Dear Dr. Slate—

Thank you for offering to consider a revised version of our manuscript entitled, ‘*Patterns of speciation and parallel genetic evolution under adaptation from standing variation*’, for publication in *Evolution Letters*. We agree that the reviews were very thorough and have made our best effort to address them carefully and comprehensively.

In some cases we followed the reviewers suggestions exactly and they invariably made for a stronger manuscript. In other cases where we did not follow the reviewers comment we provide detailed reasons why below their comment. There were a few comments where the reviewers were not sufficiently specific to allow us to understand what they meant, and we have noted these throughout the response.

Some comments from reviewers were based on misunderstandings of our article (e.g., comments #X…). Our view is that these misunderstandings are a result of a lack of clarity on our part and so we have taken the opportunity to revise the manuscript to clarify areas where reviewers may have been off the mark.

While we appreciate the offer from reviewer #3 to disclose their identity we agree (and can sense from their comments) that this is undesirable unless communication mediated via the editorial board fails. We believe we have understood their main points and if they see our responses and are convinced that we misunderstood anything we would be receptive to their making direct contact with the corresponding author.

To facilitate future correspondence we have numbered all of the comments given by each reviewer which are included in italics (occasionally abridged). Where relevant we include pasted text from the revision in red font followed by line numbers.

1. *We have received three very thorough reviews of your manuscript. The three reviewers agreed that the paper was generally well written and that it addressed important questions in evolutionary biology (and I agree with the reviewers on both points). With that said, the reviewers raised a number of very important concerns that preclude publication at this time. Addressing these concerns will require additional simulations/analyses and rather substantial revisions to the main text. Please pay particular attention to the suggestions for more exhaustive simulations (coupled with cautions against over extending the results) and general suggestions for improving the simulations themselves or summaries of the simulations. I think this paper has a good deal of potential and I look forward to receiving the revised manuscript.*

We are grateful to the associated editor for offering us the opportunity to revise our manuscript. We have conducted many additional simulations and extensive revisions to the main text, as suggested, and feel that the review process has substantially improved the paper.

***### Reviewer 1 ###***

1. *This paper addresses an important question that has recently been gaining renewed attention. It is clearly written, and generally well argued. However, it considers only a very limited range of parameters, and does not exploit much of the relevant population genetics theory - especially, on the maintenance of genetic variance under stabilising selection.  It would require very substantial revision, with a good deal of additional simulation and analysis, to be substantial enough for publication.  Evolution Letters does have a compact format, but nevertheless, it would be possible to make a much deeper and more reliable analysis within those constraints.*

In our revision we have attempted to thoroughly address this main criticism of the reviewer. We have responded to the specific comment about limited parameters in comment #7. We have responded to the points about population genetic theory as they were raised, and also conduct new simulations in a different regime of mutation-selection balance (see comment #…).

1. *How could one test any of this?  This is not easy to answer, but one would like to see some discussion of possible approaches - either experiments, or comparisons across populations, or estimates of key parameters.*

We have added a section to the discussion called ‘outstanding empirical questions’ that addresses this comment. Due to the length of this section we refer the reviewer to Lines XXX-YYY of the submitted manuscript.

1. *"parallel adaptation" is used to mean adaptation using functionally equivalent alleles. It could also mean, using functionally different alleles at the same locus, or functionally identical alleles, but with different mutational origins. It would be worth making the specific definition used here clear at the start.*

We have changed the language throughout the manuscript make this clearer. In particular, we are more careful to reference parallel ‘phenotypic’ evolution when referencing phenotypes and parallel ‘genetic’ evolution when referring to genotypes.

For our model, where there are only two possible alleles at a given locus, parallel genetic evolution is via the same locus and non-parallel genetic evolution is via different loci. In line Bolnick et al. (2018), we argue that the latter scenario represents convergence rather than parallelism. Specifically, we now write (lines XX-YY):

An infinite number of loci is assumed, such that each mutation arises at a new locus and therefore each locus can only have two alleles (‘infinite-sites’ sensu Kimura [1969]]. Accordingly, parallel genetic evolution in our model cannot occur via recurrent mutation at a given locus but only through parallel changes in allele frequencies.

1. *83 -  "populations diverge genetically by chance in the course of adapting to similar environments" is broader than "'mutation-order' speciation - divergence could also occur by drift based on standing variation. I think that it is confusing to use the term "mutation-order" speciation, when that is not necessarily what is meant.*

We have changed the language throughout our revision to emphasize that we are focussing specifically on ‘speciation via parallel natural selection’ rather than ‘mutation-order speciation’. We have kept the ‘keyword’, [but explain how our results are broader?], because we suspect our results will be of interest to biologists who study ‘mutation-order’ speciation.

1. *Fisher's geometric model is usually associated with successive  fixation of mutations.  Wouldn't stabilising selection on  an additive trait be a better term, and less likely to mislead?*

We have changed this language, although our thinking is shaped by Fisher’s model our approach is distinct. The section on Fisher’s model is no longer in the article, although we have kept the ‘keyword’ [to credit our inspiration and connect with this body of literature, which does not always consider successive fixation and does sometimes model standing genetic variation (e.g., Anciaux et al 2018 Genetics)].

1. *A very limited range of parameters are used: a fixed mutation and selection, fixed numbers at beginning and end.  Thus, it is more a demonstration that a specific point - "adaptation from standing variation forestalls speciation via parallel selection and promotes speciation via divergent selection" - may hold in principle. Yet, the abstract gives the impression that this has been established in general.*

We intentionally considered only a small set of parameter values in our original submission because preliminary investigations revealed our conclusions to be fairly general and given the compact format we wanted to focus on the main conclusions.

In our revised manuscript we follow the reviewer’s suggestion and conduct simulations over a greater range of parameter values. In particular, we independently vary dimensionality, the strength of selection [equivalently, the size of mutations], and population size [and thus the rate of input of new mutations].

This has resulted in several new figures, which we include in the supplementary materials and discuss in the main text where relevant. We find, in general, that our qualitative conclusions are unchanged by variation in parameters.

1. *Keeping the number of dimensions at 2 for most analyses seems quite unrealistic: there needs to be an examination of behaviour as m increases to very large values; this will have a strong influence on the outcome.*

In response to the reviewer’s comments we now conduct simulations in 5 and 10 dimensions, in addition to the original 2. We chose 5 and 10 following Chevin et al. (2014). We present results in the new supplementary figures (see response to comment #7).

1. *A specific issue here is that the distribution of allele frequencies is crucial to the outcome, but this is not investigated directly.  In particular, it is not clear whether, under the chosen parameters, variation  is due to rare alleles ("House of Cards"), rather than a continuum of common variants.  At minimum, the simulation results need to be compared with the substantial body of theory on this question.*

In the original submission we presented the number of derived alleles, the mean frequency of derived alleles, and the distribution of allelic effect sizes in the ancestral standing genetic variation (Figure S1). To better characterize the ancestral standing genetic variation we now also show the site frequency spectrum (Fig. SX). Together these four plots give a great amount of information about the nature of the ancestral standing genetic variation assumed in our study.

