1. *ED: Manuscript ID EVL3-18-0066.R1 entitled "Parallel genetic evolution and speciation from standing variation" which you submitted to Evolution Letters, has been read by the three original reviewers. They are all impressed with the revision, as is the AE handling your paper. Some further revisions have been suggested, but I am confident that you can address these, and that the final decision can probably be made by the AE and I, without the need to go back to reviewers. I invite you to respond to the comments appended below and revise your manuscript. I'd also like to say how impressed I was with the effort that has gone into improving the manuscript and responding to the reviewers' comments. I am confident that this work will end up as an excellent contribution to Evolution Letters.*

2. *AE: We have received three reviews of your revised manuscript. All three reviewers expressed enthusiasm for the paper and were mostly happy with the revisions. I concur on both points. With that said, all three reviewers pointed out areas where additional clarification is needed (especially see comment 2 from reviewer 2). Once these issues are addressed, I think that this paper will make an important contribution to the field. I look forward to seeing the revised paper.*

3. *44, 65 Drop "It is increasingly clear" - the importance of standing variation has been known for a long time - it is obvious from the success of artificial selection, even on small populations!*

We have made the change.

4. *80-90: A clearer distinction should be made between the two causes of hybrid unfitness: being adapted to an intermediate haitat (and hence to neither divergent environment) vs loss of fitness due to cryptic divergnece, which releases variation orthogonal to the dorection of divergence. In the results, it would have been helpful to report the two components of RI separately.*

We have revised the paragraph and made the distinction clear. However, we emphasize again that some variation improved hybrid fitness and some does not. It is accordingly difficult to parse these components separately because while lag load does always reduce fitness, there really is no such thing as 'variance load' ~~in the way the reviewer is interpreting it~~. ~~This~~ would involve detailed analysis of individual hybrids which is beyond the scope of the present article.

5. *131 - Here, only the first mode of RI is referred to (see above).*

We believe our results show that ecologically-dependent post-zygotic isolation is a function of both the lag-load and the segregation variance. So, in our view, ecologically-dependent postzygotic isolation does in fact include both 'modes' of RI. We believe that the following text, present in the previous draft, should clarify this to readers:

"Because hybrids are recombinant, hybrid fitness reflects both the effects of displacement of the mean phenotype from the optimum (the 'lag' load) and what in diploids is known as 'F_2 hybrid breakdown' (Burton et al. 2006)." Lines XXX-YYY

6. *251 - Actually there seems some confusion on this point: Isn't orthogonal variance potentially "intrinsic" divergence?? I suspect that I have in mind a different interpretation of their model - imagining a complete phenotypic space with an enormous # of dimensions - whereas the authors are thinking specifically of traits that are somehow related to some environmental*

challenge. In that case, a dimension of only 5 or 10 might make sense - but would not capture the full evolutionary process.

We are slightly troubled by 'intrinsic' divergence because in our model it is possible for any phenotype to be favoured by selection. 'Intrinsic' divergence implies ~~to me~~ that ~~it~~ is truly 'environment-independent'… we are imagining out various dimensions as representing ecological traits like for example, limb length, that are under stabilizing ~~rather than directional selection.~~ None of the dimensions affect traits like 'viability' or 'sterility' which are by definition under directional selection ~~toward an unmoving optimum~~. We have added this to the text:

"We consider our model to be one of 'extrinsic' rather than 'intrinsic' isolation because we do not consider traits such as 'gamete viability or 'development' that ~~have~~ environment-independent ~~(i.e., global) optima.~~ Rather, traits in our model are more akin to ~~organism features~~ like 'limb size' or 'beak depth', under stabilizing selection ~~with no global optimum~~."

7. *260 - The reader needs a reminder that "genetic parallelism" refers to a specific statistic defined above - the mean fraction of the # of alleles fixed in both populations,.*

We have added the reminder text.

8. *404 - An unavoidable problem with empirical studies of multivariate divergence is that only a subset of the phenotype space is observed. Thus, one cannot really estimate the "angle" between divergence vectors. The reader needs to be reminded of this.*

This is a valid point and we have added a change to make clear that any estimate of angle / delta is made with error:

"Of course, phenotypic measurements are imperfect and typically non-comprehensive, and accordingly any estimates of angle and/or distance are necessarily made with some amount of error."

