System drift and speciation

Joshua S. Schiffman[†] Peter L. Ralph[†‡]

[†]Molecular and Computational Biology, University of Southern California, Los Angeles, California 90089, U.S.A.
[‡]Departments of Mathematics and Biology & The Institute for Ecology and Evolution, University of Oregon, Eugene, Oregon 97403, U.S.A.

jsschiff@usc.edu plr@uoregon.edu

June 1, 2020

Abstract

Even if a species' phenotype does not change over evolutionary time, the underlying mechanism may change, as distinct molecular pathways can realize identical phenotypes. Here we use linear system theory to explore the consequences of this idea, describing how a gene network underlying a conserved phenotype evolves, as the genetic drift of small changes to these molecular pathways cause a population to explore the set of mechanisms with identical phenotypes. To do this, we model an organism's internal state as a linear system of differential equations for which the environment provides input and the phenotype is the output, in which context there exists an exact characterization of the set of all mechanisms that give the same input–output relationship. This characterization implies that selectively neutral directions in genotype space should be common and that the evolutionary exploration of these distinct but equivalent mechanisms can lead to the reproductive incompatibility of independently evolving populations. This evolutionary exploration, or *system drift*, is expected to proceed at a rate proportional to the amount of intrapopulation genetic variation divided by the effective population size (N_e). At biologically reasonable parameter values this could lead to substantial interpopulation incompatibility, and thus speciation, in fewer than N_e generations. This model also naturally predicts Haldane's rule, thus providing another possible explanation of why heterogametic hybrids tend to be disrupted more often than homogametes during the early stages of speciation.

Introduction

Biological systems, at nearly all levels of organization, display considerable *degeneracy* – that is systems can be structurally different yet remain functionally equivalent [Edelman and Gally, 2001]. Examples can be found across nearly all levels of biological organization from the level of the genetic code itself [**??**] all the way up to the convergent evolution of adaptive traits. Between these levels, degeneracy can be found in the *neutral networks* of nucleic acid sequences that map to an RNA secondary structure [**??**] and amino acid sequences that fold similarly [**?**], as well as in the thermodynamic stability of proteins [**?**]. Further examples include the vast space of functionally equivalent potential regulatory sequences [Hare et al., 2008], in the logic of transcriptional [Tsong et al., 2006, Matsui et al., 2015, Dalal et al., 2016, Dalal and Johnson, 2017] and neural circuits [**?**], and in developmental systems [True and Haag, 2001].

Likewise, it is well known that in many contexts mathematical models can fundamentally be *nonidentifiable* and/or *indistinguishable* – meaning that there can be uncertainty about an inferred model's parameters or even its claims about causal structure, despite access to complete and perfect data [Bellman and Åström, 1970, Grewal and Glover, 1976, Walter et al., 1984]. Models with different parameter schemes, or even different mechanics can make equally accurate predictions, but still not actually reflect the internal dynamics of the system being modeled. In control theory, where electrical circuits and mechanical systems are often the focus, it is understood that there can be an infinite number of "realizations", or ways to reverse engineer the dynamics of a "black box", even if all possible input and output experiments are performed [Kalman, 1963, Anderson et al., 1966, Zadeh and Deoser, 1976]. The inherent nonidentifiability of chemical reaction

networks is sometimes referred to as "the fundamental dogma of chemical kinetics" [Craciun and Pantea, 2008]. Finally, the field of *inverse problems* studies those cases in which, despite the existence of a theoretical one-to-one mapping between a model and behavior, tiny amounts of noise make inference problems nonidentifiable in practice [Petrov and Sizikov, 2005].

In this paper we use linear systems theory to model gene regulatory networks, linking biological degeneracy with mathematical nonidentifiability. Although our focus is on linear systems, nonidentifability is by no means unique to such frameworks, and is a general property of many nonlinear models [**?**].

System degeneracy has been argued to be necessary for natural selection [**?**], as it may function as a mechanism for biological robustness and evolvability [**?**], or manifest as *canalization* [**?**]. Degeneracy of the genetic code, and of genetic networks, for instance, can make sequences more fault-tolerant to mutations and open up pathways to adaptation (cite). Biological degeneracy can also enable the evolution of neutral and nearly neutral molecular changes [**?**], and perhaps even contribute to the formation of new species [**?**].

We are interested in the nature and evolutionary implications of degeneracy and nonidentifiability in the context of gene regulatory networks. Specifically, we would like to understand the degree of which gene regulatory network architectures are functionally equivalent and how this relates to network size and efficiency. If systems architectures are not functionally unique, can this open up neutral evolutionary paths, and do these paths manifest as *developmental system drift* [True and Haag, 2001] and contribute to speciation? First, we apply results from system theory which give an analytical description of the set of all linear gene network architectures that yield identical phenotypes. Finally, we briefly outline and apply a quantitative genetics framework to estimate the speed at which system drift can lead to speciation via reproductive incompatibility, which we develop in more detail in a companion paper [**?**].

Results

We use a model of gene regulatory networks that describes the temporal dynamics of a collection of n coregulating molecules – such as transcription factors – as well as external or environmental inputs. We write $\kappa(t)$ for the vector of n molecular concentrations at time t. The vector of m "inputs" determined exogenously to the system is denoted $u(t)$, and the vector of ℓ "outputs" is denoted $\phi(t)$. The output is merely a linear function of the internal state: $\phi_i(t) = \sum_j C_{ij}\kappa_j(t)$ for some matrix C. Since ϕ is what natural selection acts on, we refer to it as the *phenotype* (meaning the "visible" aspects of the organism), and in contrast refer to κ as the *kryptotype*, as it is "hidden" from direct selection. Although ϕ may depend on all entries of κ, it is usually of lower dimension than κ, and we tend to think of it as the subset of molecules relevant for survival. The dynamics are determined by the matrix of regulatory coefficients, A, a time-varying vector of inputs $u(t)$, and a matrix B that encodes the effect of each entry of u on the elements of the kryptotype. The rate at which the i^{th} concentration changes is a weighted sum of the concentrations as well as the input:

$$\begin{aligned} \dot{\kappa}(t) &= A\kappa(t) + Bu(t) \\ \phi(t) &= C\kappa(t). \end{aligned} \tag{1}$$

Furthermore, we always assume that $\kappa(0) = 0$, so that the kryptotype measures deviations from initial concentrations. Here A can be any $n \times n$ matrix, B any $n \times m$, and C any $\ell \times n$ dimensional matrix, with usually ℓ and m less than n. We think of the system as the triple (A, B, C), which translates (time-varying) m-dimensional input $u(t)$ into the ℓ-dimensional output $\phi(t)$. Under quite general assumptions (*e.g.* the system is linear, time-invariant and in the zero-state), we can write the phenotype as

$$\phi(t) = \int_0^t Ce^{A(t-s)}Bu(s)ds, \tag{2}$$

which is a convolution of the input $u(t)$ with the system's *impulse response*, which we denote as $h(t) := Ce^{At}B$.

In terms of gene regulatory networks, A_{ij} determines how the j^{th} transcription factor regulates the i^{th} transcription factor. If $A_{ij} > 0$, then κ_j upregulates κ_i, while if $A_{ij} < 0$, then κ_j downregulates κ_i. The i^{th}

row of A is therefore determined by genetic features such as the strength of j-binding sites in the promoter of gene i, factors affecting chromatin accessibility near gene i, or basal transcription machinery activity. The form of B determines how the environment influences transcription factor expression levels, and C might determine the rate of production of downstream enzymes. To demonstrate this approach, we apply it to construct a simple gene network in Example 1 below.

Example 1 (An oscillator)**.** *For illustration, we consider an extremely simplified model of oscillating gene transcription, as for instance is found in cell cycle control or the circadian rhythm. There are two genes, whose transcript concentrations are given by $\kappa_1(t)$ and $\kappa_2(t)$, and gene-2 upregulates gene-1, while gene-1 downregulates gene-2 with equal strength. Only the dynamics of gene-1 are consequential to the oscillator (perhaps the amount of gene-1 activates another downstream gene network). Lastly, both genes are equally upregulated by an exogenous signal. The dynamics of the system are described by*

$$\begin{aligned}\dot{\kappa_1}(t) &= \kappa_2(t) + u(t)\\ \dot{\kappa_2}(t) &= -\kappa_1(t) + u(t)\\ \phi(t) &= \kappa_1(t).\end{aligned}$$

In matrix form the system regulatory coefficients are given as, $A = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$, and $C = \left[\begin{smallmatrix} 1 & 0 \end{smallmatrix}\right]$. If the input is an impulse at time zero (a delta function), then the phenotype is equal to the impulse response:

$$\phi(t) = h(t) = \sin t + \cos t.$$

The system and its dynamics are referred to in Figure 1. We return to the evolution of such a system below.

