Response to reviewers

We are grateful to the recommender and reviewers for their detailed and constructive suggestions. Since the first round of reviews, we have also had correspondence about the preprint. As a result of this correspondence we have made two major changes to the manuscript not suggested directly by the reviewers. First, by generalizing methods for estimating gene number (especially Lande 1981) we have extended our model, so that we now allow for segregating variation within the parental lines. This entailed a more complicated derivation, and a slight change in meaning for the quantities h and p_{12}, which are now defined as probabilities, rather than as known proportions. Results are, however, more-or-less the same, once they are framed in terms of Fisherian additive effects. Moreover, the old derivation, in terms of fixed proportions, is still included in Appendix 1. Second, we have greatly extended and rearranged the second section of the manuscript, whose aim is to interpret the key quantities and relate them to divergence history. As a corollary, we have decided to remove our previous brief treatment of local adaptation and the associated Figure (old Figure 3), placing the most important material in a new Table 3. This is partly because we felt that the topic of "intrinsic vs. extrinsic isolation" needed a fuller treatment, and because it had already been made in a slightly different way by Chevin et al.(2014) (see reviewer comments below). In addition to these two major changes, we have made changes in response to all of the referee's helpful suggestions, as detailed below. We have also made Hilde Schneemann a joint lead author, due to her contribution to these revisions.

Review by Matthew Hartfield

Your manuscript has been assessed by two reviewers. Both find this a good and insightful investigation into how the form form of divergence influences reproduction isolation, and I agree. This manuscript has potential to be recommended by PCI Evolutionary Biology, but the reviewers have made several suggestions for revisions. I have made a few myself that I list below, but note that most suggestions revolve around improving the clarity of the manuscript rather than substantial methodological changes.

We are grateful for this positive response. We agree that the clarity could be been improved, and have made many changes to this end (including some not directly requested).

My additional comments below:

- I would like a bit more information on the model assumptions regarding how the two parental individuals diverge in the first place, and how they relate to the MRCA. I presume that, sometime in the past, there was an ancestral individual in a well-mixed population (the MRCA individual), whose offspring started diverging genetically, leading to the two parental individuals P1 and P2. Hence, the fixed genetic differences (of which there are D of them) are given relative to this ancestral individual. Is that the case? I also assume that the MRCA can be located anywhere in fitness space (as suggested in Figure 1), and not necessarily at the optimum (as is mostly simulated in Figure 2)? Clarifying these starting assumptions will help the reader better understand the biological process that are being modelled.

We accept that this was not clear in the previous submission. Now that we have generalized the model to allow for variable parental populations, including shared polymorphisms, the concept of MRCA is now far less relevant and has been removed from all Figures and text. We include more detailed discussion of the direction of the changes (e.g., the P1-to-P2 directions shown in Fig 1A, vs. the ancestral-to-derived direction) in the main text (especially in the discussion of Fig. 2).

For the illustrative simulations, we started with two identical and genetically uniform populations (not necessarily optimal), avoiding complications from LD, which we do not treat analytically here. This is now made clear in the main text. We note that the analytical results apply to genetic divergence however it was accrued (e.g., including the case where it involved only small frequency shifts in shared polymorphisms). The simulations are much less general, and aim to give simple illustrations of the types of divergence scenario

that lead to predictable changes in the key quantities. We make this clear in lines 245-251.

- It would be easier to understand Figure 2 if the different scenarios were explained first, before the figure is discussed in depth on pages 6-8. These scenarios are eventually explained over the subsequent few pages, but on first reading it is unclear what is being investigated and hence how they affect the quantities being studied.

This section has changed greatly, but as suggested, we do now explain the scenarios before discussing the results (lines 245-266). We have also created separate sections for directional and stabilizing selection to improve clarity.

- Eq. 26: Could you clarify what is meant by the notation 'if lnwP1 = 0 or lnwP2 = 0'? If applied literally then this equation would reduce to the absolute value of the non-zero term and does not need to be spelt out in full.

We agree this was not clear, and have written it more clearly in Appendix 1, e.g. eqs. 57-58.

- Line 312: Unclear what is meant by "heterozygosity has little impact on the results" as it does not appear that heterozygosity has been investigated in Figure 3.