9. *Some parts of the main text are difficult to understand / misleading if one hasn't looked at the Appendix text. The discussion of hyperspheres in the main text really only considers the start of the adaptive walk (e.g., "Considering two populations, each with their own hypersphere, a given allele is beneficial in both—and thus could fix in parallel via positive natural selection—if it brings a population's phenotype into the region where the two hyperspheres overlap"; Fig. 2b). They only mention very briefly that no change is expected during the course of the adaptive walk but they don't specify. More detail is needed here; at least this part from the Appendix could be in the main text: "…adaptation's effect is a shrinking of the radii of the hyperspheres (at roughly equivalent rates in the two populations if adaptation proceeds relatively deterministically). Thus, because the fraction of overlap (Eq. A1) does not depend on the radii of the hyperspheres, the fraction of overlap is expected to remain constant throughout adaptation."*

This is a good point, thank you for bringing it up. We have added the recommended passage to the text.

10. *The authors say that segregation variance is beneficial when the angle is large, and that segregation variance along the relevant axis is reduced when adaptation is from SGV, which*

makes hybrids less fit in the case of SGV compared to new mutation. This seems logical. However, Fig. 2c shows that segregation variance at large angles is actually increased in the SGV case. As far as I understand they think that this is entirely due to other axes, but I think the whole section of the Results discussing this is not very clear; in particular, they don't graphically show the variance in all dimensions (It should be possible to add what is happening in the other dimensions to some version of Fig. 4).

We think the reviewer's point is a very good one and have added a set of panels to figure 4 showing the segregation variance in two of the dimensions ~~with an optimum at 0~~, where any segregation variance is maladaptive.

11. *Line 362: "This latter effect seemed to allow alleles with more deleterious pleiotropic effects to fix during adaptation from standing variation than when adaptation was from new mutation alone (Fig. 5C)." Is it correct that the alleles don't have larger side-effects in absolute terms,* sds*but rather relative to the effect in the direction of selection? I.e., essentially the overall increased segregation variance with SGV is due to the fact that a larger number of alleles (each with pleiotropic side effects) is needed to adapt to the new optimum? If this is true maybe it would be helpful to change the wording here.*

The reviewer has it right. We added a sentence after their quoted sentence to make this clearer:

"That is, populations initiated with standing genetic variation used a greater number of alleles—each tending to have pleiotropic side-effects—to adapt to their new optimum."

12. *Loci showing "genetic parallelism" should be enriched for ones with low levels of pleiotropy. This would also be interesting to look at in real data.*

We are not sure we agree with the reviewer that our paper says this. Parallelism is a function of *s* which can similar between a large-effect allele with some pleiotropy and a small-effect allele with little pleiotropy. ~~So~~ we think this is known already. We have elected not to include this in the discussion.

13. *E.g. line 426, Jones et al. 2012 citation: I think this is a bit misleading – as far as I remember, this paper was specifically searching for signatures of parallel genetic evolution, rather than getting an estimate of how important parallelism is relative to non-parallelism.*

This is a good point. Jones et al (2012) calculated the numerator (parallelism) of our genetic parallelism metric but not the denominator (non-parallel).

14. *The authors say that SGV counteracts speciation when the angle is small, and facilitates speciation when it is large. With a small angle, the effect results from the fact that the two populations tend to fix the same alleles, so it is crucial that the two populations adapt from THE SAME standing genetic variation. However, in the large-angle case, the effect seems to result from the smaller adaptive steps and more pleiotropic side-effects for alleles from SGV, so it might not be crucial that both populations have the same SGV – any SGV should increase isolation, when compared to new mutations. Is this correct? This is relevant for predicting what would happen in natural populations because in many cases, there might be a time lag between the split into the two "parental" populations and the change of the selection pressure, so that parental populations will have different SGV when the adaptive walk starts.*

This is an excellent point and we have included a new short paragraph in the discussion saying something similar:

"We also emphasize that the reason that adaptation from standing variation affects fitness differs under parallel vs. divergent selection. Under parallel selection, standing variation's ~~effect on~~ hybrid fitness is caused by parallel genetic evolution and therefore adaptation from standing variation is most likely to have an effect if populations start with *the same* standing variation. Under divergent selection, standing variation's ~~effect on~~ hybrid fitness is not caused by genetic parallelism but rather on the characteristics of adaptive walks from standing variation (more meandering & smaller steps). Therefore, our predictions about the effect of adaptation from standing variation on hybrid fitness hold *regardless* of whether populations have the same or different initial standing variation."

