Reviewer AE:

This is a very interesting theory paper that connects the modeling and evolution of regulatory networks with speciation. While quite theoretical in nature, the authors made an effort to provide examples, and I consider this paper of very high relevance for the field.

The authors have addressed most of the previous comments from the first round of reviews, and the two new reviewers only raise minor concerns and suggestions, especially regarding the paragraph on Haldane's rule and regarding a better placement of the paper in the context of speciation research. I agree with these concerns; regarding the latter point I have three additional questions/comments, one of which was also raised in the previous round of reviews:

Thank you for the kind words and the helpful input. We have revised various sections of the paper, including the section on Haldane's rule, and have added a paragraph to the discussion about the distinction between the analysis we conduct here – essentially that of a continuous-trait quantitative model – and predictions of various models that have to do with discrete loci.

(AE.1) Gene flow *As it is described, the model only holds for allopatric populations; I would expect even a small amount of gene flow to make neutral evolution of hybrid incompatibility impossible (Gavrilets 1997, Bank et al. 2012). This is not discussed in the current version of the paper; even more interestingly, would the authors expect conditions in which emergence of incompatibility could be possible in this model also in the presence of gene flow?*

Reply: This is a very interesting question that we haven't attempted to analyze. Currently, the paper relies on general results from quantitative genetics to estimate the rate at which systems drift through the neutral space that don't depend on the details of the genetic architecture. However, the effects of gene flow depend on more things; in particular the effect sizes of mutations, something we've stayed away from. What effect sizes we might expect to be segregating is a good question (and we've added some discussion of this; (p. 14, l. 462)), but we feel that would be a topic for a separate paper (perhaps the former other half of this paper).

We have also added a brief mention that gene flow will likely complicate matters in the discussion (p. 13, l. 460).

(AE.2) Orr's Snowball *The authors state that rapid emergence of incompatibility is possible in their model. How would the accumulation of incompatibility here compare with Orr's "snowball" model, and how is this expected to depend on the complexity of the underlying networks?*

Reply: Good question: on face value, the similarity to the snowball depends entirely on the relationship between "distance to optimum" and fitness: since distance to optimum increases linearly in F_2s, if this relation is quadratic then our model will also show a quadratic increase in incompatibility. However, this seems somewhat arbitrary. It is more interesting to consider if our model displays the same underlying behavior that leads to the snowball, in which each pair of lineage-specific substitutions possibly contributes to incompatibility. This may be an empirical question (that could be examined in simulation, as noted in (p. 14, l. 462)), but at the level of analysis in our paper it seems the answer is "no": Consider two lineages moving away from each other along the neutral manifold. Each substitution moves them further apart, but the rate of increase of the distance between them is linear, and each subsequent substitution does not increase their rate of separation. We considered adding this point to the paper, but decided that it is sufficiently confusing that it would not add much.

(AE.3) Genetic architecture *The genetic architecture (e.g., interactions between cis-regulatory and coding regions) is important to determine the barrier to gene flow that an incompatibility builds. If there is a lot of short-distance interactions (physically in the genome; i.e. low recombination), they might not play a role for speciation per se because the incompatibility will often not be exposed to selection even in*

1

hybrids. In the light of our current knowledge of the genetic architecture of regulatory networks, how would this affect the conclusions from the model?

Reply: This is an interesting question. In our model system coefficients are treated as independent heritable units, and thus would represent genes spread out sufficiently far apart along the genome. Since the major source of F_2 hybrid breakdown in this model is from recombining regulatory coefficients, if the genetic architecture of these coefficients was such that the genes were close enough to not recombine, we would not see the stark pattern of hybrid breakdown that we observe in F_2 oscillators. We suspect that genetic architectures that allow for recombination, albeit at a lower rate, than for architectures in which genes are very far apart, will diminish the speed of speciation. This also leads to interesting questions about the evolution of genetic architecture; perhaps architectures in which co-regulating genes are neighbors are more robust.

We briefly mention such complications to the model in the discussion (p. 13, l. 460).

Reviewer 1:

I am basically enthusiastic about this paper, which has been influential for some time on bioRxiv. The central sections, especially e.g. (p. 3, l. 127) to (p. 6, l. 226), are very clear and novel. I do think, though that the work could be better placed in context.

Thank you. We were also excited that the paper had many readers on the bioRxiv.

