Rapid speciation despite conservation of phenotype

Joshua S. Schiffman[†] Peter L. Ralph[†‡]
[†]University of Southern California, Los Angeles, California [‡]University of Oregon, Eugene, Oregon
jsschiff@usc.edu plr@uoregon.edu

Abstract

We introduce an analytical theory to study the evolution of biological systems, such as gene regulatory networks. The evolutionary conservation of phenotype under selective and environmental stasis does not necessitate conservation of the underlying mechanism, as distinct molecular pathways can realize identical phenotypes. Here we give an exact expression for the set of all linear mechanisms with identical phenotypes, and expect evolution under neutrality to explore this set. We employ a quantitative genetic approach to model evolution under neutrality as a random process over the set of all phenotype-invariant mechanisms: only mutational tweaks to the pathway that leave the phenotype invariant are optimally fit. We show that there is never a unique linear system 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 and thus lead to the rapid formation of new species – in fewer generations than there are breeding individuals in a population.

Additional ideas to consider adding:

- hybrid vigor (need to do calculation)
- discuss linearity, linearization, and canalization in introduction

Note: in the LaTeXsource I'm putting in semantic linebreaks, so it's easy to edit and move around phrases and ideas.

Need to come up with a consistent term for "the A_{ij}"s. – "regulatory coefficients"? "genotype"?

Introduction

Bridging the gulf between an organism's genome and phenotype is a poorly understood and complex molecular machinery. Progress in a suite of biological subdisciplines is stalled by our general lack of understanding of this molecular machinery: with respect to both its function and evolution. . There does exist a growing body of data on the evolutionary histories and molecular characterizations of particular gene regulatory networks [**???**], as well as thoughtful verbal and conceptual models [**????**]. However, verbal theories are often insufficient, if not downright misleading [**?**]. This is especially pertinent given the staggering complexity and scope of contemporary research programs. This outlook necessitates the advancement of conceptual frameworks of such precision, only mathematics will suffice, as models allow the development of concrete numerical predictions.

Saying "stalled" reflects negatively on other fields - rephrase?

I don't think it reflects negatively. Even if it did, as long as it's a true statement it should be fine.

The molecular machinery, interacting with the environment, and bridging genotype to phenotype can be mathematically described as a dynamical system – or a system of differential equations [**?**]. Movement in this direction is ongoing, as researchers have begun to study the evolution of both abstract [**?????**] and empirically inspired computational and mathematical models of gene regulatory networks (GRNs) [**????????**]. If we allow the reasonable assumption that the genotype-phenotype map can be represented as a system of differential equations, we can immediately discuss its evolution and function in a much more mechanistic, yet general, manner.

I don't think the HW example will speak to our audience. Also, better to say why math is good rather than why not-math is bad.

deleted HW. I don't think we need to write an anodyne – the current sentence is true enough.

probably some more recent Wagner papers in this line as well?

In some fields that seek to fit parametric models to experimental data, such as control theory, chemical engineering, and statistics, it is well known that mathematical models can fundamentally be *unidentifiable* 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 [**???**]. Models with different parameter schemes, or even different mechanics can be equally accurate, but still not *actually* agree with what is being modelled. 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 [**???**]. In chemical engineering, those who study chemical reaction networks sometimes refer to the fundamental unidentifiability of these networks as "the fundamental dogma of chemical kinetics" [**?**]. In computer science, this is framed as the relationship among processes that simulate one another [**?**]. Although this may frustrate the occasional engineer or scientist, viewed from another angle, the concepts of unidentifiability and indistinguishability 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. In fact, evolutionary biologists who study convergent versus parallel evolution, homology, and analogy are very familiar with such functional symmetries; macroscopically identical phenotypes in even very closely related species can in fact be divergent at the molecular and sequence level [**???????**].

In this paper we outline a theoretical framework to study the evolution of biological systems. Presently, we focus solely on a neutral scenario, that is where phenotype is conserved over evolutionary time. however, this framework could be applied to a wider set of evolutionary scenarios.

We present an analytical description of the set of all linear biological systems with identical phenotypes – that is we describe the set of all gene network architectures that yield identical phenotypes, and show that all linear biological systems can, in principal, can undergo systems drift. In the neutral case, this set describes a manifold that evolution explores leaving phenotype invariant with respect to mutation, and predicts that if two populations become reproductively isolated, hybrid incompatibility can occur, despite the absence of adaptation, directional selection, or environmental change. Speciation typically occurs on timescales approximately on the order of N_e generations, where N_e is the effective population size.

Gene Networks as Linear Dynamical Systems

Organisms' phenotypes are constructred by gene by gene by environment interactions. Here we simply define the *phenotype* to be the temporal molecular dynamics directly under natural selection. The *what*, *when*, and *how much*, of an organism's molecules that are physiologically or otherwise relevant to survival.

Thus an organism's phenotype $\phi(t)$ – a vector of molecular concentrations at time t – is determined both by the structure and organization of a biological system (*e.g.* a gene regualtory network), given by the triple (A, B, C), and by its environment $u(t)$.

Such a biological system $\mathcal{S}$ is given by,

$$\mathcal{S} := \begin{cases} \dot{\kappa}(t) & = A\kappa(t) + Bu(t) \\ \phi(t) & = C\kappa(t) \end{cases} \tag{1}$$

Generally A is any real $n \times n$ matrix, B any $n \times \ell$, and C any $\ell \times n$ dimensional matrix. Although many different biological systems can be modeled with this approach, for clarity, we focus on gene regulatory networks. Each ith row of A describes the *cis*-regulatory module for gene i, and each jth entry, the specific regulatory influence of gene j on gene i. As such,

here you are jumping to evolution – haven't said yet why evolution "can explore" these – needs to say explicitly what we mean by neutral here

rather than "we expect" maybe say that the framework could be applied to non-neutral evolution

fixed

here's the definition of neutral, but it's not quite right?

how's that?

We need something here saying what we mean by "gene network architecture". Also, I don't think we claim anything for "all biological systems".

added the word "linear."

Below we mix general language with language specific to GRNs, like switching between "internal state" and "transcription factor concentration. Rather than keep making comments like "(or some other system)", we could say up top that this applies to other situations, but to make it concrete we'll talk about regulatory networks.

A_{ij} is the magnitude at which transcription factor $\kappa_j(t)$ regulates transcription factor $\kappa_i(t)$, and if $A_{ij} > 0$, we say that κ_j upregulates κ_i. If $A_{ij} < 0$, we say that κ_j down-regulates κ_i.