15. *Line 207: Unclear how* T *was defined exactly.*

We have made this clearer by giving the value for *T* used in the main text.

16. *Line 186, "Loci fixed in the ancestral population were also fixed in the parental population.": Does this mean these loci were considered in the following analyses, i.e. when calculating the extent of parallelism? I would assume that they should be discarded but this sentence seems to suggest that they were used somehow.*

Sorry this wasn't clear. The loci were indeed discarded (otherwise it would not be possible to get values of zero since alleles invariably fixed in the ancestor). We have modified the text to read:

"Loci fixed in the ancestral population were also fixed in the parental population, but were not considered when quantifying parallelism"

Reviewer: 3

17. *The authors have satisfyingly addressed my previous concerns, and have overall strengthened the points in their manuscript (which was already initially quite interesting) by further exploring robustness to several of their assumptions, and showing more results. I am now happy with this version, which I think will make a very nice contribution to Evolution Letters, and only have a few minor comments, mostly about form.*

We are glad the reviewer found our response and revisions to be of value.

18. *173: "was stable" -> "were stable"*

Made the change:

"Both the mean frequency of derived alleles and the phenotypic (genotypic) variance were stable (Fig. S9B),"

19. *174: It's a bit unclear what you mean with "Segregating derived alleles were all at unique loci". An allele is necessarily defined for a given locus anyway. Please be more specific. I think*

what you mean is that under the proper infinite site model, each polymorphic has no more than two alleles, and each allele traces back to a single mutation event.

We have added some new text similar to the reviewer's explanation (after the em dash below):

"Segregating derived alleles were all at unique loci by assumption—that is, each polymorphic locus has exactly two alleles and each derived allele can be traced back to a single mutation event."

20. *220-221: "We then we paired random individuals from the two parental populations to produce 100 recombinant haploid F1 hybrids". Does each of the 100 hybrids originate from a different pair of parents? Please specify*

We made the change, and also caught a typo (we then we):

"We then paired random individuals from the two parental populations to produce 100 recombinant haploid F_1 hybrids. Parents were sampled without replacement such that each could contribute to a maximum of one mating."

21. *338: I still think it is somewhat of a wrong terminology to speak of a "fitness valley" here, as you assess fitness separately in the environments of each parental population, for which there is a single fitness peak in both cases. In other words, the effect you describe would occur even if you assessed the hybrid load in only one of the parental environments, in which case it would make no sense to speak of a fitness valley. The reason is what you replied in your letter: the variance load is in fact not constant across environments with a Gaussian fitness peak, and phenotypic variance may increase fitness when the mean phenotype is far from the optimum. This would be a useful correction to some statements in previous papers you cite.*

OK, we take your point. We have removed the 'valley' term from the sentence:

"The general reason for this is that hybrids (but not the parents) have reduced fitness in each parental habitat as the angle of divergence increases, and some segregation variance is beneficial for mean hybrid fitness when the mean hybrid phenotype is far from either optimum (see Fig. 3C)."

Thank you for this. We believe there is some utility in maintaining the 'valley' terminology but making it more clear that the term only applies to the 'lag load' as this reviewer previously mention.