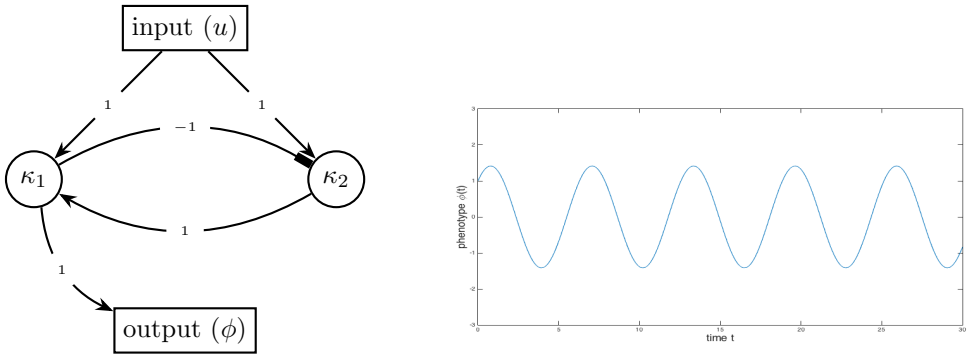

Figure 1: (Left) Diagram of the gene network in Example 1 , and (right) plot of the phenotype $\phi(t)$ against time t .

Equivalent gene networks

As reviewed above, some systems with identical phenotypes are known to differ, sometimes substantially, at the molecular level; systems with identical phenotypes do not necessarily have identical kryptotypes. How many different mechanisms perform the same function?

Two systems are equivalent if they produce the same phenotype given the same input, i.e., have the same input–output relationship. We say that the systems defined by (A, B, C) and $(\bar{A}, \bar{B}, \bar{C})$ are **phenotypically equivalent** if their impulse response functions are the same: $h(t) = \bar{h}(t)$ for all $t \geq 0$. This implies that for any acceptable input $u(t)$, if $(\kappa_u(t), \phi_u(t))$ and $(\bar{\kappa}_u(t), \bar{\phi}_u(t))$ are the solutions to equation (1) of these two systems, respectively, then

$$\phi_u(t) = \bar{\phi}_u(t) \qquad \text{for all } t \geq 0.$$

In other words, phenotypically equivalent systems respond identically for *any* input.

One way to find other systems phenotypically equivalent to a given one is by change of coordinates: if V is an invertible matrix, then the systems (A, B, C) and (VAV^{-1}, VB, CV^{-1}) are phenotypically equivalent because their impulse response functions are equal:

$$\begin{aligned} h(t) &= Ce^{At}B = CV^{-1}Ve^{At}V^{-1}VB \\ &= CV^{-1}e^{VAV^{-1}t}VB = \bar{C}e^{\bar{A}t}\bar{B} = \bar{h}(t). \end{aligned} \tag{3}$$

These "changes of coordinates" are not simply different ways of looking at the same system – if each dimension of the kryptotype corresponds to the concentration of a particular transcription factor, changing A corresponds to changing the strengths of regulatory interactions. We will even see below that interactions may change sign. However, not all phenotypically equivalent systems are of this form: systems can have identical impulse responses without being coordinate changes of each other. In fact, systems with identical impulse responses can involve interactions between different numbers of molecules, and thus have kryptotypes in different dimensions altogether.

This implies that most systems have at least n^2 degrees of freedom, where recall n is the number of components of the kryptotype vector. This is because for an arbitrary $n \times n$ matrix Z, taking V to be the identity matrix plus a small perturbation in the direction of Z above implies that moving A in the direction of $ZA - AZ$ while also moving B in the direction of ZB and C in the direction of $-CZ$ will leave the phenotype unchanged to second order in the size of the perturbation. If the columns of B and the rows of C are not all eigenvectors of A, then any such Z will result in a different system.

It turns out that in general, there are more degrees of freedom, except if the system is *minimal* – meaning, informally, that it uses the smallest possible number of components to achieve the desired dynamics. Results in system theory show that any system can be realized in a particular minimal dimension (the dimension of the kryptotype, $n_{\min}$), and that any two phenotypically equivalent systems of dimension $n_{\min}$ are related by a change of coordinates. Since gene networks can grow or shrink following gene duplications and deletions, these additional degrees of freedom can apply, in principle, to any system.

Even if the system is not minimal, results from systems theory explicitly describe the set of all phenotypically equivalent systems. We refer to $\mathcal{N}(A_0, B_0, C_0)$ as the set of all systems phenotypically equivalent to the system defined by (A_0, B_0, C_0):

$$\mathcal{N}(A_0, B_0, C_0) = \left\{(A, B, C) : Ce^{At}B = C_0e^{A_0t}B_0 \text{ for } t \geq 0\right\}. \tag{4}$$

These systems need not have the same kryptotypic dimension n, but must have the same input and output dimensions (ℓ and m, respectively).

The Kalman decomposition, which we now describe informally, elegantly characterizes this set [Kalman, 1963, Kalman et al., 1969, Anderson et al., 1966]. To motivate this, first note that the input $u(t)$ only directly pushes the system in certain directions (those lying in the span of the columns of B). As a result, different combinations of input can move the system in any direction that lies in what is known as the *reachable subspace*. Analogously, we can only observe motion of the system in certain directions (those lying in the span of the rows of C), and so can only infer motion in what is known as the *observable subspace*. The Kalman decomposition then classifies each direction in kryptotype space as either reachable or unreachable, and as either observable or unobservable. Only the components that are both reachable and observable determine the system's phenotype – that is, components that both respond to an input and produce an observable output.

Concretely, the **Kalman decomposition** of a system (A, B, C) gives a change of basis P such that the transformed system (PAP^{-1}, PB, CP^{-1}) can be written in block matrix form:

$$PAP^{-1} = \begin{bmatrix} A_{r\bar{o}} & A_{r\bar{o},ro} & A_{r\bar{o},\bar{r}\bar{o}} & A_{r\bar{o},\bar{r}o} \\ 0 & A_{ro} & 0 & A_{ro,\bar{r}o} \\ 0 & 0 & A_{\bar{r}\bar{o}} & A_{\bar{r}\bar{o},\bar{r}o} \\ 0 & 0 & 0 & A_{\bar{r}o} \end{bmatrix},$$

and

$$PB = \begin{bmatrix} B_{r\bar{o}} \\ B_{ro} \\ 0 \\ 0 \end{bmatrix} \qquad\qquad (CP^{-1})^T = \begin{bmatrix} 0 \\ C_{ro}^T \\ 0 \\ C_{\bar{r}o}^T \end{bmatrix}.$$

The n-dimensional system has been divided into subspaces of dimensions $n_{r\bar{o}} + n_{ro} + n_{\bar{r}\bar{o}} + n_{\bar{r}o} = n$, and so, for instance, $A_{r\bar{o}}$ is the $n_{r\bar{o}} \times n_{r\bar{o}}$ square matrix in the top left corner of PAP^{-1}. The impulse response of the system is given by

$$h(t) = C_{ro} e^{A_{ro}t} B_{ro},$$

and therefore, the system is phenotypically equivalent to the *minimal* system (A_{ro}, B_{ro}, C_{ro}).

This decomposition is unique up to a change of basis that preserves the block structure. In particular, the minimal subsystem obtained by the Kalman decomposition is unique up to a change of coordinates. This implies that there is no equivalent system with a smaller number of kryptotypic dimensions than the dimension of the minimal system. It is remarkable that the gene regulatory network architecture to achieve a given input–output map is never unique – 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. (However, some of these subspaces may affect how the system deals with noise.)

Note on implementation: The *reachable subspace*, which we denote by $\mathcal{R}$, is defined to be the closure of span(B) under applying A (or equivalently, the span of $B, AB, A^2B, \ldots A^{n-1}B$), and the *unobservable subspace*, denoted $\bar{\mathcal{O}}$, is the largest A-invariant subspace contained in the null space of C. The four subspaces, $r\bar{o}$, ro, $\bar{r}\bar{o}$, and $\bar{r}o$ are defined from these by intersections and orthogonal complements – ro refers to the both *reachable and observable* subspace, while $\bar{r}\bar{o}$ refers to the *unreachable and unobservable* subspace, and similarly for $\bar{r}o$ and $r\bar{o}$.

For the remainder of the paper, we interpret $\mathcal{N}$ as the neutral set in the fitness landscape, along which a large population will drift under environmental and selective stasis. Even if the phenotype is constrained and remains constant through evolutionary time, the molecular mechanism underpinning it is not constrained and likely will not be conserved.

Finally, note that if B and C are held constant – i.e., if the relationships between environment, kryptotype, and phenotype do not change – there are *still* usually degrees of freedom. The following example 2 gives the set of minimal systems equivalent to the oscillator of Example 1, that all share common B and C matrices. The oscillator can also be equivalently realized by a three-gene (or larger) network, and will have even more evolutionary degrees of freedom available, as in Figure 3.

Example 2 (All equivalent rewirings of the oscillator)**.** *The oscillator of example 1 is minimal, and so any equivalent system is a change of coordinates by an invertible matrix V. If we further require B and C to be invariant then we need $VB = B$ and $CV = C$. Therefore the following one-parameter family $(A(\tau), B, C)$ describes the set of all two-gene systems phenotypically equivalent to the oscillator:*

$$A(\tau) = \frac{-1}{\tau + 1} \begin{bmatrix} -\tau & -1 \\ 2\tau(\tau+1)+1 & \tau \end{bmatrix} \text{ for } \tau \neq -1.$$

The resulting set of systems are depicted in Figure 2.