The form of B determines precisely how the environment influences the organism – that is B filters and translates the input to the system. C filters and translates the dynamics of the system and precesily determines the output, that is, what is visible to selection. *E.g.* for a metabolic system, C_{ij} is the amount the jth metabolite u_j affects the production of the ith enzyme.

Furthermore, whereas the phenotype is a subset of molecules *visible* to selection, the *kryptotype* includes the molecular dynamics *hidden* from selection. That is $\kappa(t)$ is a vector of the system's molecular concentrations at time t, directly, indrectly, and irrelevant to survival.

This "Thus" needs expanding, e.g., "the solution to this equation is unique and given by" with a reference.

Finally, we can write the phenotype as a convolution of the system organization and the environment,

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

where we refer to $h(t) := Ce^{At}B$ as the system's *impulse response*.

Example 1 (Oscillating Gene Network: Cell Cycle Control) *Cellular division is governed by many different processes, however it is thought that its rhythm is partially controlled by oscillating gene transcription [**?**]. Here we consider a simplified model of oscillating gene transcription.*

Suppose gene-2 up-regulates the transcription of gene-1 and that gene-1 down-regulates gene-2 with equal magnitudes, whose concentrations are given by $\kappa_1(t)$ and $\kappa_2(t)$. Furthermore, suppose that only the dynamics of gene-1 are consequential to the cell cycle (perhaps the amount of gene-1 activates another downstream gene network). Lastly suppose that the production of both genes is stimulated by an impulse of a molecule present immediately after division.

If the rate each of these genes is expressed is a linear function of their concentrations, the dynamics of the system are given by

$$\dot{\kappa_1}(t) = \kappa_2(t) + u(t)$$
$$\dot{\kappa_2}(t) = -\kappa_1(t) + u(t)$$

where $\dot{\kappa}$ denotes the time derivative. The initial conditions $\kappa_1(0)$ and $\kappa_2(0)$, and the input $u(t)$ then determine the concentrations through time. If we record the regulatory coefficients in the matrix

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

and define the column vector $B = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$, then in matrix notation the dynamics are

$$\dot{\kappa} = A\kappa(t) + Bu(t).$$

Since only the dynamics of gene-1 are directly relevant to biological function, the dynamics of interest are given by

$$\phi(t) = C\kappa(t)$$

where the row vector C is defined as $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$.

Since the input is simply an impulse, its phenotype is equivalent to its impulse response

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

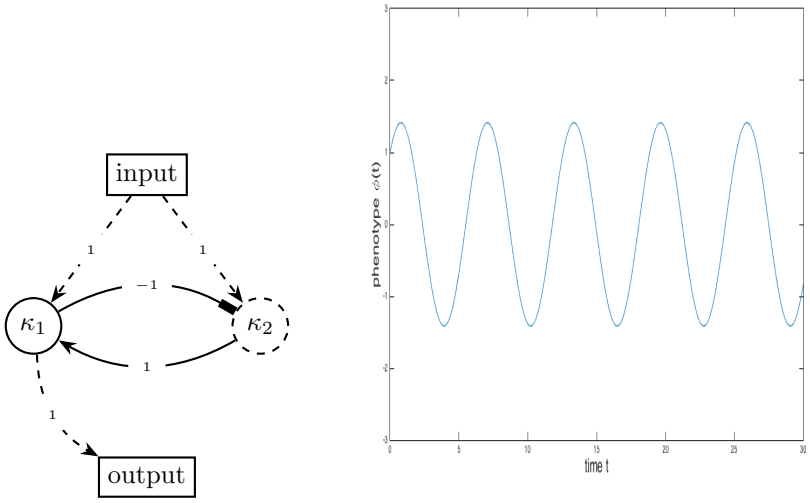

Figure 1: (Left) Graphical representation of the cell cycle control gene network, and (right) plot of the phenotype $\phi(t)$ against time t.

We return to the evolution of such a system below.

Phenotype-equivalent gene networks

Despite a symmetry in functionality or phenotype systems can often differ, sometimes substantially, at the molecular level. How many different mechanisms have the same function?

Gene regulatory networks with identical phenotypes do not necessarily have identical kryptotypes. Any linear and minimal system (a gene network) – minimal, informally meaning that the system's phenotype is achieved with the fewest possible genes – has an identical impulse response and therefore an identical phenotype given an identical input $u(t)$, up to a change of coordinates.

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

Two biological systems, $\mathcal{S} = \{A, B, C\}$, and $\bar{\mathcal{S}} = \{\bar{A} = VAV^{-1}, \bar{B} = VB, \bar{C} = CV^{-1}\}$, have the same phenotype, for all possible inputs, if they are related by a change of coordinates.

Therefore, if a system is minimal, the set of all phenotypically identical systems $\mathcal{S}(V)$ can be parameterized as,

$$\mathcal{S}(V) := \left\{ \begin{array}{ll} V\dot{\kappa}(t) & = VAV^{-1}\kappa(t) + VBu(t) \\ \phi(t) & = CV^{-1}\kappa(t) \end{array} \right. \tag{4}$$

where V is any invertible matrix and it's elements are free parameters.

Therefore varying V can tweak the relationships between genes within a regulatory network, yet preserve the phenotype. Some gene networks, however, can grown or shrink, perhaps following gene duplications and deletions, and also still preserve their phenotypes. These changes cannot be described by 4, but they are ecanpuslated by 5, below. More generally, we denote by $\mathcal{S}_n(\mathcal{S}_0)$ the set of all n-dimensional systems equivalent to $\mathcal{S}_0$:

$$\begin{aligned} \mathcal{S}_n(\mathcal{S}_0) &= \left\{(A, B, C) : Ce^{At}B = C_0e^{A_0t}B_0 \text{ for } t \geq 0\right\} \\ &= \left\{(A, B, C) : CA^rB \overline{\overline{4}} C_0A_0^rB_0 \text{ for } 1 \leq r \leq n-1\right\}. \end{aligned} \tag{5}$$

"has an identical phenotype" is not what you mean to say here

edited – does that work?

what are we calling these – not "biological system" – that's too broad

I see your point, but I don't want to bog down the paper by constantly qualifying statements – it seemsnatural to refer to biological systems, since these are, afterall dynamical systems.

is this the place to say "and only if, in the minimal dimension"?

Equivalence of the two characterizations follows from the Cayley-Hamilton theorem. Usually, the dimension n and the reference system A_0 is implicit and we write only $\mathcal{S}$.