From this figure, it is clear that allele frequencies are more akin to a continuum of alleles (Gaussian, rather than house of cards). We briefly describe this result as it relates to the emergent segregation variance and discuss Turelli, Kingman, and Slatkin & Lande. NEEDS WORK

1. *Barton (1989) derives results for the rate of speciation under stabilising selection towards a common optimum. That was for a two-allele model, but results may be similar if the simulations here are in the rare allele ("HoC") regime (which is not clear).  This needs some discussion, and ideally, quantitative comparison.*

We now discuss our results in the context of population genetic theory more thoroughly. Barton (1989) considers a case where two populations experience stabilizing selection and diverge due to drift. Eventually, the populations become so diverged that unfit recombinants are produced. The model is not directly comparable to our central analyses because in Barton’s case the populations do not undergo adaptation and since the majority of differences accumulate during and not after adaptation the individuals accumulate RI more rapidly. Under his parameters Barton finds that the reduction in fitness of the F2 is 0.1 relative to parents.

Should we conduct some simulations to compare? Can generate sims with barton’s values on p.72?

1. *The simulations are rather inefficient, in that a large # are culled in each generation. It would save electricity to draw parents in proportion to their fithess, as in a standard Wright-Fisher model.*

This is a good suggestion and has been implemented in the simulations conducted in our revision. We also hold population size constant (see #12)

1. *Populations may go extinct, or suffer severe bottlenecks. Does this happen often? It would be cleaner just to fix population size, and avoid introducing the additional process of density fluctuation. (And, more efficient - see above)*

Under the conditions we explored in the original submission, it was the case that populations rarely went extinct. This is a good suggestion and population size is now set at a fixed size.

1. *139 - Fraisse et al is not the appropriate reference here - one of Lande's early papers on multitrait stabilising selection would be more appropriate.*

We now cite Lande (1979) instead of Fraïsse.

1. *-Separate populations that remain at high population size throughout will slowly diverge over time, even with a constant optimum.  The reduction in N from 10000 to 1000 will increase the rate of divergence due to drift between alternative "adaptive peaks".  Really, one would like to know how the rate of divergence (eg measured by segregation variance) depends on population size. There are simple theoretical predictions for this, which could be used to check the simulations.*

We show in a new supplementary figure how divergence (FST and segregation variance) changes between populations that are held at an optimum for population sizes of 100, 1000, and 10000 (see Fig. SX). We find that the reduction from 1000 to 100 has a large effect, but the reduction from 10000 to 1000 does not have much of an effect. We are not sure which theoretical predictions would be most appropriate or relevant here, but would be happy to make any further comparisons recommended by the reviewer.

1. *Wright's "shifitng balance" should be cited, since the model here is essentially the same.*

We now cite Wright’s shifting balance theory when we introduce our general simulation framework.

1. *161 - The full allele frequency distribution needs to be given, not just this brief summary.  Also, make clear that the 192 alleles are necessarily all at different loci (by assumption).*

We now state plainly here that alleles are all necessarily at different loci:

Ancestral populations had, on average, a mean of … segregating alleles—each at a different locus (by assumption; Fig. S1A)—and each segregating allele was present in, on average, 7.3 % of individuals (Fig. S1B). (Lines XX-YY)

Regarding the allele frequency distribution, we have plotted this now in Fig S1D.

1. *165 - The procedure for sampling different amounts of standing variation, by sampling n alleles, seems artificial. Surely it would be more natural to sample a population of a certain size from the ancestral population? Or, vary the mutation rate in the ancestor?  Also, need to state how observables such as genetic variance relate to n, and hiow fast n increases back to equilibrium due to mutation.*

The reviewer is right to note that this approach is artificial. However, we use this artificial approach to provide us with more control while controlling for various other possibilities. If we were to take the reviewer’s first suggestion of sampling a population of a certain size, then we would have varying population sizes confounded with the quantity of standing variation (see this reviewer’s point # 12), which is undesirable. If we were to vary the ancestral mutation rate, we argue this is also unrealistic because it’s not plausible that mutation rate could differ considerably between ancestral and derived populations. In sum we believe that our sampling of *n* alleles from the ancestor accomplish a realistic feature of colonizing population—variation in how many alleles are present in the founders—in an artificial manner. If the reviewer has a specific suggestion that would not confound founder population size or have a different mutation rate between the derived and ancestral populations, we would happily implement it.

Regarding genetic variance and *n* increasing to equilibrium, we now plot the phenotypic variance of populations sampled randomly from the ancestor with varying *n* in Fig. SX.

With respect to *n* increasing to equilibrium due to mutation, regrettably we do not understand the reviewer’s meaning here. The populations are subject to strong selection after the simulation is initiated and then held at an optimum. In our Fig. S2 we plot a number of variables that change as a simulation reaches equilibrium, perhaps the answer to the reviewer’s question in there. If not we would be happy to conduct additional analyses to address this point.

1. *226 – [i]One would like to see an analysis of segregation variance in the direction of divergence, vs in orthogonal directions. [ii]Similarly, fitness can be separated into components due to deviation of the mean from the optimum, and variance around the optimum, and each of these into components associated with the axis of divergence, vs the rest.*

We do consider this solved already by Chevin et al. 2014: divergence is greater along the axis of parental divergence than orthogonal. While we agree that the fitness of hybrids can in principal be partitioned in this way, we are less clear what the partition means for speciation than a simpler comparison of mean/max hybrid fitness. Accordingly we have not made these changes but would be happy to with further clarification from the reviewer about the reasons to do so.

We have made some changes in response to the second point and now compare fitness consequences due to ‘displacement’ and those caused by variance (i.e., ‘lag’ load); see response to Comment #45, and new Fig. 4A. (dashed line).

1. *Fig 2b doesn't show the intercept at distance 0 - yet, there will be some rate of divergence at d=0.  Also, it would help to give the value of d relative to something measurable, such as the genetic sd or the fitness loss at distance d.*

We originally decided not to plot the value at *d* = 0 because we thought it could be the case that very high amounts of standing variation could cause lowest segregation variance if the results are random and would distract from the biology (this is because there would be very low variance among ‘n’ values whereas selection causes a particular ‘n’ to be deterministically lowest for d > 0). We now plot the full distribution. Instead of plotting the calculated value of *n* that minimizes segregation variance, we now plot the estimated value of *n* that maximizes parallelism (the global maximum of a loess fit).

We now describe the *x*-axis of this plot as the number of mutational standard deviations away from the origin. For example *d* = 0.1 is 1 mutation size SD, and *d* of 1 is 10 mutation size SDs. We make this clear in the caption of Fig. 2 (Lines XX-YY).

1. *376 - There is a hint here at interesting results for higher dimensions - why are these not presented in full?*

In our original submission we were attempting to keep our results to a scale conducive for an *Evolution Letters* article. In response to comments from this reviewer and others (see comment #7), we now present results for a 2, 5, and 10 dimensions. In the main text, we also present all the main results for five dimensions—instead of the original two—that make the interesting results more obvious. Because the results are qualitatively consistent across dimensions (see new supplemental figures) we still present the numerical results for only *m* = 5.

1. *428 - What could a dimensionality of 10 for bacteria actually mean??*

This is a very good question. Practically, it would mean that there are 10 ‘independent’ traits. That is if you were to measure all of the ‘traits’ in a bacterial population (by, for example, measuring the number of transcripts of all genes; see Rockman 2012 The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*), the number of principal components explaining appreciable variation would be 10.