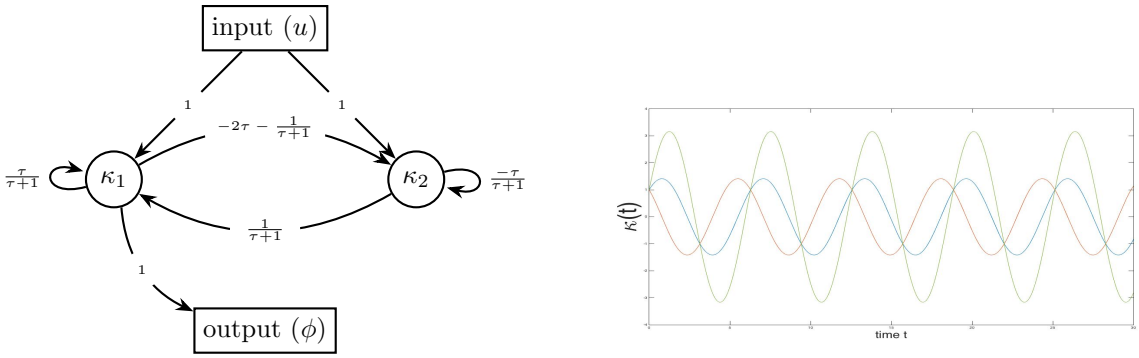

Figure 2: (Left) $A(\tau)$, the set of all phenotype-equivalent cell cycle control networks. (Right) Gene-1 dynamics (blue) for both systems $A(0)$ and $A(-2)$ are identical, however, $A(0)$ gene-2 dynamics (orange) differ from $A(-2)$ (green).

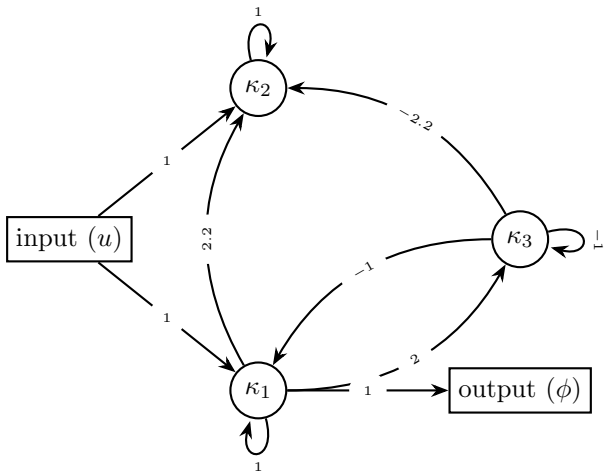

Figure 3: A possible non-minimal three-gene oscillator, phenotypically equivalent to $A(\tau)$, the systems in Examples 1 and 2.

Sexual reproduction and recombination Parents with phenotypically equivalent yet differently wired gene networks may produce offspring with dramatically different phenotypes. If the phenotypes are significantly divergent then the offspring may be inviable or otherwise dysfunctional, despite both parents being well adapted. If this is consistent for the entire population, we would consider them to be separate species, in accord with the biological species concept [Mayr, 2000].

First, we must specify how sexual reproduction acts on these systems. Suppose that each of a diploid organisms' two genomes encodes a set of system coefficients with the same kryptotype dimension. We assume that a diploid which has inherited systems (A', B', C') and (A'', B'', C'') from its two parents has phenotype determined by the system that averages these two, $((A' + A'')/2, (B' + B'')/2, (C' + C'')/2)$.

Each genome an organism inherits is generated by meiosis, in which both of its diploid parents recombine their two genomes, and so an F_1 offspring carries one system copy from each parent, and an F_2 is an offspring of two independently formed F_1s. If the parents are from distinct populations, these are simply first– and second–generation hybrids, respectively.

Exactly how the coefficients (i.e., entries of A, B and C) of a haploid system inherited by an offspring from her diploid parent are determined by the parent's two systems depends on the genetic basis of any variation in the coefficients. Thanks to the randomness of meiotic segregation, the result is random to the extent that each parent is heterozygous for alleles that affect the coefficients. Since the i^{th} row of A summarizes how each gene regulates gene i, and hence is determined by the promoter region of gene i, the elements of a row of A tend to be inherited together, which will create covariance between entries of the same row. It is, however, a quite general observation that the variation seen among recombinant systems is proportional to the difference between the two parental systems.

Offspring formed from two phenotypically identical systems do not necessarily exhibit the same phenotype as both of its parents – in other words $\mathcal{N}$, the set of all systems phenotypically equivalent to a given one, is not, in general, closed under averaging or recombination. If sexual recombination among systems drawn from $\mathcal{N}$ yields systems with divergent phenotypes, populations containing significant diversity in $\mathcal{N}$ can carry genetic load, and isolated populations may fail to produce hybrids with viable phenotypes.

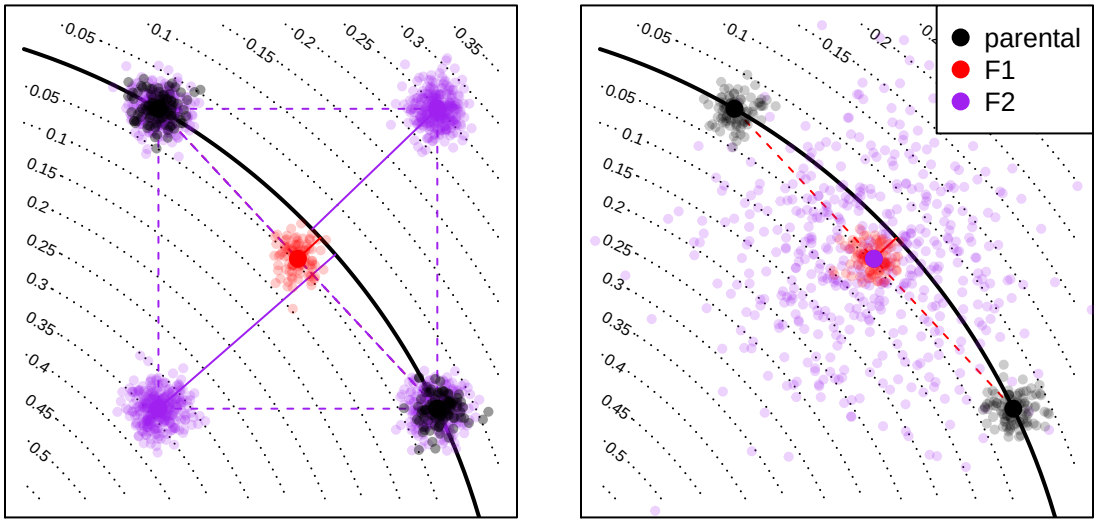

Figure 4: A conceptual figure of the fitness consequences of hybridization: axes represent system coefficients (i.e., entries of A); the line of optimal system coefficients is down in black; dotted lines give phenotypic distances to the optimum. A pair of parental populations are shown in black, along the optimum; a hypothetical population of F_1s are shown in red, and the distribution of one type of F_2 is shown in purple (other types of F_2 are not shown; some would be a similar distance to the other side of the optimal set). The distribution of F_2 hybrids is appropriate for mixed homozygotes if both traits have a simple, one-locus genetic basis, but there is variation within each population at that locus.

Hybrid incompatibility

Two parents with the optimal phenotype can produce offspring whose phenotype is suboptimal if the parents have different underlying systems. How quickly do hybrid phenotypes break down as genetic distance between parents increases? We will quantify how far a system's phenotype is from optimal using a weighted difference between impulse response functions. Suppose that $\rho(t)$ is a nonnegative weighting function, $h_0(t)$ is the *optimal* impulse response function and define the "distance to optimum" of another impulse response function to be

$$D(h) = \left(\int_0^\infty \rho(t) \|h(t) - h_0(t)\|^2 dt \right)^{1/2}. \tag{5}$$

In practice, we take $\rho(t) = \exp(-t/4\pi)$, so that fitness is determined by the dynamics of the system over a few multiples of 2π, but not longer. Consider reproduction between a parent with system (A, B, C) and another displaced by distance ϵ in the direction (X, Y, Z), i.e., having system $(A + \epsilon X, B + \epsilon Y, C + \epsilon Z)$. We assume both are "perfectly adapted" systems, i.e., having impulse response function $h_0(t)$, and their offspring has impulse response function $h_\epsilon(t)$. A Taylor expansion of $D(h_\epsilon)$ in ϵ is explicitly worked out in Appendix A, and shows that the phenotype of an F_1 hybrid between these two is at distance proportional to ϵ^2 from optimal, while F_2 hybrids are at distance proportional to ϵ. This is because an F_1 hybrid has one copy of each parental system, and therefore lies directly between the parental systems (see Figure 4) – the parents both lie in $\mathcal{N}$, which is the valley defined by D, and so their midpoint only differs from optimal due to curvature of $\mathcal{N}$. In contrast, an F_2 hybrid may be homozygous for one parental type in some coefficients and homozygous for the other parental type in others; this means that each coefficient of an F_2 may be equal to either one of the parents, or intermediate between the two; this means that possible F_2 systems may be as far from the optimal set, $\mathcal{N}$, as the distance between the parents. The precise rate at which the phenotype of a hybrid diverges depends on the geometry of the optimal set $\mathcal{N}$ relative to segregating genetic variation.