To make sense of this we need to say how C and B change with n.

Two systems, even if in different dimensions, can have identical phenotypes if they are in $\mathcal{S}_n$. Further, there is no unique triple (A, B, C) for the mapping $u \mapsto \phi$. This implies that the gene regulatory network architecture per phenotype/environment pair is never unique – that is *all* gene regulatory networks can drift/rewire. This set can be precisley defined and completely parameterized using the *Kalman decomposition* A.

Example 2 (All Phenotypically Equivalent Cell Cycle Control Networks) *The set of all two-gene regulatory networks phenotypically equivalent to the cell cycle control network in 1, where only A can vary, is given by*

$$A(\tau) = \begin{bmatrix} 1 & 0 \\ \tau & 1-\tau \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{\tau}{\tau-1} & \frac{-1}{\tau-1} \end{bmatrix} \qquad \forall\ \tau \neq 1$$

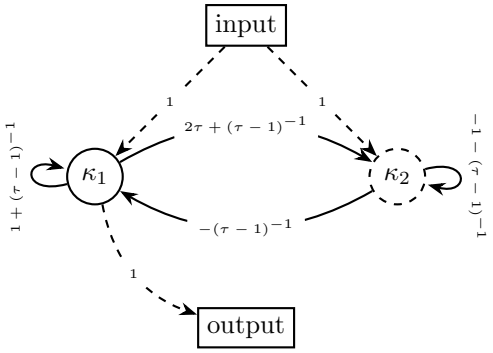

Figure 2: The set of all phenotype-equivalent cell cycle control networks, $A(\tau)$. $A(\tau) = \{\mathcal{S}_2 : B = B_0,\ C = C_0\}$.

Despite the phenotypic equivalence of all instantiations of $A(\tau)$, the kryptotypes, vary as a function of τ. Gene-1 dynamics (blue) are equivalent for network architectures $A(0)$ and $A(2)$, however the dynamics of gene-2 (orange) differ with τ.

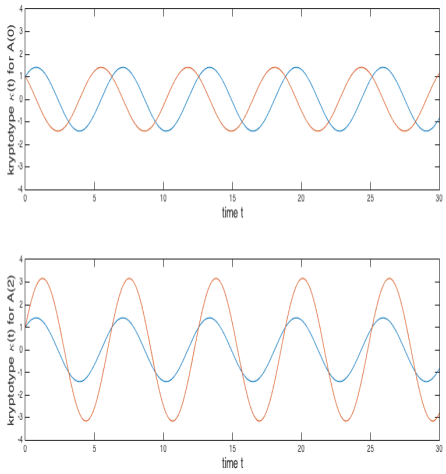

Figure 3: Gene-1 (blue) and gene-2 (orange) dynamics for $A(0)$ (top) and $A(2)$ (bottom). Both (top and bottom) gene-1 dynamics are given by $\kappa_1 = \sin(t) + \cos(t)$, and gene-2 by $\kappa_2 = \cos(t) - \sin(t)$ (top) and $\kappa_2' = \cos(t) + 3\sin(t)$ (bottom).

Sexual reproduction and speciation A diploid organism contains two gene network copies: one from each parent. As such, a diploid's system dynamics is computed by averaging the coefficients of both systems. To reproduce, each organism passes on a recombined haploid gene network to its offspring. Haploid gene networks swap system matrix rows randomly between its own two networks. Therefore the phenotypic dynamics typical of a first generation (F_1) cross between two allopatric populations will be determined simply by averaging it's two parental genomes. However, in second generation hybrid (F_2) crosses, first new haploid systems will be formed by recombination – in the process shuffling and combining regulatory coefficients from allopatric populations – and then brought together to form a diploid.

Example 3 (Hybrid Incompatibility in an Oscillating Gene Network) *Here we compare the phenotypes for F2 hybrids formed by crossing oscillators* $A(2)$ *with* $A(2.01)$, $A(2.1)$, *and* $A(2.5)$ *(B and C are the same as above). Each A is phenotypically identical (*$\phi(t) = \sin(t) + \cos(t)$*), however some of the hybrids exhibit markedly different dynamics.*

$$A(2) = \begin{bmatrix} 2 & -1 \\ 5 & -2 \end{bmatrix} \qquad A(2.01) = \begin{bmatrix} 2 - \frac{1}{101} & -1 + \frac{1}{101} \\ 5 + \frac{1}{99} & -2 + \frac{1}{101} \end{bmatrix}$$

$$A(2.1) = \begin{bmatrix} 2 - \frac{1}{11} & -1 + \frac{1}{11} \\ 5 + \frac{6}{55} & -2 + \frac{1}{11} \end{bmatrix} \qquad A(2.5) = \begin{bmatrix} 2 - \frac{1}{3} & -1 + \frac{1}{3} \\ 5 + \frac{2}{3} & -2 + \frac{1}{3} \end{bmatrix}$$

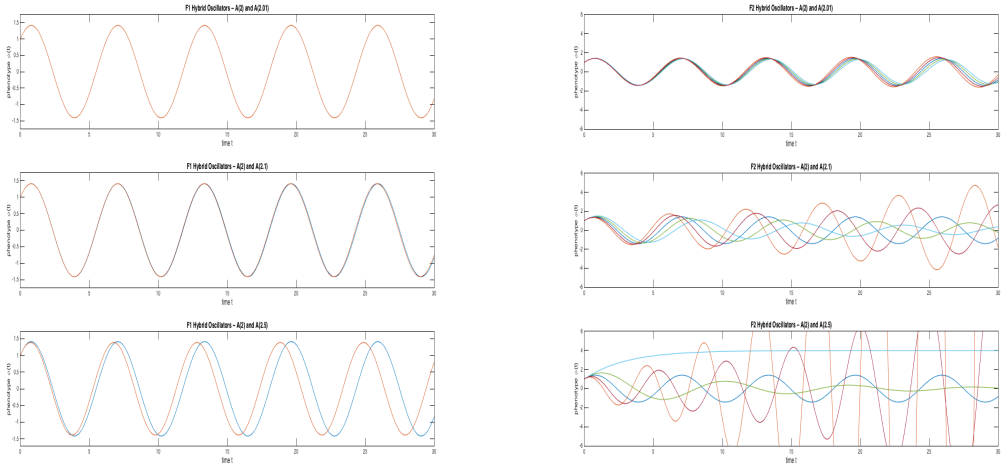

Figure 4: F_1 (left) and F_2 (right) hybrids crossing $A(2)$ with $A(2.01)$ (top), $A(2.1)$ (middle), and $A(2.5)$ (bottom). Note the difference in scale on the y-axis. F2 hybrids display more phenotypic divergence than F1s, on average. Further, some F2s completely fail to oscillate, as seen in an $A(2.5)$ F2 (light blue). *make axis labels bigger*

$A(2)$ and $A(2.1)$ differ in regulatory interaction strengths by 5.09%. If intra-population regulatory variation is approximately 5%, then this level of divergence is expected after $\sqrt{\frac{t}{N_e}} = 1$ generations.