Reviewer: 2

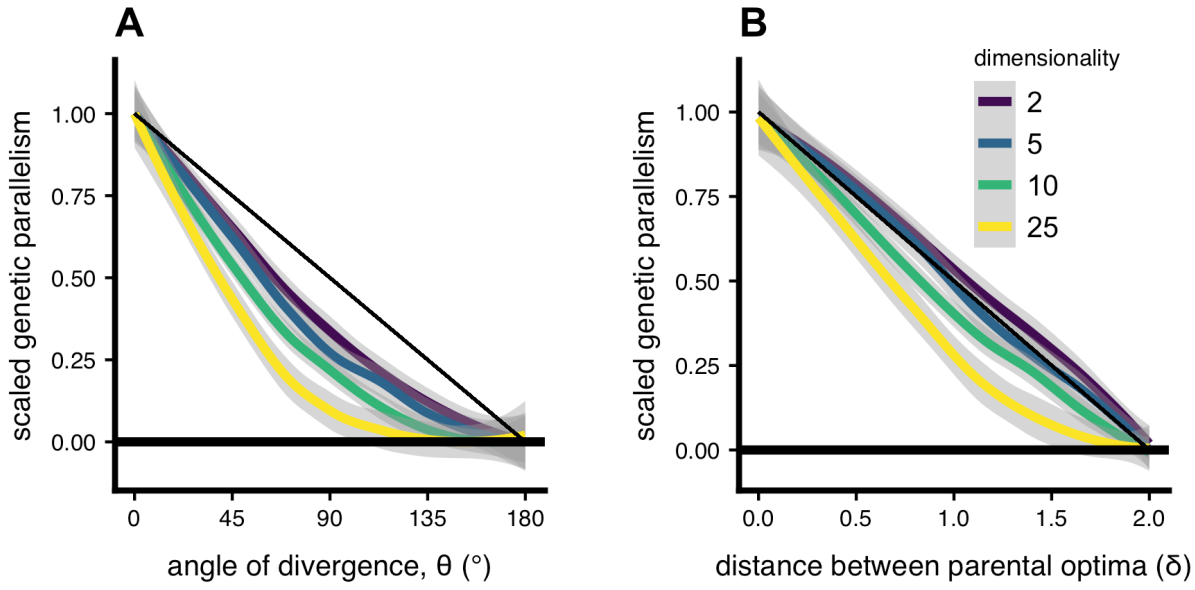
22. *The authors have done very extensive revisions and should be commended for their responses to the major comments that myself and other reviewers raised. They have made nice extensions and added new results, and I find the manuscript to be greatly improved and more clear than before. The figures showing parallelism explicitly as the proportion of same alleles are a very nice addition. They also pointed out a stupid mistake I made in my review (comment #1 below). I am very supportive of this manuscript eventually being published in Evolution Letters, as I think it is discussing a very important question, but I think there is still one major important issue to resolve.*

    In particular, strong statements are being made about simulation results regarding dimensionality and non-linearity of change in parallelism that either are not shown clearly

enough to warrant a strong statement, or in some cases, the simulations seem to not agree with the conclusions, which is problematic (see major comment #2). This is important because empirical studies of one trait that is parallel will commonly involve traits that have adapted in a non-parallel fashion, but that go undetected. This non-detection is more likely to happen for high dimensional adaptation (because not everything can be measured), and based on the analytical predictions in Figure 2A, it is also more likely to have a strong effect on results in such cases. Thus, we might see much less parallelism in genome scans in such high-dimensional cases, relative to low dimensionality traits where non-detection is less likely, and where the effect of any non-parallelism on genomic signatures of parallelism is reduced (at least, according to Figure 2A). This is a really interesting prediction, because many environments might be highly complex -- I like the example of stickleback adaptation to freshwater in the Discussion, as many dimensions of the environment such as parasitism, etc. might vary among lakes, and thereby greatly reduce parallelism. But it is critical to show that this prediction holds --- that higher dimensional environments actually do cause much faster decreases in parallelism, and this is why I think that nailing down point #2 is so important. This is presented in the discussion as a main finding (in the very first sentence), so I think it is important to be clear on this point. I would also note that this prediction is probably highly contingent on the assumption of universal pleiotropy, and that highly modular traits may show much more parallelism and lack of an effect of dimensionality, if modularity corresponds to the dimensions in fitness space that are parallel (point #4).

We greatly appreciate the reviewer holding us to account on strong statements that we made which were not ~~necessarily~~ supported by the data presented.

We believe that some of the ways the raw data ~~are~~ presented might have obscured some of the patterns. We present a couple of new figures in the supplementary materials that we believe demonstrate complete support for the prediction that parallelism decreases with dimensionality. We include the figure here to facilitate review. The difference between this figure and the figures referenced by the reviewer is that the scale of the y-axis is quite different between panels. We have scaled the y-axis here and also include new simulations in 25-dimensional space.



We absolutely agree that many processes will limit the application of this prediction to real data. We note that we did mention that our results were specific to a case of universal pleiotropy. We elaborate on ore in a revision, and write:

~~We also assumed universal pleiotropy and it would be valuable to extend our model to incorporate modularity.~~

23. *Major comment 1*
Thank your for the careful description of the hyperspheres and why they still overlap as divergence proceeds (my previous point #4), and I'm sorry for missing this obvious point. I knew this at some point in the past but for some reason was thinking about it backwards in the present – you were completely correct.