We spend more time describing dimensionality and the general importance in the article in the ‘genotype to phenotype’ section. We hope that this addresses concerns about what dimensionality actually means and how it could be measured. The new text reads:

We focus on the case of five phenotypic dimensions (*m* = 5) in the numerical analyses presented in the main text. Practically, dimensionality can be thought of as ‘organism complexity’ (Orr 2000) and quantified as the number of principal component axes describing the phenotype if all traits could be measured (Tenaillon 2014). Empirical estimates typically range between 1 and 10 for model microorganisms and nematodes (Tenaillon 2014) but is almost certainly higher in taxa such as arthropods and vertebrates where estimates are lacking. Future studies might consider using standardized high-throughput phenotyping of ‘arbitrary traits’ (e.g., the number of transcripts for all genes in a genome in a laboratory cross; Rockman 2012) to begin estimating dimensionality. (Lines XX-YY)

**### Reviewer: 2 ###**

1. *The authors have tackled an important question in evolutionary biology: how does adaptation to similar vs. divergent environments affect parallelism and reproductive isolation. The manuscript is clearly written and uses a nice design for the simulations, and I think it could make an excellent contribution to the literature. I think showing how parallelism scales with initial standing variation and examining whether parallelism decreases faster than linearly with angle and with dimensionality are both very important questions to resolve. However, I have some important concerns about the interpretation, generality, and accuracy of results that need to be addressed and some suggestions for expanding and clarifying these results. Most important is the issue of whether the decrease in parallelism is faster than linear with divergence in optimum. I have heard second-hand from people that have seen this work presented at a conference that this was an exciting component of it, so it is critical to be crystal clear on this point. On this broad point, I am concerned about a potentially inaccurate statement in the appendix about equation A1 (point #4), and whether it is accurate to extrapolate beyond the assumptions of this derivation, which forms the basis for the analytical results. Regardless of whether the interpretation of the analytical results needs to be scaled back, there are several ways that the analysis of the simulations should be reconsidered to strengthen our understanding here: analyzing the data using Euclidean distance instead of angle (point #5), assessing whether the curved fitness function contributes to non-linearity by rerunning simulations with a linear fitness function (point #5), and considering an alternative measure of parallelism rather than segregation variance (points #1 & #2). Finally, the effect of scaling of segregation variance with n in the main simulations (figure 2A) is quite slight and I'm wondering about the biological significance of this (point #6), and this might be more clear with an alternative metric of parallelism such as F2 hybrid breakdown or proportion of parallel alleles. It is possible that some of these above points stem from my misunderstanding of the way the model works, and if this is the case it might help to clarify it to avoid this confusion.*

We are grateful to the reviewer for their detailed comments on our article. Regarding the reviewer’s main concern we hope that our detailed answer in response to item # X resolves this issue. It is somewhat counter-intuitive and we followed the exact same thought process when we first were developing our models before realizing that it really is only ‘angle’ that matters (if both populations have the same initial distance to the optimum and approach their respective optima at the same rate). We have responded to all queries in detail below.

1. *It appears that "segregation variance" is just capturing phenotypic variance that occurs in the F1 generation? (lines 216-218). F2 hybrid breakdown is also interesting and should be considered here, as this is greatly increased under high amounts of genetic redundancy. Perhaps I don't understand something here about how these hybrids were made (and if so, please ignore these comments and clarify the presentation of the methods), but it would seem to me that F2 hybrid breakdown will scale much more strongly with non-parallelism than F1 segregation variance. Also F2 hybrid breakdown will tend to be greatly exacerbated as further time progresses and populations under stabilizing selection independently fix combinations alleles that keep them on their local optimum. By contrast, segregation variance is only affected by the mean of the two parental phenotypes, and can be predicted independently of the actual genetic details. See Chevin et al. 2014:*

Because we are considering a haploid model, the F1s actually do present ‘hybrid breakdown’. This is because F2 hybrid breakdown is a result of genetic segregation, which occurs in F2s in diploids and F1s in haploids. We think this comment stems from a lack of clarity on our part and so in our revised methods we have expanded on the relevance of segregation variance in haploid F1s. So since the reviewer suggested that ‘hybrid breakdown’ will scale with non-parallelism, and our metric does this, we believe our metric is relevant.

Regarding segregation variance being ‘greatly exacerbated as further time progresses’, we show in Fig. S6 that this generally is not the case. Rather, it expands rapidly during adaptation and then plateaus (although we do note it will always tick upward at a slow rate due to populations fixing alternative compensatory mutations).

Fundamentally, segregation variance is not causally and solely predicted by the parental mean phenotypes because we allow for genetic parallelism. Chevin et al. 2014 considered only unique mutations and therefore in their model the genetic details do not matter, but they do in ours. We have tried to make this more clear in the methods:

Because hybrids are recombinant, hybrid fitness reflects both the effects of displacement of the mean phenotype from the optimum and what in diploids is known as ‘F2 hybrid breakdown’ (Burton et al. 2006). (Lines XX-YY)

We do not include the reviewer’s detailed explanation of transgressive segregation and F2 hybrid breakdown in this response letter, but note that, using their scheme, under parallel evolution where two populations fix the same number of alleles (assume 3) in chevin’s model the populations are necessarily +++--- and ---+++. Whereas in our model the populations can actually be +++--- & +++--- or any combination in between (e.g., +++--- & +-++--).

We fully acknowledge that a diploid model would behave differently than a haploid model and did try to explain this a bit in our original submission. We now elaborate further, saying (lines xx-yy)…

1. *Regardless of whether I understood how you calculated segregation variance, it seems that a missing step in the results is to relate the proportion of parallel loci to segregation variance, as the latter is being used to assess the degree of parallelism without direct links being made, and there is lots of talk of parallelism without explicit results showing it. Even if it appears intuitively simple that this scales in some cases, this should be clearly shown, as many people are now studying the extent of parallelism in the genomic basis of adaptation. I would like to see some results shown for what proportion of loci are actually parallel vs. non-parallel. There are a variety of metrics that have been developed for analyzing this (Pearson's r, Jaccard, etc), or you could simply report the proportion.*

We hope that our response to query #23 has fully resolved the possible misunderstanding.

That said, we agree with the reviewer that it would be worthwhile to illustrate the proportion of parallel loci. We explain the new approach on Lines XX-YY:

And we plot the result in the main text (Fig. X).