(1.1) Empirical evidence for the (minor?) role of drift in speciation: *(p. 12, l. 399) discusses previous discussion about the role of drift in speciation, but they focus on theory. I think that some strong arguments for the primacy of selection in speciation have been empirical (bottlenecking Drosophila populations and testing for RI, MK-tests on "speciation genes" etc.). Some of the earlier work was summarized in the last chapter of Coyne and Orr 2004. (p. 13, l. 424) seems to ignore a lot of this work, and so I would rephrase.*

Reply: This is an important point. We have added references to the experimental speciation literature, around (p. 12, l. 399).

(1.2) Additive quantitative trait models *The previous version of this manuscript contained a standard quantitative trait model. I think it was a good idea to remove this material, but I still think that this modelling approach (including Fisher's geometric model) needs to be discussed a little more here. In the introduction, for example, it seems important to state clearly that "kryptotypes" (many-to-one genotype-to-phenotype maps) are also a feature of these standard models, and that previous authors have shown how phenotypically identical populations can evolve RI, in some cases via drift alone (e.g., Mani and Clarke 1990; Rosas et al. 2010; Fraisse et al. 2016 and others).*

Martin 2014 argued that the results of these models and "systems-biology" type models are likely to approximate one another under some conditions, and that could be mentioned.

Reply: We agree that this is an important thing to discuss. We think that this is fairly well covered in the Introduction (paragraph at (p. 2, l. 80)), and cite Rosas et al. [2010] and Fraïsse et al. [2016] in that section – we have tried to make the connection to previous literature complete without being overwhelming. We have also added a reference to Mani and Clarke [1990] at (p. 12, l. 410), where we think it's better put in context, and added further discussion of the possible connections to Fisher's geometric model in the Discussion (p. 14, l. 462), including a citation of Martin [2014] (which is very relevant; thanks!).

(1.3) RI due to kryptotypes not involving drift *The section beginning on (p. 13, l. 447) seems quite important to me, but it seems to downplay an important point. Populations might diverge genetically, or in their kryptotypes, but not in their phenotypes, if they experience fluctuating selection. This goes back the*

Wright-Fisher debates, and has been discussed in the context of speciation by, e.g., Mani and Clarke 1990, Barton 2001 and later authors. I think that the empirical evidence for fluctuating selection is also quite strong (e.g Bell 2010), and so it seems more important to discuss that, e.g., selection for robustness.

Reply: We agree that the earlier draft of the paper downplayed this point more than it should have; we have added discussion on the role of selection: see the paragraph at (p. 13, l. 417).

(1.4) Interpretation of the Haldane's Rule result The abstract refers to "another possible explanation" of Haldane's Rule, and the explanation given is said to be quite different from the "dominance theory", because "it derives from the nature of segregation variance". I wonder whether this is not misleading. My view is that the sort of model explored here (like additive quantitative trait model used by Barton 2001) naturally generates the sort of dominance relations required for the "dominance theory" to hold. Unless the authors strongly disagree, I would modify the abstract and (p. 10, l. 290) to (p. 10, l. 300), or at least provide a fuller clarification for why the explanation is different from that discussed by Barton 2001.

Reply: Good point; on closer consideration we have decided to remove the reference to the dominance theory (p. 10, l. 300), and have modified the introduction to say "a concrete explanation" rather than "another explanation". More generally, we have added a paragraph to the discussion about the connection between our (quantitative and essentially continuous) model and predictions based on the effects of single mutations on fitness. (p. 14, l. 462)

(1.5) (very minor) (p. 12, l. 370): As you say, F2 hybrid breakdown has been known about for a very long time. As such, the single citation used feels a bit arbitrary.

Reply: Removed the citation.

Reviewer 2:

The manuscript explores how regulatory changes yielding the same phenotype may lead to hybrid incompatibilities and speciation. The authors model gene regulatory networks as sets of linear equations, where environmental inputs are integrated, giving rise to specific phenotypic outputs. Using Kalman decomposition, the authors identify the subset of such systems that produce the same phenotype and thus, characterize how a given system may be modified without affecting its phenotypic response. Resorting to a simple oscillatory network as an example, the authors show that phenotypically identical parents can generate offspring with considerably different phenotypes, resulting in hybrid incompatibilities if the set of equivalent phenotypes is not closed under the average. The phenotypic distance between parents and offspring is proportional to the square of the genetic distance between parents for F1 hybrids and directly proportional to the genetic distance between parents for F2 hybrids.

