

# Rapid speciation despite conservation of phenotype

Joshua S. Schiffman<sup>†</sup>      Peter L. Ralph<sup>†‡</sup>

<sup>†</sup>University of Southern California, Los Angeles, California      <sup>‡</sup>University of Oregon, Eugene, Oregon  
jsschiff@usc.edu      plr@uoregon.edu

## Abstract

It is known that even if a species' phenotype has remained unchanged over evolutionary time, the underlying mechanism may have changed, since distinct molecular pathways can realize identical phenotypes. In this paper, we use quantitative genetics and linear systems theory to study how a gene network underlying a conserved phenotype evolves as genetic drift of small mutational tweaks to the molecular pathways cause a population to explore the set of mechanisms with identical phenotypes. In this setting we treat organisms as “black boxes” for which the environment provides input and the phenotype is the output, and there exists an exact characterization of the set of all mechanisms that give the same input-output relationship. We show that in this situation, there is never a unique architecture for any phenotype and that the evolutionary exploration of these distinct and mutationally connected mechanisms can lead to the rapid accumulation of hybrid incompatibilities between allopatric populations. We estimate that in reasonably numerous species, this process could thus the formation of new species over a number of generations proportional to the effective population size. This model also provides a natural explanation for Haldane's rule that appears mechanistically distinct from other proposed mechanisms.

*Additional ideas to consider adding:*

- speciation literature
- discuss linearity, linearization, and canalization in introduction
- “On the origin of species not by means of natural selection”

## Introduction

A complex molecular machinery translates an organism's genome into the characteristics on which natural selection acts, her phenotype. Attaining a general understanding of the functioning and evolution of this molecular machinery is an overarching goal of many subdisciplines of biology. For example, there is a growing body of data on evolutionary histories and molecular characterizations of particular gene regulatory networks [Jaeger, 2011, Davidson and Erwin, 2006, Israel et al., 2016], as well as thoughtful verbal and conceptual models [True and Haag, 2001, Pavlicev and Wagner, 2012, Weiss and Fullerton, 2000, Edelman and Gally, 2001]. Mathematical models of both particular regulatory networks and the evolution of such systems in general can provide guidance where intuition fails, and thus has the potential to discover general principles in the organization of biological systems and provide concrete numerical predictions [Servedio et al., 2014].

The dynamics of the molecular machinery and its interactions with the environment can be mathematically described in various ways as a dynamical system [Jaeger et al., 2015]. Movement in this direction is ongoing, as researchers have begun to study the evolution of both abstract [Wagner, 1994, 1996, Siegal and Bergman, 2002, Bergman and Siegal, 2003, Draghi and Whitlock, 2015] and empirically inspired computational and mathematical models of gene regulatory networks, [e.g. Mjolsness et al., 1991, Jaeger et al., 2004, Kozlov et al., 2012, 2015, 2014, Crombach et al., 2016, Wotton et al., 2015, Chertkova et al., 2017]. 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, even with 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 equally accurately predict the observed behavior of a given physical system, but still not actually reflect the internal dynamics of the system. 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 on the

“black box” are performed [Kalman, 1963, Anderson et al., 1966, Zadeh and Deoser, 1976]. The fundamental nonidentifiability of chemical reaction networks is sometimes referred to as “the fundamental dogma of chemical kinetics” [Craciun and Pantea, 2008]. In computer science, this is framed as the relationship among processes that simulate one another [Van der Schaft, 2004]. Finally, the field of *inverse problems* studies those cases where, even if a one-to-one mapping between model and behavior is possible in theory, even tiny amounts of noise can make inference problems nonidentifiable in practice.

Although nonidentifiability may frustrate the occasional engineer or scientist, viewed from another angle, these concepts can provide a starting point for thinking about externally equivalent systems – systems that evolution can explore, so long as the parameters and structures can be realized biologically. These functional symmetries manifest in convergent and parallel evolution, as well as *developmental system drift*: the observation that macroscopically identical phenotypes in even very closely related species can in fact be divergent at the molecular and sequence level [True and Haag, 2001, Tanay et al., 2005, Tsong et al., 2006, Hare et al., 2008, ?, Dalal et al., 2016, Dalal and Johnson, 2017].

EDIT IN THESE REFS? The literature is filled with detailed observations of molecular systems and their diversity. *Diversity doesn’t imply system drift – only if the diverse systems are homologous*. There are examples of significant diversity in the networks underlying processes such as circadian rhythm [Sancar, 2008], cell cycle control [Cross et al., 2011, Kearsey and Cotterill, 2003], pattern formation *cite?*, and metabolism [Lavoie et al., 2009, Martchenko et al., 2007, Christensen et al., 2011, Hartl et al., 2007, Alam and Kaminsky, 2013].

AND THESE? Over the last several years, several different computational approaches have been applied to study reproductive incompatibility and speciation. ? simulated the evolution of a transcription factor and its binding site using a thermodynamic model. Their simulations suggest that the language by which a transcription factor recognizes its binding site can change, and potentially lead to hybrid incompatibility when allopatric populations employ divergent readout languages. This study, despite looking at gene regulation, does not analyze overall gene network architecture – as we do here – it only looks at the expression level of a single gene. Furthermore, they report reproductive isolation primarily following directional selection for a change in expression levels in each allopatric population; the evidence for reproductive isolation following balancing selection is much weaker. Johnson and Porter 2000 did not observe any hybrid fitness declines under stabilizing selection – only under directional selection. Khatari et al, Tulchinsky et al, and Porter et al, all study hybrid incompatibility from a transcription factor/binding site interaction perspective, not from an overall network architecture perspective. Palmer and Feldman only see hybrid incompatibility in constant environments if the parental populations are relatively poorly adapted initially. Otherwise hybrids between two allopatric populations have fairly high fitnesses. MAKE SURE TO REF DMI LIT SOMEWHERE and Barton’s paper.

In this paper we outline a theoretical framework to study the evolution of biological systems that can be described by systems of differential equations, such as gene regulatory networks. We study the scenario where the optimal phenotype remains constant over evolutionary time, so that despite strong stabilizing selection for phenotype, neutral drift in the underlying genotype is possible. Results from systems theory provide an analytical description of the set of all linear gene network architectures that yield identical phenotypes, which gives concrete expectations for how any such system can, in principal, undergo system drift and rewiring. Even with stabilization selection, a population will explore the set of all possible phenotypically equivalent gene networks. Consequentially, two populations isolated for a sufficiently long period of time, will likely produce inviable hybrids, despite the absence of adaptation, directional selection, or environmental change.

## Methods

We study an abstract model of temporal dynamics of the concentrations of a collection of  $n$  coregulating molecules within an organism, that may also be affected by temporally varying signals from the environment. We write  $\kappa(t)$  for the vector of  $n$  molecular concentrations at time  $t$ . There are also  $m$  “inputs” determined exogenously to the system, denoted  $u(t)$ , and  $\ell$  “outputs”, 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  $\phi$  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  $\kappa$  as the *kryptotype*, as it is “hidden” from direct selection. Although  $\phi$  may depend on all entries of  $\kappa$ , it is usually of lower dimension than  $\kappa$ , 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  $i^{\text{th}}$  concentration changes is a weighted sum of the concentrations of the other 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,  $B$  a  $n \times m$ , and  $C$  any  $\ell \times n$  dimensional matrix, with usually  $\ell$  and  $m$  less than  $n$ . We think of the system as the triple  $\mathcal{S} = (A, B, C)$ , which translates (time-varying)  $m$ -dimensional input  $u(t)$  into the  $\ell$ -dimensional output  $\phi(t)$ . Under quite general assumptions, we can write the phenotype as

$$\phi(t) = Ce^{At}\kappa(0) + \int_0^t Ce^{A(t-s)}Bu(s)ds,\tag{2}$$

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

Although many different biological systems can be modeled with this approach, for clarity, we focus on gene regulatory networks. In this interpretation,  $A_{ij}$  determines how the  $j^{\text{th}}$  transcription factor regulates the  $i^{\text{th}}$  transcription factor. If  $A_{ij} > 0$ , then  $\kappa_j$  upregulates  $\kappa_i$ , while if  $A_{ij} < 0$ , then  $\kappa_j$  downregulates  $\kappa_i$ . 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 be the rate of production of a downstream enzyme (although other arrangements could be made).

Here we have assumed that the system is linear, and begins from the “zero” state ( $\kappa(0) = 0$ ). Of course, neither of these are necessarily true for real systems, but the dynamics of most nonlinear systems can be approximated locally by a linear systems near most points. Furthermore, the ease of analyzing linear systems makes this an attractive place to start.

**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. Suppose there are two genes, whose transcript concentrations are given by  $\kappa_1(t)$  and  $\kappa_2(t)$ , and that gene-2 upregulates gene-1 and that gene-1 downregulates gene-2 with equal strength. Furthermore, suppose that only the dynamics of gene-1 are consequential to the oscillator (perhaps the amount of gene-1 activates another downstream gene network). Lastly suppose that the production of both genes is 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,  $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ ,  $B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ , and  $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$ .

Suppose the input is an impulse at time zero (a delta function), and so its phenotype is equal to its 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.

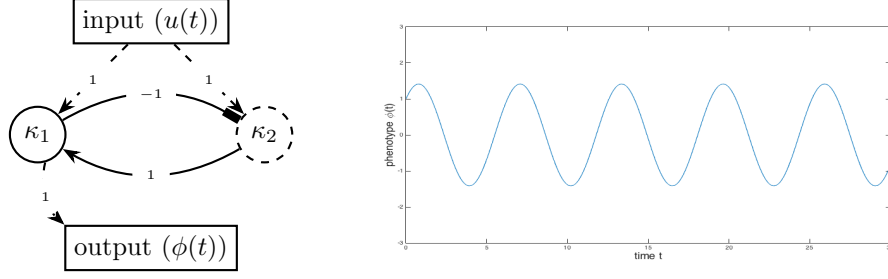


Figure 1: (Left) Graphical representation of the cell cycle control gene network, and (right) plot of the phenotype  $\phi(t)$  against time  $t$ .

## 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

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

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:

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

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. If  $A$  is invertible, for instance, 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.

Some gene networks, however, can grow or shrink, perhaps following gene duplications and deletions, and also still preserve their phenotypes. More generally, 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  $(A_0, B_0, C_0)$ . Concretely, this is

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

These systems need not have the same kryptotypic dimension  $n$ , but must have the same input and output dimensions ( $\ell$  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 columns 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 respond to an input and components that 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  $(PAP^{-1}, PB, CP^{-1})$  has the following 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

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

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  $(A_{ro}, B_{ro}, C_{ro})$ . (Here the subscript  $ro$  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  $r\bar{o}$ .)

Any two minimal systems are related by a change of coordinates, and so the minimal subsystems obtained by the Kalman decomposition are unique up to a change of coordinates. In particular, this implies that there is no equivalent system with a smaller number of kryptotypic dimensions than the dimension of the minimal system. It is also remarkable to note 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.

*Note on implementation:* The *reachable subspace*, which we denote by  $\mathcal{R}$ , is defined to be the closure of  $\text{span}(B)$  under applying  $A$ , 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  $nro$  are defined from these by intersections and orthogonal complements.

For the remainder of the paper, we interpret  $\mathcal{N}$  as the phenotypically neutral landscape, wherein 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. These correspond to distinct genetic networks that perform indistinguishable functions. 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.

**Example 2** (All Phenotypically Equivalent Oscillators). *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  $CV = C$ . Solving these equations, we find that a one-parameter family  $(A(\tau), B, C)$  describes the set of all two-gene systems phenotypically equivalent to the oscillator, where*

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

178 The resulting set of systems, and their dynamics, are depicted in Figure 2

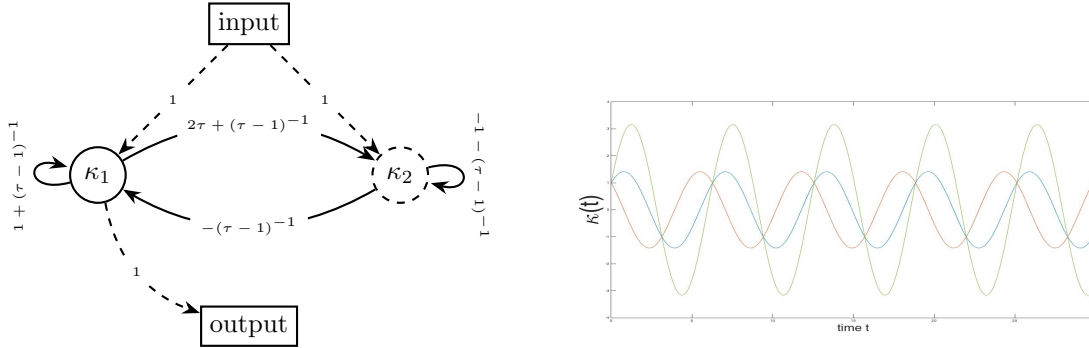


Figure 2: (Left)  $A(\tau)$ , the set of all phenotype-equivalent cell cycle control networks. (Right) Gene-1 dynamics (blue) for both systems  $A(0)$  and  $A(2)$  are identical, however,  $A(0)$  gene-2 dynamics (orange) differ from  $A(2)$  (green).

179 **Sexual reproduction and recombination** Parents with phenotypically equivalent, yet different, gene  
180 networks may produce offspring with dramatically different phenotypes. If the phenotypes are significantly  
181 divergent then the offspring may be inviable or otherwise have very low fitnesses, despite both parents being  
182 well adapted. If this is consistent for the entire population, we would consider them to be separate species.

183 Diploid organisms have two copies of the genome, each of which encodes a set of system coefficients.  
184 We assume that a diploid who has inherited systems  $(A', B', C')$  and  $(A'', B'', C'')$  from her mother and  
185 father, respectively, has phenotype determined by the system that averages these two,  $((A' + A'')/2, (B' +$   
186  $B'')/2, (C' + C'')/2$ .

187 Each of the copies of the genome an organism inherits from her parents are generated from their two  
188 copies of the genome by meiosis, in which the diploid parent recombines her two genomes. We will assume  
189 that each coefficient (i.e., entry of  $A$ ,  $B$  or  $C$ ) is determined by a single nonrecombining locus, so that each  
190 coefficient in the system produced by meiosis is an independent random choice of the two parental coefficients.  
191 With these definitions, an  $F_1$  offspring of two individuals carries one system copy from each parent, and an  
192  $F_2$  is the offspring of two independently formed  $F_1$  individuals. If the parents are from distinct populations,  
193 these are first- and second-generation hybrids, respectively.

194 This is a simplification: since the  $i^{\text{th}}$  row of  $A$  summarizes how each gene regulates gene  $i$ , and hence  
195 is determined by the promoter region of gene  $i$ , we would actually expect the elements of a row of  $A$  to  
196 tend to be inherited together. Similarly, we expect in practice heritable variation in each coefficient to be  
197 determined by more than one locus – but this may be a reasonable approximation.

198 The offspring of two systems with the same phenotype may not have the same phenotype as the parents  
199 – in other words  $\mathcal{N}$ , the set of all systems phenotypically equivalent to a given one, is not, in general, closed  
200 under averaging or recombination. Next we discuss how this fact can contribute to hybrid incompatibility  
201 and genetic load. If sexual recombination among systems drawn from  $\mathcal{N}$  yields systems with divergent  
202 phenotypes, populations containing significant diversity in  $\mathcal{N}$  can carry genetic load, and isolated populations  
203 may fail to produce hybrids with viable phenotypes.



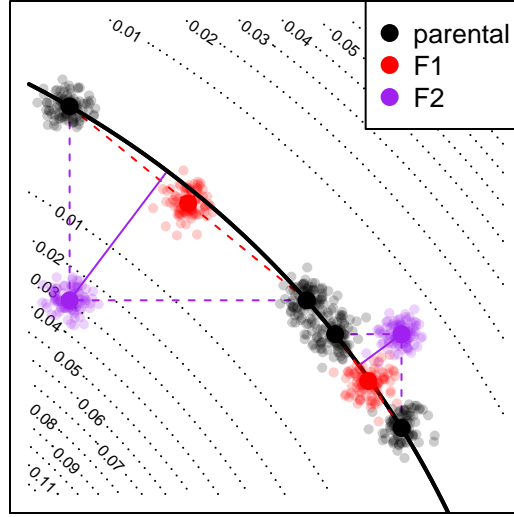


Figure 3: A conceptual figure of the fitness consequences of hybridization: axes represent system coefficients (i.e., entries of  $A$ ); the line of optimal system coefficients is down in black; dotted lines give phenotypic distances to the optimum. Two pairs of parental populations are shown in black, along the optimum; a hypothetical population of  $F_1$ s are shown for each in red, and the distribution of one type of  $F_2$  is shown in purple (other types of  $F_2$  are not shown). Solid lines depict the distance of the  $F_2$  to optimum.

**Hybrid incompatibility** Two parents with the optimal phenotype can produce offspring whose phenotype is suboptimal if the parents have different underlying systems. This leads to the question: How quickly does the hybrid phenotype break down with increasing distance between the parents? To quantify this, we will measure how far a system’s phenotype is from optimal using a weighted difference between impulse response functions. Suppose that  $\rho(t)$  is a nonnegative, smooth, square-integrable weighting function, suppose that  $h_0(t)$  is the *optimal* impulse response function and define the “distance to optimum” of another impulse response function to be

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

Consider reproduction between a parent with system  $(A, B, C)$  and another displaced by distance  $\epsilon$  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  $\epsilon$  is explicitly worked out in Appendix F, and shows that the phenotype of an  $F_1$  hybrid between these two is at distance proportional to  $\epsilon^2$  from optimal, while  $F_2$  hybrids are at distance proportional to  $\epsilon$ . This is because an  $F_1$  hybrid has one copy of each parental system, and therefore lies directly between the parental systems (see Figure 3) – 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 – in Figure 3, this is depicted as the angle of the black line (the optimal set) with respect to the coordinates.

**Example 3** (Hybrid Incompatibility in the Oscillator). *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*

219  $\tau = 2$  has phenotype  $\phi_{F_1}(t) = e^t$ , rather than  $\phi(t) = \sin t + \cos t$ . However, the coefficients of these two  
 220 parental systems differ substantially, probably more than would be realistically observed between diverging  
 221 populations. In figure 4 we compare the phenotypes for  $F_1$  and  $F_2$  hybrids between more similar parents,  
 222 and see increasingly divergent phenotypes as the difference between the parental systems increases. (In this  
 223 example, the coefficients of  $A(2 + \epsilon)$  differ from those of  $A(2)$  by an average factor of  $1 + \epsilon/2$ ; such small  
 224 differences could plausibly be caused by changes to promoter sequences.) This divergence is quantified in  
 225 Figure 5, which shows that mean distance to optimum phenotype of the  $F_1$  and  $F_2$  hybrid offspring between  
 226  $A(2)$  and  $A(2 + \epsilon)$  increases with  $\epsilon^2$  and  $\epsilon$ , respectively.

227 The coefficients in  $A$  – i.e., the regulatory coefficients – differ between parents by only a few percent  
 228 (around 0.5% for  $\tau = 2.01$  and 5% for  $\tau = 2.1$ ). This is well within the amount of regulatory coefficient  
 229 variation we expect to find segregating within real populations (discussed further below). For these small  
 230 values of  $\epsilon$ , hybrid phenotypes remain relatively stable, consistent with the idea that natural selection will  
 231 allow such intrapopulation variation.

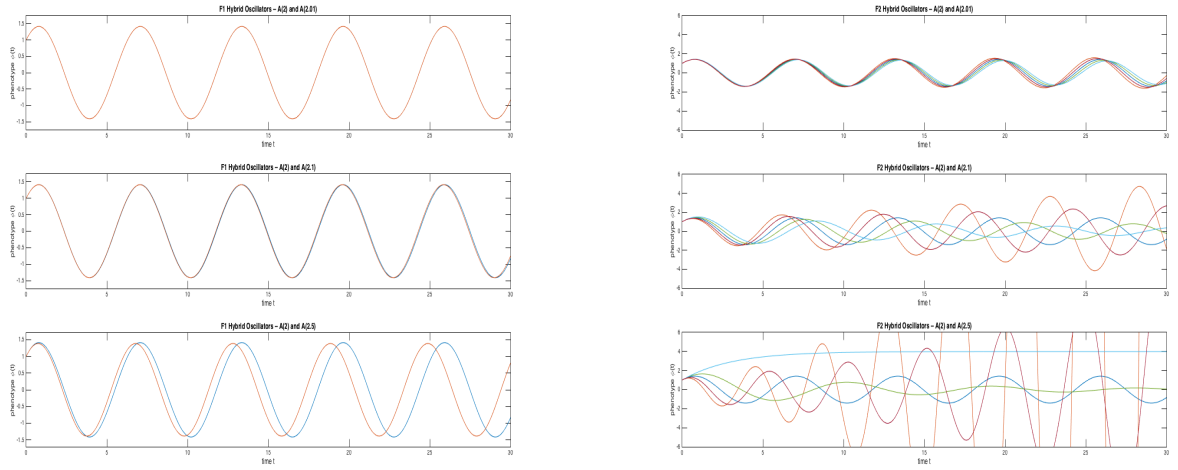


Figure 4: **(left)** Phenotypes of  $F_1$  hybrids between an  $A(2)$  parent and, top-to-bottom, an  $A(2.01)$ , an  $A(2.1)$ , and  $A(2.5)$  parent. Parental phenotypes ( $\sin t + \cos t$ ) are shown in blue, and hybrid phenotypes in orange. **(right)** Phenotypes of  $F_2$  hybrids between the same set of parents, with parental phenotype again in blue. Different lines correspond to different  $F_2$ s produced by recombination; note that some completely fail to oscillate. *make axis labels bigger*





Figure 5: **(left)** Phenotype distance to optimum,  $D$ , using  $\rho(t) = \exp(-t/4\pi)$  for  $F_1$  hybrids between an  $A(2)$  and an  $A(2 + \epsilon)$  parent. **(right)** Same, but for  $F_2$  offspring (one line per type of  $F_2$ ). *make this  $D$  not  $D^2$  and say which line is the overall average  $F_2$ .*

**Haldane's rule** *Maybe something like dosage compensation drift could also lead to Haldane's rule. That is, if the dosage compensation can either be: upregulate the X by 2 in male flies, versus in another species: downregulate the X in females by 1/2, and all possible combinations inbetween?* 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]. 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 are  $(\begin{bmatrix} A_1 & A_2 \\ X_1 & X_2 \end{bmatrix}, \begin{bmatrix} A_1 & A_2 \\ X_1 & X_2 \end{bmatrix})$  has phenotype determined by  $A = \begin{bmatrix} A_1 & A_2 \\ X_1 & X_2 \end{bmatrix}$ , thanks to *Drosophila*-like dosage compensation mechanism (the X is twice upregulated in heterogametes relative to homogametes), while a female homozygous for the haplotype  $\begin{bmatrix} \bar{A}_1 & \bar{A}_2 \\ \bar{X}_1 & \bar{X}_2 \end{bmatrix}$ , has phenotype determined by  $A = \begin{bmatrix} \bar{A}_1 & \bar{A}_2 \\ \bar{X}_1 & \bar{X}_2 \end{bmatrix}$ . An  $F_1$  male offspring of these two will have phenotype determined by  $\begin{bmatrix} (A_1 + \bar{A}_1)/2 & (A_2 + \bar{A}_2)/2 \\ \bar{X}_1 & \bar{X}_2 \end{bmatrix}$ . 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  $F_2$ s.

This mechanism is apparently distinct from the currently favored “dominance”, “faster-X”, and “faster-male” theories [Orr, 1997, Coyne and Orr, 1998] to explain Haldane's rule, and makes several different predictions than the dominance theory as formulated by Turelli and Orr [1995]. *what are they?* Turelli and Orr [1995] employ a population genetic model which strictly requires alleles causing hybrid incompatibility on the X chromosome not to exhibit additive effects, that is the *dominance coefficient* (typically  $h$  in classic population genetics) must not be equal to  $\frac{1}{2}$ . In their model, if alleles are strictly additive, that is neither dominant nor recessive, then male and female  $F_1$  hybrids will drop in fitness at the exact same rates, and therefore defy Haldane's rule.

## System drift and the accumulation of incompatibilities

Thus far we have shown that many distinct molecular mechanisms can realize identical phenotypes and that these mechanisms may fail to produce viable hybrids. This begs the question: does evolution shift molecular mechanisms fast enough to be a significant driver of speciation? To approach this question, we explore a general quantitative genetic model in which a population drifts stochastically near a set of equivalent and

optimal systems due to the action of recombination, mutation, and demographic noise. Although this is motivated by the results on linear systems above, the quantitative genetics calculations are more general, and only depend on the presence of genetic variation and a continuous set of phenotypically equivalent systems.

We will suppose that each organism’s phenotype is determined by her vector of coefficients, denoted by  $x = (x_1, x_2, \dots, x_L)$ , and that the corresponding fitness is determined by the distance of her phenotype to optimum. The optimum phenotype is unique, but is realized by many distinct  $x$  – those falling in the “optimal set”  $\mathcal{N}$ . The phenotypic distance to optimum of an organism with coefficients  $x$  is denoted  $D(x)$ . In the results above,  $x = (A, B, C)$  and  $D(x)$  is given by equation (5). determines a system (this is a system  $(A, B, C)$  above) with fitness determined by its phenotypic distance  $D(x)$  from an optimum (this is analogous to  $\mathcal{N}$  above). Concretely, we define the fitness at  $x$  to be  $\exp(-D(x)^2)$ . We will assume that in the region of interest, the map  $D$  is smooth and that we can locally approximate the optimal set  $\mathcal{N}$  as a quadratic surface. As above, an individual’s coefficients are given by averaging her parentally inherited coefficients and adding random noise due to segregation. Concretely, we use the *infinitesimal model* for reproduction [?] – the offspring of parents at  $x$  and  $x'$  will have coefficients  $(x + x')/2 + \varepsilon$ , where  $\varepsilon$  is a Gaussian displacement due to random assortment of parental alleles.

**System drift** We work with a randomly mating population of effective size  $N_e$ . If the regulatory coefficient population variation has standard deviation  $\sigma$  in a particular direction, since subsequent generations resample from this diversity, the population mean coefficient will move a random distance of size  $\sigma/\sqrt{N_e}$  per generation, simply because this is the standard deviation of the mean of a random sample [?]. Selection will tend to restrain this motion, but movement along the optimal set  $\mathcal{N}$  is unconstrained, and so we expect the population mean to drift along the optimal set like a particle diffusing. The amount of variance in particular directions in coefficient space depends on constraints imposed by selection and correlations between the genetic variation underlying different coefficients (the  $G$  matrix [?]). It therefore seems reasonable to coarsely model the time evolution of population variation in regulatory coefficients as a “cloud” of width  $\sigma$  about the population mean, which moves as an unbiased Brownian motion through the set of network coefficients that give the optimal phenotype.

There will in general be different amounts of variation in different directions; to keep the discussion intuitive, we only discuss  $\sigma_N$ , the amount of variation in “neutral” directions (i.e., directions along  $\mathcal{N}$ ), and  $\sigma_S$ , the amount of variation in “selected” directions (perpendicular to  $\mathcal{N}$ ). The other relevant scale we denote by  $\gamma$ , which the scale on which distance to phenotypic optimum changes as  $x$  moves away from the optimal set,  $\mathcal{N}$ . Concretely,  $\gamma$  is  $1/(\frac{d}{du} D(x + uz))$  with respect to  $u$  where  $x$  is optimal and  $z$  is a “selected” direction perpendicular to  $\mathcal{N}$ . With these parameters, a typical individual will have a fitness of around  $\exp(-(\sigma_S/\gamma)^2)$ . Of course, there are in general many possible neutral and selected directions; we take these values to be representative of the possible directions.

**Hybridization** The means of two allopatric populations each of effective size  $N_e$  separated for  $T$  generations will be a distance roughly of order  $2\sigma_N\sqrt{T/N_e}$  apart along  $\mathcal{X}$ . (Consult figure 3 for a conceptual diagram.) A population of  $F_1$  hybrids has one haploid genome from each, whose coefficients are averaged, and so will have mean system coefficients at the midpoint between their means. Each  $F_2$  hybrid will be homozygous for one parental allele on average at half of the loci in the genome, so the distribution of  $F_2$ s will have mean at the average of the two populations, but will have higher variance. The variance of  $F_2$ s can be shown to increase linearly with the square of the distance between parental population means under models of both simple and polygenic traits. This is suggested by figure 3 and shown in Appendix C. *connect to Barton etc polygenic adaptation lit* Concretely, we expect the population of  $F_1$ s to have variance  $\sigma_S^2$  in the selected direction (the same as within each parental population), but the population of  $F_2$ s will have variance of order  $\sigma_S^2 + 4\omega\sigma_N^2T/N_e$ , where  $\omega$  is a factor that depends on the genetic basis of the coefficients. In a model of  $p$  traits in which the optimal set  $\mathcal{N}$  has dimension  $q$ , using the polygenic model of appendix C,  $\omega = (p - q)/8$ . If each trait is controlled by a single locus, as in figure 3, the value is similar.

What are the fitness consequences? A population of  $F_2$ s will begin to be substantially less fit than the

parentals once they differ from the optimum by a distance of order  $\gamma$ , i.e., once  $\sqrt{4\omega T/N_e} \approx \gamma/\sigma_N$ . This implies that hybrid incompatibility among  $F_2$ s should appear on a time scale of  $N_e(\gamma/\sigma_N)^2/(4\omega)$  generations. The  $F_1$ s will not suffer fitness consequences until the hybrid mean is further than  $\gamma$  from the optimum; as shown in appendix D (and suggested by figure 3), this deviation of the mean from optimum grows with the square of the distance between the parental populations, and so we expect fitness costs in  $F_1$ s to appear on a time scale of  $N_e^2$  generations.

For a more concrete prediction, suppose that the distribution among hybrids is Gaussian. A population whose trait distribution is Gaussian with mean  $\mu$  and variance  $\sigma$ , has mean fitness

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{-\frac{x^2}{2\gamma^2}} dx = \sqrt{\frac{1}{1+\sigma^2/\gamma^2}} \exp\left\{-\frac{\mu^2}{\gamma^2} \left(\frac{1}{1+\sigma^2/\gamma^2}\right)\right\}. \quad (6)$$

This assumes a single trait, for simplicity; the multivariate case is done in appendix E. A population of  $F_2$ s will have, as above, variance  $\sigma^2 = \sigma_S^2 + 4\omega\sigma_N^2 T/N_e$ . The mean diverges with the square of the distance between the parentals with a speed depending on the local geometry of the optimal set (calculated in appendix F), so we set  $\mu = c_\mu \gamma T/N_e$ . The mean fitness in parental populations is as in equation 6 with  $\mu = 0$  and  $\sigma = \sigma_S$ . This implies that if we define  $\mathcal{F}_2(T)$  to be the mean relative fitness among  $F_2$  hybrids between two populations separated by  $T$  generations, (i.e., the mean fitness divided by the mean fitness of the parents) then *figure out how dimensions come in here*

$$\mathcal{F}_2(T) = \left(1 + \frac{4\omega(\sigma_N/\gamma)^2}{(1+(\sigma_S/\gamma)^2)} \frac{T}{N_e}\right)^{-1/2} \exp\left\{-\left(c_\mu \frac{T}{N_e}\right)^2 \left(\frac{1}{1+(\sigma_S/\gamma)^2 + 4\omega(\sigma_N/\gamma)^2 T/N_e}\right)\right\}. \quad (7)$$

If each of the  $q$  selected directions acts independently, the drop in fitness will be  $\mathcal{F}_2(T)^q$  (the expression for the correlated cases is given in Appendix E). We discuss the implications of this expression in the next section.

**Speciation rates under neutrality** Above, equation (7) shows how fast hybrids become inviable as the time that the parental populations are isolated increases; what does this tell us about speciation rates under neutrality? From equation (7) we can observe that time is always scaled in units of  $N_e$  generations, the population standard deviations are always scaled by  $\gamma$ , and the most important term is the rate of accumulation of segregation variance,  $4\omega(\sigma_N/\gamma)^2$ . All else being equal, this process will lead to speciation more quickly in smaller populations and in populations with more neutral genetic variation (larger  $\sigma_N$ ). These parameters are related – larger populations generally have more genetic variation – but since these details depend on the situation, we leave these separate.

How does this prediction depend on the system size and constraint? If there are  $p$  trait dimensions, constrained in  $q$  dimensions, and if  $\omega$  is proportional to  $p-q$ , then the rate that  $F_2$  fitness drops is, roughly,  $(1+4(p-q)CT/n_e)^{-q/2} \propto q(p-q)$ , where  $C$  is a constant. This suggests that both degree of constraint and number of available neutral directions affect the speed of accumulation of incompatibilities. However, note that in real systems, it is likely that  $\gamma$  also depends on  $p$  and  $q$ . *revisit with sims*

Suppose in a large, genetically diverse population, the amount of heritable variation in the neutral and selected directions are roughly equal ( $\sigma_N \approx \sigma_S$ ) but the overall amount of variation is (weakly) constrained by selection ( $\sigma_N \approx \gamma$ ). If so, then the first term of equation (7) is  $1/\sqrt{1+2\omega T/N_e} \approx 1 - \omega T/N_e$ . If also  $\omega = 1$ , then, for instance, after  $0.1N_e$  generations the average  $F_2$  fitness has dropped by 10% relative to the parentals.

Consider instead a much smaller, isolated population whose genetic variation is primarily constrained by genetic drift, so that  $\sigma_N \approx \sigma_S \ll \gamma$ . Setting  $a = (\sigma_N/\gamma)^2$  to be small, the fitness of  $F_2$ s is  $\mathcal{F}_2 \leq 1/\sqrt{1+4\omega a T/N_e} \approx 1 - 2\omega a T/N_e$ . Hybrid fitness seems to drop more slowly in this case in figure 6, but since time is scaled by  $N_e$ , so speciation may occur *faster* than in a large population. However, at least in some models [?], in small populations at mutation-drift equilibrium the amount of genetic variance ( $\sigma_N^2$ ) is proportional to  $N_e$ , which would compensate for this difference, perhaps even predicting the rate of decrease of hybrid fitness to be *independent* of population size for small populations.

In the other direction, consider large metapopulations (or isolated “species complexes”) among which heritable variation is strongly constrained by selection (i.e., there is substantial recombination load), so that  $\sigma_S \approx \gamma$  but  $\sigma_N/\gamma$  is large. Then the fitness of  $F_2$ s is  $\mathcal{F}_2 \leq 1/\sqrt{1 + 2\omega aT/N_e} \approx 1 - \omega aT/N_e$ , and could be extremely rapid if  $a$  is large.

For instance, in a population of one million organisms that has 10 generations per year (a drosophilid species, perhaps) under the “large population” scenario of Figure 6A, system drift would lead to a substantial fitness drop of around 10% in  $F_2$  hybrids in only 10,000 years. This drop may be enough to induce evolutionary reinforcement of reproductive isolation. If one thousand of these organisms is isolated (perhaps on an island, as in Figure 6B), then a similar drop could occur in around 120 years. On the other hand, if the population is one of several of similar size that have recently come into secondary contact after population re-expansion, the situation may be similar to that of Figure 6C with  $N_e = 10^6$ , the same drop could occur after 1,100 years. However, hyperdiverse populations of this type may not be stable on these time scales.

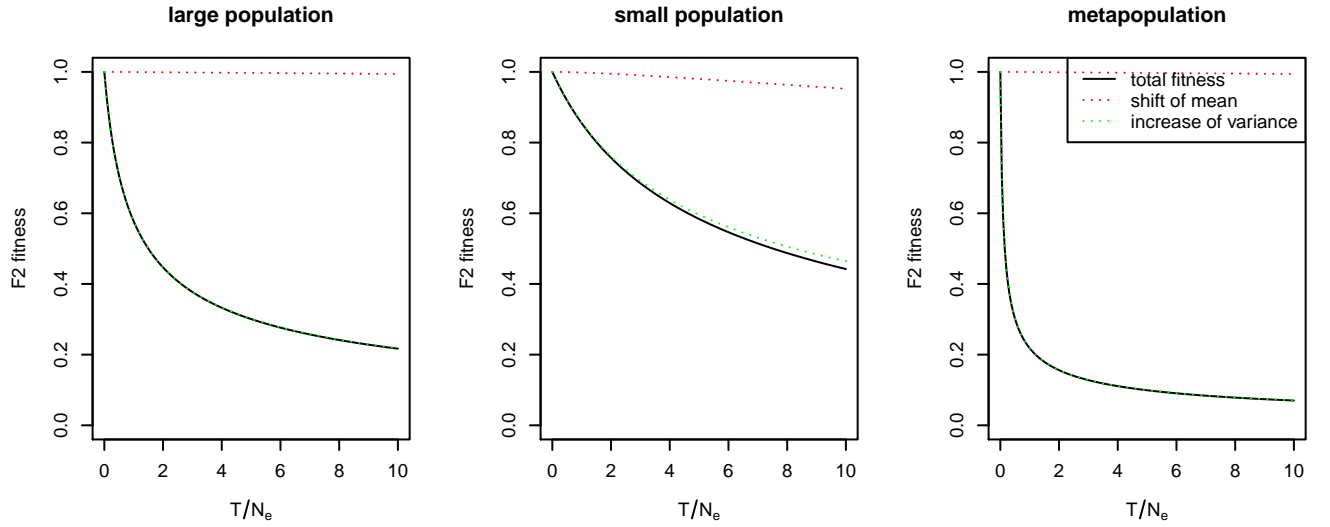


Figure 6: Mean drop if  $F_2$  fitness relative to parental species, with  $\omega = 1$  and (A)  $\sigma_N^2 = \sigma_S^2 = \gamma^2$  (B)  $\sigma_N^2 = \sigma_S^2 = 0.1\gamma^2$  (C)  $0.1\sigma_N^2 = \sigma_S^2 = \gamma^2$ . The total fitness is from equation (7), and is the product of the “shift of mean” and “increase of variance” terms (the exponential and the square root, respectively). Note that, as we assume, the contribution of the shift in mean is small relative to that of increased variance.

## 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 [?]. 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  $\mathcal{S}$  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 be occupied for transcription to begin), and variation

in expression levels measures changes in steady-state concentration (our  $\kappa_i$ ) rather than the *rate* of change. Nonetheless, [Kasowski et al. \[2010\]](#) found differential occupancy in 7.5% of binding sites of a transcription factor (p65) between human individuals. [Verlaan et al. \[2009\]](#) showed that cis-regulatory variation accounts for around 2–6% of expression variation in human blood-derived primary cells, while [Lappalainen et al., 2013](#) found that human population variation explained about 3% of expression variation. Taken together, this suggests that variation in the entries of  $\mathcal{S}$  may be on the order of at least a few percent between individuals of a population – doubtless varying substantially between species and between genes.

The impact of regulatory variation ( $\sigma$ ) above depends only on its magnitude relative to selection ( $\gamma$ ), but it seems reasonable that these two quantities are of the same magnitude.

## Discussion

Above we synthesize concepts and tools from molecular and evolutionary biology with tools from control theory to study the evolution of a mechanistic model of the molecular genotype-phenotype map under stabilizing selection. This model allows us to analytically describe the processes discussed verbally as phenogenetic drift [\[Weiss and Fullerton, 2000\]](#) and developmental system drift [\[True and Haag, 2001\]](#). In this context, the Kalman decomposition [\[Kalman, 1963\]](#) implies that nearly all systems are nonidentifiable, and gives an analytical description of all phenotypically equivalent gene networks. System nonidentifiability, rather than being a computational nuisance, implies the existence of axes of genetic variation that are not constrained by selection. The independent movement of separated populations along these axes can lead to reduction in hybrid viability, and hence speciation, at a speed that depends on effective population size and amount of genetic variation. In this model, at biologically reasonable parameter values, system drift is a significant – and often rapid – driver of speciation. This may at first be surprising because hybrid inviability appears as a consequence of recombining different, yet functionally equivalent, mechanisms.

Consistent with empirical observation [cite](#), we see that the fitnesses of  $F_2$  hybrids drop at a much faster rate than those of  $F_1$ s, since  $F_1$  and  $F_2$  phenotypes diverge linearly and at the square root of time relative to parentals, respectively. [refer to Turelli here](#) 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  $F_1$ s show variation similar to that seen in  $F_2$ s.

Is there evidence that this is actually occurring? System drift and network rewiring has been inferred across the tree of life [\[Dalal and Johnson, 2017, Johnson, 2017\]](#), and there is often significant regulatory variation segregating within populations [\[?\]](#). [cite sergey?](#) Transcription in hybrids between closely related species with conserved transcriptional patterns can also be divergent [\[Haerty and Singh, 2006, Maheshwari and Barbash, 2012, Coolon et al., 2014, Michalak and Noor, 2004\]](#). Furthermore, in cryptic species complexes (e.g., sun skinks [\[Barley et al., 2013\]](#)), strongly reproductively incompatible species may be nearly morphologically indistinguishable. [check the skinks have RI!! revisit](#) Of course the basic assumptions of this model will be violated in practice (constant selective pressures, etc.), however, this model can function as a “neutral null” description of gene network evolution (a strategy advocated for in [Lynch \[2007\]](#), [Koonin \[2016\]](#)). If this model succeeds in describing actual system evolution, it can be inferred that the mechanism underlying species formation does not require the inclusion of adaptation or changes in selection to explain the emergence of hybrid incompatibility.

**The origin of species not by means of natural selection?** Although, 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 for the origin of Dobzhansky–Muller incompatibilities by neutral genetic drift have been proposed [\[?\]](#) and under stabilizing selection [\[Fierst and Hansen, 2009\]](#), previous authors have argued that neutral processes are likely too slow to be a significant driver of speciation [\[Nei et al., 1983, Seehausen et al., 2014\]](#). Using simulations, [Porter and Johnson \[2002\]](#) demonstrated the accumulation of hybrid incompatibilities under directional, but not

stabilizing selection, and Palmer and Feldman [2009] observed the appearance of incompatibilities in suboptimally adapted populations in a constant environment. This, in light of the few known incompatibilities, has lead 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 [?]. In contrast, the model we develop here, suggests that even under strictly neutral conditions, and in the absence of adaptation, system drift can lead to speciation and perhaps at a rate fast enough to play a major role in species formation across the tree of life. Our results show that hybrids, under neutral processes, break down as a function of genetic distance, and despite the present model’s myriad simplifications and assumptions, it may, in part, explain the observed broad patterns of reproductive isolation, such as the observed consistent relationship between molecular divergence and genetic isolation across diverse taxa and ecologies by Roux et al. [2016], and the suggestively nonadaptive, clocklike speciation patterns observed by Hedges et al. [2015] (as diversification is not observed to be accelerated following mass extinctions). While the incompatibility of any one given network within an organism may be small, we note that organisms are made up of many different networks, and that the cumulative impact of system drift across all of these may be substantial. Further, even in populations experiencing directional selection, networks underlying conserved parts of the phenotype can still experience neutral drift and harbor incompatibilities.

**Nonlinearity** 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 qualitatively similar results. Furthermore, nonidentifiability (which implies the existence of neutral directions) is often found in practice in moderately complex models of biological systems. *need citation*

**Peter: Quant gen discussion** Maybe compare to Fierst and Hansen [2009]?

Do we talk about the  $G$  matrix and mutational correlation somewhere?

Note that islands or bottlenecks leave large  $\sigma_N$  but small  $N_e$  so may go fast.

What is  $N_e$ ? In a metapopulation depends on migr rate: classic Turelli paper.

This has implications for genetic load also!!

hominids; neanderthal load.

## Acknowledgements

We would like to thank Sergey Nuzhdin, Stevan Arnold, Erik Lundgren, and Hossein Asgharian for valuable discussion.

## References

- Md Kausar Alam and Susan GW Kaminskyj. Aspergillus galactose metabolism is more complex than that of saccharomyces: the story of gal7 and gal1. *Botany*, 91(7):467–477, 2013. 2
- 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. 2, 5
- 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. 13
- Richard Ernest Bellman and Karl Johan Åström. On structural identifiability. *Mathematical biosciences*, 7(3-4):329–339, 1970. 1



- Aviv Bergman and Mark L Siegal. Evolutionary capacitance as a general feature of complex gene networks. *Nature*, 424(6948):549–552, 2003. 1
- Aleksandra A. Chertkova, Joshua S. Schiffman, Sergey V. Nuzhdin, Konstantin N. Kozlov, Maria G. Samsonova, and Vitaly V. Gursky. In silico evolution of the drosophila gap gene regulatory sequence under elevated mutational pressure. *BMC Evolutionary Biology*, 17(1):4, 2017. ISSN 1471-2148. doi: 10.1186/s12862-016-0866-y. URL <http://dx.doi.org/10.1186/s12862-016-0866-y>. 1
- Ulla Christensen, Birgit S Gruben, Susan Madrid, Harm Mulder, Igor Nikolaev, and Ronald P de Vries. Unique regulatory mechanism for d-galactose utilization in aspergillus nidulans. *Applied and environmental microbiology*, 77(19):7084–7087, 2011. 2
- 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. 13
- 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. 9, 13
- Gheorghe Craciun and Casian Pantea. Identifiability of chemical reaction networks. *Journal of Mathematical Chemistry*, 44(1):244–259, 2008. 2
- 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. 1
- Frederick R Cross, Nicolas E Buchler, and Jan M Skotheim. Evolution of networks and sequences in eukaryotic cell cycle control. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 366(1584):3532–3544, 2011. 2
- 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. 2, 13
- 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. 2
- Eric H Davidson and Douglas H Erwin. Gene regulatory networks and the evolution of animal body plans. *Science*, 311(5762):796–800, 2006. 1
- Jeremy Draghi and Michael Whitlock. Robustness to noise in gene expression evolves despite epistatic constraints in a model of gene networks. *Evolution*, 69(9):2345–2358, 2015. 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. 13, 14
- 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
- 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. 13
- J BS Haldane. Sex ratio and unisexual sterility in hybrid animals. *Journal of genetics*, 12(2):101–109, 1922. 9

494 Thomas F. Hansen and Emilia P. Martins. Translating between microevolutionary process and macroevolu-  
495 tionary patterns: The correlation structure of interspecific data. *Evolution*, 50(4):1404–1417, 1996. ISSN  
496 00143820, 15585646. URL <http://www.jstor.org/stable/2410878>. 22, 27

497 Emily E Hare, Brant K Peterson, Venky N Iyer, Rudolf Meier, and Michael B Eisen. Sepsid even-skipped  
498 enhancers are functionally conserved in drosophila despite lack of sequence conservation. *PLoS Genet*, 4  
499 (6):e1000106, 2008. 2

500 Lukas Hartl, Christian P Kubicek, and Bernhard Seiboth. Induction of the gal pathway and cellulase genes  
501 involves no transcriptional inducer function of the galactokinase in hypocrea jecorina. *Journal of Biological*  
502 *Chemistry*, 282(25):18654–18659, 2007. 2

503 S Blair Hedges, Julie Marin, Michael Suleski, Madeline Paymer, and Sudhir Kumar. Tree of life reveals  
504 clock-like speciation and diversification. *Molecular biology and evolution*, 32(4):835–845, 2015. 14

505 Jennifer W Israel, Megan L Martik, Maria Byrne, Elizabeth C Raff, Rudolf A Raff, David R McClay, and  
506 Gregory A Wray. Comparative developmental transcriptomics reveals rewiring of a highly conserved gene  
507 regulatory network during a major life history switch in the sea urchin genus *heliocidaris*. *PLoS Biol*, 14  
508 (3):e1002391, 2016. 1

509 Johannes Jaeger. The gap gene network. *Cellular and Molecular Life Sciences*, 68(2):243–274, 2011. 1

510 Johannes Jaeger, Svetlana Surkova, Maxim Blagov, Hilde Janssens, David Kosman, Konstantin N Kozlov,  
511 Ekaterina Myasnikova, Carlos E Vanario-Alonso, Maria Samsonova, David H Sharp, et al. Dynamic control  
512 of positional information in the early drosophila embryo. *Nature*, 430(6997):368–371, 2004. 1

513 Johannes Jaeger, Manfred Laubichler, and Werner Callebaut. The comet cometh: evolving developmental  
514 systems. *Biological theory*, 10(1):36–49, 2015. 1

515 Alexander D Johnson. The rewiring of transcription circuits in evolution. *Current Opinion in Genetics &*  
516 *Development*, 47:121–127, 2017. 13

517 R. E. 1930-(Rudolf Emil) Kalman, Peter L. Falb, and Michael A. Arbib. *Topics in mathematical system*  
518 *theory*. McGraw-Hill, New York, 1969. ISBN 0754321069. 5

519 Rudolf Emil Kalman. Mathematical description of linear dynamical systems. *J.S.I.A.M Control*, 1963. 2, 5,  
520 13

521 M. Kasowski, F. Grubert, C. Heffelfinger, M. Hariharan, A. Asabere, S. M. Waszak, L. Habegger, J. Ro-  
522 zowsky, M. Shi, A. E. Urban, M. Y. Hong, K. J. Karczewski, W. Huber, S. M. Weissman, M. B. Gerstein,  
523 J. O. Korbel, and M. Snyder. Variation in transcription factor binding among humans. *Science*, 328(5975):  
524 232–235, April 2010. 13

525 Stephen E Kearsey and Sue Cotterill. Enigmatic variations: divergent modes of regulating eukaryotic dna  
526 replication. *Molecular cell*, 12(5):1067–1075, 2003. 2

527 Eugene V Koonin. Splendor and misery of adaptation, or the importance of neutral null for understanding  
528 evolution. *BMC biology*, 14(1):114, 2016. 13

529 Konstantin Kozlov, Svetlana Surkova, Ekaterina Myasnikova, John Reinitz, and Maria Samsonova. Modeling  
530 of gap gene expression in *Drosophila Kruppel* mutants. *PLoS Computational Biology*, 2012. 1

531 Konstantin Kozlov, Vitaly Gursky, Ivan Kulakovskiy, and Maria Samsonova. Sequence-based model of gap  
532 gene regulatory network. *BMC Genomics*, 2014. 1

533 Konstantin Kozlov, Vitaly V Gursky, Ivan V Kulakovskiy, Arina Dymova, and Maria Samsonova. Analysis  
534 of functional importance of binding sites in the *Drosophila* gap gene network model. *BMC Genomics*,  
535 2015. 1

- Russell Lande. Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences*, 78(6):3721–3725, 1981. URL <http://www.pnas.org/content/78/6/3721.abstract>. 22
- Tuuli Lappalainen, Michael Sammeth, Marc R Friedländer, Peter ACt 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. 13
- Hugo Lavoie, Hervé Hogues, and Malcolm Whiteway. Rearrangements of the transcriptional regulatory networks of metabolic pathways in fungi. *Current opinion in microbiology*, 12(6):655–663, 2009. 2
- Michael Lynch. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences*, 104(suppl 1):8597–8604, 2007. 13
- 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. 13
- Mikhail Martchenko, Anastasia Levitin, Herve Hogues, Andre Nantel, and Malcolm Whiteway. Transcriptional rewiring of fungal galactose-metabolism circuitry. *Current Biology*, 17(12):1007–1013, 2007. 2
- 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. 13
- Eric Mjolsness, David H Sharp, and John Reinitz. A connectionist model of development. *Journal of theoretical Biology*, 152(4):429–453, 1991. 1
- Masatoshi Nei, Takeo Maruyama, and Chung-I Wu. Models of evolution of reproductive isolation. *Genetics*, 103(3):557–579, 1983. 13
- H Allen Orr. Haldane’s rule. *Annual Review of Ecology and Systematics*, 28(1):195–218, 1997. 9
- H Allen Orr, John P Masly, and Daven C Presgraves. Speciation genes. *Current opinion in genetics & development*, 14(6):675–679, 2004. 14
- Michael E Palmer and Marcus W Feldman. Dynamics of hybrid incompatibility in gene networks in a constant environment. *Evolution*, 63(2):418–431, 2009.
- Mihaela Pavlicev and Gunter P Wagner. A model of developmental evolution: selection, pleiotropy and compensation. *Trends in Ecology & Evolution*, 2012. 1
- Adam H Porter and Norman A Johnson. Speciation despite gene flow when developmental pathways evolve. *Evolution*, 56(11):2103–2111, 2002. 13
- 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. 14
- Aziz Sancar. The intelligent clock and the rube goldberg clock. *Nature structural & molecular biology*, 15(1):23–24, 2008. 2
- Dolph Schluter. Evidence for ecological speciation and its alternative. *Science*, 323(5915):737–741, 2009. 14
- 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. 13
- Maria R Servedio, Yaniv Brandvain, Sumit Dhole, Courtney L Fitzpatrick, Emma E Goldberg, Caitlin A Stern, Jeremy Van Cleve, and D Justin Yeh. Not just a theory: the utility of mathematical models in evolutionary biology. *PLoS Biol*, 12(12):e1002017, 2014. 1

577 Mark L Siegal and Aviv Bergman. Waddington’s canalization revisited: developmental stability and evolu-  
578 tion. *Proceedings of the National Academy of Sciences*, 99(16):10528–10532, 2002. [1](#)

579 K. Stefflova, D. Thybert, M. D. Wilson, I. Streeter, J. Aleksic, P. Karagianni, A. Brazma, D. J. Adams,  
580 I. Talianidis, J. C. Marioni, P. Flicek, and D. T. Odom. Cooperativity and rapid evolution of cobound  
581 transcription factors in closely related mammals. *Cell*, 154(3):530–540, August 2013. [12](#)

582 Amos Tanay, Aviv Regev, and Ron Shamir. Conservation and evolvability in regulatory networks: the  
583 evolution of ribosomal regulation in yeast. *Proceedings of the National Academy of Sciences of the United*  
584 *States of America*, 102(20):7203–7208, 2005. [2](#)

585 John R True and Eric S Haag. Developmental system drift and flexibility in evolutionary trajectories.  
586 *Evolution & development*, 3(2):109–119, 2001. [1](#), [2](#), [13](#)

587 Annie E Tsong, Brian B Tuch, Hao Li, and Alexander D Johnson. Evolution of alternative transcriptional  
588 circuits with identical logic. *Nature*, 443(7110):415–420, 2006. [2](#)

589 Michael Turelli and H Allen Orr. The dominance theory of haldane’s rule. *Genetics*, 140(1):389–402, 1995.  
590 [9](#)

591 AJ Van der Schaft. Equivalence of dynamical systems by bisimulation. *IEEE transactions on automatic*  
592 *control*, 49(12):2160–2172, 2004. [2](#)

593 Dominique J Verlaan, Bing Ge, Elin Grundberg, Rose Hoberman, Kevin CL Lam, Vonda Koka, Joana Dias,  
594 Scott Gurd, Nicolas W Martin, Hans Mallmin, et al. Targeted screening of cis-regulatory variation in  
595 human haplotypes. *Genome research*, 19(1):118–127, 2009. [13](#)

596 Andreas Wagner. Evolution of gene networks by gene duplications: a mathematical model and its implica-  
597 tions on genome organization. *Proceedings of the National Academy of Sciences*, 91(10):4387–4391, 1994.  
598 [1](#)

599 Andreas Wagner. Does evolutionary plasticity evolve? *Evolution*, pages 1008–1023, 1996. [1](#)

600 Eric Walter, Yves Lecourtier, and John Happel. On the structural output distinguishability of parametric  
601 models, and its relations with structural identifiability. *IEEE Transactions on Automatic Control*, 29(1):  
602 56–57, 1984. [1](#)

603 Kenneth M Weiss and Stephanie M Fullerton. Phenogenetic drift and the evolution of genotype–phenotype  
604 relationships. *Theoretical population biology*, 57(3):187–195, 2000. [1](#), [13](#)

605 Karl R Wotton, Eva Jiménez-Guri, Anton Crombach, Hilde Janssens, Anna Alcaine-Colet, Steffen Lemke,  
606 Urs Schmidt-Ott, and Johannes Jaeger. Quantitative system drift compensates for altered maternal inputs  
607 to the gap gene network of the scuttle fly *megascelia abdita*. *Elife*, 4:e04785, 2015. [1](#)

608 Lotfi A Zadeh and Charles A Deoser. *Linear system theory*. Robert E. Krieger Publishing Company  
609 Huntington, 1976. [2](#)

## 610 Examples

611 **Example 4** (Metabolic network). Consider an organism that can metabolize two different sugars (present  
612 at logarithmic concentrations  $u_1$  and  $u_2$ ), with two enzymes (at log concentrations  $\phi_1$  and  $\phi_2$ ). Further  
613 suppose that the second sugar is preferred, and that depending on  $u_1$  and  $u_2$  the organism will synthe-  
614 size an appropriate  $\phi_1$  and  $\phi_2$ . Furthermore, assume this system contains at least two transcription fac-  
615 tors, whose log concentrations are  $\kappa_1, \kappa_2, \dots, \kappa_n$ . Minimally such a system may have the architecture,  
616  $S_{\min}(U) = (U \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} U^{-1}, U, U^{-1})$ . We can find alternative equivalent systems by changing coordinates

617  $(U \rightarrow U')$  or, more generally, by applying the Kalman decomposition A. To illustrate that phenotypic  
 618 invariance does not require dimensional invariance, we apply the Kalman decomposition to  $\mathcal{S}_{\min}$  to find  
 619  $\mathcal{S}(D_{1-3}, V) = (V \begin{bmatrix} D_1 & D_2 \\ 0 & A \end{bmatrix} V^{-1}, V \begin{bmatrix} D_3 \\ B \end{bmatrix}, [0 \ C] V^{-1})$ , both of which are in  $\mathcal{N}$  ( $V$  can be any 4-dimensional and  
 620 invertible matrix,  $D_{1-3} \in \mathbb{R}^{2 \times 2}$ , and  $(A, B, C) \in \mathcal{S}_{\min}$ ).

## 621 A Kalman Decomposition

**Definition 1** (Phenotypic equivalence of systems). Let  $(\kappa(t), \phi(t))$  and  $(\bar{\kappa}(t), \bar{\phi}(t))$  be the solutions to (??) with coefficient matrices  $(A, B, C)$  and  $(\bar{A}, \bar{B}, \bar{C})$  respectively, and both  $\kappa(0)$  and  $\bar{\kappa}(0)$  are zero. The systems defined by  $(A, B, C)$  and  $(\bar{A}, \bar{B}, \bar{C})$  are **phenotypically equivalent** if

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

Equivalently, this occurs if and only if

$$h(t) = \bar{h}(t) \quad \text{for all } t \geq 0,$$

622 where  $h$  and  $\bar{h}$  are the impulse responses of the two systems.

One way to find other systems equivalent to a given one is by change of coordinates (“algebraic equivalence”): if  $V$  is an invertible matrix, then the systems  $(A, B, C)$  and  $(VAV^{-1}, VB, CV^{-1})$  have the same dynamics because their transfer functions are equal:

$$CV^{-1}(zI - VAV^{-1})^{-1}VB = CV^{-1}V(zI - A)^{-1}V^{-1}VB = C(zI - A)^{-1}B.$$

623 However, the converse is not necessarily true: systems can have identical transfer functions without being  
 624 changes of coordinates of each other. In fact, systems with identical transfer functions can involve interactions  
 625 between different numbers of molecular species.

626 The set of all systems phenotypically equivalent to a given system  $(A, B, C)$  is elegantly described using  
 627 the Kalman decomposition, which also clarifies the system dynamics? tells us a lot about how it works?  
 628 *or something* To motivate this, first note that the input  $u(t)$  only directly pushes the system in directions  
 629 lying in the span of the columns of  $B$ . As a result, different combinations of input can move the system in  
 630 any direction that lies in the *reachable subspace*, which we denote by  $\mathcal{R}$ , and is defined to be the closure of  
 631  $\text{span}(B)$  under applying  $A$  (or equivalently, the span of  $B, AB, A^2B, \dots, A^{n-1}B$ ). Analogously to this, we  
 632 define the *observable subspace*,  $\mathcal{O}$ , to be the closure of  $\text{span}(C^T)$  under applying  $A$ . (Or:  $\mathcal{O}$  is the largest  
 633  $A$ -invariant subspace contained in the null space of  $C$ ; and  $\mathcal{R}$  is the largest  $A$ -invariant subspace contained  
 634 in the image of  $B$ .)

635 If we define

- 636 1. The columns of  $P_{r\bar{o}}$  are an orthonormal basis for  $\mathcal{R} \cap \bar{\mathcal{O}}$ .
- 637 2. The columns of  $P_{ro}$  are an orthonormal basis of the complement of  $\mathcal{R} \cap \bar{\mathcal{O}}$  in  $\mathcal{R}$ .
- 638 3. The columns of  $P_{\bar{r}o}$  are an orthonormal basis of the complement of  $\mathcal{R} \cap \bar{\mathcal{O}}$  in  $\bar{\mathcal{O}}$ .
- 639 4. The columns of  $P_{\bar{r}\bar{o}}$  are an orthonormal basis of the remainder of  $\mathbb{R}^n$ .

If we then define

$$P = [ P_{r\bar{o}} \mid P_{ro} \mid P_{\bar{r}o} \mid P_{\bar{r}\bar{o}} ],$$

then

$$P^T P = \left[ \begin{array}{c|c|c|c} I & 0 & 0 & 0 \\ \hline 0 & I & U & 0 \\ \hline 0 & V & I & 0 \\ \hline 0 & 0 & 0 & I \end{array} \right].$$

Check this. Can we get  $U = V = 0$ ?

The following theorem can be found in SOME REFERENCE.

**Theorem 1** (Kalman decomposition). *For any system  $(A, B, C)$  with corresponding Kalman basis matrix  $P$ , the transformed system  $(PAP^{-1}, PB, CP^{-1})$  has the following form:*

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

and

$$\hat{B} = PB = \begin{bmatrix} B_{r\bar{o}} \\ B_{ro} \\ 0 \\ 0 \end{bmatrix},$$

and

$$\hat{C} = CP^{-1} = \begin{bmatrix} 0 & C_{ro} & C_{\bar{r}\bar{o}} & 0 \end{bmatrix}.$$

The transfer function of both systems is given by

$$H(z) = C_{ro}(zI - A_{ro})^{-1}B_{ro}.$$

In the latter case, we say that the system is *minimal* – there is no equivalent system with a smaller number of species. Note that this says that any two equivalent minimal systems are changes of basis of each other.

Since any system can be put into this form, and once in this form, its transfer function is determined only by  $C_{ro}$ ,  $A_{ro}$ , and  $B_{ro}$ , therefore, the set of all equivalent systems are parameterized by the dimension  $n$ , the choice of basis ( $P$ ), the remaining submatrices in  $\hat{A}$ ,  $\hat{B}$ , and  $\hat{C}$  (which are unconstrained), and an invertible transformation of  $\text{span}(P_{ro})$ , which we call  $T_{ro}$ .

**Theorem 2** (Parameterization of equivalent systems). *Let  $(A, B, C)$  be a minimal system.*

(a) *Every equivalent system is of the form given in Theorem 1, i.e., can be specified by choosing a dimension,  $n$ ; submatrices in  $\hat{A}$ ,  $\hat{B}$ , and  $\hat{C}$  except for  $A_{ro} = A$ ,  $B_{ro} = B$ , and  $C_{ro} = C$ ; and choosing an invertible matrix  $P$ .*

(b) *conjecture: The parameterization is unique if  $P$  is furthermore chosen so that each  $P_x$  other than  $P_{ro}$  is a projection matrix, and that*

$$0 = P_x^T P_y$$

*for all  $(x, y)$  except  $(ro, \bar{r}\bar{o})$ .*

*Another way of saying it: pick the  $\mathcal{R}$  and  $\bar{\mathcal{O}}$  subspaces, that must intersect in something of the minimal dimension; then let  $P$  be the appropriate basis?*

In some situations we may be interested in only “network rewiring”, where  $A$  changes while  $B$  and  $C$  do not. For instance, if all non-regulatory functions of each molecule are strongly constrained, then  $C$  cannot change. Likewise, if responses of each molecule to the external inputs are not changed by evolution, then  $B$  does not change.



## A.1 Neutral directions from the Kalman decomposition

The Kalman decomposition above says that any system  $(A, B, C)$  can be decomposed into  $(P, \hat{A}, \hat{B}, \hat{C})$  so that

$$\begin{aligned} A &= P^{-1} \hat{A} P \\ B &= P^{-1} \hat{B} \\ C &= \hat{C} P, \end{aligned}$$

and we know precisely how we can change these to preserve the transfer function:

1.  $P \rightarrow P + \epsilon Q$  as long as the result is still invertible,
2.  $\hat{A} \rightarrow A + \epsilon X$  as long as  $X$  is zero in the correct places,
3.  $\hat{B} \rightarrow B + \epsilon Y$  as long as  $Y$  is zero in the correct places,
4.  $\hat{C} \rightarrow C + \epsilon Z$  as long as  $Z$  is zero in the correct places.

By taking  $\epsilon \rightarrow 0$ , these tell us the local directions we can move a system  $(A, B, C)$  in. All statements below are up to first order in  $\epsilon$ , omitting terms of order  $\epsilon^2$ .

First, since  $(P + \epsilon Q)^{-1} = P^{-1} + \epsilon P^{-1} Q P^{-1}$ , modifying  $P \rightarrow P + \epsilon Q$  changes

$$\begin{aligned} A &\rightarrow A + \epsilon P^{-1} \hat{A} Q - \epsilon P^{-1} Q P^{-1} \hat{A} P \\ &= A + \epsilon (A P^{-1} Q - P^{-1} Q A), \\ B &\rightarrow B - \epsilon P^{-1} Q B \\ C &\rightarrow C + \epsilon C P^{-1} Q. \end{aligned}$$

Since  $P$  is invertible and  $Q$  can be anything (if  $\epsilon$  is small enough), this allows changes in the direction of an arbitrary  $W$ :

$$\begin{aligned} A &= A + \epsilon (A W - W A), \\ B &\rightarrow B - \epsilon W B \\ C &\rightarrow C + \epsilon C W. \end{aligned}$$

Then,  $\hat{A} \rightarrow A + \epsilon X$  does

$$A \rightarrow A + \epsilon P^{-1} X P$$

and  $\hat{B} \rightarrow B + \epsilon Y$  does

$$B \rightarrow B + \epsilon P^{-1} Y$$

and  $\hat{C} \rightarrow C + \epsilon Z$  does

$$C \rightarrow C + \epsilon Z P.$$

These degrees of freedom look like they depend on  $P$ , which is not unique, but for any two choices of  $P$  there are corresponding choices of  $X$  that give the same actual change in  $A$  (and likewise for  $Y$  and  $Z$ ).

Therefore, this gives us an upper bound on the number of degrees of freedom, in terms of the dimensions of the blocks in the Kalman decomposition ( $n_{ro}$  etc) and the dimensions of  $B$  and  $C$  ( $n_B$  and  $n_C$  respectively): namely, for  $W$ ,  $X$ ,  $Y$ , and  $Z$  respectively:

$$n^2 + (n_{r\bar{o}} + n_{ro} n_{\bar{r}o} + n_{\bar{r}\bar{o}} (n_{\bar{r}\bar{o}} + n_{\bar{r}o}) + n_{\bar{r}o}^2) + n_B n_{r\bar{o}} + n_C n_{\bar{r}\bar{o}}.$$

However, some of these may be redundant. For instance, changing  $P$  in the direction of a  $Q$  that satisfies both  $A P^{-1} Q = P^{-1} Q A$  and  $C P^{-1} Q = 0$  is equivalent to changing  $B$  by  $Y = Q B$ .

## B Meiotic recombination in linear systems

Recombination is performed by taking two analogous system components from  $\mathcal{S}$  and  $\mathcal{S}'$  and randomly swapping rows or columns. *E.g.* gamete systems  $(A'', B'', C'')$ , produced by the diploid  $\{\mathcal{S}, \mathcal{S}'\}$ , are formed by recombining (randomly swapping rows or columns) between two, possibly distinct, systems  $\mathcal{S} = (A, B, C)$  and  $\mathcal{S}' = (A', B', C')$  such that,

$$\mathcal{S}'' = \begin{pmatrix} A'' & = MA + (I - M)A', \\ B'' & = MB + (I - M)B', \\ C'' & = CM + C'(I - M) \end{pmatrix}$$

where  $M$  is a diagonal matrix where each diagonal element is a Bernoulli random variable ( $M_{ii} = 0$  or  $1$  with equal probability, and  $M_{ij} = 0$  if  $i \neq j$ ). If systems are different dimensions the smaller system elements can be augmented with 0s (*e.g.*  $\begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} B \\ 0 \end{bmatrix}, \begin{bmatrix} C & 0 \end{bmatrix}$ ).

## C Genetic drift with a multivariate trait

For completeness, we provide a brief exposition of how a population evolves due to genetic drift with a quantitative genetics model, as in Lande [1981] or Hansen and Martins [1996]. These do not directly model underlying genetic basis, but developing a more accurate model is beyond the scope of this paper.

Suppose that the population is distributed in trait space as a Gaussian with covariance matrix  $\Sigma$  and mean  $\mu$ , whose density we write as  $f(\cdot; \Sigma, \mu)$ . Selection has the effect of multiplying this density by the fitness function and renormalizing, so that if expected fitness of  $x$  is proportional to  $\exp(-\|Lx\|^2/2)$ , then the distribution post-selection has density at  $x$  proportional to  $f(x; \Sigma, \mu) \exp(-\|Lx\|^2/2)$ . By the computation below (“Completing the square”), the result is a Gaussian distribution with covariance matrix  $(\Sigma^{-1} + L^T L)^{-1}$  and mean  $(\Sigma^{-1} + L^T L)^{-1} \Sigma^{-1} \mu$ .

After selection, we have reproduction: suppose this occurs as in the infinitesimal model [?], so that each offspring of parents with traits  $x$  and  $y$  is drawn independently from a Gaussian distribution with mean  $(x + y)/2$  and covariance matrix  $R$ . Here,  $R$  is the contribution of “segregation variance”. If  $\tilde{\Sigma} = (\Sigma^{-1} + L^T L)^{-1}$  is the covariance matrix of the parents post-selection, then the distribution of offspring will again be Gaussian, with mean equal to that of the parents and covariance matrix  $\tilde{\Sigma}/2 + R$ .

In summary, a generation under this model modifies the mean ( $\mu$ ) and covariance matrix ( $\Sigma$ ) of a population as follows:

$$\begin{aligned} \mu &\mapsto \mu' = (\Sigma^{-1} + L^T L)^{-1} \Sigma^{-1} \mu \\ \Sigma &\mapsto \Sigma' = \frac{1}{2}(\Sigma^{-1} + L^T L)^{-1} + R. \end{aligned}$$

What measures are stable under this transformation? The condition  $\mu = \mu'$  reduces to  $\Sigma L^T L \mu = 0$ ; if we assume  $R$  and therefore  $\Sigma$  are of full rank, then this happens if and only if  $\mu$  is in the null space of  $L$ , i.e., if  $\mu$  lies in a neutral direction. The condition  $\Sigma' = \Sigma$  can also be solved, at least numerically. After rearrangement, it reduces to  $\Sigma L^T L \Sigma + (I/2 - R L^T L) \Sigma = R$ . We can find a more explicit description if we assume that  $x^T L^T L x = \sum_{i=1}^k x_i^2$ , i.e., that selection only cares about the first  $k$  coordinates, and then with no interactions between traits. If so, the condition  $\Sigma' = \Sigma$  can be written in block form as

$$\begin{bmatrix} \Sigma_{11}^2 + (I/2 - R_{11})\Sigma_{11} & \Sigma_{11}\Sigma_{12} + (I/2 - R_{11})\Sigma_{12} \\ \Sigma_{12}^T \Sigma_{11} + \Sigma_{12}^T/2 - R_{12}^T \Sigma_{11} & \Sigma_{22} - R_{12}^T \Sigma_{12} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} \\ R_{12}^T & R_{22} \end{bmatrix}.$$

The first equation,  $\Sigma_{11}^2 + (I/2 - R_{11})\Sigma_{11} = R_{11}$ , can be solved with the quadratic formula:

$$\Sigma_{11} = (R_{11} - I/2 + Q)/2$$

for any  $Q$  that commutes with  $R_{11}$  and is a solution to  $Q^2 = (R_{11} - I/2)^2 + 4R_{11}$ . Since we need  $\Sigma$  to be positive definite, we take the solution with positive eigenvalues. Given  $\Sigma_{11}$ , the remaining components are

$$\begin{aligned}\Sigma_{12} &= (\Sigma_{11} + I/2 - R_{11})^{-1} R_{12} \\ \Sigma_{22} &= R_{12}^T \Sigma_{12} + R_{22}.\end{aligned}$$

691 Importantly, the mean  $\mu$  does not affect either how the covariance matrix moves, or its stable shape.

Above we have described the *expected* motion of the mean and covariance.. However, random resampling will introduce noise. Suppose that a population of  $N$  individuals behaves approximately as described above. By the above, we may expect that the covariance matrix stays close to a constant value  $\Sigma$ , computed from  $R$  and  $L$  as above, so that we need only consider motion of the mean,  $\mu$ . Since we take a sample of size  $N$  to construct the next generation, the next generation's mean is drawn from a Gaussian distribution with mean  $\mu'$  and covariance matrix  $\Sigma/N$ . Defining  $\Gamma = (I - (I + \Sigma L^T L)^{-1})$ , this can be written as

$$\mu' - \mu = \Gamma \mu + \epsilon / \sqrt{N},$$

where  $\epsilon$  is a multivariate Gaussian with mean zero and covariance matrix  $\Sigma$ . Let  $\mu(k)$  denote the mean in the  $k^{\text{th}}$  generation, and suppose that  $\mu$  differs from optimal by something of order  $1/N$ : if  $\nu(t) = N\mu(tN)$ , then the previous equation implies that as  $N \rightarrow \infty$ , in the limit  $\nu$  solves the Itô equation

$$d\nu(t) = \Gamma \nu(t) dt + \Sigma^{1/2} dW(t),$$

where now  $W(t)$  is a multivariate white noise. This has an explicit solution as a multivariate Ornstein-Uhlenbeck process:

$$\nu(t) = e^{-t\Gamma} \nu_0 + \int_0^t e^{-(t-s)\Gamma} \Sigma^{1/2} dW(s).$$

The asymptotic variance of this process in the direction  $z$  is

$$\lim_{t \rightarrow \infty} \text{Var}[\nu(t) \cdot z] = \int_0^\infty z^T e^{-s\Gamma} \Sigma e^{-s\Gamma} z ds, \quad (8)$$

692 which is infinite iff  $\Gamma z = 0$ , which occurs iff  $Lz = 0$ . In other words, population mean trait values lie away  
693 from the optimal set by a Gaussian displacement of order  $1/N$  with a covariance matrix given by equation  
694 (8).

695 *Now write what this means intuitively.*

**Completing the square** First note that if  $A$  is symmetric,

$$(x - y)^T A (x - y) = x^T A (x - 2y) + y^T A y,$$

and so if  $B$  is also symmetric and  $A + B$  is invertible,

$$\begin{aligned}(x - y)^T A (x - y) + x^T B x &= x^T (A + B) (x - 2(A + B)^{-1} A y) + y^T A y \\ &= (x - (A + B)^{-1} A y)^T (A + B) (x - (A + B)^{-1} A y) \\ &\quad + y^T A y - y^T A^T (A + B)^{-1} A y.\end{aligned}$$

Therefore, by substituting  $A = \Sigma^{-1}$  and  $B = L^T L$ ,

$$\frac{f(x; \Sigma, y) \exp(-x^T L^T L x / 2)}{\int f(z; \Sigma, y) \exp(-z^T L^T L z / 2) dz} = f(x; (\Sigma^{-1} + L^T L)^{-1}, (\Sigma^{-1} + L^T L)^{-1} \Sigma^{-1} y).$$

## Evolution of segregation covariance

The description above does not completely describe how two diverging populations interact, since the amount of *segregation variance*, quantified by  $R$ , will not stay constant. To get an idea of how this might change, suppose that a multivariate trait is determined by  $L$  unlinked, biallelic loci, and that the  $i^{\text{th}}$  locus has two alleles with additive effects  $\pm x_i$ , so that begin homozygous for the  $+$  allele contributes  $+2x_i$  to the trait. For simplicity, we will neglect the effects of selection. *fixup the below to be actually multivariate* If the  $+$  allele at locus  $i$  is at frequency  $p_i$  in a population, then the mean and genetic variance of the trait in a diploid population with random mating is *the mean should be  $\sum_i x_i(2p_i - 1)$*

$$m = 2 \sum_i p_i x_i$$

$$s^2 = 2 \sum_i p_i (1 - p_i) x_i^2.$$

Segregation variance between two parents depends on the loci at which either are heterozygous, and each locus contributes independently since alleles are additive. If the alleles are at Hardy-Weinberg proportions, then since segregation is a fair coin flip, a heterozygous locus contributes  $x_i^2/4$  to the variance, and so the *mean* segregation variance, averaging across parents, is

$$R_0(p) = \frac{1}{2} \sum_i p_i (1 - p_i) x_i^2.$$

On the other hand, if the second parent came from a distinct population with frequencies  $q_i$  (an  $F_1$  hybrid), this would be

$$R_1(p, q) = \frac{1}{4} \sum_i p_i^2 (1 - p_i)^2 x_i^2 + \frac{1}{4} \sum_i q_i^2 (1 - q_i)^2 x_i^2$$

$$= (R_0(p) + R_0(q))/2.$$

697 If we assume that the populations are at equilibrium,  $R_0(p) \approx R_0(q)$ , and so  $R_1(p, q) \approx R_0(p)$ .

Now consider an  $F_2$  hybrid, where both parents are  $F_1$  and so each heterozygous at locus  $i$  with probability  $p_i(1 - q_i) + (1 - p_i)q_i$ . Then

$$R_2(p, q) = \frac{1}{4} \sum_i (p_i(1 - q_i) + (1 - p_i)q_i) x_i^2.$$

Suppose that the two populations are slightly drifted from each other, with frequency difference  $p_i - q_i = 2\epsilon_i$ . Then,

$$p(1 - q) + p(1 - q) = (u + \epsilon)(1 - u + \epsilon) + (u - \epsilon)(1 - u - \epsilon)$$

$$= 2u(1 - u) + 2\epsilon^2.$$

If the frequencies have evolved neutrally in unconnected, Wright-Fisher populations of effective size  $N$  for  $t$  generations from a common ancestor with allele frequency  $u$ , then  $\epsilon$  has mean zero and variance roughly  $u(1 - u)t/N$ . Still assuming the populations are at stationarity, so that  $R_0$  is constant between the two, and taking the frequencies  $p_i$  as a proxy for the ancestral frequencies  $u_i$ , this implies that we expect

$$R_2 \approx R_0 + \frac{1}{2} \sum_i p_i (1 - p_i) x_i^2 t / N$$

$$= \left(1 + \frac{t}{N}\right) R_0.$$

On the other hand, the expected squared difference in trait *means* here is

$$4 \sum_i p_i (1 - p_i) x_i^2 t / N = 8R_0 t / N. \quad (9)$$

698 This implies that under this model, segregation variance in  $F_2$ s between two populations is roughly increased  
699 by a factor of 1/8 of the difference between their means.

## 700 D Away from the optimum

Let two points on  $\mathcal{N}$  be  $x_1$  and  $x_2$ , let  $\bar{x} = (x_1 + x_2)/2$ , and let  $z = (x_2 - x_1)/2$ . Then with  $\nabla D$  and  $\nabla^2 D$  the first and second derivatives of  $D$ , respectively, Taylor expanding about  $x_1$  and  $x_2$  finds that

$$\begin{aligned} D(\bar{x}) &= D(x_1) + \nabla D(x_1) \cdot z + \frac{1}{2} z^T \nabla^2 D(x_1) z + O(\|z\|^3) \\ &= D(x_2) - \nabla D(x_2) \cdot z + \frac{1}{2} z^T \nabla^2 D(x_2) z + O(\|z\|^3). \end{aligned}$$

Now, since  $D(x_1) = D(x_2) = \nabla D(x_1) = \nabla D(x_2) = 0$  and

$$\begin{aligned} \nabla D(x_2) &= \nabla D(x_1) + 2z^T \nabla^2 D(x_1) + O(\|z\|^2), \quad \text{and} \\ \nabla^2 D(x_2) &= \nabla^2 D(x_1) + O(\|z\|), \end{aligned}$$

adding together the two equations above and dividing by two gets that

$$D(\bar{x}) = \Phi_0 - \frac{3}{2} z^T \nabla^2 D(x_1) z + O(\|z\|^3).$$

## 701 E Gaussian load

Suppose that a population has a Gaussian distribution in  $d$ -dimensional trait space with mean  $\mu$  and covariance matrix  $\Sigma$ , and that fitness of an individual at  $x$  is  $\exp(-\|Lx\|^2/2)$ . Then, completing the square as above with  $A = \Sigma^{-1}$ ,  $y = \mu$ , and  $B = L^T L$ , and defining  $Q = (\Sigma^{-1} + L^T L)^{-1}$ ,

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}^n \det(\Sigma)^{1/2}} \int e^{-\frac{1}{2} x^T \Sigma^{-1} x} e^{-\frac{1}{2} x^T L^T L x} dx \\ &= \frac{1}{\sqrt{2\pi}^n \det(\Sigma)^{1/2}} \int e^{-\frac{1}{2} (x - Q \Sigma^{-1} \mu)^T Q^{-1} (x - Q \Sigma^{-1} \mu)} dx \\ & \quad \times e^{\mu^T (I - \Sigma^{-1} Q) \Sigma^{-1} \mu} \\ &= \sqrt{\frac{\det(Q)}{\det(\Sigma)}} \exp \{ \mu^T (I - \Sigma^{-1} Q) \Sigma^{-1} \mu \} \\ &= \sqrt{\frac{1}{\det(\Sigma) \det(\Sigma^{-1} + L^T L)}} \exp \{ \mu^T (I - (I + L^T L \Sigma)^{-1}) \Sigma^{-1} \mu \} \end{aligned}$$

Now suppose that  $\Sigma = \sigma^2 I$  and  $L = I/\gamma$ . Then,

$$\begin{aligned} \sqrt{\frac{1}{\det(\Sigma) \det(\Sigma^{-1} + L^T L)}} &= \sqrt{\frac{1}{\sigma^{2d} (1/\sigma^2 + 1/\gamma^2)^d}} \\ &= \frac{1}{(1 + (\sigma/\gamma)^2)^{d/2}}. \end{aligned}$$

Also,

$$\begin{aligned} (I - (I + L^T L \Sigma)^{-1}) \Sigma^{-1} &= \frac{1}{\sigma^2} (1 - (1 + (\sigma/\gamma)^2)^{-1}) I \\ &= \frac{1}{\gamma^2} \frac{1}{(1 + (\sigma/\gamma)^2)} I \end{aligned}$$

## F Differentiating the fitness function

*Add refs to sensitivity analysis.*

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  $h_0(t) = C_0 e^{A_0 t} B_0$  a representative of the optimal set:

$$\begin{aligned} D(A, B, C) &:= \int_0^\infty \rho(t) |h_A(t) - h_0(t)|^2 dt \\ &:= \int_0^\infty \rho(t) |C e^{A t} B - C_0 e^{A_0 t} B_0|^2 dt \\ &= \int_0^\infty \rho(t) \operatorname{tr} \left\{ (C e^{A t} B - C_0 e^{A_0 t} B_0)^T (C e^{A t} B - C_0 e^{A_0 t} B_0) \right\} dt \\ &= \int_0^\infty \rho(t) \operatorname{tr} \left\{ (C e^{A t} B - C_0 e^{A_0 t} B_0) (C e^{A t} B - C_0 e^{A_0 t} B_0)^T \right\} dt. \end{aligned} \tag{10}$$

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

$$\frac{d}{du} e^{(A+uZ)t} \Big|_{u=0} = \int_0^t e^{As} Z e^{A(t-s)} ds, \tag{11}$$

we have that

$$\begin{aligned} \frac{d}{du} D(A + uZ, B_0, C_0) \Big|_{u=0} &= 2 \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 \left( \int_0^t e^{As} Z e^{A(t-s)} ds \right) B_0 B_0^T (e^{At} - e^{A_0 t})^T C_0^T \right\} dt \\ &= 2 \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 \left( \int_0^t e^{As} Z e^{A(t-s)} ds \right) B_0 (h_A(t) - h_0(t))^T \right\} dt \end{aligned} \tag{12}$$

and, by differentiating this and supposing that  $A$  is on the optimal set, i.e.,  $h_A(t) = h_0(t)$ , (so wolog  $A = A_0$ ):

$$\begin{aligned} \mathcal{H}^{A,A}(Y, Z) &:= \frac{1}{2} \frac{d}{du} \frac{d}{dv} D(A_0 + uY + vZ, B_0, C_0) \Big|_{u=v=0} \\ &= \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 \left( \int_0^t e^{A_0 s} Y e^{A_0(t-s)} ds \right) B_0 B_0^T \left( \int_0^t e^{A_0 s} Z e^{A_0(t-s)} ds \right)^T C_0^T \right\} dt. \end{aligned} \tag{13}$$

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

$$H_{ij, k\ell}(A) := \mathcal{H}(\Delta_{ij}, \Delta_{k\ell}). \tag{14}$$

We could use this to compute the gradient of  $D$ , or 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 can diagonalize  $A = U \Lambda U^{-1}$ . Then

$$\int_0^t e^{As} Z e^{A(t-s)} ds = \int_0^t U e^{\Lambda s} U^{-1} Z U e^{\Lambda(t-s)} U^{-1} ds \tag{15}$$



Now, notice that

$$\int_0^t e^{s\lambda_i} e^{(t-s)\lambda_j} ds = \frac{e^{t\lambda_i} - e^{t\lambda_j}}{\lambda_i - \lambda_j}. \quad (16)$$

Therefore, defining

$$X_{ij}(t, Z) = (U^{-1} Z U)_{ij} \frac{e^{t\lambda_i} - e^{t\lambda_j}}{\lambda_i - \lambda_j} \quad (17)$$

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

$$\int_0^t e^{As} Z e^{A(t-s)} ds = U X(t, Z) U^{-1}. \quad (18)$$

Following on from above, we see that if  $Z = \Delta_{k\ell}$ , then

$$X_{ij}^{k\ell}(t) = \frac{e^{t\lambda_i} - e^{t\lambda_j}}{\lambda_i - \lambda_j} (U^{-1})_{\cdot k} U_{\ell \cdot}, \quad (19)$$

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

$$H_{ij,k\ell}(A) = \int_0^\infty \rho(t) \operatorname{tr} \{ C U X^{ij}(t) U^{-1} B B^T (U^{-1})^T X^{k\ell}(t)^T U^T C^T \} dt. \quad (20)$$

This implies that

$$D(A_0 + \epsilon Z) \approx \epsilon^2 \sum_{ijk\ell} H^{ij,k\ell} Z_{ij} Z_{k\ell} \quad (21)$$

and so

$$D(A_0 + \epsilon Z) \approx \epsilon^2 \sum_{ijk\ell} H^{ij,k\ell} Z_{ij} Z_{k\ell} \quad (22)$$

By section C, 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  $\sigma^2$  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  $D$ . First, in  $B$ :

$$\begin{aligned} \mathcal{H}^{B,B}(Y, Z) &:= \frac{1}{2} \frac{d}{du} \frac{d}{dv} D(A_0, B_0 + uY + vZ, C_0)|_{u=v=0} \\ &= \frac{1}{2} \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 e^{tA_0} \frac{d}{du} \frac{d}{dv} (uY + vZ)(uY + vZ)^T |_{u=v=0} e^{tA_0^T} C_0^T \right\} dt \\ &= \frac{1}{2} \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 e^{tA_0} (YZ^T + ZY^T) e^{tA_0^T} C_0^T \right\} dt. \end{aligned} \quad (23)$$

Next, in  $C$ :

$$\begin{aligned} \mathcal{H}^{B,B}(Y, Z) &:= \frac{1}{2} \frac{d}{du} \frac{d}{dv} D(A_0, B_0, C_0 + uY + vZ)|_{u=v=0} \\ &= \frac{1}{2} \int_0^\infty \rho(t) \operatorname{tr} \left\{ B_0 e^{tA_0^T} \frac{d}{du} \frac{d}{dv} (uY + vZ)^T (uY + vZ) |_{u=v=0} e^{tA_0} B_0 \right\} dt \\ &= \frac{1}{2} \int_0^\infty \rho(t) \operatorname{tr} \left\{ B_0 e^{tA_0^T} (YZ^T + ZY^T) e^{tA_0} B_0 \right\} dt. \end{aligned} \quad (24)$$

730 Now, the mixed derivatives in  $B$  and  $C$ :

$$\begin{aligned}\mathcal{H}^{B,C}(Y, Z) &:= \frac{1}{2} \frac{d}{du} \frac{d}{dv} D(A_0, B_0 + uY, C_0 + vZ)|_{u=v=0} \\ &= \int_0^\infty \rho(t) \operatorname{tr} \left\{ Y e^{tA_0^T} C_0^T Z e^{tA_0} B_0 \right\} dt.\end{aligned}\tag{25}$$

731 In  $A$  and  $B$

$$\begin{aligned}\mathcal{H}^{A,B}(Y, Z) &:= \frac{1}{2} \frac{d}{du} \frac{d}{dv} D(A_0 + uY, B_0 + vZ, C_0)|_{u=v=0} \\ &= \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 \left( \int_0^t e^{sA_0} Y e^{(t-s)A_0} ds \right) B_0 Z^T e^{tA_0} C_0 \right\} dt,\end{aligned}\tag{26}$$

732 and finally in  $A$  and  $C$ :

$$\begin{aligned}\mathcal{H}^{A,C}(Y, Z) &:= \frac{1}{2} \frac{d}{du} \frac{d}{dv} D(A_0 + uY, B_0, C_0 + vZ)|_{u=v=0} \\ &= \int_0^\infty \rho(t) \operatorname{tr} \left\{ C_0 \left( \int_0^t e^{sA_0} Y e^{(t-s)A_0} ds \right) B_0 B_0 e^{tA_0} Z \right\} dt.\end{aligned}\tag{27}$$