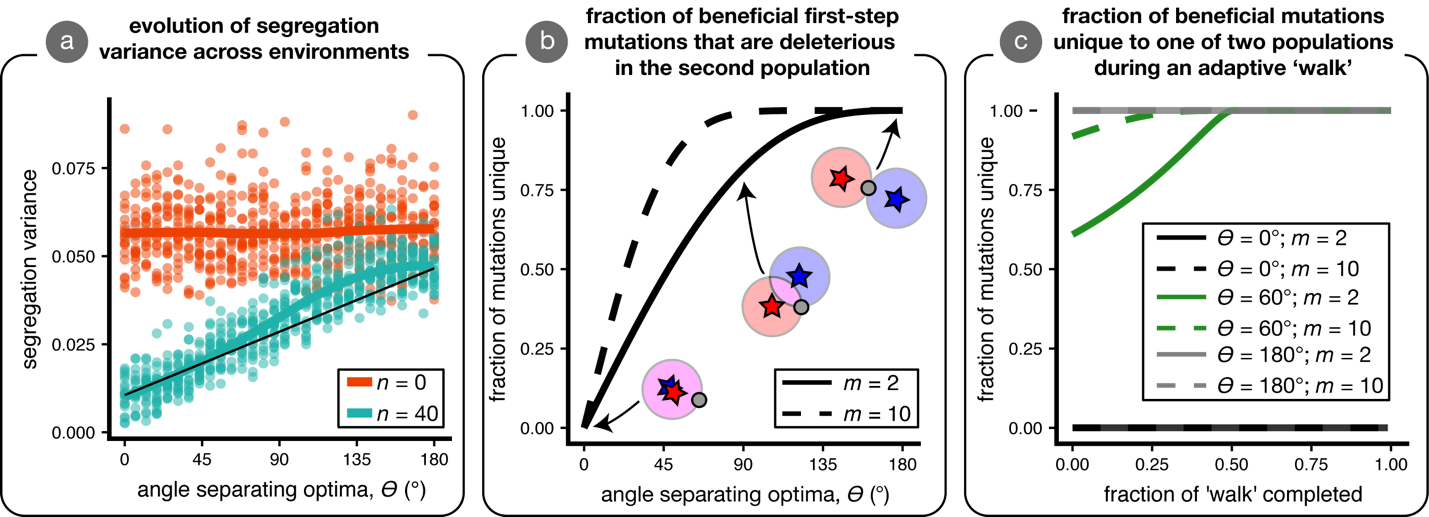
1. *I do not understand where equation A1 comes from and why an incomplete regularized beta function is relevant here. More description of this derivation would be helpful, preferably with a more descriptive formula that could be used without going to a textbook. Also, on line 682, the sentence appears to relate to the formula that is a function of cos on line 681, and m does not enter into this formula, but is mentioned as "written as a function of m,...".*

This will be for MATT.

1. *I think the statement on lines 722-724 is incorrect:  
     
   "In fact, because the fraction of overlap (Eq. A1) does not depend on the radii of the hyperspheres, the fraction of non-overlap is expected to remain constant throughout adaptation as long as ϴ is constant"*

*I think this statement is only true when the two populations are at the same point in phenotypic space, at the outset of the simulations. As adaptation progresses, the radius of each hypersphere shrinks, and the absolute magnitude of overlap will necessarily decrease. Imagine a case where theta is some intermediate value, say 90 degrees. There is some overlap at the outset, but eventually, once the populations are vanishingly close to their respective optima, their respective hyperspheres become very very small. It is impossible that there would be any overlap in the hyperspheres at this point, so the proportion of non overlap would be 1 (and the change in this proportion would have been gradual as adaptation progressed). Thus, the ratio of overlap WILL change as a function of the amount of adaptation (distance of the mean phenotype along the line from origin to optimum). I think this misunderstanding occurs because equation A1 assumes that both populations are at the origin, but it does not apply to cases when some evolution has occurred and the populations are part way towards their respective optima. I think this whole paragraph should be deleted as it is unhelpful, and it should just be stated that more complicated models are necessary for the case where the populations have started to diverge, which could be part of the next paragraph.  
  
Given that this is incorrect, it is important to be clear where these results are discussed on lines 302-309, as I do not think it is correct that "this pattern is expected to remain constant over the course of an adaptive walk and regardless of whether populations have the same or different initial phenotype" (line 307-309), or at least if this is true, equation A1 does not demonstrate this because it is only an accurate description when both populations have the same phenotype, at the outset. It is fine if the analytical model only applies to the initial state of the simulations, as the simulations can show what happens for more complex cases (it would be nice to simulate a more complex trait with m = 10).*

We thank the reviewer for this comment but believe the reviewer is mistaken in their assessment that our interpretation is incorrect. We begin our response by illustrating the concept the reviewer is describing analytically. See the figure below:

**a**

fraction overlap

In this figure, we illustrate the fraction of overlap (as in the original submission’s Figure 3B) over the course of an adaptive walk. As the reviewer notes, it changes and necessarily goes to zero for angles other than zero, and is especially rapid for higher dimensions.

However, the reviewer is not correct that this matters for genetic parallelism. To make this clear, imagine two populations that start at different locations in phenotype space, for example , a population with very large limbs, and a population with very short limbs such that the phenotype of population one for limb length is a value of ‘5’ and the phenotype of population two is a value of ‘1’. Imagine if the ‘red’ population below is long-limbed and yellow is short-limbed (X

**🡨 SHORT | Limb length | 🡪 LONG**

If selection were to favour an increase in limb length for both populations, say to increase it in value by 0.5 such that the optimal phenotypes are 5.5 and 1.5 respectively, a mutation (blue arrow) that increases limb size by 0.25 would be favoured in BOTH populations, even though there is no ‘overlap’. This is because when assessing whether a mutation is favoured in both populations each population must be projected onto the same space.

**🡨 SHORT | Limb length | 🡪 LONG**

We hope that this pictorial and verbal augment has convinced the reviewer that their concern is not a problem for our model.

1. *lines 285-288: if this "faster-than-linear" result is potentially partially caused by scaling between angles and Euclidean distance, why not plot it as a function of Euclidean distance between the optima, and more clearly show the effect you are aiming to demonstrate? Distance between the optima rises faster at first and then drops off as theta increases, so it might be linear when re-plotted as a function of this distance. The same thing applies to line 303. Does the effect of dimensionality on the deviation from a linear relationship also arise because of how angles work in higher dimensions?*

We decided to plot angle rather than Euclidean distance for several reasons, and have not made the suggested change in the main text of the article.

At low dimensions what the reviewer says is true. However, at high dimensions this scaling is less important. We include a new paragraph to explain this. We write:

To summarize the above: genetic parallelism decreases rapidly with *θ* for two reasons. The first reason is that optima move away from each other most quickly (as *θ* increases) when *θ* is small (the ‘*angle’* effect). This is not affected by dimensionality because two fixed points are equally distant from one another if they are on a line (e.g., [-1], [1]) as when they are contained within a 3D volume (e.g., [-1,0,0], [1,0,0]). The second reason that parallelism rapidly decreases with *θ* is that the reduction in the fraction of overlap between round objects (circles, spheres, hyperspheres) moving away from each other along a line (where *δ* changes at a constant rate) occurs most rapidly at first and slows as the distance between their centres (*δ*) increases (the ‘*round’* effect). This reduction in overlap occurs more rapidly in higher-dimensional space because for all values of *δ* except 0 (where overlap is complete; *θ* = 0°) and 2*d* (where overlap is zero; *θ* = 180°) overlap is less in higher dimensions. Therefore, at small dimensions the ‘angle’ effect is primarily responsible for the rapid decrease in parallelism with *θ*, but the ‘round’ effect is increasingly important as dimensionality increases (see Fig. SX). (Lines XX-YY).

While fully acknowledging that Euclidean distance is an important and complementary way to view the results, we would like to keep the focus on angle rather than Euclidean distance in our article for three reasons. First, angle is constant across values of ‘*d*’ and therefore comparable, whereas Euclidean distance is not. While it’s possible to standardize Euclidean distance between the optima (𝛿) between 0 and 2*d* (0° and 180°), our second reason for wanting to keep the focus on angle is that we believe that 𝛿 is generally less intuitive to the average reader. We do plot the results with 𝛿 in the supplementary material. Last, empirical biologists use angle and not 𝛿, and so keeping the y-axis as theta will make our work more accessible to empiricists. We include all the raw data associated with this article and so any interested reader can explore angle and Euclidean distance.