Systems drift and the accumulation of incompatibilities

At any given time, there will be a range of regulatory coefficients present in the population due to segregating genetic variants. Over many generations, even if selective pressures do not change, this range of networks will shift as recombination, mutation, and demographic noise create new alleles and shift allele frequencies. How much variation do we expect to find within a population? Is this range limited by available variation or kept in check by selection? How fast will a population explore the space of equivalent networks? In this section we explore informally a general model for this situation, in which a population drifts stochastically near a set of equivalent, optimal systems. We work with a population of effective size N_e.

Suppose that a set x of coefficients that determine a system (this is A above), produce a phenotype $\Phi(x)$ (the time course of $\phi(t)$). There is an optimal phenotype Φ_0, and a set $\mathcal{X}$ of "optimal" coefficients that produce this phenotype. Fitness depends on distance to the optimal phenotype – we will write the "distance" between phenotypes ϕ and ψ as $d(\phi, \psi)$, measured so that the fitness of an organism with coefficents x is $\mathcal{F}(x) = \exp(-d(\Phi(x), \Phi_0)^2)$. We will assume that the map Φ is smooth and that the optimal set $\mathcal{X}$ is locally isomorphic to $\mathbb{R}^m$.

need to add γ^2 here?

say that better

Offspring. Individuals are diploid; we assume that each haploid genome determines a set of coefficients, and the individual's coefficients are the average of her two haploid values (no dominance). This implies that the diploid population variance, σ^2, is one-half the haploid variance. Each new gamete is produced from the parent's two haploid copies; for simplicity we assume that the gamete inherits a random choice of one of the two parental copies, and so

σ remains constant, up to a $1/N_e$ term. A more general model including segregation variance [**?**] would result in the same qualitative conclusions. *but put this in the appendix?*

System drift If the variation within a population of some coefficient has standard deviation σ, then 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 the standard deviation of the mean of a random sample [**?**]. Selection will tend to restrain this motion, but mean movement along the optimal set $\mathcal{X}$ is unconstrained. The amount of variance in particular directions in coefficient space depend on constraints imposed by selection and the covariance between genetic variation between different coefficients (the G matrix [**?**]). For instance, if the variation is due to *cis*-regulatory variants, the genetic basis of each *row* of A likely lies within a few kilobases of tightly linked sequence, across which a population may carry only a few common haplotypes. However, covariance due to transiently assembled haplotypes is not expected to be stable over long periods of time – a common *cis*-regulatory haplotype of transcription factor k with particularly strong binding to both i and j (leading to positive covariance between A_{ik} and A_{jk}) is no more likely to appear than one with strong binding to i but particularly weak binding to j (negative covariance). (Such transient covariances may well increase the variance of the per-generation change in network mean, however [**?**].)

To obtain a general quantitative picture, we need to know σ_N and σ_S, the standard deviations of coefficient variation along and perpendicular to $\mathcal{X}$ respectively, and γ, the scale on which phenotype changes moving away from $\mathcal{X}$. Concretely, γ is the inverse of the derivative of $d(\Phi(x+uz), \Phi(x))$ with respect to u for $x \in \mathcal{X}$ and z perpendicular to the tangent space at x. With these parameters, a typical individual will have a fitness of around $\exp(-(\sigma_S/\gamma)^2)$.

since it moves at rate $\sigma\sqrt{T/N_e}$ isn't saying $\sigma/\sqrt{N_e}$ *per generation* not that clear?

Hybridization By the arguments above, the means of two allopatric populations separated for T generations will be a distance of order $2\sigma_N\sqrt{T/N_e}$ apart along $\mathcal{X}$. 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, and variance equal to σ. Each F_2 hybrid will be homozygous for one parental allele on averge at half of the loci in the genome, so the distribution of F_2s will have mean at the average of the two populations, as before, but variance equal to $\sigma^2 + z^2/2$, where z is the distance between the parental populations. These are depicted in figure 5.

improve figure by putting labels on from the following Suppose that two populations have drifted independently to differ by z, and that z is of the same order as σ but is smaller than γ. The mean F_1 is the average of the parental means, and since the first-order terms in the Taylor series vanish, has phenotype differing from the optimum by a distance of order $\|z\|^2$ (see appendix C). The mean F_2 is the same, but the standard deviation is of order z, so that up to lower order terms, while the typical fitness of an individual in the original population is $\mathcal{F}_0 = \exp(-(\sigma_s/\gamma)^2)$; of an F_1 is $\mathcal{F}_1/\mathcal{F}_0 = \exp(-(c_1\sigma_N^2 T/N_e)^2)$; and of an F_2 is $\mathcal{F}_2/\mathcal{F}_0 = \exp(-T/(N_e\gamma^2))$.

Assume that selection acts directly against organisms whose regulatory coefficients are not on the optimum manifold. Further assume that selection constrains this diversity to be, on average, a distance of m from optimum. We would expect $\sigma_S = m$. Furtheremore, two individuals within a population may both be on the optimum manifold, but at different points, on average of σ_N. Since individuals a distance of σ_N apart will produce F_2s off of the manifold of $\mu = \frac{\sigma_N}{2}\sin 2\theta$, recombination load may constrain σ_N such that $\mu \leq m = \sigma_S$ (where θ is the slope of the manifold – as θ approaches 0 or 1, $\sigma_N \to \infty$.). We can estimate $\sigma_N = \frac{2\sigma_S}{\sin 2\theta}$.

I put F1s and F2s here, and $\mathcal{I}$, but with less explanation than below. Merge somehow.

In progress, see paragraph above.

I think you mean σ^2?

Can you clarify why $z^2/2$? Is this because it's averaging two distributions?

