SUMMARY of Modular environmental pleiotropy of genes involved in local adaptation to climate despite physical linkage

The authors present an important advance in how we think about and measure pleiotropy using co-association networks and galaxy plots. As a geneticist studying pleiotropy and someone who is deeply interested in the genotype to phenotype map, I am very intrigued by these results and novel methods. The figures are beautiful and well thought out. The section in the introduction about the debate surrounding how to define traits (cold, dry, cold-dry) is among the clearer parts of the paper. A major impact of the paper is the 'co-association network' approach, which solves the aforementioned debate by finding clusters of genes that define relevant traits (or in this case, relevant environmental responses).

A major problem with the paper is that it is often unclear and important sections are difficult to follow or seemingly left out. Because the authors are studying pleiotropy, they necessarily have to deal with multiple complicated topics at once, including (in their words), "the modularity of the architecture (number of distinct climate factors), overlap among modules, and physical linkage among loci". There are many times when I am not sure which of these topics is being addressed. Some of this confusion comes from terminology, for example, genetic architecture usually refers to genome structure or linkage, but occasionally the authors also use architecture to refer to the number of distinct climate factors. The word modularity can also be very tricky because it can refer to the number of distinct climate modules or the number of distinct genetic modules that respond to these climates. The terminology in this manuscript should be used more carefully.

The first sentence of the abstract is another good example of the unclearness: "Physical proximity among alleles shaped by different sources of selection is a fundamental aspect of genetic architectures critical for predicting their evolution". I did not understand this sentence at all. In addition to terminology concerns (e.g. genetic architecture is not yet defined), there are major grammatical errors that prevent a reader from discerning the object and subject of the sentence (e.g. are the allele frequencies shaped by selection, or is their proximity shaped by selection? which of these are you trying to predict?). Further, this sentence summarizes many different, difficult concepts simultaneously and it is unclear which is the main focus. It would be better to start with one central concept, and then step through the next concepts one at a time. i.e. "Are there modules of genes that respond to particular environmental changes? This is important to understand for reasons X, Y, and Z. We might be able to answer this question by looking at [describe data]. But physical linkage among loci can obscure this signal…"

Given this paper has the potential to demonstrate high impact findings, it is worth the time to restructure the manuscript, stepping through each topic more slowly and making clear to the reader (perhaps with numbered lists in text) what are the difficulties in measuring the extent of pleiotropy and how this method deals with those difficulties. Sometimes, instead of relying on words like "pleiotropy," perhaps it would improve clarity to talk about "genetic changes that affect multiple environments", and the genotype-phenotype map. Instead of staying so close to the terminology, and to the literature regarding modular pleiotropy, I wonder if the authors might use less technical terms to explain the motivation for this work, and its significance to a broader audience.

In sum, this manuscript presents a novel approach to studying the genotype-phenotype map, and also a novel way of thinking about and defining phenotype. I recommend the authors re-structure the paper, stepping through its important contributions one at a time to improve clarity.

INTRODUCTION

This sentence in the introduction indicates an omission of the literature: “Although there is emerging agreement that organisms have modular organization of genes in their effects on phenotypes.” There does not seem to be agreement at all and I think the authors should acknowledge this. The cited paper from Boyle et al 2017 did not acknowledge this and authors apologized over Twitter. There are many papers by Hill (1,2) and by Rockman (3) that do not support the modular pleiotropy model. Further, these are many “reply to” papers that highlight just how controversial this topic remains (1,4,5).

The introduction is a little bit unfocused. The authors review a great deal of literature and terminology, but the main message of the paper is a little lost. I think this is a difficult task because so many issues are tackled in this introduction. A single sentence early on explaining a single, simple important contribution might help a reader focus.

There seems to have been another study that generated some important data used in the current study. I don’t understand, even after re-reading the paragraph many times, what these data look like. This seems problematic, as these data are key to the current study. What are “candidates”? Are these “candidate loci” or “candidate trees”? How many are there? Why is convergent adaptation relevant here? This paragraph should start off with a comment about pleiotropy, and how these data will allow authors to study pleiotropy. Not all details from this dataset need to be described, but the details relevant to the insights about modularity and pleiotropy should be described here. How do these data differ from those used in previous studies of pleiotropy? Does this feature of the data give you more power?

Authors sometimes omit the word “frequencies” from the phrase “allele frequencies”. Authors should be clearer when they are talking about an allele itself versus its measured frequency across environments.

How does a co-association network analysis “characterize pleiotropy and linkage of the architecture?” What features does it use? Is this a new approach? What are its benefits? What novel insights might it provide? I understand that this is the introduction, so you don’t want to give away the results. But you must state something about this novel approach and why it is a good idea.

METHODS:

This is very difficult for anyone to read who does not study plants.

What is a seedlot?

How do you have needle tissue? I thought you only have seeds (“seed”lot)

Did you grow these trees in different environments? Or did you sample trees from different environments and calculate how allele frequency changes across environments? This part of the methods is really unclear.

Why did you obtain samples and extract sequence information if Yeaman et al 2016 already did this. Do you have different samples? What is the difference between your work and the Yeaman study?

The co-expressed genes section is also missing relevant details. What 7 climates? How many plants per cluster? How did you choose the plants in each cluster?

RESULTS:

In the first paragraph, again it is not very clear what is relevant about these top candidates in terms of pleiotropy or modularity. Why are you studying them? You provide information about how many there are before really explaining why they are in the paper.

In paragraph 2, again the motivation is not clear. The co-association method is described nicely in the methods. But I am left to infer what these SNP-environment associations mean and why you might want to cluster them.

At the end of the co-association network section, perhaps add a final sentence to summarize the observation. Which of your hypotheses was supported and what does this mean about selection and about pleiotropy?

I think the section on PCA is appropriate to include as it finally explains some of the benefits of co-association networks over other approaches.

The section on benefits and caveats of co-association networks also seems beneficial to include. However, it focuses on yet another impact of this study, identifying high confidence SNPs. It does not focus much on pleiotropy or modularity, these words are not even used. It would be very beneficial to pick a single contribution of this paper and continually come back to it in the text. It is not necessarily confusing to include multiple contributions and insights within a single manuscript, but one in particular should be highlighted as the main focus. Otherwise the paper reads as unfocused and the motivation/impact is unclear.

In sum, the co-association analysis is a novel and important idea. I hope with some re-structuring and decisions about how to focus the manuscript this idea will reach a broader audience.

1. Hill WG, Zhang X-S. Assessing pleiotropy and its evolutionary consequences: pleiotropy is not necessarily limited, nor need it hinder the evolution of complexity. Nat Rev Genet. Nature Publishing Group; 2012 Apr 1;13(4):296–6.

2. Hill WG, Zhang X-S. On the Pleiotropic Structure of the Genotype–Phenotype Map and the Evolvability of Complex Organisms. Genetics. Genetics; 2012 Mar 1;190(3):1131–7.

3. Rockman MV. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution. 2012 Jan;66(1):1–17.

4. Paaby AB, Rockman MV. Pleiotropy: what do you mean? Reply to Zhang and Wagner. Trends Genet. Elsevier; 2013 Jul;29(7):384.

5. Wagner GP, Zhang J. Universal pleiotropy is not a valid null hypothesis: reply to Hill and Zhang. Nat Rev Genet. Nature Publishing Group; 2012 Apr 1;13(4):296–6.