Example 3 (Hybrid incompatibility: misregulation due to system drift)**.** *Offspring of two equivalent systems from Example 2 can easily fail to oscillate. For instance, the F_1 offspring between homozygous parents at $\tau = 0$ and $\tau = -2$ has phenotype $\phi_{F_1}(t) = e^t$, rather than $\phi(t) = \sin t + \cos t$. However, the coefficients of these two parental systems differ substantially, probably more than would be observed between diverging populations. In figure 5 we compare the phenotypes for F_1 and F_2 hybrids between more similar parents, and see increasingly divergent phenotypes as the difference between the parental systems increases. (In this example, the coefficients of $A(\epsilon)$ differ from those of $A(0)$ by an average factor of $1 + \epsilon/2$; such small differences could plausibly be caused by changes to promoter sequences.) This divergence is quantified in Figure 6, which shows that mean distance to optimum phenotype of the F_1 and F_2 hybrid offspring between $A(0)$ and $A(\epsilon)$ increases with ϵ^2 and ϵ, respectively.*

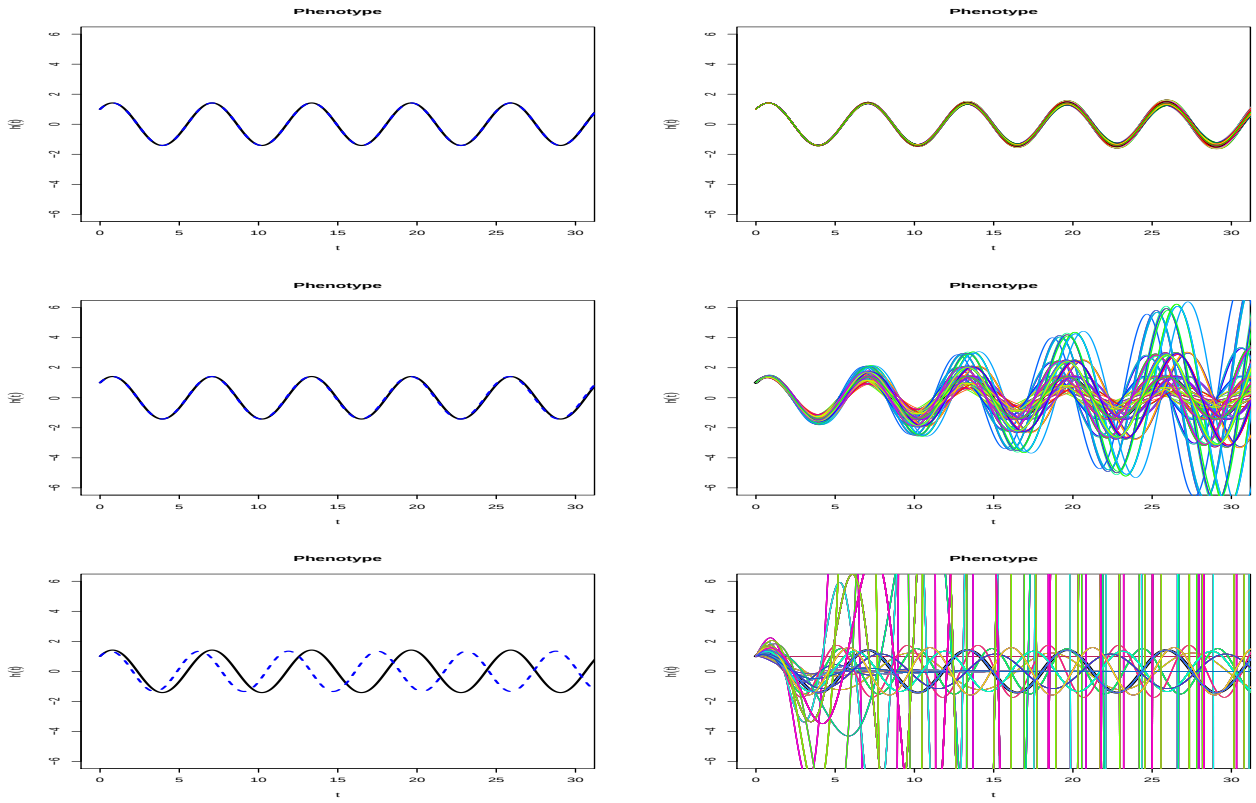

Figure 5: **(left)** Phenotypes of F_1 hybrids between a homozygous $A(0)$ parent and, top-to-bottom, homozygous $A(1/100)$, $A(1/10)$, and $A(1/2)$ parents; parental coefficients differ by around 0.5%, 5%, and 25% respectively. Parental phenotypes ($\sin t + \cos t$) are shown in solid black, and hybrid phenotypes in dashed blue. **(right)** Phenotypes of all $3^4 = 81$ possible F_2 hybrids between the same set of parents, with parental phenotype again in black. Different colored lines correspond to different F_2 hybrids; note that some completely fail to oscillate.

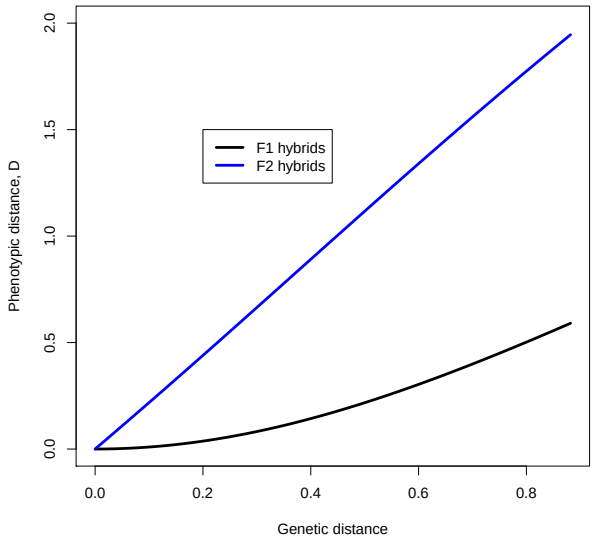

Figure 6: Mean hybrid phenotypic distance from optimum computed with equation (5), using $\rho(t) = \exp(-t/4\pi)$ for F_1 (black) and F_2 (blue) hybrids between $A(0)$ and $A(\epsilon)$ parent oscillators. Genetic distance is computed as $\left(\sum_{ij}(A_{ij}(0) - A_{ij}(\epsilon))^2\right)^{1/2}$.

Haldane's rule This model naturally predicts Haldane's rule, the observation that if only one hybrid sex is sterile or inviable it is likely the heterogametic sex (e.g., the male in XY sex determination systems) [Haldane, 1922, Orr, 1997]. For example, consider an XY species with a two-gene network where the first gene resides on an autosome and the second gene on the X chromosome. A male whose pair of haplotypes is $\left(\left[\begin{smallmatrix} A_1 & A_2 \\ \cdot & \cdot \end{smallmatrix}\right], \left[\begin{smallmatrix} A_1 & A_2 \\ X_1 & X_2 \end{smallmatrix}\right]\right)$ has phenotype determined by $A = \left[\begin{smallmatrix} A_1 & A_2 \\ X_1 & X_2 \end{smallmatrix}\right]$, if dosage compensation upregulates heterogametes by a factor of two relative to homogametes (as with *Drosophila*), while a female homozygous for the haplotype $\left[\begin{smallmatrix} \bar{A}_1 & \bar{A}_2 \\ \bar{X}_1 & \bar{X}_2 \end{smallmatrix}\right]$, has phenotype determined by $A = \left[\begin{smallmatrix} \bar{A}_1 & \bar{A}_2 \\ \bar{X}_1 & \bar{X}_2 \end{smallmatrix}\right]$. An F_1 male offspring of these two will have its phenotype determined by $\left[\begin{smallmatrix} (A_1+\bar{A}_1)/2 & (A_2+\bar{A}_2)/2 \\ \bar{X}_1 & \bar{X}_2 \end{smallmatrix}\right]$. If both genes resided on the autosomes, this system would only be possible in an F_2 cross. More generally, if the regulatory coefficients for a system are shared between the sex and one or more autosomal chromosomes, F_1 males are effectively equivalent to purely autosomal-system F_2 hybrids, and recall that F_2s are significantly less fit on average than F_1s (see Figure 6). Although many alleles will be dominant if the phenotype–fitness relationship is convex, the underlying mechanism does not depend on the *dominance theory* [Turelli and Orr, 1995] to explain Haldane's rule: instead, it derives from the nature of segregation variance.