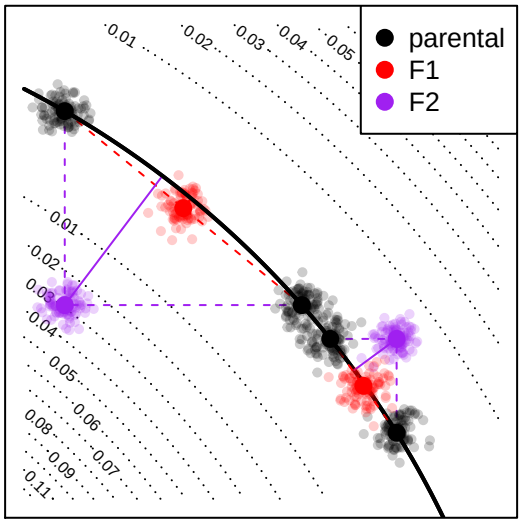

Figure 5: A conceptual figure of the fitness consequences of hybridization: axes represent system coefficients (i.e., entries of A); the line of optimal system coefficients is down in black; dotted lines give phenotypic distances to the optimum. Two pairs of parental populations are shown in black, along the optimum; a hypothetical population of F_1s 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. *Should show all types of F2? would be messy.*

Parameter estimates for real systems

To translate these results into real predictions, we need to know the strength of stabilizing selection on the phenotype, and the amount (and structure) of heritable variation in the genotype. These are known at best only roughly [**?**], so we aim for order-of-magnitude estimates.

We quantify (roughly) the amount of heritable variation by σ^2, the genetic variance present in a population in a typical entry of A. The coefficient A_{ij} 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 [**?**]. 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 A simultaneously. On the other hand, a diverse set of buffering mechanisms are thought to contribute to phenotypic stability in the presence of substantial molecular noise [**??**], suggesting that substantial variation in the micro-scale dynamics we consider here may be necessary to produce relevant phenotypic effects downstream.

replace "micro-scale" with sthg else or discuss earlier

The amount and structure of this standing variation is established over long time scales by many factors, including mutation-selection balance, shifts in the phenotypic optimum, and/or spatial variation in the optimum [**?**]. Quantitative genetics models of mutation-selection balance predict precise levels and structure of standing variation [**???**], but it is unclear how

well these predictions match reality [**?**] and how much they are expected to change over time [**?**]. However, empirical work allows us to estimate at least the rough magnitude of variation. Differences in A_{ij} due to a sequence change are hard to measure, but variation in both transcription factor binding site occupancy and expression levels (e.g., cis-eQTL) have been measured in various systems. However, 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 κ_i) rather than the *rate* of change. Nonetheless, **?** found differential occupancy in 7.5% of binding sites of a transcription factor (p65) between human individuals. **?** showed that cis-regulatory variation accounts for around 2–6% of expression variation in human blood-derived primary cells, while [**?**] found that human population variation explained about 3% of expression variation. Taken together, this suggests that variation in the entries of A may be on the order of 1% between individuals of a population – doubtless varying substantially between species and between genes.

Get some data from at least one other species in here!

It seems certain that selection in most species is not so strong that intra-population variation is strongly deleterious, so that if β is the typical scale on which selection acts, then $\beta > \sigma$. However, a range of studies have found evidence for weak stabilizing selection on regulatory SNPs and cis-eQTL. For instance, **?** found evidence that large-effect regulatory mutations are weakly selected against in Drosophila. This suggests that the strength of selection on phenotype is sufficient to weakly constrain regulatory variation, so that perhaps σ and β are relatively close. This is as would be expected if available variation is held in check by mutation–selection balance rather than genetic drift. A conservative estimate would be that $\beta = 5\sigma$; taking $\sigma = .01$ as above, this suggests that changes in phenotype of 5% are sufficient to effect a noticeable drop in fitness.

do we call this β or u?

β, u is input.

find them

(others?)

BUT σ IS VARIATION IN A NOT PHENOTYPE

We have guessed that within a population, the entries of A vary by around 1%, at least for networks whose function is strongly constrained.

The rate of speciation under neutrality in allopatric populations

separate out here (or above) the difference in genotype of hybrids from the optimum set, the difference in phenotype, and the difference in fitness.

F_2s between allopatric populations will have an average fitness relative to their parents of,

$$\exp\left(-\left(\frac{2\sigma_N}{\beta}\sqrt{\frac{T}{N_e}}\right)^2\right) \tag{6}$$

where β is the strength of selection against phenotypic divergence.

If the strength of selection is only a few multiples ξ_1 of the standing phenotypic diversity within a population σ_S, such that $\beta = \xi_1\sigma_S$, and recombination load constrains the standing kryptotype diversity σ_N to be only a few multiples ξ_2 of σ_S, then F_2 fitness will be,

$$\exp\left(-\left(\xi\sqrt{\frac{T}{N_e}}\right)^2\right) \tag{7}$$

where $\xi := 2\xi_2/\xi_1$. This suggests that if a population contains as much phenotypic diversity as allowed by selection and if σ_N is on the order of σ_S, reproductively isolated populations can speciate rapidly – on timescales shorter than N_e generations.

Discussion

The complexity of biological systems has limited our understanding of their function and evolution. Above we outline an approach, a first step, towards untangling this complexity

in reference to function and evolution. This methodology borrows successfully applied tools from engineering and aims to synthesize these with the concepts and tools of molecular and evolutionary biology.

Theoretical models in evolution and population genetics often lack the molecular details of physiology or of the genotype-phenotype map. Here, we offer a tractable and simple model which includes these missing features. Further, we provide, in clear mathematical language, an analytical description of phenomena hitherto discussed verbally and conceptually (phenogenetic drift [**?**], developmental systems drift [**?**], biological degeneracy [**?**], *etc.*). The tractability and relative simplicity of this exposition enables the interested biologist to work out by hand, if desired, the dynamics of a genetic system, as well as perturbations to the system – an attribute not likely to be found in less tractable models and simulations.

this is a claim we are the first to do something. best not to make these claims (and we aren't the first to make math models of these)

removed the word "only"

probably not desired

:p

We have suggested an interpretation of system identification: to see it as an evolutionarily neutral manifold, and not simply a computational nuisance. We have demonstrated a method to analytically determine the set of all phenotypically invariant gene networks; by a simple change of coordinates in the minimal configuration, or more generally by applying the Kalman decomposition in higher dimensions. Further, we emphasize that evolution proceeds through this high dimensional space as stochastic coordinate transformation, constrained by sexual reproduction and selection. This set is explored over evolutionary time when phenotype is conserved, and can lead to a diverse set of consequences, including the accumulation of Dobzhansky-Muller incompatibilities. We emphasize that these incompatibilities are a consequence of recombining different, yet functionally equivalent, mechanisms.

check original Bateson/DM papers to see if this accords with those defns

Furthermore, using a quantitative genetic approach, we estimated that a genetically variable population will drift in neutral system space at a rate determined by its intra-population variation and its effective population size. Because mechanistically distinct yet phenotypically equivalent biological systems can fail to produce viable hybrids, we predict allopatric populations to accumulate genetic incompatibilities at a rate on the order of N_e under reasonable population genetic parameter estimates. Additionally we see second-generation hybrid fitness plummet much faster than that of first-generation hybrids. This is a consequence of combining our mechanistic model with a quantitative genetic one: we observe that F_1 phenotypes diverge quartically, and F_2 phenotypes quadratically, with evolutionary time. This result is also consistent with Haldane's rule; that if only one hybrid sex is inviable or sterile it is likely the heterogametic sex. The consistency comes from gene networks localized to the sex chromosomes functioning as an F_2 hybrid cross within a diploid F_1 heterogamete as there is only one sex chromosome.

