the authors define terms that have a more general meaning in a narrow way.

Transcriptional responses fall under a subset of possible plastic responses that an organism can express, so we decided to use the more specific term "transcriptional plasticity."

Instead, can the authors try to describe things more simply?
For example, this sentence is so confusing: "We also compared the modularity of plastic responses to climate (as measured by co-expression networks) to the modularity of evolved genetic responses to climate (as measured by co-association networks)." Why not stick to the basics and instead say: "We also compared co-expression networks to co-association networks." I do not see any need to use the word plasticity here.

We used the suggested wording here.

18. The conclusions in lines 590 – 597 seems very strong, and not necessarily justified

by the data. Perhaps tone down the language here to make these suggestions rather than conclusions. Also, see major comment #2 regarding definition of modularity. It seems in the discussion the authors finally expand on this concept. Maybe this should come earlier.

We made some minor changes to the wording and rearranged some sentences. The language is speculative and the paragraph ends with only a suggestion that “these patterns warrant further investigation.” We also clarified the use of the term “module” at the beginning of the paragraph.

THANK YOU FOR THESE VERY HELPFUL COMMENTS! They greatly improved the accessibility of the manuscript.