The speed of speciation

We have shown that system drift can lead to speciation in principle, but is it rapid enough to be an important factor in practice? In other words, after what period of time would we expect the fitness of hybrids between two allopatric populations to be substantially lower than the parentals? The population mean of an unconstrained quantitative trait with genetic variance V_G in a population with effective size N_e will move in t generations a random amount whose variance is tV_G/N_e [**?**]. The mean difference between two such populations has twice the variance. Although this mean difference is along neutral directions, we would in many cases expect the range of variation among F_2s in *all* directions to be of the same order as the differences between the populations, as depicted in Figure 4. This suggests that, naively, two such

populations that have been separated for t generations will produce F_2 offspring that differ from optimal by an amount proportional to $\sqrt{tV_G/N_e}$. Since we assume they are at a local figness optimum, without much loss of generality we can assume that fitness is locally quadratic, and so F_2 fitness decays linearly in time: proportionally to tV_G/N_e – fastest in small, diverse populations.

It is useful to think in more detail about the assumptions in the rough argument above. The key aspect is how population differences in neutral directions (along the fitness ridge) translate to segregation variance in F_2s in selectively constrained directions. To move the the system (the A matrix) a given distance generally involves moving many individual interaction coefficients (the entries A_{ij}). The movements must be coordinated, for the population to stay near the fitness ridge. However, mixing elements between systems that have made independent sets of coordinated changes to remain on the fitness ridge is unlikely to produce a a set of coordinated changes; and the resulting system could move away from the ridge in almost any direction.

Genetic variation in empirical regulatory systems

What is known about the key quantity above, the amount of heritable variation in real regulatory networks? The coefficient A_{ij} from the system (1) measures how much the rate of net production of i changes per change in concentration of j. It is generally thought that regulatory sequence change contributes much more to inter- and intraspecific variation than does coding sequence change affecting molecular structure [Schmidt et al., 2010]. In the context of transcription factor networks this may be affected not only by the binding strength of molecule j to the promoter region of gene i but also the effects of other transcription factors (e.g., cooperativity) and local chromatin accessibility [Stefflova et al., 2013]. For this reason, the mutational target size for variation in A_{ij} may be much larger than the dozens of base pairs typically implicated in the handful of binding sites for transcription factor j of a typical promoter region, and single variants may affect many entries of $\mathcal{N}$ simultaneously.

Variation in binding site occupancy may overestimate variation in A, since it does not capture buffering effects (if for instance only one site of many needs to be occupied for transcription to begin), and variation in expression level measures changes in steady-state concentration (our κ_i) rather than the *rate* of change. Nonetheless, these measures likely give us an idea of the scale of variability. It has been shown that between human individuals, there is differential occupancy in 7.5% of binding sites of a transcription factor (p65) [Kasowski et al., 2010]. It has also been inferred that cis-regulatory variation accounts for around 2–6% of expression variation in human blood-derived primary cells [Verlaan et al., 2009], and that human population variation explained about 3% of expression variation [Lappalainen et al., 2013]. Allele-specific expression is indicative of standing genetic *cis*-regulatory variation; allele-specific expression in 7.2–8.5% of transcripts of a flycatcher species has been observed [Wang et al., 2017], as well as allele-specific expression in 23.4% of genes studied in a baboon species [Tung et al., 2015]. Taken together, this suggests that variation in the entries of A may be on the order of at least a few percent between individuals of a population – doubtless varying substantially between species and between genes.

Discussion

In this paper, we use tools from linear system theory and quantitative genetics to study the evolution of a mechanistic model of the genotype-phenotype map, in which the phenotype is subject to stabilizing selection. In so doing, we provide an explicit model of phenogenetic drift [Weiss and Fullerton, 2000] and developmental system drift [True and Haag, 2001]. In this context, the Kalman decomposition [Kalman, 1963] gives an analytical description of all phenotypically equivalent gene networks. This description shows that the space of functionally equivalent network architectures increases with the square of a network's size, and that this space increases further if networks grow larger than absolutely necessary – that is use more interacting components than the most efficient potential architectures. In this framework, even minimal gene network architectures – efficient architectures that contain only the requisite number of interacting parts, are not structurally unique with respect to function. Functionally equivalent architectures are often related by

continuous parameter changes, suggesting that equivalent networks might be mutationally connected, and that there exist axes of genetic variation unconstrained by natural selection. The independent movement of separated populations along these axes by genetic drift can lead to a significant reduction in hybrid viability, and thus precipitate speciation, at a speed dependent on the effective population size and the amount of genetic variation. In this model, at biologically reasonable parameter values, system drift is a significant – and possibly rapid – driver of speciation. This may be surprising because hybrid inviability appears as a consequence of recombining different, yet functionally equivalent, mechanisms, and since species are often defined by their unique adaptations or morphologies.

Consistent with empirical observation of hybrid breakdown (e.g., Plötner et al. [2017]) and transgressive segregation [**?**], we see that the fitnesses of F_2 hybrids drop at a much faster rate than those of F_1s. Another natural consequence of the model is Haldane's rule, that if only one F_1 hybrid sex is inviable or sterile it is likely to be the heterogametic sex. This occurs because if the genes underlying a regulatory network are distributed among both autosomes and the sex chromosome, then heterogametic F_1s show variation (and fitnesses) similar to that seen in F_2 hybrids.

Is there evidence that this is actually occurring? System drift and network rewiring has been inferred across the tree of life [Wotton et al., 2015, Crombach et al., 2016, Dalal and Johnson, 2017, Johnson, 2017, Ali et al., 2017], and there is often significant regulatory variation segregating within populations. Transcription in hybrids between closely related species with conserved transcriptional patterns can also be divergent [Haerty and Singh, 2006, Maheshwari and Barbash, 2012, Coolon et al., 2014, Michalak and Noor, 2004, Mack and Nachman, 2016], and hybrid incompatibilities have been attributed to cryptic molecular divergence underlying conserved body plans [Gavin-Smyth and Matute, 2013]. Furthermore, in cryptic species complexes (e.g., sun skinks [Barley et al., 2013]), genetically distinct species may be nearly morphologically indistinguishable.

The origin of species not by means of natural selection? As classically formulated, the Dobzhansky-Muller model of hybrid incompatibility is agnostic to the relative importance of neutral versus selective genetic substitutions [Coyne and Orr, 1998], and plausible mechanisms have been proposed whereby Dobzhansky–Muller incompatibilities could originate under neutral genetic drift [Lynch and Force, 2000] or stabilizing selection [Fierst and Hansen, 2009]. The same holds for the "pathway model" [Lindtke and Buerkle, 2015], which is closer to the situation here. However, previous authors have argued that neutral processes are likely too slow to be a significant driver of speciation [Nei et al., 1983, Seehausen et al., 2014]. This has led some to conclude that hybrid incompatibility is typically a byproduct of positive selection [Orr et al., 2004, Schluter, 2009] or a consequence of genetic conflict [Presgraves, 2010, Crespi and Nosil, 2013], two processes that typically act much more rapidly than genetic drift. However, our calculations suggest that even under strictly neutral processes, hybrid fitness breaks down as a function of genetic distance rapidly enough to play a substantial role in species formation across the tree of life. This is consistent with broad patterns such as the relationship between molecular divergence and genetic isolation seen by Roux et al. [2016], and the clocklike speciation rates observed by Hedges et al. [2015].

Neutral processes are certainly not the only drivers of speciation. All of these forces – adaptive shifts, conflict, and network drift – are plausible drivers of speciation, and may even interact. Many of our observations carry over to models of directional selection – for instance, rapid drift along the set of equivalent systems could be driven by adaptation in a different, pleiotropically coupled system. Or, reinforcement due to local adaptation might provide a selective pressure that speeds up system drift. Furthermore, while the fitness consequences of incompatibility in any one given network may be small, the cumulative impact of system drift across the many different networks an organism relies on may be substantial. It remains to be seen how the relative strengths of these forces compare.

The dimensionality of trait space *I didn't think that our results depend on the kryptotype being in a higher dimensional space than the phenotype. Any coordinate change on (A,B,C) will rewire the system. So as long as none of the system matrices are constrained, any minimal system should have an infinite number of realizations. Unless my system theory is rusty?*

We have focused on examples of single traits (where the phenotype is one-dimensional), but phenotypes under selection are often high-dimensional, and different traits often share a genetic basis. Even in networks where the phenotype and kryptotype are of the same dimension, system theory shows us that there will always be available degrees of freedom as specific system realizations are only unique up to a change of coordinates. Some phenotypes, however, require kryptotypic dimensions to be larger than that of the phenotype. For instance, there exists minimal realizations (*e.g.* the oscillator in Example 1) where the dimension of the kryptotpye is larger than that of the phenotype, implying that for these phenotypic dynamics to be realized, the kryptotype dimension *has* to be larger than the dimension of the phenotype. Of course we see many more degrees of freedom in nonminimal networks, but it is not yet clear if evolution commonly arrives at such structures. There are some hints that this may be occuring: compared to the molecular circuitry of the cyanobacteria circadian rhythm, the human clock appears to be highly and unneccessarily complex [Sancar, 2008]. The true "dimensionality of trait space" has been debated at some length in the literature [**?**]. Whatever it is, however, it seems likely to be highly degenerate, even under universal pleiotropy [**??**], XXX

Nonlinearity and model assumptions Of course, real regulatory networks are not linear dynamical systems. Most notably, physiological limits put upper bounds on expression levels, implying saturating response curves. It remains to be seen how well these results carry over into real systems, but the fact that most nonlinear systems can be locally approximated by a linear one suggests our qualitative results may hold more generally. Furthermore, nonidentifiability (which implies the existence of neutral directions) is often found in practice in moderately complex models of biological systems [Gutenkunst et al., 2007, Piazza et al., 2008].