I read through Orr's review of those papers, which included excerpts. It seems like the DMI definition is very general and this accords.

refer to Turelli here

awkward phrasing

We also suggest that gene networks may not always use their components parsimoniously as network size tends to ratchet up in the absence of strong selection against extra parts. Although we leave this question unexplored, this phenomena may lead to insights on evolvability and developmental innovation. Lastly, we show that hybrid gene networks break down as function of genetic distance, and may, in part, explain broad patterns of reproductive isolation among diverse phyla [**?**].

probably shouldn't claim this is "unexplored"

I meant unexplored within this paper.Changed wording. Does that work?

As Richard Levins opined, models in population biology face a trade-off among precision, realism, and generality [**?**]. As Levins expects, any tractable and general model, such as the present one under discussion, will have limitations. Most notable is linearity. It is often stated that life is not linear. This is often true, however, many of the ideas developed here should be generalizable to nonlinear cases (multi-linear systems, say). Further, we see this as a necessary first step in the direction of more life-like nonlinear evolutionary systems theory. Depending on an actual biological system's particularities, its (potential) non-linearity, may buffer or exacerbate effects elucidated in this paper, such as the acquisition of Dobzansky-Muller incompatibilities.

this paragraph isn't very clear

Should I delete it? I guess it's sort of just a vague "yeah we know this method has shortcomings".

This theoretical framework can easily be applied to other interesting questions in evolutionary biology not tackled presently: such as the evolution of linkage, the necessity of network complexity (does evolution tend towards Rube Goldberg or parsimonious network

organization?), evolvability, structure/function inference, and intra-population context dependency of mutational effects, as well as many others.

Acknowledgements

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

Examples

Example 4 (Metabolic network) *Consider an organism that can metabolize two different sugars s_1 and s_2 (present at logarithmic concentrations u_1 and u_2 respectively), with enzymes e_1 and e_2 (with log concentration denoted as ϕ_1 and ϕ_2). Further suppose that one of the sugars s_2 is the preferred energy source (perhaps it contains significantly more energy than the other sugar or is otherwise more efficiently metabolized). The organism may have a gene regulatory network $\mathcal{S}$ that can deploy a situation specific metabolic strategy. That is depending on both u_2 and u_1 the organism will synthesize an appropriate ϕ_1 and ϕ_2. Furthermore, consider this system to contain at least two transcription factors, whose log concentrations are given by $\kappa_1, \kappa_2, \ldots \kappa_n$.*

Concentrations are "logarithmic", you mean that $u_1(t)$ = log(concentration of molcule 1 at time

I think you argue below that the optimal response is a complete switch from one enzyme to the other in the presence of the second sugar? Should say so here.

Minimally such a system may have the architecture,

$$\mathcal{S}_{\min} = \begin{cases} \dot{\kappa}(t) &= \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix}(t) + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}(t) \\ \phi(t) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix}(t) \end{cases}$$

$\mathcal{S}_{\min}$ is minimal as its reachability and observability matrices both have rank $= 2$, rank$(\mathcal{R})$ = rank$(\mathcal{O}) = 2 = n \implies$ min.

The impulse response matrix of this system is,

$$h(t) = \begin{bmatrix} 1 & -t \\ 0 & 1 \end{bmatrix}.$$

The phenotype is,

$$\phi(t) = Ce^{At}\kappa(0) + \int_0^t h(s-t)u(s)ds$$

($\kappa(0)$ can be set to something like $[-10, -10]$, assuming the transient transcription factor and enzyme concentrations in the organism are typically quite low).

We can see that changing coordinates on this system (with any invertible $V \in \mathbb{R}^{2\times 2}$) will find new system architectures, as shown before, however, we can also apply the Kalman decomposition to find systems in different dimensions – that is systems that employ more than 2 transcription factors, yet have the same external dynamics/phenotype. This may happen if a system is co-opted from another, or may be a consequence of gene duplication and deletion.

Ah – this is much simpler if we let B and C change as well!

$\widehat{\mathcal{S}}$ is 4-dimensional biological system,

where did $\widehat{\mathcal{S}}$ come from? need a lead-in here.

$$\widehat{\mathcal{S}} = \left\{ \widehat{A} = V \begin{bmatrix} X_1 & X_2 \\ 0 & A_{\min} \end{bmatrix} V^{-1}; \quad \widehat{B} = V \begin{bmatrix} X_3 \\ B_{\min} \end{bmatrix}; \quad \widehat{C} = \begin{bmatrix} 0 & C_{\min} \end{bmatrix} V^{-1} \right\}$$

where V is any invertible 4×4 matrix and X_j is any 2×2 matrix.

So for example, some system can be wired as follows, and still be input-output equivalent

the point here is not clear.

to the minimal metabolic system $\mathcal{S}_{\min}$*:*

$$A' = \begin{bmatrix} 1.6923 & 1.5385 & -2.6154 & -2 \\ 0.8462 & 0.7692 & -1.3077 & -1 \\ 1.2692 & 1.1538 & -1.9615 & -1.5 \\ 0.4231 & 0.3846 & -0.6538 & -0.5 \end{bmatrix}$$

$$B' = \begin{bmatrix} 4 & 4 \\ 2 & 2 \\ 3 & 3 \\ 1 & 3 \end{bmatrix}$$

$$C' = \begin{bmatrix} 0.2692 & 0.1538 & 0.0385 & -0.5 \\ -0.4231 & -0.3846 & 0.6538 & 0.5 \end{bmatrix}$$

Despite the present example consisting of a minimal 2×2 *and a non-minimal* 4×4 *system, any* n*-dimensional system can be constructed using this method – applying a change of coordinates to the Kalman decomposition – to construct a mechanistically different system with identical phenotypic dynamics. Depending on the specifics of a system being modeled, one may have to take care to restrict the free parameter values and network architectures to be biologically appropriate.*

what are you thinking of as being a potential problem here? say so explicitly.

