Reviewer AE:

> I have received the evaluations of two reviewers, and I have read the paper myself. First, sorry for the long time it took to review your manuscript; it was reviewed on the timescale of mathematicians (for Reviewer2), but both reviewers provided thorough evaluations of your work (which in my opinion is better than a quick but superficial review).

Thanks for the apology, but given the delay in our resubmission, we can hardly complain, at this point.

> Your manuscript is quite long already, and the reviewers' suggestions of modifications and clarification, which need to be implemented, will make the manuscript even longer. I however share the reviewers' opinion that the manuscript's two parts are only loosely related. My suggestion is therefore to publish them separately. Although connecting the systems biology and the quantitative genetics parts is indeed an exciting endeavor, this is not really achieved in the current version of the manuscript, and it would be more profitable to first better describe each part separately. Regarding the connexion between the two parts, please also pay particular attention to R2's first specific comment about dimensionality.

> This may be because I am more familiar with Fisher's geometric model than systems biology models, but it seems to me that the first part is more novel, and should therefore deserve your attention first, should you follow my suggestion of publishing the two parts separately.

Thanks for the careful attention to the paper. We have decided to go with this suggestion, and have split the paper in two; this paper is the "systems biology" part of it. We've left in reviewer comments regarding the "Fisher's geometric model" portion below, commenting that the relevant bit "has been removed".

(AE.1) Format: *Please add line numbers to your manuscript; it is straightforward with the lineno package in LaTeX. Please also do not increase text width too much, as this decreases legibility (LaTeX's default settings already optimize the number of characters per line). Finally, be careful about not confusing citep and citet (a lot of missing parentheses around citations in the second part).*

Reply: Apologies! We've added line numbers and corrected improperly formatted citations.

(AE.2) all figures: *Please ensure that all axes legends and labels are big enough and not distorted (same of arrow labels on the diagrams). Please also make sure that the lines are thick enough to be visible.*

Reply: We have regenerated figures to address this.

(AE.3) Above eq (2): *"Under quite general assumptions": quickly mention them?*

Reply: Good idea. The formula is quite general, and a more precise statement of the assumptions would veer too far afield (perhaps needing definitions of function spaces), but we've added a hopefully clarifying comment. (p. 3, l. 114).

(AE.4) eq (2): *Do you need the first term on the rhs given that you assume that $\kappa(0) = 0$?*

Reply: Ah, good point. We've removed it. (p. 3, l. 114)

(AE.5) Example 1: *"and so its phenotype" whose phenotype? (the subject of the sentence is "the input")*

Reply: Fixed. (p. 3, l. 134)

(AE.6) p5, Note on implementation *"the closure of span(B)" may not be understandable by most of Evolution's readers; please explain what this means.*

Reply: We've added a clarifying note (p. 6, l. 199) but we're hoping that readers not familiar with linear algebra will realize that they can skip this "note on implementation" without loss of continuity.

(AE.7) Figures 1-3: *In figures 1 and 3, there are two types of arrow heads (+ and - effects), but not in Figure 2, probably because the sign of the effect depends on the value of τ. Maybe use a different type of arrow heads for interactions that may change signs, to avoid the confusion with positive effects?*

Reply: We think that adding a third type of arrowhead may only add to the confusion, so now we are using only the standard arrowhead.

(AE.8) p7, Hybrid incompatibility: *Please provide a rationale for choosing a particular weight ρ.*

Reply: Good suggestion; added. (p. 8, l. 255)

(AE.9) p11, 2nd paragraph *Try to better relate these different studies to yours (e.g., what you add, what they do and you do not). In addition to R1's suggestions, consider Weinreich and Knies [2013] and Blanquart and Bataillon [2016].*

Reply: This part has been removed.

(AE.10) p12 Hybridization: *What is $\mathcal{X}$?*

Reply: This part has been removed.

(AE.11) p15 *"The importance of including neutral directions in these models, which is not usually done" – Do not line of isofitness correspond to neutral directions, and aren't these neutral directions already included in those models?*

Reply: This part has been removed.

Reviewer 1:

> The ms combines two parts that are only loosely linked: one "systems biology" part and a "popgen" part. The combination of both parts is appealing, primarily because we lack such combined approaches and it is not easy to come up with an adequate framework. Below, I will discuss both parts in turn.

This manuscript is mostly the "systems biology" part, but we think that the potential applications to "popgen" are still reasonably visible. (And, we plan to resubmit the "popgen" part separately.)