1. *Another thing to consider is that you have used a curved fitness function (Gaussian), which may induce non-linearity in some way (hard to imagine how, but it's worth discounting this as a supp mat figure).*

The analytical derivation does not include any assumptions about the fitness function and therefore does not affect the observed nonlinearity. What this asks is simply: what proportion of possible beneficial mutations that are beneficial in one population (i.e., have a positive selection coefficient) are also beneficial in a second population?

The reviewer is absolutely correct that assumptions about the fitness landscape affect the probability that a mutation will fix in parallel given that it favoured in both populations. This is because—assuming stabilizing selection—the fitness landscape affects how beneficial a given mutation is but not whether or not it is beneficial.

We do not aim to provide a comprehensive theory of parallel genetic evolution here and assumptions about the fitness landscape would certainty have to be factored in to do so. As would assumptions about the distribution of fitness effects of new mutations. This will be a valuable follow-up to the present article.

In any case, we now show the results similar to what we plot in Fig. 3A and Fig. 4B, but for a linear and Gaussian fitness function. The qualitative patterns are similar (see Fig. SX.)

1. *lines 258-262: while I agree that this is intuitively what should be expected to occur, the results do not really show this clearly. The increase in segregation variance for n > green arrow in Figure 2A is quite small. I would suggest showing this more clearly with the proportion of parallelism in the actual alleles contributing to divergence and variance? (point #2). Given the small magnitude of the increase in segregation variance with higher "n", I am concerned about the biological importance of conclusions being drawn from Figure 2B, which I think is the basis for statements about parallelism decreasing at very high and redundant initial levels of standing variance? I think this would be more clear if a different response variable was used instead of segregation variance. I suspect, for example, that F2 hybrid breakdown would increase much more strongly with n with high redundancy.*

We thank the reviewer for this suggestion. We have included a new panel with the proportion of mutations that fix in parallel in two populations (See response to comment #X)

Eventually, the segregation variance and fraction of non-parallelism for n > green arrow will approach that which is observed under divergent selection, but this depends on a number of parameters such as the strength of selection, population size, and dimensionality. We plot some other cases in new supplementary figures (see Fig. SX).

1. *Instead of number of mutations in Figure 2B, could you express this as phenotypic standard deviation in the initialized population, or as this sd as a ratio of total distance to evolve?*

Phenotypic standard deviation of the initialized population could work but we have instead taken a suggestion from reviewer 1 that is in line with this reviewer’s second point (see response to comment #1). We note that we now plot the relationship between *n* and the phenotypic segregation variance (proportional to SD) as a supplementary figure (Fig. SX).

1. *Line 187: the angle between the trajectories towards the optima (theta) is not ever measured in practice as it is impossible to know what is optimal in nature, but this can be guessed at from  measurements of realized phenotypic differences and their angle of divergence. However, it is possible that empirical measurements of the angle of phenotypic divergence don't capture the true angle between the optima, if some pleiotropic costs prevent a population from reaching the optimum.*

It is of course true that the phenotype of a population may not be optimal. However we note that it is actually the realized angle of divergence that affects parallelism rather than the fundamental angle of divergence. In the reviewer’s terms: it is the ‘angle of phenotypic divergence’ that matters, and not the ‘true angle between optima’. In our simulations, these invariably overlap to a substantial degree to many simplifying assumptions such as low dimensionality, free recombination between all loci, etc. So we argue that this is not a problem for our analysis. We note this issue in our methods when describing the utility of *θ*.:

We note that empirical estimates of *θ* from phenotypic data capture the realized divergence and not necessarily the angle between selective optima. However the realized evolutionary trajectory reflects which alleles did or did not fix and therefore which may have evolved in parallel. In addition, the observation that selection is typically weak in natural populations (Kingsolver et al. 2001) suggests that most populations are fairly close to an optimum. (Lines XX-YY)

1. *line 202: if the populations continue to diverge indefinitely (and this divergence presumably contributes to further reproductive isolation), then this isn't really an "equilibrium".*

While this is technically true, it would take a substantial amount of time for our conclusions to change appreciably. See Fig. S6 in original submission: the lines at *t* = 2000 are effectively horizontal. We have made a change to the text in our revision:

Genetic divergence between populations increases indefinitely after they reach their optima, but their mean phenotypes are stable and the phenotypic segregation variance changes little as time progresses. Therefore, our conclusions refer to quasi-equilibrium conditions rather than transient states and are unaffected by standing variation’s influence on the speed of adaptation. (Lines Xx-YY)

1. *line 225: I think you mean "initial standing variation (n)" here?*

Fixed.

1. *For Figure 2B, I think that each point would correspond to the green arrow in 2A? This could be stated more clearly.*

We have revised the caption to make this clearer.

1. *Line 681: Is it perhaps simpler to represent this formula as 2\*d \* sin(theta/2)? Then there are no exponents and it is more directly connected to a simple algebraic representation whereby (theta/2) is the angle, the distance from midpoint between optima to one optimum is "opposite" and d is the hypotenuse.*

The reviewer is correct, we’ve re-written the formula.

1. *Figure 3B: I think this is using equation A1? If so, it should state this.*

We now reference equation A1 in the Fig. 3 caption.

1. *Given the cursory description of the derivation of the equation A1, which is much more important, there is too much time spent on more obvious stuff later in the appendix (e.g. lines 705-714 could be deleted)*

While we appreciate the reviewer’s suggestion to remove the text we can’t help but feel that the misunderstanding in query #26 could have been remedied by a clearer explanation here. Therefore rather than delete the text we have taken the opportunity to revise it extensively.

1. *I think the loci are probably additive for phenotype but this should be stated somewhere in the methods as this greatly affects the dynamics (mentioned in passing on line 407).*

We note that we mentioned this on line 130 of original submission, but have made it more clear in the revised text by adding the following in the first methods paragraph:

Our simulations consider a case of stabilizing selection a multivariate phenotype determined by numerous additive loci. (Lines XX-YY)

***### Reviewer 3 ###***

1. *…your assumption of infinite site model (146-147) seems to imply that there cannot be the typical mutation-selection balance, where an allele is recurrently introduced by mutation and removed by selection, since there is no recurrent mutation at any locus. Then what does mutation-selection balance (160 and later) mean in your model? What determines the frequencies reached by each allele in the ancestral population? How can alleles that appear only once by mutation reach appreciable frequencies unless they are positively selected? And how do your results compare to models that focused on mutation-selection drift equilibrium for allelic frequencies under a fixed optimum (such as Barton 1989, Genet Res)? This information should appear somewhere in your ms.*

In our simulations, when we say ‘mutation-selection balance’ we mean that mutations are being introduced into the population (via *de novo* mutation) at the same rate as they are being removed (via selection). We now make this clearer and write:

Ancestral populations reached mutation-selection-drift balance such that the rate of acquisition of new mutations was balanced by the rate of loss of mutations that arose in earlier generations and the mean frequency of alleles was stable (see Fig. S1). (Lines XX-YY)

Mutations can rise to high frequency via drift because ancestral selection is weak and the population is finite. In addition, mutations are positively selected in the ancestor if they compensate for deleterious alleles that previously rose to high frequency via drift. We now specify this:

Some alleles reach high frequency likely because they compensate for deleterious mutations that reached high frequency due to drift (Hartl and Taubes 1996; Orr 2005). (Lines XX-YY)

Our results generate the same qualitative patterns as plotted in Fig. 7 of Barton 1989 Genet Res (thank you for suggesting this paper); compare to Fig. S6 in our original submission. Specifically, Barton found that…

1. *Then when you write that you run simulations “with different amounts of ancestral standing variation (via n)” (247-248, and also in the methods), it is unclear whether n is a parameter that you set (ie you choose to pick only n of the segregating alleles), or whether it emerges natural from your random draw of phenotypes (but then some values of n would presumably be missing). Please specify.*

We mean that *n* is a parameter we set: we pick only *n* of the segregating alleles. We now make this more clear in the methods:

Founder populations were fixed for all alleles that were fixed in the ancestor, and the amount of standing variation was determined by sampling a pre-determined number—the parameter *n*—of segregating alleles from the ancestor. Populations adapted from only new mutation when *n* was set to 0 and otherwise contained some ancestral standing variation. This admittedly artificial sampling procedure gives us more control over the amount of shared standing genetic variation in the adapting populations, removing noise and thus allowing us to better see the effects of shared variation (Lines xx-yy).