Finally, despite our model's precise separation of phenotype and kryptotype, this relationship in nature may be far more complicated as aspects of the kryptotype may be less "hidden" than we currently assume, and the neutral network changes we describe here may only be nearly neutral. For instance, attributes excluded from the phenotype as modeled here, ignore the potential energy costs associated with excessively large (non-minimal) kryptotypes, as well as the relationship between a specific network architecture and robustness to mutational, transcriptional, or environmental noise. More precise modeling will require better mechanistic understanding not only of biological systems, but also the nature of selective pressures and genetic variation in these systems.

Acknowledgements

We would like to thank Sergey Nuzhdin, Stevan Arnold, Michael Turelli, Patrick Phillips, Erik Lundgren and Hossein Asgharian for valuable discussion. We would also like to thank Nick Barton, Sarah Signor, Todd Parsons, Joachim Hermisson for very helpful comments on the manuscript. Work on this project was supported by funds from the Sloan Foundation and the NSF (under DBI-1262645) to PR.

References

Sammi Ali, Sarah Signor, Konstantin Kozlov, and Sergey Nuzhdin. Quantitative variation and evolution of spatially explicit morphogen expression in Drosophila. *bioRxiv*, page 175711, 2017. 12

BDO Anderson, RW Newcomb, RE Kalman, and DC Youla. Equivalence of linear time-invariant dynamical systems. *Journal of the Franklin Institute*, 281(5):371–378, 1966. 1, 4

Anthony J Barley, Jordan White, Arvin C Diesmos, and Rafe M Brown. The challenge of species delimitation at the extremes: diversification without morphological change in philippine sun skinks. *Evolution*, 67(12): 3556–3572, 2013. 12

Richard Ernest Bellman and Karl Johan Åström. On structural identifiability. *Mathematical biosciences*, 7 (3-4):329–339, 1970. 1

François Blanquart and Thomas Bataillon. Epistasis and the structure of fitness landscapes: Are experimental fitness landscapes compatible with Fisher's geometric model? *Genetics*, 203(2):847–862, 2016. ISSN 0016-6731. doi: 10.1534/genetics.115.182691. URL http://www.genetics.org/content/203/2/847. 3

Joseph D Coolon, C Joel McManus, Kraig R Stevenson, Brenton R Graveley, and Patricia J Wittkopp. Tempo and mode of regulatory evolution in Drosophila. *Genome research*, 24(5):797–808, 2014. 12

Jerry A Coyne and H Allen Orr. The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 353(1366):287–305, 1998. 12

Gheorghe Craciun and Casian Pantea. Identifiability of chemical reaction networks. *Journal of Mathematical Chemistry*, 44(1):244–259, 2008. 2

Bernard Crespi and Patrik Nosil. Conflictual speciation: species formation via genomic conflict. *Trends in Ecology & Evolution*, 28(1):48–57, 2013. 12

Anton Crombach, Karl R Wotton, Eva Jiménez-Guri, and Johannes Jaeger. Gap gene regulatory dynamics evolve along a genotype network. *Molecular biology and evolution*, 33(5):1293–1307, 2016. 12

Chiraj K Dalal and Alexander D Johnson. How transcription circuits explore alternative architectures while maintaining overall circuit output. *Genes & Development*, 31(14):1397–1405, 2017. 1, 12

Chiraj K Dalal, Ignacio A Zuleta, Kaitlin F Mitchell, David R Andes, Hana El-Samad, and Alexander D Johnson. Transcriptional rewiring over evolutionary timescales changes quantitative and qualitative properties of gene expression. *eLife*, 5:e18981, 2016. 1

Gerald M Edelman and Joseph A Gally. Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, 98(24):13763–13768, 2001. 1

Janna L Fierst and Thomas F Hansen. Genetic architecture and postzygotic reproductive isolation: evolution of Bateson-Dobzhansky-Muller incompatibilities in a polygenic model. *Evolution*, 2009. 12

Jackie Gavin-Smyth and Daniel R Matute. Embryonic lethality leads to hybrid male inviability in hybrids between *Drosophila melanogaster* and *D. santomea*. *Ecology and Evolution*, 3(6):1580–1589, 2013. 12

M Grewal and K Glover. Identifiability of linear and nonlinear dynamical systems. *IEEE Transactions on Automatic Control*, 21(6):833–837, Dec 1976. doi: 10.1109/TAC.1976.1101375. 1

Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, 3(10):e189, 2007. 13

Wilfried Haerty and Rama S Singh. Gene regulation divergence is a major contributor to the evolution of Dobzhansky–Muller incompatibilities between species of Drosophila. *Molecular Biology and Evolution*, 23 (9):1707–1714, 2006. 12

JBS Haldane. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, 12(2):101–109, 1922. 10

Thomas F. Hansen and Emilia P. Martins. Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data. *Evolution*, 50(4):1404–1417, 1996. ISSN 00143820, 15585646. URL http://www.jstor.org/stable/2410878. 18

Emily E Hare, Brant K Peterson, Venky N Iyer, Rudolf Meier, and Michael B Eisen. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. *PLoS Genetics*, 4(6):e1000106, 2008. 1

S Blair Hedges, Julie Marin, Michael Suleski, Madeline Paymer, and Sudhir Kumar. Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, 32(4):835–845, 2015. 12

Alexander D Johnson. The rewiring of transcription circuits in evolution. *Current Opinion in Genetics & Development*, 47:121–127, 2017. 12

Rudolf E. Kalman. Mathematical description of linear dynamical systems. *J. SIAM Control*, 1963. 1, 4, 11

Rudolf E. Kalman, Peter L. Falb, and Michael A. Arbib. *Topics in mathematical system theory*. McGraw-Hill, New York, 1969. ISBN 0754321069. 4

M. Kasowski, F. Grubert, C. Heffelfinger, M. Hariharan, A. Asabere, S. M. Waszak, L. Habegger, J. Rozowsky, M. Shi, A. E. Urban, M. Y. Hong, K. J. Karczewski, W. Huber, S. M. Weissman, M. B. Gerstein, J. O. Korbel, and M. Snyder. Variation in transcription factor binding among humans. *Science*, 328(5975): 232–235, April 2010. 11

Tuuli Lappalainen, Michael Sammeth, Marc R Friedländer, Peter AC 't Hoen, Jean Monlong, Manuel A Rivas, Mar Gonzalez-Porta, Natalja Kurbatova, Thasso Griebel, Pedro G Ferreira, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501(7468):506–511, 2013. 11

Dorothea Lindtke and C Alex Buerkle. The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution*, 69(8):1987–2004, 2015. 12

Michael Lynch and Allan G Force. The origin of interspecific genomic incompatibility via gene duplication. *The American Naturalist*, 156(6):590–605, 2000. 12

Katya L Mack and Michael W Nachman. Gene regulation and speciation. *Trends in Genetics*, 2016. 12

Shamoni Maheshwari and Daniel A Barbash. Cis-by-trans regulatory divergence causes the asymmetric lethal effects of an ancestral hybrid incompatibility gene. *PLoS Genetics*, 8(3):e1002597, 2012. 12

Takeshi Matsui, Robert Linder, Joann Phan, Fabian Seidl, and Ian M Ehrenreich. Regulatory rewiring in a cross causes extensive genetic heterogeneity. *Genetics*, 201(2):769–777, 2015. 1

Ernst Mayr. The biological species concept. *Species concepts and phylogenetic theory: a debate. Columbia University Press, New York*, pages 17–29, 2000. 6

Pawel Michalak and Mohamed AF Noor. Association of misexpression with sterility in hybrids of *Drosophila simulans* and *D. mauritiana*. *Journal of Molecular Evolution*, 59(2):277–282, 2004. 12

Masatoshi Nei, Takeo Maruyama, and Chung-I Wu. Models of evolution of reproductive isolation. *Genetics*, 103(3):557–579, 1983. 12

H Allen Orr. Haldane's rule. *Annual Review of Ecology and Systematics*, 28(1):195–218, 1997. 10

H Allen Orr, John P Masly, and Daven C Presgraves. Speciation genes. *Current Opinion in Genetics & Development*, 14(6):675–679, 2004. 12

Y.P. Petrov and V.S. Sizikov. *Well-posed, ill-posed, and intermediate problems with applications*, volume 49. Walter de Gruyter, 2005. 2

Matthew Piazza, Xiao-Jiang Feng, Joshua D Rabinowitz, and Herschel Rabitz. Diverse metabolic model parameters generate similar methionine cycle dynamics. *Journal of Theoretical Biology*, 251(4):628–639, 2008. 13

Björn Plötner, Markus Nurmi, Axel Fischer, Mutsumi Watanabe, Korbinian Schneeberger, Svante Holm, Neha Vaid, Mark Aurel Schöttler, Dirk Walther, Rainer Hoefgen, et al. Chlorosis caused by two recessively interacting genes reveals a role of RNA helicase in hybrid breakdown in *Arabidopsis thaliana*. *The Plant Journal*, 2017. 12

Daven C Presgraves. The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, 11 (3):175–180, 2010. 12

Camille Roux, Christelle Fraisse, Jonathan Romiguier, Yoann Anciaux, Nicolas Galtier, and Nicolas Bierne. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, 14 (12):e2000234, 2016. 12

Aziz Sancar. The intelligent clock and the rube goldberg clock. *Nature structural & molecular biology*, 15 (1):23–24, 2008.