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*

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

Equivalently, this occurs if and only if

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

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:

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

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

The set of all systems phenotypically equivalent to a given system (A, B, C) is elegantly described using the Kalman decomposition, which also clarifies the system dynamics? tells us a lot about how it works? To motivate this, first note that the input $u(t)$ only directly pushes the system in directions 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 the *reachable subspace*, which we denote by $\mathcal{R}$, and is defined to be the closure of span(B) under applying A (or equivalently, the span of $B, AB, A^2B, \ldots A^{n-1}B$). Analogously to this, we define the *observable subspace*, $\mathcal{O}$, to be the closure of span(C^T) under applying A. (Or: $\bar{\mathcal{O}}$ is the largest A-invariant subspace contained in the null space of C; and $\mathcal{R}$ is the largest A-invariant subspace contained in the image of B.)

or something

If we define

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

If we then define

$$P = \left[\begin{array}{c|c|c|c} P_{r\bar{o}} & P_{ro} & P_{\bar{r}o} & P_{\bar{r}\bar{o}} \end{array} \right],$$

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:*

$$\widehat{A} = PAP^{-1} = \left[\begin{array}{cccc} A_{r\bar{o}} & A_{r\bar{o},ro} & A_{r\bar{o},\bar{r}\bar{o}} & A_{r\bar{o},\bar{r}o} \\ 0 & A_{ro} & 0 & A_{ro,\bar{r}o} \\ 0 & 0 & A_{\bar{r}\bar{o}} & A_{\bar{r}\bar{o},\bar{r}o} \\ 0 & 0 & 0 & A_{\bar{r}o} \end{array} \right],$$

and

$$\widehat{B} = PB = \left[\begin{array}{c} B_{r\bar{o}} \\ B_{ro} \\ 0 \\ 0 \end{array} \right],$$

and

$$\widehat{C} = CP^{-1} = \left[\begin{array}{cccc} 0 & C_{ro} & C_{\bar{r}\bar{o}} & 0 \end{array} \right].$$

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 $\widehat{A}$, $\widehat{B}$, and $\widehat{C}$ (which are unconstrained), and a invertible transformation of 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 $\widehat{A}$, $\widehat{B}$, and $\widehat{C}$ except for $A_{ro} = A$, $B_{ro} = B$, and $C_{ro} = C$; and choosing an invertible matrix P.*

(b) *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 $(ro, \bar{r}\bar{o})$.

conjecture:

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.

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?

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

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

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

$$\begin{aligned} A &\to A + \epsilon P^{-1}\hat{A}Q - \epsilon P^{-1}QP^{-1}\hat{A}P \\ &= A + \epsilon\left(AP^{-1}Q - P^{-1}QA\right), \\ B &\to B - \epsilon P^{-1}QB \\ C &\to C + \epsilon CP^{-1}Q. \end{aligned}$$

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

$$\begin{aligned} A &= A + \epsilon\left(AW - WA\right), \\ B &\to B - \epsilon WB \\ C &\to C + \epsilon CW. \end{aligned}$$

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

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

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

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

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

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

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 $AP^{-1}Q = P^{-1}QA$ and $CP^{-1}Q = 0$ is equivalent to changing B by $Y = QB$.

B Genetic drift with a multivariate trait

For completeness, we provide a brief argument of how the population mean moves under genetic drift with a quantitative genetics model, as in **?** or **?**. These ignore details of the underlying genetic basis, but developing a more accurate model is beyond the scope of this paper.

Completing the square First note that

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

and so

$$\begin{aligned}(x-y)^T A (x-y) + x^T B x &= x^T (A+B) \left(x - 2(A+B)^{-1} A y\right) + y^T A y \\ &= \left(x - (A+B)^{-1} A y\right)^T (A+B) \left(x - (A+B)^{-1} A y\right) + (\text{terms that don't depend on } x).\end{aligned}$$

Therefore, if $f(x; \Sigma, y)$ is the density of a Gaussian with mean y and covariance matrix Σ then substituting $A = \Sigma^{-1}$ and $B = U^{-1}$ above,

$$\frac{f(x;\Sigma,y) f(x;U,0)}{\int_x f(z;\Sigma,y) f(z;U,0) dz} = f(x; (\Sigma^{-1} + U^{-1})^{-1}, (\Sigma^{-1} + U^{-1})^{-1} \Sigma^{-1} y).$$

Now suppose that the population is distributed in genotype space as a Gaussian with covariance matrix Σ and mean y. Selection has the effect of multiplying this density by the fitness function and renormalizing, so that if expected fitness of x is proportional to $f(x; U, z)$ then the above argument shows that the next generation will be sampled from a Gaussian distribution with covariance matrix $(\Sigma^{-1} + U^{-1})^{-1}$ and mean $z + (\Sigma^{-1} + U^{-1})^{-1} \Sigma^{-1} (y - z)$ Taking a sample of size N to construct the next generation will produce something close to this but with a slightly (stochastically) deviating mean. The next generation's mean is drawn from a Gaussian distribution with mean with covariance matrix $(\Sigma^{-1} + U^{-1})^{-1}/N$ and mean $z + (\Sigma^{-1} + U^{-1})^{-1} \Sigma^{-1} (y - z)$.

Roughly, what is this doing? Suppose that the population mean differs from the optimum by ϵ, that $\Sigma = \sigma^2 I$ and $U = I/\beta^2$ (so, stablizing selection happening on a distance scale of β). Then the population mean gets closer to the optimum on average, moving to $\epsilon/(1+\sigma^2\beta^2)$ and adds noise of size $(1/\beta)\sigma/\sqrt{N\sigma^2 + N1/\beta^2}$. At equilibrium, these two movements will be of the same order, so that ϵ is of order $(\sigma/\sqrt{N})\sqrt{1+\sigma^2\beta^2}$.