(1.1) *First note that the concept of "nonidentifiability" is closely related to themes that have been discussed in evolutionary biology for a long time, under names like "mutational robustness", "network neutrality", "canalization", "redundancy" etc. In particular, if you characterize your model by "many distinct (and mutationally connected) molecular pathways can realize identical phenotypes", note that this essentially describes what has been called a "neutral network" and studied under this name in many articles. Also concrete developmental networks have been studied in this context (eg, von Dassow etal 2000, Nature 406:188–192). Discussion of this literature - and more importantly discussion of terms and notions used here relative to terms that have been used elsewhere - is largely missing. I won't be able to summarize all this here, but an older review is the Evolution Perspective piece "Evolution and detection of genetic robustness" (de Visser et al 2003). You can work work backward and forward from there.*

Reply: Thanks for the additional suggestions! We've reworked the introduction, adding these (and many other) additional references into the (large and growing) literature. (p. 1, l. 46)We hope it's more representative now.

(1.2) *Related to this: I find it non-intuitive to start the introduction with the notion of "nonidentifiability", which is a consequence rather than a cause, the cause being what you later call "phenotypic equivalence" and what had been discussed under other names in other articles.*

Reply: Good point. In the introdcution we now introduce nonidentifiability after discussing general cases of phenotypic equivalence. (p. 2, l. 50)

(1.3) *Still related: You give long lists of references in the introduction, but do not provide the reader with any information about specific contributions. A bit more would be helpful (starting with True and Haag coined the term "developmental system drift", etc).*

Reply: We've tried to explain things better in the revised Introduction. (p. 1, l. 46)However, we find assigning specific contributions to be difficult because similar ideas appear in many disparate parts of the literature.

(1.4) 4. *Your ms does not simply assume that there is a large "neutral network" underlying a phenotype, but suggests a mechanistic underpinning to create this neutral space. Since much of the appeal of the first part of the paper is connected to this fact, I'd like to see some more in-depth discussion. (suggestions to follow)*

Reply: Yes, that's right! We've followed the suggestions (see below), and have added additional discussion to the Discussion.

(1.5) (mechanistic models) *You should probably mention somewhere that mechanistic models of neutral spaces exist in the (quite different) context of RNA folding (papers by Hofacker, Fontana, etal)*

Reply: Added (p. 1, l. 41).

(1.6) (mechanistic models) *For development, Andreas Wagner once suggested a discrete version of a linear model (iterated matrix multiplication). Variants of this have later been used by others (eg Draghi and GP Wagner, I think). How is your model related to these approaches?*

Reply: This is a natural quesiton. We've addressed it after introducing the model. (p. 3, l. 123)

(1.7) (mechanistic models) *You model development as a linear system and this assumption is essential for all further steps (the explicit solution of the dynamics and the Kalman decomposition). For me, the only justification of a linear model (other than mathematical convenience) is local approximation. Do you agree? This will be important for the second part of the ms, see below.*

Reply: We do agree, and have clarified this in the Discussion. (p. 14, l. 438)

(1.8) (mechanistic models) *You gain a lot of "neutrality" by the assumption that the dimension of the "krypotype" is larger than the dimension of the trait. If the trait is the expression of a gene at the top of some pathway, this is certainly true. However, if we include further traits that are affected by the same genes (i.e., pleiotropy) this is no longer clear. The space of "all traits under selection" of an organism is awfully high-dimensional and pleiotropy is wide-spread (even if we do not believe in "omnigenic" models). It is not clear to me from what kind of data we could learn more about these dimensions and I do not expect an answer in your ms, but the issue deserves more discussion. Currently, "pleiotropy" is not even mentioned in the ms.*

4

Reply: Counterintuitively, it is not necessary to have the kryptotype have a higher dimension than the phenotype to have degrees of freedom. The precise details depend on what you want to treat as fixed or evolvable – for instance, should the output matrix C be subject to drift? – but we don't think it's controversial to suppose that there's a bunch of genes whose expression levels aren't the direct target of selection. We have added discussion of this in the Discussion. (p. 14, l. 398)

(1.9) Introduction: *"Genotypes encoding identical phenotypes can even persist stably within a species"* - *if there is population structure, I suppose.*

Reply: Yes, that's right; we've clarified this. (p. 2, l. 87)

(1.10) Introduction: *"It is not a new observation that there is often more than one way to do the same thing, and that this may lead to speciation" - Isn't this what people would call a neutral Dobzhansky-Muller incompatibility?*