Dolph Schluter. Evidence for ecological speciation and its alternative. *Science*, 323(5915):737–741, 2009. 12

D Schmidt, M D Wilson, B Ballester, P C Schwalie, G D Brown, A Marshall, C Kutter, S Watt, C P Martinez-Jimenez, S Mackay, I Talianidis, P Flicek, and D T Odom. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, 328(5981):1036–1040, May 2010. doi: 10.1126/science.1186176. URL https://www.ncbi.nlm.nih.gov/pubmed/20378774. 11

Ole Seehausen, Roger K Butlin, Irene Keller, Catherine E Wagner, Janette W Boughman, Paul A Hohenlohe, Catherine L Peichel, Glenn-Peter Saetre, Claudia Bank, Åke Brännström, et al. Genomics and the origin of species. *Nature Reviews Genetics*, 15(3):176–192, 2014. 12

K. Stefflova, D. Thybert, M. D. Wilson, I. Streeter, J. Aleksic, P. Karagianni, A. Brazma, D. J. Adams, I. Talianidis, J. C. Marioni, P. Flicek, and D. T. Odom. Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell*, 154(3):530–540, August 2013. 11

John R True and Eric S Haag. Developmental system drift and flexibility in evolutionary trajectories. *Evolution & Development*, 3(2):109–119, 2001. 1, 2, 11

Annie E Tsong, Brian B Tuch, Hao Li, and Alexander D Johnson. Evolution of alternative transcriptional circuits with identical logic. *Nature*, 443(7110):415–420, 2006. 1

Jenny Tung, Xiang Zhou, Susan C Alberts, Matthew Stephens, and Yoav Gilad. The genetic architecture of gene expression levels in wild baboons. *eLife*, 4:e04729, 2015. 11

Michael Turelli and H Allen Orr. The dominance theory of Haldane's rule. *Genetics*, 140(1):389–402, 1995. 10

Dominique J Verlaan, Bing Ge, Elin Grundberg, Rose Hoberman, Kevin CL Lam, Vonda Koka, Joana Dias, Scott Gurd, Nicolas W Martin, Hans Mallmin, et al. Targeted screening of cis-regulatory variation in human haplotypes. *Genome research*, 19(1):118–127, 2009. 11

Eric Walter, Yves Lecourtier, and John Happel. On the structural output distinguishability of parametric models, and its relations with structural identifiability. *IEEE Transactions on Automatic Control*, 29(1): 56–57, 1984. 1

Mi Wang, Severin Uebbing, and Hans Ellegren. Bayesian inference of allele-specific gene expression indicates abundant cis-regulatory variation in natural flycatcher populations. *Genome Biology and Evolution*, 9(5): 1266–1279, 2017. 11

Daniel M. Weinreich and Jennifer L. Knies. Fisher's geometric model of adaptation meets the functional synthesis: Data on pairwise epistasis for fitness yields insights into the shape and size of phenotype space. *Evolution*, 67(10):2957–2972, 2013. doi: 10.1111/evo.12156. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.12156. 3

Kenneth M Weiss and Stephanie M Fullerton. Phenogenetic drift and the evolution of genotype–phenotype relationships. *Theoretical Population Biology*, 57(3):187–195, 2000. 11

Karl R Wotton, Eva Jiménez-Guri, Anton Crombach, Hilde Janssens, Anna Alcaine-Colet, Steffen Lemke, Urs Schmidt-Ott, and Johannes Jaeger. Quantitative system drift compensates for altered maternal inputs to the gap gene network of the scuttle fly *Megaselia abdita*. *eLife*, 4:e04785, 2015. 12

Lotfi A Zadeh and Charles A Deoser. *Linear system theory*. Robert E. Krieger Publishing Company Huntington, 1976. 1

A Local expansion of the fitness surface

Suppose that $\rho(t) \geq 0$ is a weighting function on $[0, \infty)$ so that fitness is a function of $L^2(\rho)$ distance of the impulse response from optimal. With $h_0(t) = C_0 e^{tA_0} B_0$ a representative of the optimal set:

$$\begin{aligned} D(A,B,C)^2 &:= \int_0^\infty \rho(t) \left| h_A(t) - h_0(t) \right|^2 dt \\ &:= \int_0^\infty \rho(t) \left| Ce^{At}B - C_0 e^{A_0 t} B_0 \right|^2 dt \\ &= \int_0^\infty \rho(t) \operatorname{tr} \left\{ \left(Ce^{At}B - C_0 e^{A_0 t} B_0 \right)^T \left(Ce^{At}B - C_0 e^{A_0 t} B_0 \right) \right\} dt \\ &= \int_0^\infty \rho(t) \operatorname{tr} \left\{ \left(Ce^{At}B - C_0 e^{A_0 t} B_0 \right) \left(Ce^{At}B - C_0 e^{A_0 t} B_0 \right)^T \right\} dt, \end{aligned} \tag{6}$$

where $\operatorname{tr} X$ denotes the trace of a square matrix X. How does this change as we perturb about (A_0, B_0, C_0)? First we differentiate with respect to A, keeping $B = B_0$ and $C = C_0$ fixed. Since

$$\frac{d}{du} e^{(A+uZ)t}|_{u=0} = \int_0^t e^{As} Z e^{A(t-s)} ds, \tag{7}$$

we have that

$$\begin{aligned} \frac{d}{du} D(A + uZ, B_0, C_0)^2|_{u=0} &= 2 \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 \left(\int_0^t e^{As} Z e^{A(t-s)} ds \right) B_0 B_0^T \left(e^{At} - e^{A_0 t} \right)^T C_0^T \right\} dt \\ &= 2 \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 \left(\int_0^t e^{As} Z e^{A(t-s)} ds \right) B_0 \left(h_A(t) - h_0(t) \right)^T \right\} dt \end{aligned} \tag{8}$$

and, by differentiating this and supposing that A is on the optimal set, i.e., $h_A(t) = h_0(t)$, (so without loss of generality, $A = A_0$):

$$\begin{aligned} \mathcal{H}^{A,A}(Y,Z) &:= \frac{1}{2} \frac{d}{du} \frac{d}{dv} D(A_0 + uY + vZ, B_0, C_0)^2|_{u=v=0} \\ &= \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 \left(\int_0^t e^{A_0 s} Y e^{A_0 (t-s)} ds \right) B_0 B_0^T \left(\int_0^t e^{A_0 s} Z e^{A_0 (t-s)} ds \right)^T C_0^T \right\} dt. \end{aligned} \tag{9}$$

The function $\mathcal{H}$ will define a quadratic form. To illustrate the use of this, suppose that B and C are fixed. By defining Δ_{ij} to be the matrix with a 1 in the (i,j)th slot and 0 elsewhere, the coefficients of the quadratic form is

$$H_{ij,k\ell}(A) := \mathcal{H}(\Delta_{ij}, \Delta_{k\ell}). \tag{10}$$

We could use this to get the quadratic approximation to D near the optimal set. To do so, it'd be nice to have a way to compute the inner integral above. Suppose that we diagonalize $A = U \Lambda U^{-1}$. Then

$$\int_0^t e^{As} Z e^{A(t-s)} ds = \int_0^t U e^{\Lambda s} U^{-1} Z U e^{\Lambda (t-s)} U^{-1} ds \tag{11}$$

Now, notice that

$$\begin{aligned}\int_0^t e^{s\lambda_i} e^{(t-s)\lambda_j} ds &= \frac{e^{t\lambda_i} - e^{t\lambda_j}}{\lambda_i - \lambda_j} && \text{if} \quad i \neq j \\ &= t e^{t\lambda_i} && \text{if} \quad i = j\end{aligned} \tag{12}$$