We accept this was unclear. This whole section has now been removed (see above)

- In the simulations, it is unclear what is meant when selection and dominance effects have 'vanishing' parameters (on lines 455 and 458). Perhaps it would be better to write out the distributions in full?

We accept this was unclear. We have altered the text, and now refer only to the standard beta distribution, with a specified mean and variance.

Best regards,
Matthew Hartfield

Review by Luis-Miguel Chevin

In this manuscript, De Sanctis et al investigate how the mode of divergence between parental populations affects the fitness of their hybrids across environments. The mode of divergence encompasses the phenotypic direction of evolution from the most recent common ancestor (MRCA), as well as the phenotypic effects of mutations that fixed, both in terms of magnitude (additive effects) and within-locus interactions (dominance effects). In line with previous work on this question, starting from Barton (2001) and followed up by a series of papers in the last decade (including several by their group), they model adaptation using Fisher's geometrical model (FGM), whereby fitness depends on squared phenotypic distances from the optimum at multiple traits. Using this common baseline, they rely on combinatorics to generalize and extend previous findings about the way evolutionary trajectories of parental populations influence the fitness of their hybrids.

One of their main insights is that all outcomes regarding hybrids fitness, including the relative contributions of intrinsic vs extrinsic (environment-dependent) isolation, essentially depend on two evolutionary metrics: the total amount of evolutionary change, which sums the squared magnitudes of all mutation effects, and the net effect of evolutionary change, which is the squared magnitude of the overall change (which itself sums all mutation effects). These two metrics, which can be applied to additive and dominance effects, summarize complementary aspects of the evolutionary trajectories of parental populations. When combined with summaries of the genetic composition of hybrids (hybrid index, proportion of heterozygous sites, etc), they are sufficient to predict their expected (log) fitness. The authors provide complementary interpretations of their results in terms of geometry in phenotypic space, and of selection coefficients and epistasis.

The work presented in this ms is an interesting and useful extension of previous theory on this topic, providing insightful general explanations that clarify and unify earlier findings, as well as exploring a larger diversity of evolutionary scenario. The analysis seems solid, therefore I have no major criticisms, but rather suggestions for improvement of presentation.

We are grateful for this positive and constructive response, and for the accurate summary of the submission's contents.

First, some arguments made here appeared in simpler forms in earlier work, which could perhaps be credited a bit more for completeness. For instance, the difference between the squared total phenotypic change ("net effect of evolutionary change here") vs sum of squared changes ("total amount of evolutionary change"), and how this relates to fitness epistasis and the angle between mutations, was a key step in the derivations and overall argument in Chevin et al (2014), although perhaps a bit cryptically (between eq. 3 and appendix S1 of that paper, in a non-isotropic model where distances are less directly meaningful and are thus translated into fitness effects), and restricted to a scenario of directional selection akin to that explored in the last section and Figure 3 of the current ms. Some results about mutation-selection-drift equilibrium around a constant optimum, and the partitioning between intrinsic and extrinsic isolation, also seemed quite similar to points made here, after accounting for the fact that the "total amount of evolutionary change" is very much related to the segregation variance (including regarding the contrasted effects of a few large vs many small mutations, in fig 3 of the present ms). This is not to say that the results here are not novel or insightful, but rather that it may be worthwhile relating them a bit more explicitly to work that touched on very similar issues in a less general way, so as to clarify the improvements made here.

We agree that there was a severe failure to sufficiently acknowledge previous work in our submission. We apologize to the reviewer for this omission. Neglected past work included Lande (1981), and Chevin et al (2014), which we cited, but did not properly or adequately acknowledge. We now cite these papers throughout the main text. The expression of results in terms of fitness effects is now relegated to Appendix 1, partly because it is difficult to apply with segregating variation and applies only with a quadratic fitness landscape, but also because – as we now properly acknowledge – this part was especially derivative of Chevin et al (2014). For the same reason, the discussion of intrinsic/extrinsic and the old Fig. 3 has been removed.