I am glad our explanation came across clearly, and, having made the same mistake ourselves, we hope that the revision you encouraged us to make will help future readers.

24. *Major comment 2*
I am still confused about the "faster than linear decrease" with dimensionality, and I would like to see some more clarity about how this depends upon both theta (angle) and delta (distance between optima) for the simulation results. This has the potential to be an important contribution of the manuscript, if clarified. It is critical to differentiate predictions (ie. due to the overlap in the hypersphere), where things are pretty clear, from observations of simulation results (i.e. observed amount of parallelism), where things are a bit unclear.

~~We have removed 'faster than linear' throughout the manuscript. More detailed responses below.~~

THETA:
In some cases, we might expect to see pronounced non-linearity (given Figure 2A), but we do not in fact see a whole lot of non-linearity in the simulations. Rather than showing non-linearity depends on m, Figure S3 suggests the opposite. Panels G, H, and I vary only in their dimensionality, but the amount of non-linearity is strikingly similar, and panel D (m=2) appears even more non-linear than panel F (m=10). The evidence here does not seem congruent with the predictions in a robust way, contrary to what is stated on line 292. There is perhaps a bigger effect of dimensionality on non-linearity in the decrease in parallelism for panels G, H, and I in Figure S2, but it's not nearly as pronounced as Figure 2A would suggest. How do you explain this? It is nice to see that the shape of the fitness function doesn't have a pronounced effect on the non-linearity of change in parallelism (Figure S8). But why are the simulation results about parallelism not agreeing with your predictions? It is clear that there is a more pronounced decrease in fitness variance for the standing variation cases with higher dimensionality (Figures S4-S6), but this appears to be also linear with theta.

As discussed above in response to query #22 we believe that some of the variation in scaling of the y-axis obscures the patterns. In addition, with the higher population size simulations we had to run them with fewer replicates because they unfortunately are very computationally intensive. In response to this concern, we have re-plotted the same data from Fig. S2 which we think are the best because there is little drift and yet they are fast enough that we can run many replicate simulations. The 'scaled' y-axis is directly comparable to the analytical model shown in Fig. 2B. From this figure we are fairly confident that the reviewer will be convinced of the agreement between our analytical model and the simulation results.

25. *DELTA:*
    I appreciate that the authors have added many more results about delta, as this is a quantity that is interesting for theoreticians. Unfortunately, I also find this issue a bit opaque where the manuscript discusses delta instead of theta. For example, on line 268: "This rapid decrease in genetic parallelism also occurs when the phenotypic distance between optima is used as the independent variable, although we note that non-linearity is only appreciable in higher dimensions (see Fig. S14)", but S14 doesn't show this directly: it is not clearly shown how the overlap in the hypersphere decreases with delta at different dimensionalities (prediction), nor is it shown that the increased dimensionality causes a more rapid decrease in parallelism with delta (observation). Figure S14A does show a non-linear relationship between theta and delta (black line), and also shows that the relationship between theta and fraction of overlap is non-linear, so we can extrapolate that the relationship between delta and the amount of overlap must also decrease faster with higher dimensionality, but it would be easier to see if this was shown directly (although this is a minor point). More importantly, the relationship in Figure S14B looks quite barely different than linear, and only one dimension parameter is shown (presumably m = 5?). In the caption of S14, on line 934 it is stated that "Accordingly, the nonlinearity emerges even when considering delta but only appreciably when considering higher dimensions (m > 5)", but this is stated not shown. Given the above lack of congruence between Figure 2 and Figure S3 (and to a lesser extent S1 and S2), I am concerned about this issue and think that if non-linearity is being discussed, it must be clarified. I realize that it has been de-emphasized in the current version, but I'm curious as to why we don't see an effect that is predicted by your analytical model in Figure 2A, and as mentioned above, I think it is very important to be clear about this.

We hope that the most important aspect of this criticism will be addressed by our new Figure XB, which we hope clearly shows that the non-linearity in parallelism with delta is indeed present and more appreciable in higher dimensions. If there is anything else the reviewer or editor thinks would be beneficial, we would be happy to include it.