This is an interesting paper that nicely bridges core concepts of systems biology and evolution. The examples provided are very "visual", making quite clear the idea of "equivalent systems" and will greatly assist readers not familiar with Kalman decomposition. The figures are both informative and elegant, and the discussion aptly covers model validity and makes an effort to relate this work to standard speciation models. In addition, the authors have adequately addressed the comments of previous reviewers. Overall, this manuscript constitutes an appropriate addition to Evolution.

We recommend this paper for publication, but we have some suggestions that we consider would improve the clarity of this paper for a broader biological audience.

Thank you. We are happy that you liked the examples and their accompanying visuals.

(2.1) Haldane's rule *Our primary concern with the text regards Haldane's rule. The authors state that a male F1 offspring has a phenotype that could usually only be obtained in an F2 individual if all the involved genes were autosomal. Although this is true, it does not look like an arbitrary F2 individual but one with a very particular structure. This type of individual is like F1 in all autosomal genes but equal to the parental genotype in the genes present in the sex chromosome. It is not clear to us that this should necessarily generate a worse outcome than a female individual.*

The F2 individuals only perform worse than F1 on average, and looking at figure 4 (left), we can see that some F2 individuals actually perform better than the F1 individuals.

Reply: Yes, that is correct. On average, F_2s will have lower fitness than F_1s. As the reviewer points out, some F_2s could have higher fitnesses than F_1s. In general, hybrid fitnesses will depend on the specific geometry of the optimal set. As such, depending on the particular geometry, as well as the parental distances, Haldane's rule might not always occur. The key point we would like to communicate is that when F_2 hybrid breakdown *can* occur (i.e. due to geometry and distance), and system coefficients are distributed across the autosomes and sex chromosomes, Haldane's rule will manifest. We have added some clarifying text (p. 10, l. 300).

(2.2) *(p. 2, l. 69): This paragraph summarises the main body of work performed in this paper. Thus, consider moving it to the end of the Introduction (i.e. switching places with the last paragraph).*

Reply: Thank you for this suggestion, but we would prefer to keep the current paragraph order. We agree that the second-to-last paragraph does provide a nice summary of the work, however, after trying out the alternate ordering, we decided that the current order (which first discusses system theory, followed by quantitative genetics) is more consistent with the ordering of the rest of the text.

(2.3) *(p. 2, l. 80) : "...more than one way to do the same thing..." sounds a bit too colloquial. Consider changing it to "...more than one way to accomplish the same outcome..." or some other alternative phrasing.*

Reply: Thank you for the suggestion, but we would prefer to keep the sentence as written. We understand your concern, but in order to make this paper easy to read, we have avoided using acronyms and tried to use simple language wherever possible, so long as the meaning remains clear.

(2.4) *(p. 5, l. 179): "The Kalman decomposition then classifies each direction in kryptotype space as either reachable or unreachable". Is there a biological motivation for considering unreachable directions in kryptotype space as degrees of freedom of the neutral space?*

Reply: A natural question; we've added some additional discussion of what the subspaces represent. (p. 6, l. 198)

(2.5) *(p. 5, l. 194): "both the change of basis used to obtain the decomposition and, once in this form, all submatrices other than A_{ro}, B_{ro}, and C_{ro} can be changed without affecting the phenotype, and so represent degrees of freedom". A minor question on generality - does Kalman decomposition require that the coefficient matrices are time-invariant?*

Reply: Yes, the Kalman decomposition is for linear time-invariant systems. We now note this (p. 5, l. 164).

(2.6) Recombination *(p. 7, l. 227): In the section "Sexual reproduction and recombination", we think that the recombination process is not clearly explained, making it more challenging to read the text from this point. We suggest that the authors clarify how exactly recombination is implemented in the model. Perhaps an example would help the reader to have a better picture of the process. This example could be inline, like the one introduced in the section on Haldane's rule.*

Reply: Good point; this was unclear. We have clarified. (p. 7, l. 237)

(2.7)   *(Figure 4 and Figure 5): We suggest dividing the figure into sub-figures (e.g. A and B). Furthermore, we suggest adding "F1" and "F2" labels to the left and right plots in Figure 5, respectively, as well as the values of epsilon. This way, the reader would be immediately informed of what is changing.*

Reply: Done.