1. *And when you write that a “randomly generated population was duplicated and used to found both adapting populations”, does this mean that you didn't use different random samples of genotypes for each adapting population? Please make this very explicit.*

The reviewer has it right: we used the exact same samples of genotypes for each adapting population. We now make this more clear: there are no ‘founder effects:

This randomly generated population was duplicated and used to found both adapting populations which accordingly possessed the exact same sample of genotypes. Therefore, we consider a case of adaptation from standing variation where there are no founder effects. This biologically unrealistic scenario thus reduces stochastic effects, allowing us to better observe the direct effects of shared standing genetic variation (lines xx-yy)

1. *I was a bit confused by your report of a strength of stabilizing selection of s\_anc=0.01 (158), while the value in Table 1 is s = 1 for the adapting populations. At first I thought it was a mistake, then I realized it means that you use 100-times weaker selection in the ancestral population than in adapting populations. I imagine weak selection in the ancestor was used to allow sufficient (nearly neutral) standing variation to be maintained prior to adaptation. But then you need to make it very clear that you assumed that the strength of selection is dramatically increased in the adapting populations (essentially all selection coefficient are multiplied by 100); and provide some justification for why this assumption was made. If I am right that s = 100\*s\_anc, then I think this is a third important feature of the new optima, alongside the other two that you list (181).*

The reviewer is right that we simulated very weak selection in order to generate sufficient standing variation. We made this explicit in our revision:

We made the assumption of weak stabilizing selection in the ancestor because it allowed for the accumulation of appreciable standing variation that was used to found the focal populations. (Lines XX-YY):

We also specify explicitly that we assumed selection in the ancestral environment to be quite strong relative to the ancestor, and in response to comments from another reviewer we conduct simulations for several values of *s* (see Fig. SX).

1. *It is not entirely clear (at least in the main text, e.g. 276) which aspect of divergence you keep constant when you change the angle. I imagine it’s the radius (=divergence with ancestor). This needs to be specified very clearly, so that the reader understands how the angle determines the phenotypic difference between parental populations.*

The reviewer is correct; the radius (parameter *d*) is held constant as the angle changes. We make this more clear in the results and discussion section before moving into describing the simulation results:

Therefore, the value of *θ* is what determines phenotypic differences between parental populations in these simulations (as d is held constant). (Lines XX-YY).

1. *In fact, your focus on the angle as the main description of divergence is meaningful here because you assume that each parental population diverges equally from the ancestor, such that both parental populations lie on the same sphere centered on the ancestor. But this is a special case, and more generally one population may have diverged much more than the other. At one (quite plausible) extreme of this scenario, a single population may have diverged, while the other has stayed in the ancestral environment with the same optimum. In any case, when divergence from ancestor differs across parental populations, the divergence between parents, degree of non-overlap, segregation variance, and so on, are not only determined by the angle.*

This is an excellent point and we fully acknowledge that this is a special case where we vary theta holding d constant. We note this in our revision (Line XXX). It is possible to vary the angle and distance independently. Along with the reviewer’s suggestion we were inspired to solve the case at the other extreme: what happens when one population diverges in any direction and the other population is held constant (this, as the reviewer notes is varying d with constant theta).

Mathematically, this is a difficult problem that has fortunately already been solved. We want to solve for the volume of an ‘*n*-ball’. We explain this in our revised appendix (lines x to x) and present the results in a new supplementary figure, Fig. SX.

We also note that the situation considered here is a special case where the total quantity of adaptive divergence (*d*) in each derived population is equal and held constant while the angle (*θ*) between them varies. It is possible to imagine a different case where the angle between populations is held constant but the degree of divergence (d) changes. Indeed, equation A1 is solvable for any combination of *d* and *θ*. In the supplementary material (Fig. SX) we show how genetic parallelism is expected to decrease for populations undergoing parallel evolution (*θ* = 0°) but with different degrees of adaptive divergence (i.e., varying *d*). (Lines XX-YY).

1. *One of your main results is that when the angle of divergence is large, hybrid fitness in the parental environments is lower when parental population adapt from standing genetic variation (as compared to de novo mutations), even though segregation variance is lower for the former (352-363, 441-442, Figure 4). This is quite interesting, but I don't think your explanation for this result (358-363) is the correct one. There is no disruptive selection per se in your model (unlike implied by the mention of "fitness valley"), since fitness is assayed separately in each of the parental environments, which both include a single optimum phenotype. Instead, what happens under divergent selection (large theta, towards 180°C) is that, in addition to the segregation variance load, there is also a load caused by deviation of the mean phenotype from the optimum in the parental environment (so-called lag load, or mismatch load). The reason why this interacts with segregation variance is that, as the phenotypic variance increases, the fitness landscape (log mean fitness against mean trait) becomes broader, which reduces the lag load, for a given mismatch of the mean hybrid phenotype from the parental optimum. And reciprocally, for a given angle (which fully determines the mean mismatch in your model) decreasing the variance load (as occurs under adaptation from standing variation) should increase the lag load. Then whether overall relative fitness of hybrids increases or decreases when adapting from standing variance depends on how the lag load compares to the variance load, which itself depends (for a given angle theta) on the divergence of parental populations from their ancestor.*

*An important consequence with respect to your conclusions is that statements like: “reduced segregation variance […] reduces mean hybrid fitness under divergent selection” (441-442) is probably not always true. Whether this holds depends on the width of the fitness peak and the distance of the mean hybrid phenotype to the optimum. I think your statement only holds when the deviation of the mean hybrid phenotype from the parental optimum is larger than 1/sqrt(s), i.e. the width of the fitness peak. This can be proved using a well-known result (going back at least to Lande 1976) for models of Gaussian stabilizing selection, namely that the effective strength of stabilizing selection on the mean phenotype is, using your notation, s'= 1/(1/s + P), where P is the phenotypic variance. Note that s' is a decreasing function of P, so increasing the phenotypic variance reduces the fitness cost of maladaptation of the mean phenotype (but increases the variance load). The lag load is then easily computed using s' and the deviation of the mean hybrid phenotype from the optimum (in the parental environment), which is just half the final phenotypic divergence between parental populations. Note also that in the context of your model, for a given angle this divergence between parental populations depends on their divergence from their ancestor (radius of the sphere). So in summary, there should be a condition on the radius of the sphere relative to the strength of selection s, which determines whether the fitness of hybrids is higher or lower when adapting from standing variation.*

*Lastly on this issue, I think it would be more satisfying in terms of explanation to partition these two components of hybrid fitness (lag load and variance load), as only the former is strictly caused by segregation variance (constant across environments, and thus described as intrinsic load in Chevin et al 2014), while the latter is caused by environment-specific deviations of the mean phenotype from the optimum (extrinsic load). It is an interesting point that segregation variance also modulates the extrinsic effect of mean deviations from the optimum (as explained above), but I would find the analysis of the process much more illuminating if these two components were partitioned.*

We thank the reviewer for this detailed explanation and we have made some changes in response. First respectfully disagree that the explanation of a ‘fitness valley’ is misplaced. Imagine two sympatric beetle populations that specialize on willow or birch, respectively. Hybrids may not be able to perform well on either. Thus even if individuals can choose which environment (tree) to reside in, hybrids may have lower mean fitness than either parent.