Reply: That'd be one way to do it – and, it turns out, Bateson thought BDMIs would probably be neutral; we've added this. (p. 2, l. 82) (The quote from Bateson [1909] is *"Now if the sterility of the cross-bred be really the consequence of the meeting of two complementary factors, we see that the phenomenon could only be produced among the divergent offspring of one species by the acquisition of at least two new factors; for if the acquisition of a single factor caused sterility the line would then end. Moreover each factor must be separately acquired by distinct individuals, for if both were present together, the possessors would by hypothesis be sterile. And in order to imitate the case of species each of these factors must be acquired by distinct breeds. The factors need not, and probably would not, produce any other perceptible effects; they might, like the colour-factors present in white flowers, make no difference in the form or other characters."*)

(1.11) Results: *(Eq. 1) development as a linear system: has this been done before (probably yes?) - If so, references?*

Reply: We wouldn't be surprised if it had, but we're not aware of it – other papers all seem to take the discrete-time approach as in Wagner [1994].

(1.12) Results: *"Of course, neither of these are necessarily true for real systems" - delete "necessarily"?*

Reply: This sentence has been removed.

(1.13) Eq. 3: *What does a coordinate change mean biologically? In our JC, we discussed for a while whether this is just a different parametrization of the exact same biological thing (like transformation to principle components). I think I now understand that it is more, but this should be explained better.*

Reply: Biologically, a coordinate change may produce an unrecognizably different system (repression changed to activation or the like); we've added a hopefully clarifying sentence. (p. 4, l. 152)

(1.14) *"Since gene networks can grow or shrink following gene duplications and deletions, these additional degrees of freedom can apply in principle to any system." - are there examples of equivalent gene networks with a different size due to gene duplication or loss?*

Reply: We're not aware of any biological examples of "equivalent" networks in more than one species at all, but we think it's not clear whether we understand the dynamics of *any* gene regulatory network well enough to determine equivalence, so this is an absence of evidence, not evidence of absence. (There are plenty of examples of homologous networks with missing genes, but there's so many uncertainties we don't think bringing this up will clarify things.)

(1.15) Kalman decomposition: *Define the sub-matrices directly when first used.*

Reply: Good point; we now say "in block matrix form" and add more explanation. (p. 5, l. 187)

(1.16) Figure 2: Colors are hardly visible

Reply: We've regenerated figures to improve clarity. (p. 6, l. 218)

> **(II)** The popgen part. This part models the "system drift", which is nothing else but neutral drift of a population along a high-fitness ridge in an epistatic landscape (conceptual figure 4), right? If two populations in allopatry drift in different directions, this can lead to hybrid incompatibilities, which are uncovered upon secondary contact. Fitness is modeled by the weighted distance from the optimal impulse response function (eq 5). This is a natural assumption.

Reply: We've removed the "popgen" part, but some of the comments below still apply.

(1.17) For system drift, you assume that "Selection will tend to restrain this motion, but movement along the optimal set N is unconstrained, and so we expect the population mean to drift along the optimal set like a particle diffusing." I see two major problems with this view. Both lead to slower divergence and therefore run against your conclusions. (further points to follow)

Reply: We have removed this part.

(1.18) (drift along a ridge) In the presence of epistasis, evolution on a neutral network (a high-fitness ridge) is *not* due to drift alone, but also affected by (weak second order) selection in favor of mutational robustness / genetic canalization. In contrast to what you write, diffusion on the set of network coefficients corresponding to the optimal phenotype is not unbiased - even if the optimal phenotypes do indeed all have the same fitness. Instead, selection drives the population to "thicker" parts of the network where the mean fitness of the population (including a cloud of mutants) is higher that on a narrow ridge. This is the basis of the evolution of robustness/canalization. It is possible to account for this effect, see Hermisson et al. [2003], Álvarez-Castro et al. [2009], or also Rice [1998], Nimwegen et al. [1999]. Note that epistasis is necessary for the neutral evolution of incompatibilities, which is what you are aiming for.

Reply: Thanks for pointing out this important omission. We have added mention of this (and these citations), including a paragraph to the discussion. (p. 14, l. 419)

(1.19) (drift along a ridge) a second problem results from the fact that evolution on a high-fitness ridge often requires coordinated changes at many loci. Take the oscillator system that you use as an example: simultaneous changes at two genes are required to maintain the phenotype. This leads to a phenomenon called "adaptive inertia" (see Baatz and Wagner [1997] and Álvarez-Castro et al. [2009]), which effectively slows down the movement along the ridge considerably. This problem applies, in particular, in a "house-of-cards" mutation regime when where rarely two mutations occur together on the same haplotype. In small populations, it typically requires that populations drift through shallow fitness valleys. While this is possible, it slows down the process.