Therefore, defining

$$\begin{aligned}X_{ij}(t, Z) &= \left(U^{-1} Z U\right)_{ij} \frac{e^{t\lambda_i} - e^{t\lambda_j}}{\lambda_i - \lambda_j} && \text{if} \quad i \neq j \\ &= \left(U^{-1} Z U\right)_{ii} t e^{t\lambda_i} && \text{if} \quad i = j\end{aligned} \tag{13}$$

moving the U and U^{-1} outside the integral and integrating we get that

$$\int_0^t e^{As} Z e^{A(t-s)} ds = U X(t, Z) U^{-1}. \tag{14}$$

This implies that

$$D(A_0 + \epsilon Z)^2 \approx \frac{1}{2}\epsilon^2 \int_0^\infty \rho(t) \operatorname{tr}\left\{C U X(t, Z) U^{-1} B B^T (U^{-1})^T X(t, Z)^T U^T C^T\right\} dt. \tag{15}$$

To compute the $n^2 \times n^2$ matrix H, we see that if $Z = \Delta_{k\ell}$, then

$$\begin{aligned}X_{ij}^{k\ell}(t) &= (U^{-1})_{\cdot k} U_{\ell \cdot} \frac{e^{t\lambda_i} - e^{t\lambda_j}}{\lambda_i - \lambda_j} && \text{if} \quad i \neq j \\ &= (U^{-1})_{\cdot k} U_{\ell \cdot} t e^{t\lambda_i} && \text{if} \quad i = j\end{aligned} \tag{16}$$

where $U_{k\cdot}$ is the kth row of U, and so

$$H_{ij,k\ell}(A) = \int_0^\infty \rho(t) \operatorname{tr}\left\{C U X^{ij}(t) U^{-1} B B^T (U^{-1})^T X^{k\ell}(t)^T U^T C^T\right\} dt. \tag{17}$$

This implies that

$$D(A_0 + \epsilon Z)^2 \approx \frac{1}{2}\epsilon^2 \sum_{ijk\ell} H_{ij,k\ell}(A_0) Z_{ij} Z_{k\ell}. \tag{18}$$

By section **??**, if we set $\Sigma = \sigma^2 I$ and $U = H$, then a population at $A_0 + Z$ experiences a restoring force of strength $(I + \sigma^2 H^{-1})^{-1} Z$ (treating Z as a vector and H as an operator on these). If σ^2 is small compared to H^{-1} then this is approximately $-\sigma^2 H^{-1} Z$. This suggests that the population mean follows an Ornstein-Uhlenbeck process, as described (in different terms) in Hansen and Martins [1996].

More generally, B and C may also change. To extend this we need the remaining second derivatives of D^2. First, in B:

$$\begin{aligned}\mathcal{H}^{B,B}(Y, Z) &:= \frac{1}{2}\frac{d}{du}\frac{d}{dv} D(A_0, B_0 + uY + vZ, C_0)|_{u=v=0} \\ &= \frac{1}{2}\int_0^\infty \rho(t) \operatorname{tr}\left\{C_0 e^{tA_0} \frac{d}{du}\frac{d}{dv}(uY + vZ)(uY + vZ)^T|_{u=v=0} e^{tA_0^T} C_0^T\right\} dt \\ &= \frac{1}{2}\int_0^\infty \rho(t) \operatorname{tr}\left\{C_0 e^{tA_0} \left(YZ^T + ZY^T\right) e^{tA_0^T} C_0^T\right\} dt.\end{aligned} \tag{19}$$

Next, in C:

$$\begin{aligned}\mathcal{H}^{B,B}(Y, Z) &:= \frac{1}{2}\frac{d}{du}\frac{d}{dv} D(A_0, B_0, C_0 + uY + vZ)|_{u=v=0} \\ &= \frac{1}{2}\int_0^\infty \rho(t) \operatorname{tr}\left\{B_0 e^{tA_0^T} \frac{d}{du}\frac{d}{dv}(uY + vZ)^T(uY + vZ)|_{u=v=0} e^{tA_0} B_0\right\} dt \\ &= \frac{1}{2}\int_0^\infty \rho(t) \operatorname{tr}\left\{B_0 e^{tA_0^T} \left(YZ^T + ZY^T\right) e^{tA_0} B_0\right\} dt.\end{aligned} \tag{20}$$

Now, the mixed derivatives in B and C:

$$\begin{aligned}\mathcal{H}^{B,C}(Y,Z) &:= \frac{1}{2}\frac{d}{du}\frac{d}{dv}D(A_0, B_0+uY, C_0+vZ)|_{u=v=0} \\ &= \int_0^\infty \rho(t)\,\mathrm{tr}\left\{Ye^{tA_0^T}C_0^T Ze^{tA_0}B_0\right\}dt.\end{aligned} \tag{21}$$

In A and B

$$\begin{aligned}\mathcal{H}^{A,B}(Y,Z) &:= \frac{1}{2}\frac{d}{du}\frac{d}{dv}D(A_0+uY, B_0+vZ, C_0)|_{u=v=0} \\ &= \int_0^\infty \rho(t)\,\mathrm{tr}\left\{C_0\left(\int_0^t e^{sA_0}Ye^{(t-s)A_0}ds\right)B_0Z^Te^{tA_0}C_0\right\}dt,\end{aligned} \tag{22}$$

and finally in A and C:

$$\begin{aligned}\mathcal{H}^{A,C}(Y,Z) &:= \frac{1}{2}\frac{d}{du}\frac{d}{dv}D(A_0+uY, B_0, C_0+vZ)|_{u=v=0} \\ &= \int_0^\infty \rho(t)\,\mathrm{tr}\left\{C_0\left(\int_0^t e^{sA_0}Ye^{(t-s)A_0}ds\right)B_0B_0e^{tA_0}Z\right\}dt.\end{aligned} \tag{23}$$

Together, numerical computation of these expressions, along with estimates of genetic covariance within a population, allow precise predictions of evolutionary dynamics of a particular system. The approximation should be good as long as the second-order Taylor approximation holds.

Joshua Schiffman
and Peter Ralph

Resubmission Cover Letter
Evolution

June 1, 2020

To the Editor(s) –

We are pleased to submit a revision of our manuscript,

Sincerely,

Joshua Schiffman and Peter Ralph

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).

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.

(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: 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:

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

Reply: Thank you for the suggestion. We now mention some (p. 2, l. 82). *I think the general assumptions are time invariance and linearity? Any others we should mention?*

(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. 2, l. 82)

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

Reply: Fixed. (p. 3, l. 99)

(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. 5, l. 164)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 (p. 6, l. 183). *we could add a cirlce arrowhead for the unknown direction, but I think its simpler to just get rid of the rectangular negative arrowhead originally in Figure 3*

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

Reply: Good suggestion; added. (p. 8, l. 218)

(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.
>
> General remark: line numbers are very helpful for reviewing.
>
> **(I)** The first (systems biology) part makes use of the concept of "nonidentifiability" of chemical reaction networks and develops these concepts in the context of a genotype to phenotype map. In particular, this map is analyzed for the case where it can be modeled as a linear system.

(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: We've reworked the introduction and included a paragraph on the phenonmenon of biological degeneracy, which unites many of these concepts. (p. 1, l. 34)

(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. 47)

(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 reworked the introduction and included more detail about these references. (p. 1, l. 34)

(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:

(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. 30).

(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:

(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:

(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.*

Reply: We have added discussion of this in the Discussion. (p. 12, l. 356)

(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. (p. ??, l. ??)

(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:

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

Reply:

(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. 117)

(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:

(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. 152)

(1.16) Figure 2: *Colors are hardly visible*

Reply:

> **(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: *I'm leaving these points because they might apply to whatever handwaving we replace the quant gen stuff with.*

(1.17) *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 epsilon. 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:

(1.18) *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:

(1.19) (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. Amer. Nat. 161, 708-734; Alvarez-Castro et al, TPB 75 (2009) 109-122, or also Rice, 1998, Evolution 52, 647-656; van Nimwegen, etal 1999, PNAS 96:9716–9720. Note that epistasis is necessary for the neutral evolution of incompatibilities, which is what you are aiming for.*

Reply:

(1.20) (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, TPB 51, 49-66 and Alvarez-Castro etal . TPB 75, 109-122), 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:

(1.21) (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:

(1.22) *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, Genetics 191, 845–863 for a proof). This also means that even weak gene flow will counteract the process described in the ms.*

Reply:

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

Reply:

(1.24) *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: That's a good point; we resort to a Taylor expansion because we couldn't think of any other clear way to get quantitative estimates. However, our simulations XXX.

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.

(2.1) end of p6: *"we assume ... $((A + A')/2, (B + B')/2, (C + C')/2)$" . It is claimed in p4 (right before the last paragraph) 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. 6, l. 190)

(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 bi- modal 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: XXX Suggestion: change this figure to have two figures of the same type, side-by-side, one with a simple genetic basis (like this one, but with two purple clouds) and another with a more quantitative trait.

(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. 7, l. 214)(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. 222)XXX TODO EDIT THAT APPENDIX XXX

(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: Yamagushi and Iwasa, several articles by Gavrilets et al. (I think several citations are missing here), and Mirò Pina and Schertzer.*

Reply: