



Rapid Speciation via Parallel, Directional Selection on Regulatory Genetic Pathways

NORMAN A. JOHNSON* AND ADAM H. PORTER*

*Department of Entomology and Graduate Program in Organismic & Evolutionary Biology,
University of Massachusetts, Amherst, MA 01003, U.S.A.*

(Received on 1 November 1999, Accepted in revised form on 3 April 2000)

Regulatory genetic pathways are ubiquitous in organisms and play a central role in the realization of the phenotype during development. We explored the proposition that these pathways can provide a plausible source of the epistatic variation that has been implicated in the evolution of postzygotic reproductive isolation. We modeled gene regulation as a matching function between the product of one locus and the promoter site of the next locus in the pathway, with binding strength determining the amount of product. When the phenotype is subject to parallel selection in a pair of independent populations, we find that the fitnesses of F_1 and F_2 hybrids often drop to very low values as the populations respond in genetically different and incompatible ways. The simulations support the predictions of the analytical models. Hybrid fitness reduction occurs more often as the number of loci in the pathway increases, and as the binding site interactions become more complex. Less hybrid fitness reduction is seen when the populations start with imperfect binding in the pathway. In contrast, when we constructed the phenotype without gene regulation using multiplicative rules, isomorphic to the additive phenotype commonly assumed in evolutionary models, we found no appreciable F_1 fitness reduction and only slight F_2 fitness reduction. The interaction of genetic drift and mutation, even at very high rates, did not reduce hybrid fitness at all on the time-scales we considered. Clearly, the evolution of regulatory genetic pathways can play an important role in speciation, but much more empirical information is needed on the effect of allelic variability in regulatory site interactions before this role is fully understood.

© 2000 Academic Press

1. Introduction

According to the widely accepted allopatric model of speciation, geographically separated populations accumulate genetic differences because evolutionary forces act independently on them, and members of the different populations eventually become unable to produce viable, fertile hybrids (Dobzhansky, 1937; Mayr, 1963;

Coyne, 1992; Wu & Palopoli, 1994; Gavrilets, 1997, 1999). One form of reproductive isolation, prezygotic isolation, occurs when fewer than expected hybrid zygotes are produced, and may be due to such factors as differences in mating song and other aspects of courtship behavior (Noor & Aquadro, 1998; Wells and Henry, 1998) or conspecific sperm precedence (e.g. Bella *et al.*, 1992; Robinson *et al.*, 1994; Howard *et al.*, 1998). Reproductive isolation can also occur post-zygotically, after the production of hybrid zygotes, via reduced hybrid viability (Wade &

* Authors to whom correspondence should be addressed.
E-mail: njohnson@ent.umass.edu, aporter@ent.umass.edu

Johnson, 1994; Carvajal *et al.*, 1996; Orr *et al.*, 1997; Gadau *et al.*, 1999) or fertility (Coyne, 1984; Orr *et al.*, 1987; Wu & Palopoli, 1994).

Post-zygotic differs from pre-zygotic reproductive isolation and many other traits in that it requires epistasis, non-additive genetic interactions at different loci, in nearly all cases (Dobzhansky, 1937; Muller, 1940; Wu & Palopoli, 1994; Orr, 1995; Turelli & Orr, 1995; Whitlock *et al.*, 1995; Gavrilets, 1997, 1999; Johnson, 2000). Empirical studies have further shown that the pattern of these gene interactions is often complex (Palopoli & Wu, 1994; Cabot *et al.*, 1994; Perez & Wu, 1995; Carvajal *et al.*, 1996; Naveira & Masdie, 1998; Johnson, 2000). Even the molecularly characterized *Odysseus* (*Ods*) gene, conferring male sterility in hybrids between *Drosophila simulans* and *D. mauritiana*, is by itself not sufficient for reproductive isolation. It requires the co-introgression of neighboring segments to cause complete sterility (Perez *et al.*, 1993; Perez & Wu, 1995; Ting *et al.*, 1998). However, despite the recent progress in determining the genetic architecture of hybrid fitness reduction in model systems, we know little about the evolutionary dynamics of the genetic changes that can result in hybrid dysfunction.

Although gene interaction has long been implicated in the evolution of post-zygotic isolation, the evolutionary implications of the role of gene regulation in the production of the phenotype has remained largely unexplored. Gene regulation itself is a central topic of developmental genetics research, with an emphasis on molecular characterization of how these genes interact in signaling pathways (e.g. Baker *et al.*, 1994; Morisato & Anderson, 1995; Dowell *et al.*, 1998; Lorenz & Heitman, 1998; Fostier *et al.*, 1998). Each interaction has unique properties and molecular complexity is the rule rather than the exception (Gerhart & Kirschner, 1997; Yamamoto *et al.*, 1998). Even so, several basic principles have emerged (Losick, 1998). First, gene products are nearly always regulated by transcription factors, which bind to promoter sites or other sites nearby or to other transcription factors. These factors may either enhance or reduce transcription. Second, all of these regulatory interactions depend critically upon the conformation and chemical composition of these

molecules as well as their concentrations (e.g. Phillips, 1994). Finally, regulatory pathways, especially those involved in development, are typically organized into linear and branching cascades (e.g. Morisato & Anderson, 1995) with later steps depending upon what occurred in earlier steps. The overriding principle arising from the study of gene regulation during development is that gene interaction is of essence: the phenotype is created not only from the structural properties of the individual genes but also in very large part by the interactions among them (Gibson, 1996; Raff, 1996; Gerhart & Kirschner, 1997).

We propose that regulated genetic pathways are a biologically realistic way to provide the complex epistatic gene interaction seen in empirical studies of hybrid fitness reduction. Here we investigate the plausibility of this proposition. We find that reproductive isolation, in the form of severely reduced hybrid fitness, evolves under a wide variety of conditions, in contrast to the case where the phenotype is constructed using standard rules of additive or multiplicative gene effects employed by evolutionary geneticists.

2. The Model

2.1. PHENOTYPIC EXPRESSION OF SIMPLE LINEAR REGULATORY PATHWAYS

The products of loci in regulatory pathways typically are transcription factors that bind to the promoter or other associated enhancer and repressor sites of the genes they regulate (e.g. Lewin, 1997). We therefore considered each locus in our simplified regulatory pathway to have both regulatory and product sites. We studied the simplest case of a linear pathway of n loci, labeled alphabetically, with the product of the A locus binding to a promoter site on the B locus, and so forth down the pathway (Fig. 1). The phenotype is assessed at the end of the pathway based upon the strength of binding at each step (see below).

2.2. MODELING ALLELIC STATES

The binding strengths of regulatory gene products depend upon a multitude of details involving the molecular "fit" between transcriptional factors and the regulated sites to which they bind.

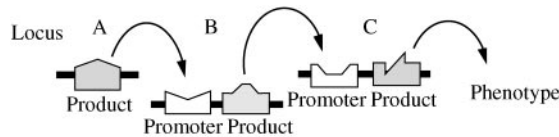


FIG. 1. A linear developmental pathway. Each locus contains a regulatory and a product site. Binding of the product of locus A to the regulatory element of locus B determines the amount of B's gene product that is produced, and so on until the amount of final product is obtained. The phenotype is determined by the amount of final product. Binding strength is assessed by the degree of matching between product and regulatory element (see text and Fig. 2).

These details include the conformation of the transcription factor protein, the presence and conformations of their repressor and enhancer proteins, other chromatin-associated proteins, and the DNA sequence of the binding site (e.g. Reinberg *et al.*, 1998). Rather than modeling this in all its complexity, we represent the allelic states of regulatory and product sites using decimal numbers, and the molecular "fit" as the extent to which these numbers match. Biologically, this is analogous to matching principal components of the factors determining binding efficacy. In the simplest case, the allelic states of regulatory and product sites are represented by a single decimal number each. We refer to this as the one-dimensional case. To assess the role of greater complexity in the binding site interactions, we also studied the n -dimensional case, using sets of n decimal numbers to represent the states of each of the product and regulatory sites. We allow mutations to change these allelic states at a given rate (u) to new values, with mutational effect sizes that follow a Gaussian distribution with zero mean and arbitrary variance. In all cases, our loci were unlinked and no recombination was permitted between the product and regulatory sites within alleles.

2.3. MODELING BINDING STRENGTH

We determine the strength of binding by the relative degree of match between the product of locus j [hereafter $product(j)$] and the regulatory element of locus $(j + 1)$ [hereafter $reg(j + 1)$]. The strength of binding between adjacent loci on the pathway, $\beta(j, j + 1)$, is assumed to follow a Gaussian distribution, scaled so that a perfect

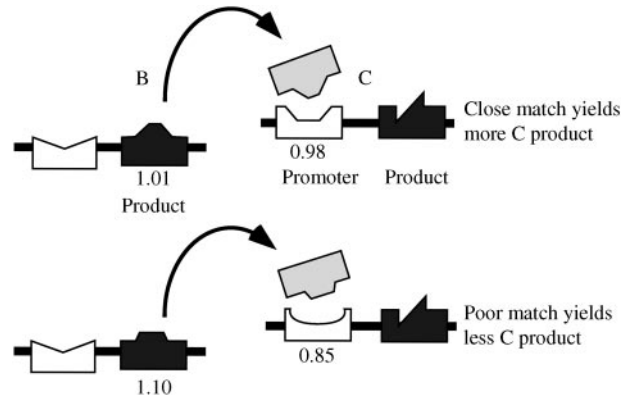


FIG. 2. An example of simple regulatory interactions in a developmental pathway, where the product of locus B binds to the regulatory element of locus C to determine the amount of C product synthesized. Product and regulatory sites of loci B and C were seeded with arbitrary numbers (shown) representing their allelic states and could mutate to new states. The strength of binding is inversely related to the degree of mismatch between these numerical values. In the top frame, the degree of mismatch is $x = 1.01 - 0.98 = 0.03$, a relatively low value. Binding strength in this case is $\exp(-0.03^2) = 99.9\%$, and this is taken as equal to the amount of C product synthesized. In the bottom frame, the mismatch is greater at $x = 1.10 - 0.85 = 0.25$. Binding here is weaker at $\exp(-0.25^2) = 93.9\%$, and the amount of C product is correspondingly less.

match produces $product(j + 1)$ constitutively. We represented this as

$$\beta(j, j + 1) = \exp \left\{ - [product(j) - reg(j + 1)]^2 \right\} \quad (1)$$

for the one-dimensional case, yielding a maximum at 1.0 with perfect matching. Figure 2 gives a worked example of the one-dimensional case for the step where locus j is B and locus $(j + 1)$ is C. In the two-dimensional case, $product(j)$ and $reg(j + 1)$ are each represented by a pair of decimal numbers, denoted as $product(j, 1)$, $product(j, 2)$ and $reg(j + 1, 1)$, $reg(j + 1, 2)$, respectively. Our general equation for binding strength for an arbitrary number of binding dimensions is

$$\beta(j, j + 1) = \exp \left\{ - \sum_c [product(j, c) - reg(j + 1, c)]^2 \right\}, \quad (2)$$

where c indexes the dimensions of each allelic state. In the diploid cases we studied, $product(j, c)$ and $reg(j + 1, c)$ were each calculated as the average of the allelic states of the two chromosomal copies before being introduced into the binding strength calculation. The binding strength is thus an average of the binding properties of the different allelic combinations for heterozygous individuals.

2.4. CALCULATING THE PHENOTYPE

Binding strength at each regulatory site determines the amount of gene product that the regulated locus produces and this is compounded along the regulatory pathway. We therefore determined the phenotype P as the arithmetic product of binding strengths along the pathway,

$$P = \prod_{j=1}^{n-1} \beta(j, j+1). \quad (3)$$

The phenotype we studied is thus entirely produced by gene regulation, and can be seen as the amount of gene product expressed by the final locus in the pathway. The allelic states of the promoter site of the first locus and the product site of the last locus are irrelevant when the phenotype is defined this way, and they were not used.

We also studied a phenotype without gene regulation, constructed from a simple multiplicative model. Here each gene copy was described by a single decimal number and the phenotype calculated as the product of these values. This multiplicative phenotype is biologically equivalent to an additive phenotype, used widely in evolutionary models, if the logarithms of the allelic values are added.

2.5. SELECTION ON THE PHENOTYPE

We studied the evolution of regulated pathways in changing and in constant environments. In changing environments, the optimal phenotype P_{opt} declined linearly (i.e. the populations experienced directional selection). In the constant environments, the optimal phenotype remained unchanged (i.e., they experienced stabilizing selection). In all our analyses, population pairs

experienced identical selection conditions; we did not study divergent selection.

2.6. FITNESS FUNCTION

Fitness follows a Gaussian distribution, centered on the optimal phenotype with arbitrary variance, scaled to a maximum of 1.0 at the optimum. Denoting the phenotype of an individual as P_i , the optimal phenotype as P_{opt} , and the standard deviation of the fitness function as Ω , the fitness of an individual is

$$W_i = \exp \left[- \left(\frac{P_i - P_{opt}}{\Omega} \right)^2 \right]. \quad (4)$$

The parameter Ω is a scaling constant, and is only relevant in comparison with the mutational effect size and rate of change of P_{opt} .