We also think it’s important to emphasize that ‘variance load’ in the strictest sense does not always exist in our model. Variance load does exist when the hybrid population is on the optimum and there is therefore no lag load. Therefore any and all variance is ‘load’. However, under divergent selection (actually any angle other than 0°), as the reviewer accurately notes, whether segregation variance is beneficial or a ‘load’ depends on the amount of segregation variance.

Our new simulations occur under conditions that allow us to evaluate the reviewer’s suggestion that segregation variance improves mean fitness under divergent selection only when *d* > 1/sqrt(*s*). Holding *d* at 1, we use an *s of* 0.1, 1, and 10 which gives *d* < 1/sqrt(*s*), *d* = 1/sqrt(*s*) and *d* > 1/sqrt(*s*). Therefore if this holds we would expect our result to not occur under weak selection and occur most strongly under strong selection. Our new simulations reveal that … FILL IN AFTER SIMS FINISH RUNNING

With respect to partitioning ‘lag’ load from ‘variance’ load we agree with the reviewer but make one small difference: we think it would be most useful to plot the ‘lag’ load against the observed hybrid fitness, which differs from the ‘lag’ load only because of variance (which may or may not be a ‘load’). Accordingly, we plot the observed lag load of each for the relevant conditions in Fig. 4A. Essentially, for each simulation we calculate the mean hybrid phenotype and then determine its fitness. See the dashed line in the revised Fig. 4A.

We note that in our original submission we do say that, as the reviewer says, our conclusion about segregation variance and mean hybrid fitness under divergent selection is not expected to be universal – at higher dimensions greater segregation variance can be deleterious (line 373-377). We revised our manuscript and reference new simulations conducted in response to reviewer at higher dimensions… We hope that our additional simulations under a wide range of parameters …

1. *About the contribution of the angle of divergence to parallel evolution from standing variation, the whole story is in fact probably more complicated than your explanations (289-309), as it also depends on the sequence of mutations that fix.*

*Even when the angle of divergence is sufficiently small that there is effectively perfect overlap (same set of genes have beneficial alleles in both populations), which alleles end up contributing to adaptation may differ deterministically between populations. Imagine two alleles (at two different loci) that are both beneficial in the two adapting populations, but with different selection coefficients across populations. These selection coefficients, together with the initial frequencies of the alleles, determine which allele fixes first in each population, changing the selective pressures on alleles at other loci, and thereby the overall genetic architecture of adaptation.   
  
More generally, the selection coefficient of a given mutation affecting a trait changes in time as the mean background trait value evolves in response to selection (e.g. Chevin & Hospital 2008 Genetics), so which allele starts spreading first has consequences for the genetic architecture of further adaptation, and the probability of parallel genetic evolution. This surely contributes quantitatively to your results, in addition to the more qualitative definition of sets of genes with alleles that are beneficial in both populations (proportion of overlap).*

The reviewer is absolutely correct that variation among populations in the order of fixation contributes quantitatively to our results. This is one of the reasons why we observe considerable variation under parallel evolution in the segregation variance under parallel evolution (original submission, Fig. 2A) despite a lack of founder effects.

Regarding parallelism of alleles with positive but different selection coefficients in each of two populations, we acknowledge this implicitly in the original submission lines 310:314

We avoided a more thorough explanation in our submitted manuscript due to space constraints but now include several sentences explaining the reviewer’s point:

Even under conditions that appear ‘ideal’ for genetic parallelism (i.e., *θ* near 0°; no founder effects) there is still quite a range of variation (observe high variation of points around black fit line in Fig. 2A). This is the result of many factors, including stochastic variation in which mutations fix in different populations. Non-parallelism in the early stages of adaptation leads to differences in the selective co-efficient of the same alleles in both populations, further promoting non-parallelism. (Lines XX-YY)

*Minor points*

1. *25: “speciation via environment-specific hybrid fitness”. This wording is a bit strange, rather "reduced hybrid fitness", as in the introduction.*

We have made the change here and throughout the manuscript.

1. *77-88 (and more generally in the introduction and the whole ms): please make it clear that you here only focus on post-zygotic isolation, which is but one mechanism contributing to ecological speciation.*

We have made changes throughout the manuscript and for clarity of presentation here do not indicate line numbers.

1. *115-116: this sequence is misleading, as the model was originally developed by Fisher (1930) mostly to argue conceptually in favor of micro-mutationism, rather than “to make inferences”.*

We have removed discussion of Fisher’s geometric model from this section.

1. *139: if bold denotes a vector/matrix, then s should not be bold, as it is a scalar.*

We have made the changes.

1. *145: “Each mating pair is a diploid individual” reads awkward. Rather, the mating pair produces a zygote that undergoes meiosis, leading to haploid offspring.*

We have revised this section to remove mention of the ‘diploid zygote’ and instead focus on haploid ‘mating pairs’.

1. *200: “Populations continue to diverge indefinitely”: Can you be more precise about this? Surely populations don't diverge indefinitely at the \*phenotypic level\* after they've reached their optima, but they can keep diverging at the genetic level.*

The reviewer is correct that we were referring to genetic divergence. we have made this more clear:

Genetic divergence between populations increases indefinitely after they reach their optima, but their mean phenotypes are stable and the phenotypic segregation variance changes little as time progresses (see Fig. SX, SY). (Lines XX-YY) (see comment #? From reviewer 2).

1. *201-202: “our conclusions refer to equilibrium conditions” this is slightly confusing, as you just wrote that populations do not reach an equilibrium, but continue to accumulate divergence, albeit at a small rate. Or perhaps here you mean at the phenotypic level only? Please specify*

See response to comment #X. Both genetic and phenotypic divergence is low when populations have reached their optimum, and like models by Barton (2001) and Chevin (2014) among others, RI evolves primarily during the ‘adaptation’ phase of a simulation and not during evolution at a peak. We now write the following to make our meaning re: equilibrium clear:

Therefore, our conclusions about phenotypic segregation variance and hybrid fitness refer to quasi-equilibrium conditions rather than transient states and are unaffected by standing variation’s influence on the speed of adaptation. (Lines XX-YY)

1. *280-281: Probably worth mentioning that this result with no standing variance agrees with Chevin et al (2014), eg their figure 2.*

Absolutely correct. We have revised the text to read

Several key results emerge from these simulations. First, in agreement with the findings of Chevin et al. (2014; their Fig. 2), segregation variance is not affected by the angle of divergencewhen there is no standing variation (Lines XX-YY):