Reply: Again, this is an important point covered in additional discussion. (p. 14, l. 419)

(1.20) (drift along a ridge) The relevance of both effects could be studied by simulations in a simple example (eg the oscillator that is used as an illustration in the ms anyway).

Reply: We agree, and hope to do this in our follow-up for this part of the paper.

(1.21) If speciation due to accumulation of incompatibilities does not occur in allopatry, but under (even weak) gene flow, some degree of positive selection is always needed (Bank et al. [2012] for a proof). This also means that even weak gene flow will counteract the process described in the ms.

6

Reply: Also a good point - we've added mention of this as well. (p. 14, l. 419)

(1.22)   Population isolates and genetic load: Isn't this exactly "founder effect speciation"?

Reply: We have removed this part.

(1.23)   Your analysis of the fitness loss in F1 and F2 hybrids rests on a local Taylor expansion of the fitness landscape. This is adequate given that the underlying linear network is also only locally valid (see above). However, you then apply this to a discussion of hybrid incompatibility (Haldane's rule etc). We are thus interested in very large fitness costs for hybrids. This does not seems to be compatible: to get hybrid incompatibility or sterility, you need to have epsilon sufficiently large (below equation 5). But then ϵ^2 is no longer smaller than ϵ. In other words: it seems to me that for the discussion of hybrid incompatibilities you apply your model of the first part beyond the local range where it is valid.

Reply: This is a good point, but tricky since it's not clear what "small" is here. A fitness difference of 0.2 is quite strong selection, for instance, although it is a fairly small number in some ways. Furthermore, the same picture holds in the opposing cases of a single locus or many loci, illustrated in Figure 4. In any case, we now rely only on the conceptual idea rather than the numerics of this estimate.

Reviewer 2:

The present paper is divided into two parts. In the first part, the authors propose a framework in which distinct genetic architectures can produce the same phenotype. (In the fitness landscape terminology, this would correspond to the existence of an evolutionary ridge in a fitness landscape.) The starting point is a nice analogy with linear systems theory. The authors highlight the fact that in general, two distinct linear (differential) systems can respond identically for any input; i.e., different systems (genotypes) can always produce the same output (phenotype) given the same input (environment). Further, the set of equivalent systems can be nicely characterized through Kalman decomposition, thus providing a nice characterization of level sets in the underlying fitness landscape. All along the first part of the paper, this analogy is well illustrated using a simple (yet quite nice) example of oscillating gene transcription. I especially like Fig. 5 where it is shown that even if F_1 hybrids have a phenotypic response close to their parents, F_2 hybrids can behave in a drastically different manner.

In the second part of the paper, motivated by the previous results, the authors explore a general quantitative genetics model in which populations can drift stochastically near a set of equivalent and optimal systems. Since the optimal set (or evolutionary ridge) is not closed under averaging or recombination, two isolated populations can drift apart and accumulate enough genetic differences so that they do not produce any viable offspring. Using some heuristics, several expressions are derived to quantify the accumulation of genetic incompatibilities.

Overall, I think the paper is well written. The framework developed in the first part of the paper is very interesting and that the analogy with system theory is quite enlightening. I am somehow less convinced by the second part. First, I find the arguments a bit sketchy (see below) and not so easy to follow. Secondly, it is not entirely clear to me what is the main contribution of this part compared to previous works. It seems to me that the main result is somehow contains in the fact that the variance (or "segregation variance") of an F_2 population is given by

$$\sigma_S^2 + 4\omega\sigma_N^2 T/N_e$$

which was already derived Slatkin and Lande according to the authors (except for the explicit expression of ω, but again I am a little bit confused by the arguments derived in the appendix).

In summary, I think this paper could be a nice contribution to Evolution. However, I am also convinced that the second part of the paper would require more work, or more explicit reference

to previous works (equation, section etc.) if the authors do not want to re-derive already existing formula.

Thanks for the encouraging words. As discssed above, we've separated out the second (less convincing) part, and the manuscript now only covers the first part.