Regarding the model per se, perhaps I haven't read it carefully enough, but after several attempts I have been unable to find a proper description of what y represents in the $g(x, y)$ functions (and S_{xy} in the methods). Since this function is a crucial component of hybrid fitness, it is important to be very explicit about it, to avoid any confusion or wrong interpretations. From my reading and previous understanding of these models, I concluded that x and y correspond to mutations that originate from the two different parental backgrounds, but I can't be sure about that.

x and y are just algebraic placeholder variables used to define functions. We have now simplified notation throughout the manuscript (including removing the function $g(x, y)$), and placeholders of this kind appear only in eqs. 10-11 of the main text.

Another point of definition that was not very clear to me was about the angle theta between substitution vectors. The first time this topic of orientation is brought up in the ms, I found it quite confusing to read (below eq. 16) that "we have $\cos(\theta) = -1$ when two substitutions point in the same phenotypic direction (such that theta = pi), [...] and cos(theta) = 1 for substitutions that point in opposite directions", because this seems to contradict basic trigonometry, if theta is the angle between two vectors. Only when looking more carefully at fig 1B did I realize that theta was in fact not the angle between the two vectors under consideration, but rather pi minus this angle (ie, the supplementary angle, as stated much later in the ms). It is unclear to me why theta was defined in this way, and I think this can be confusing to many readers used to applying geometric arguments to phenotypic spaces. If the goal was just to eliminate a minus sign

before the scalar products, at the moment this seems to be done at the expense of the intuitive meaning, which is regrettable since intuitive interpretation seems to be one of the key goals of this paper. I would find it simpler to keep the minus sign in the formulas and rely on the actual angle between vectors, such that these formulas would state that something about the relative orientation of mutations is subtracted from the rest. If not using this option, then the authors should at least state clearly that theta is not the angle between vectors as usually defined, but the supplement to that angle, and provide an intuitive explanation for why that is (as it does not seem sufficient to just write that "the negative sign comes from the need to take the supplementary angle due to the directionality of the vectors").

We agree that this was confusing. We have changed the definition of θ (see new Figure 1B) as suggested.

Minor comments
22 : "Genomic and phenotypic differentiation between populations is a major cause" $\rightarrow$ "are"?

Fixed. Thanks.

44-45: "strong assumptions about the distribution of the fixed effects (e.g., normality, universal pleiotropy, and independence among traits)". Normality is often assumed for the phenotypic distribution of mutation effects, but not for fixed effects; and at least one model of hybrid fitness in FGM (the one I know best, cited above) did allow for correlations among traits.

Agreed. This negative remark should have been directed solely at a previous paper from our own group. We have corrected the text to clarify (lines 64-67).

120-121: "describes interactions between heterospecific alleles in different states". Perhaps state here (if relevant) that you are referring to additive-by-additive or additive-by-dominance epistasis for fitness?

We have removed this sentence, and include the relevant equations only in Appendix 1, which now deals with fixed parental populations.

Eqs below 402 (methods): A key implicit step in this derivation (which is central to the results in the ms) is the requirement that the probability for a locus to be in each genetic state (heterozygote or each homozygote) does not covary with allelic phenotypic effect. I would find it useful to state this explicitly for completeness, and to explain what this implies: probably something quite general like fair segregation, or rather phenotype-independent segregation?

We hope that we have made our assumptions (which are slightly different in this submission – see above) clearer in the revised text. As requested, we have added "Results in the main text treat h and p_{12} as probabilities determined by the crossing scheme, and which apply to all loci independent of their allelic effects".

Eq below 409: For those curious about that (including me!), could you explain where this product of binomial coefficients comes from?

We agree that this should have been explained. This section is now in Appendix 1, and we have added a short explanation: "given some n, D, p_{12} and p_2, we can choose any Dp_{12} out of D sites which are heterozygous, and any Dp_2 out of the remaining $D(1-p_{12})$ sites to be homozygous for the allele from the second parental population".

Figure S1: I can't help noticing a substantial effect of phenotypic dimensionality n, mostly on the net effect m(2a), and to a lesser extent on the total amount M(2a) of evolutionary change (comparing triangles with circles), and thus on hybrid fitness. I tend to think that this is worth mentioning briefly in the ms, but

this may be a personal bias!

We chose not to emphasize the effects of n in the intiial submission, partly because it had been well covered by previous work, and partly because Roze and Blanckaert (2014) suggested that the dependencies on n might change size with mutation rate. However, we accept that we should have said more, and we now include additional material (see especially the new Equation 26, Figure 3 and Appendix 2).