> *SUMMARY of point#2: For both delta and theta, either a more non-linear decrease or a faster linear rate of decrease in parallelism with higher dimensionality would be biologically interesting, but I don't see this clearly from the simulations.*

We are grateful to the reviewer for their thoughtful comment. ~~As the reviewer notes we made an attempt to de-emphasize the 'faster-than-linear' language in the article.~~ As stated above, we hope the new simulations and presentation of the results address this main concern.

26. *Medium important comment 1*
    I like the discussion about Figure 5, in terms of adaptive walks being more "meandering" when there is standing variation. I would expect that such patterns are quite susceptible to drift, as very slightly beneficial new mutations would not rise above the drift threshold for a new mutation (s ~ 1/(2Ne)) but might contribute more readily when present as standing variation, as this threshold is less stringent for weakly beneficial alleles at higher initial frequencies. Perhaps some of the reason for lack of clear results in comment #2 is attributable to this?

~~We hope that the comment above makes it clear that the analytical model does hold generally~~. ~~It is beyond the scope of this paper to calculate the selection coefficients of mutations when they increase in frequency but we suspect the reviewer is right that more weakly selected mutations are~~

~~able to fix when present at high-ish initial frequencies in the standing variation. Indeed, this what we meant when cited the articles on this subject including 'Soft Sweeps'. Supporting this point further the fact that alleles fixed from simulations with standing variation have higher pleiotropy, and thus have a lower *s* threshold.~~

27. *Medium important comment 2*
    I would suggest that the most interesting extension to your work would be to consider non-universal pleiotropy. Modularity seems common in real biological traits, so this might greatly affect the amount of parallelism, and could mitigate some of the decrease in parallelism expected for high dimensionality traits from your prediction in Figure 2A. Not necessary to incorporate this here, but some comment would be useful to provide context.

We have added a sentence to the discussion singling out modularity as a key focus for future work. The excerpt from the text is included in response to query #22.

28. *I really like Figure 4, I think it really helps explain your results. But perhaps you should pick a case that actually has the dark green elipse become bigger than the light green elipse, so that the reversal in relative fitness from low to high theta that you show in Figure 3B is represented?*

~~We take the reviewer's point here. We think that removing the points from the panels and also showing the third and fourth trait dimension will make this much more clear.~~

29. *line 35: I think "more ecologically-dependent" would be more clear, as higher makes me think more fit.*

We've removed this sentence from the abstract.

30. *line 242: seems strange that phenotypic variance is near zero, as abundant genetic variation in quantitative traits is the norm rather than the exception. Is this a consequence of strong selection and/or weak mutation?*

It's a consequence of both together. Genetic variation in quantitative traits is high in nature because of many factors, such as frequency-dependent selection and ~~a constant environment~~. Since our simulations lack these ~~two key~~ factors ~~in the maintenance of genetic variation (and many others)~~ low genetic variation is the expected result.

31. *line 933: "For a given value of delta..." is not a complete sentence. I think it should say "for a given value of theta, delta is invariant with dimensionality"*

Good catch! We have made the change.

32. *Figure S17 states that standing genetic variance increases segregation variance relative to de novo mutations more at higher dimensionalities. But Standing variance appears to be relatively constant, and the biggest change appears to be in the reduction of segregating variance in the de novo mutation lines (green lines). I think this result is much more ambiguous than portrayed.*

The reviewer is right that we did not sufficiently highlight the decrease in mean 'per-dimension' segregation variance with dimensionality. We have added a mention to this result in the caption.

We were specifically focussing on the difference between the dark and light green lines and have clarified our description to accurately reflect the patterns in the figure panels.

33. *line 303-304: "Therefore, the extent of genetic parallelism from standing variation might be higher than what might be expected in a case where adaptation is from de novo mutation alone." This is consistent with predictions of many other papers (e.g. McPherson + Nuismer, Lee + Coop, and others).*

We don't think these two papers explicitly generate this hypothesis of comparing SGV to DNM. ~~We have decided to remove this sentence from the manuscript, however, since it was a bit speculative to be in the results section.~~

34. *I think figure 5 (and the similar one in the SI) is averaging across all angles of divergence? Are there any interesting patterns at the different angles? (not necessary to delve into this in the paper...I'm just curious).*

~~The reviewer is correct that the figure is averaging across simulations conducted at a variety of angles. However, the metrics reported are all for just a single population which accordingly does not have an 'angle' (because 'angle' only makes sense when talking about a pair of populations).~~