C Away from the optimum

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

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

Now, since $\Phi(x_1) = \Phi(x_2) = \Phi_0$ and

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

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

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

D Differentiating the fitness function

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 A_0 a representative of the optimal set:

$$\begin{aligned}D(A) &:= \int_0^\infty \rho(t)\left|h_A(t) - h_{A_0}(t)\right|^2 dt\\ &:= \int_0^\infty \rho(t)\left|Ce^{At}B - Ce^{A_0 t}B\right|^2 dt\\ &= \int_0^\infty \rho(t)\left|C\left(e^{At} - e^{A_0 t}\right)B\right|^2 dt\\ &= \int_0^\infty \rho(t)C\left(e^{At} - e^{A_0 t}\right)BB^T\left(e^{At} - e^{A_0 t}\right)^T C^T dt\end{aligned} \tag{8}$$

How does this change with A? Since

$$\frac{d}{du}e^{(A+uZ)t}|_{u=0} = \int_0^t e^{As}Ze^{A(t-s)}ds, \tag{9}$$

we have that

$$\begin{aligned}\frac{d}{du}D(A+uZ)|_{u=0} &= 2\int_0^\infty \rho(t)C\left(\int_0^t e^{As}Ze^{A(t-s)}ds\right)BB^T\left(e^{At} - e^{A_0 t}\right)^T C^T dt\\ &= 2\int_0^\infty \rho(t)C\left(\int_0^t e^{As}Ze^{A(t-s)}ds\right)B\left(h_A(t) - h_{A_0}(t)\right)^T dt\end{aligned} \tag{10}$$

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

$$\begin{aligned}\mathcal{H}(Y,Z) &:= \frac{1}{2}\frac{d}{du}\frac{d}{dv}D(A_0 + uY + vZ)|_{u=v=0}\\ &= \int_0^\infty \rho(t)C\left(\int_0^t e^{A_0 s}Ye^{A_0(t-s)}ds\right)BB^T\left(\int_0^t e^{A_0 s}Ze^{A_0(t-s)}ds\right)^T C^T dt.\end{aligned} \tag{11}$$

Here $\mathcal{H}$ is the quadratic form underlying the Hamiltonian. By defining Δ_{ij} to be the matrix with a 1 in the (i,j)th slot and 0 elsewhere, the coefficients of the quadratic form is

$$H_{ij,k\ell}(A) := \mathcal{H}(\Delta_{ij}, \Delta_{k\ell}). \tag{12}$$

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

$$\int_0^t e^{As} Z e^{A(t-s)} ds = \int_0^t U e^{\Lambda s} U^{-1} Z U e^{\Lambda(t-s)} U^{-1} ds \tag{13}$$

Now, notice that

$$\int_0^t e^{s\lambda_i} e^{(t-s)\lambda_j} ds = \frac{e^{t\lambda_i} - e^{t\lambda_j}}{\lambda_i - \lambda_j}. \tag{14}$$

Therefore, defining

$$X_{ij}(t, Z) = \left(U^{-1} Z U\right)_{ij} \frac{e^{t\lambda_i} - e^{t\lambda_j}}{\lambda_i - \lambda_j} \tag{15}$$

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

$$\int_0^t e^{As} Z e^{A(t-s)} ds = U X(t, Z) U^{-1}. \tag{16}$$

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

$$X_{ij}^{k\ell}(t) = \frac{e^{t\lambda_i} - e^{t\lambda_j}}{\lambda_i - \lambda_j} (U^{-1})_{\cdot k} U_{\ell \cdot}, \tag{17}$$

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

$$H_{ij,k\ell}(A) = \int_0^\infty \rho(t) C U X^{ij}(t) U^{-1} B B^T (U^{-1})^T X^{k\ell}(t)^T U^T C^T dt. \tag{18}$$

This implies that

$$D(A_0 + \epsilon Z) \approx \epsilon^2 \sum_{ijk\ell} H^{ij,k\ell} Z_{ij} Z_{k\ell} \tag{19}$$

and so

$$D(A_0 + \epsilon Z) \approx \epsilon^2 \sum_{ijk\ell} H^{ij,k\ell} Z_{ij} Z_{k\ell} \tag{20}$$

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

E Hybrid Vigor: Unit Circle approach

If a fitness optimum is at the origin $(0,0) \in \mathbb{R}^2$, and two populations have drifted away from the optimum by a distance of $r = 1$, then F_1s will have an average fitness of,

$$\begin{aligned} \mathbb{E}\left[d(F_1)\right] &= \int_0^{2\pi} \frac{1}{4\pi} \left\| (1,0) + (\cos\theta, \sin\theta) \right\| d\theta \\ &= \frac{2}{\pi} \end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[d(F_2)\right] &= \int_0^{2\pi}\int_0^{2\pi} \frac{1}{4\pi^2}\sum_i^2\sum_j^2\sum_k^2\sum_l^2 \frac{1}{32}\left\|(\cos(\theta_i)+\cos(\theta_k),\sin(\theta_j)+\sin(\theta_l))\right\| d\theta_1 d\theta_2 \\ &= 0.6705\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[d(F_{2_{\text{hap}}})\right] &= \sqrt{\int_0^{2\pi}\int_0^{2\pi} \frac{1}{8\pi^2}\cos(\theta_1)^2+\sin(\theta_2)^2 d\theta_1 d\theta_2} \\ &= \cos\left(\frac{\pi}{2}\right) = \frac{\sqrt{2}}{2}\end{aligned}$$

F Hybrid Vigor: Probability

$$\begin{aligned}X &\sim \mathcal{N}(x,\sigma^2) \\ Y &\sim \mathcal{N}(y,\sigma^2) \\ x &\sim \mathcal{N}(0,\eta^2) \\ y &\sim \mathcal{N}(0,\eta^2) \\ Z &= \frac{X+Y}{2} \to \mathbb{E}\left[|z|^2\right] \\ Z &= \frac{x+y+U+V}{2} \\ \mathbb{E}\left[|X|^2\right] &= \mathbb{E}\left[|x+U|^2\right] = \mathbb{E}\left[(x+U)^T(x+U)\right] \\ &= \mathbb{E}\left[x^Tx + U^Tx + x^TU + U^TT\right] \\ &= \mathbb{E}\left[x^Tx\right] + \mathbb{E}\left[U^TU\right] \\ &= d\eta^2 + d\sigma^2 \qquad d = \text{dimension} \\ X &\sim \mathcal{N}(0,\sigma^2+\eta^2) \\ Z &\sim \mathcal{N}\left(0,\frac{1}{4}(2\eta^2+2\sigma^2)\right) \\ &\to \mathbb{E}\left[|Z|^2\right] \\ &= \frac{d}{2}\left(\eta^2+\sigma^2\right) \\ &= \frac{1}{2}\mathbb{E}\left[|X|^2\right]\end{aligned}$$