Review by Juan Li

The manuscript "How does the mode of evolutionary divergence affect reproductive isolation?" is an explicit, well-written manuscript on the evolution of reproductive isolation (RI) under Fisher's geometric model. This paper aimed to investigate the connections between genotype and phenotype to explain the fitness of hybrids, from which we could gain an insight into the strength of reproductive isolation. Here, they presented the total amount (intrinsic, related to genotype divergence) and net effect (extrinsic, related to phenotype distribution) of evolutionary change caused by additive and dominance effects and their interactions. Using simulations, they showed the distribution of additive and dominance effects under different modes of evolutionary divergence, which result in predictable consequences for the hybrid fitness. They also compared the effect of intrinsic and extrinsic isolation on speciation. Within this scope, they articulate the evolutionary role of large- and weak- effect substitutions on adaptative trajectory and speciation. Based on these conclusions, they generalized their model to the effect of gene flow between species. I enjoyed reading this paper, and I found that the authors balanced the considerable number of equations and their evolutionary interpretation very well. Overall, this is interesting research and presented clearly.

My comments are extremely limited. I offer a few questions to be clarified. Hopefully, the authors could make their theoretical work not limited to being read by theoreticians but more understandable for a broader readership.

We are grateful for this positive and constructive response, and have tried to improve clarity throughout the manuscript (while also increasing the generality of the previous results).

Paragraph "Dominance effects" (line 188) – I do not understand the complementary information from additive and dominance effects. In both this paragraph and the discussion (Line 367-Lin 376), the overall dominance reflects the position of the MRCA on the graph, however, it is not related to the additive effect, like, Scenario I and II. Should I expect some explanations from g(a,d) eq. 22? It would be great if the authors could make this connection more clear in a revised manuscript.

We have greatly expanded and restructured this whole section, as suggested, including new simulation scenarios and new Figures 2-3.

Figure 3 – 1. I am aware that dominance is assumed to be 0 in the main text. It would be good to spell this out as well in the legend. 2. I can roughly understand the y-axis label, (E(ln wH))/m(2a) in panels A and B, to compare with M(a)/m(a). However, I do not understand why to intuitively scale log fitness by m(2a), i.e., the biological explanation.

We apologize for lack of clarity here. Figure and section have been removed as discussed above.

Equation 29—It is not obvious for me to get eq. 29 from eq. 19. It would be great if they authors could explain this in a bit more detail in a revised manuscript or provide citations if it has already been analyzed in other papers.

As requested, we have added more explicit derivation here (now eq. 19-20).

Line 243—Should Figure 1C be Figure 1B? Figure 1C demonstrates F1 and mid-parent, whereas Line 243 pointed to two parental phenotypes.

Apologies, changed (one panel now removed).

Section 1.3.3 – I might misunderstand the purpose of this section. I want to communicate more on eq. 26. I understand that epistasis was defined in eq. 25 from eq. 1 and the fitness effect was an approximation. In this section, dominance was not considered because only homozygous substitutions are present in the background. Meanwhile, the whole manuscript focused on the fitness of hybrids (RI). I am confused with the dominance effect when using the epistatic components in m(2a). Here, the epistatic effect could come from any non-linear genotype-phenotype-fitness map. The epistasis could also be generated on the evolutionary trajectory, meaning a similar epistatic factor is in m(2d). It would be helpful if the authors could explain the meaning and relationship of epistasis in hybrids and when two homozygous substitutions are present in one genome.

We accept that this section was both unclear, and derivative of previous work (see above). It is now removed from main text, and rewritten with greater clarity in Appendix 1. The measure of fitness epistasis used in the Appendix, ϵ, is the pairwise epistasis between two homozygous substitutions. As we discuss in the Appendix, equivalent results for heterozygous substitutions are best expressed in terms of $a + d$, not dominance deviations alone, and so they are not easily relatable to the quantities that appear in eq. 13.

Line 548 – This citation seems incomplete. It might be worth checking the reference list.

Thanks, fixed.

Overall, I found this an interesting paper that will undoubtedly stimulate people to think about the strength of reproductive isolation and the evolutionary trajectory to RI during adaptation and speciation.

We are grateful for these generous comments.