1. *283-284: Here and about the fraction of non-overlap (fig 3B), it's unclear what you mean by "increases faster than linearly". This statement implies a positive second derivative (the local slope is increasing, such that the function increases faster than a linear function), but what we see on fig 3A-B is exactly the opposite: the slope is maximal at theta = 0, and then decreases towards 0 as theta increases towards 180°.  Instead your point seems to be that the green curve is above the straight line the joins its end values (black line in fig 3A), so a more accurate description would have something to do with concavity of the function (e.g:*[*https://en.wikipedia.org/wiki/Concave\_function*](https://en.wikipedia.org/wiki/Concave_function)*)*

Thank you, we have revised the text. We believe it is correct to say that the curve approaches zero more rapidly than the linear expectation, and the point of comparison is that this occurs even more rapidly with increasing angle in higher dimensions:

Second, when there is standing variation, parallel genetic evolution rapidly decreases toward zero as the angle of divergence increases. (Lines XX-YY)

1. *285-286: “a given change in theta causes a larger change in the Euclidean distance between optima, delta, when theta is smaller”. Can you please expand on how this is likely to affect segregation variance? At the moment this is quite implicit, so it's hard to judge the validity of your argument.*

Please see detailed response to query #27—reviewer #2 had a similar comment.

1. *295 and below: “mutually beneficial” sounds strange, as it somewhat implies reciprocity, while what you mean here is simply that the mutation is beneficial in both populations. A term like “globally” (as in Bierne 2010 Evolution) or “overall” would seem more appropriate.*

We have revised the text to just say that a mutation is ‘beneficial in both populations’, as suggested. The new text reads:

This rapid decrease in parallelism results because small environmental differences among populations correspond to steep reductions in the fraction of alleles that are beneficial in both environments, and occurs more rapidly as organism complexity (i.e., trait dimensionality) increases. (Lines XX-YY)

1. *336-337: “reduces segregation variance because individuals recover a more intermediate phenotype as a greater number of smaller effect alleles contribute to character divergence”. This verbal explanation can be made more quantitative very easily: if phenotypic divergence d from the ancestor is caused by n loci with equal effects, then the contribution of these loci to the segregation variance is proportional to: n (d/n)^2 =d^2/n, so it is inversely proportional to the number of loci involved. This also holds qualitatively with loci of unequal effects.*

This is absolutely true—and the reason we cited the original work by Castle. While we appreciate the reviewer’s quantitative approach we do not feel it would be more helpful than their precise verbal explanation. Accordingly, we have changed our cumbersome explanation into something like what the reviewer said:

This, in turn, reduces segregation variance because segregation variance is inversely related to the number of loci involved in character divergence (Castle 1921; Lande 1981; Barton et al. 2017). (Lines XX)

1. *339-340: Yes, this is an important point, which probably contributes a lot here. That adaptation from standing variation generally involves alleles with weaker effects has also been shown for quantitative traits by Matuszewksi et al (2015 Genetics), albeit in a slightly different ecological context (continuously changing optimum). Probably worth citing here.*

We now cite Matuszewksi et al (2015) here to support this point.

1. *364: shouldn't there be a heading for Discussion here?*

In our manuscript we merged the results and discussion section because we did not find it conducive to our article to present all results before discussing them. We are more than happy to change the structure of our article if the editorial board or reviewers has any suggestions, but we are not sure what works best.

1. *388-389: “we assume that the ancestral standing variation is random with respect to the new environment.” I find this statement somewhat vague: random with respect to what? Ancestral standing variation is maintained at mutation-selection balance around the optimum in the ancestral  environment, so it has some predictable properties (e.g. selection coefficient) with respect to the new environment. What you seem to have in mind here are genes that are introgressed back into the ancestral population from other derived populations with environments similar to the newly founded ones. So instead of random you could just say that alleles from standing variance have never been exposed to selection in the new environment in your model.*

We have made the change suggested by the reviewer:

Third, we assume that the alleles in the ancestral standing variation have never been exposed to selection in parental environments. (Lines XX-YY).

1. *414-415: “Since segregation variance in our study directly reflects genetic divergence between populations”. Not exactly, segregation variance also depends on the effect size of mutations at traits: the same genetic divergence (in terms of number of mutations fixed) can lead to quite different segregation variances.*

This is true, but within a set of simulations from standing variation they are expected to be correlated. Some new results on genetic parallelism corroborate this point. That said, we have weakened the language from ‘directly reflects’ to ‘are associated’:

Since our results on genetic non-parallelism and segregation variance (e.g., Fig 3A & B) are both associated with genetic divergence, these results could correlate with the strength and/or number of intrinsic incompatibilities (Chevin et al. 2014). (lines xx-yy).

We also note that segregation variance is highly correlated with our metric of non-parallelism, which is agnostic to mutational effects (see new Fig. SX).

1. *415-416: “our results of segregation variance across environments (Fig. 3) could have implications for the evolution of intrinsic barriers”.  Note that this connection with more classical, environment-independent DM incompatibilities, was already made explicitly by Chevin et al (2014).*

Absolutely—we have added a citation to Chevin et al. (2014) here.

1. *Figure 3a: what does the black line represent here?*

Thanks for pointing this out: the black line is the 1:1 line meant to facilitate a comparison as compared to ‘linear’. We include this is the revised caption of Fig. 3A.

1. *Figure 3b: red and pink are almost indistinguishable*

We’ve updated the colour of the overlapping region to purple in our revision.

1. *629: “relationship between the initial fraction of mutations …” and what? “Angle of divergence” is probably missing from the sentence*

Fixed.

1. *Fig 4b: any idea why the change from positive to negative occurs at (or near) 90°?*

There’s nothing special about 90°, there’s just a certain point where segregation variance on average causes individuals to approach the optimum more than deviate. It will depend on *d*, *m*, and assumptions about the fitness landscape. We hope our new simulations plotted in several supplementary figures illustrate this well.

1. *713: “very sufficiently small mutations” reads awkward*

We have changed the text to just ‘small mutation’ (Lines XX-YY):

1. *741: does fig S1C represent effect sizes on fitness or traits? In the latter case, why are all values positive?*

This is the mutation effect size on traits, the Euclidean distance of a mutational vector. We have specified this in the revised caption for Fig. S1. (Lines XX)

1. *Fig S3b: what is the black line here?*

See response to #64. We. have clarified this in the caption:

1. *762: I don’t understand the use of “respectively” here*

We have deleted ‘respectively’ from the sentence.

1. *768: “identical properties”. You mean identical distributions of phenotypic and fitness effects. Another important property is their initial frequency, which you here set at 0.1.*

We have changed the text to specify that, in this case, mutations in the SGV have effect sizes that are the same as mutations arising *de novo*.

**References cited:**

Burton, R. S., C. K. Ellison, and J. S. Harrison. 2006. The Sorry State of F 2 Hybrids: Consequences of Rapid Mitochondrial DNA Evolution in Allopatric Populations. Am. Nat., doi: 10.1086/509046.

Hartl, D. L., and C. H. Taubes. 1996. Compensatory nearly neutral mutations: Selection without adaptation.

Kimura, M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61:893–903.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. Am. Nat. 157:245–61.

Orr, H. A. 2000. Adaptation and the cost of complexity. Evolution 54:13–20.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Rockman, M. V. 2012. The QTN program and the alleles that matter for evolution: All that’s gold does not glitter.

Tenaillon, O. 2014. The utility of Fisher’s geometric model in revolutionary genetics. Annu. Rev. Ecol. Evol. Syst. 45:179–201.