(2.1) end of p6: *"we assume ... $((A + A')/2, (B + B')/2, (C + C')/2)$" . It is claimed (p. 7, l. 226) that two kryptotypes need not have the same dimension. In this case, the previous sum does not make sense, right?*

Reply: Good point; added that caveat. (p. 7, l. 226)

(2.2) p7: fig 4. *"The distribution of F_2 ... homozygotes". I do not really grasp the meaning of this sentence. More importantly, I find the purple cloud of dots (F_2 population) quite confusing. It seems to me that the purple populations should be a cloud of points concentrated around a point sitting on the red dotted line, i.e., a Gaussian distribution with the mean at the average of the two parental populations. According to the authors, the distribution is bimodal with a peak below the optimal set (as displayed on the figure) and another distinct peak sitting on the other side of the optimal. In fact, if I understand the computations of the second part correctly, the red and purple means should coincide, but the red variance should simply be greater.*

Reply: We have improved Figure 4 and made the accompanying discussion more concrete to hopefully better demonstrate the point.

(2.3) p7: end of the page. *Why do you need ρ to be square integrable?*

Reply: Good point – clearly some assumptions jointly on the long-time behavior of $h(t)$ and ρ are required for D to be finite, but just saying that ρ is square integrable doesn't suffice; we've removed the caveat. (p. 8, l. 252)(It is always well-defined, anyhow, since the integrand is nonnegative.)

(2.4) p8: *"A Taylor expansion of $D(h_\epsilon)$ .. ". It is not clear to me at all. Could you provide some extra explanation (e.g., in the appendix)?*

Reply: We have put back in the somewhat lengthy Appendix on this subject that we'd removed before submission. (p. 8, l. 259)

(2.5) p9. Fig 5. *In the left panel, it seems to me that the two parents are homozygotes since there is a single F_1 possible offspring (dashed blue curve). It would be worth being more explicit. For the right panels, could you be more explicit on the number of curves. If I understand correctly, there are 16 possibilities due to recombination (2 per entries of the matrix) since F_1 individuals are heterozygotes. Is that correct? Finally, the labels on the y axis are not easy to read.*

Reply: We've added "homozygous" to the caption, and mentioned that there are $3^4 = 81$ possible F_2s – this is because it only matters if, at each of the four matrix entries, the offspring is heterozygous, homozygous for parent 1, or homozygous for parent 2.

(2.6) p11. Paragraph system drift. *"move a random distance σ". What is σ? I think it should be σ_N (the std deviation in the direction of the evolutionary ridge) to be consistent with the assumption that the population drifts along the optimal set. I believe that this is what is assumed thereafter. Also, the sentence "It therefore seems .... as cloud of points of width σ" is not very accurate, since the covariance matrix is not the identity.*

Reply: This part has been removed.

(2.7) p 11. *Approximating the optimal set $\mathcal{N}$ by a quadratic surface should only be accurate if we look at the genetic divergence at small time scales. This should be at least mentioned.*

Reply: This part has been removed.

(2.8) end of p11. *$1/(\frac{d}{du}D(x+uz))$ should be evaluated at $u=0$.*

Reply: This part has been removed.

(2.9) p12. Third paragraph *$\sqrt{4\omega T/N_e} \sim \gamma/\sigma_N$. I guess the underlying assumption here is that $\sigma_S \ll \sqrt{4\omega\sigma_N^2 T/N_e}$?*

Reply: This part has been removed.

(2.10) p12. before eq. 7. *$\mu = c_\mu \gamma T/N_e$. Why is μ proportional to γ?*

Reply: This part has been removed.

(2.11) Fig 7. *I have one important issue with this figure (and the assumptions of the underly- ing quantitative genetic model). If one wants to be consistent with the assumption that parental populations drift along the evolutionary ridge, I think one would need to assume that selection is strong enough to constraint the mean of the population on the surface. This would presumably require that $\sigma S/\gamma \ll 1$. First, I think this assumption (or something alternative to that) should be made explicit in the text. Secondly, it seems to me that this assumption is not satisfied for panel A and C: under the range of parameters proposed by the authors, the parental populations could easily drift away from the optimal set, and in particular, the heuristics derived in the main text would not be satisfied.*

Reply: This part has been removed.

(2.12) *Finally, it would be worth mentioning several old and recent works relating genetic drift to speciation: Yamaguchi and Iwasa [2013], several articles by Gavrilets et al. (I think several citations are missing here), and Pina and Schertzer [2019].*

Reply: Thank you for the references. We've added citations to the Introduction. (p. 2, l. 78)