3. Model Analysis

The conditions above yield evolutionary dynamics that are difficult to model comprehensively. However, with a few simplifying assumptions, we are able to make predictions about the distributions of F_1 hybrid fitness outcomes for the simplest cases and generate qualitative predictions about how these distributions should be affected in more complex cases.

3.1. MODEL FOR TWO LOCI AND ONE-DIMENSIONAL BINDING

The simplest case we studied is that of two interacting loci (A and B) each described by single numbers at their product and promoter sites, so that binding strength is determined by matching numbers along a single dimension. Consider the situation where the initial allelic values of the product and the promoter exactly match so that the initial optimal phenotype is 1.0, and selection favors a monotonically decreasing phenotypic score that is a result of looser binding. There are two, mutually exclusive routes by which selection may produce looser binding:

Route 1. The allelic value of the A-locus product becomes less than that of the B-locus regulatory site [$product(A) < reg(B)$].

Route II. The allelic value of the A-locus product site becomes greater than that of the B-locus regulatory site [$product(A) > reg(B)$].

If one population responds by Route I and the other responds by Route II, then the F_1 hybrids will maintain a phenotype near 1.0 and will have progressively lower fitness as the optimal phenotype decreases. But if each population responds along the same route, the F_1 hybrids will have phenotypes similar to their parents and close to the optimal phenotype. Consequently, the F_1 hybrids will have high fitness. The probability of obtaining F_1 fitness reduction should be $P_1Q_2 + P_2Q_1$, where P_j is the probability that the population j will respond via Route I and Q_j is the probability that it will respond via Route II ($P_j + Q_j = 1$). If the probabilities of responding are the same for each of the populations, then $P_1 = P_2 = P$ and $Q_1 = Q_2 = Q$, and the probability of obtaining F_1 fitness reduction is $2PQ$.

If the parental populations both start with phenotypes equal to 1.0 (perfect matching), Routes I and II should be equally likely, $P_j = Q_j = 0.5$, whereupon $2PQ = 0.5$. Thus, we would expect strong F_1 hybrid fitness reduction in half of these trials and negligible F_1 hybrid fitness reduction in the other half.

The expected distribution of F_1 hybrid fitnesses for the two-locus, two-dimensional binding case is derived in Appendix A, and this is generalized to the multi-locus, multi-dimensional case in Appendix B. In general, the proportion of F_1 hybrids having very low fitness increases with the number of loci and the complexity of the binding-site interactions.

4. Simulation

4.1. CONDITIONS

Two diploid, ideal-hermaphroditic populations of equal size ($N = 250$ unless specified otherwise) were created with identical, arbitrary genotypes. Each generation, each individual was assessed for fitness and survived if its fitness exceeded a random decimal number between 0 and 1. This was followed by reproduction, with each remaining parent producing 2 ova fertilized by random

gametic associations and the resulting zygotes sent to a pool of offspring. These parents died and the population was refilled to capacity using randomly chosen zygotes, whereupon the life cycle repeated. Under these rules, a population risked extinction if its mean fitness dropped below 0.5 for an extended period of time. The two populations were allowed to evolve independently under identical conditions for a prescribed number of generations (usually $t = 2000$). In changing environments, the optimal phenotype declined steadily from $P_{opt} = 1.0$ to $P_{opt} = 0.5$ in 2000 generations, a rate of $1/4000$ per generation. In the fitness model, we used $\Omega = 0.05$ as the standard deviation of fitness [see eqn (4)] in our replicates. Under these conditions, populations were not at risk of extinction. Thus, the populations experienced directional selection that moved the optimal phenotype 10 standard deviations in fitness during 2000 generations. In constant environments, the populations were held under stabilizing selection at $P_{opt} = 1.0$ using the same Gaussian fitness function [eqn (4)] as in the changing-environment cases.

Mutation—We varied the mutation rate (u) according to the number of loci in the pathway (n) such that the genomic mutation rate was held approximately constant at 3×10^{-3} (see Table 1). Under normal conditions, the lowest mutation rate used was $u = 3 \times 10^{-4}$ (when $n = 10$ loci). We also examined runs using lower and higher mutation rates, with corresponding changes in the population size or the rate at which the optimal phenotype changed (see Selection, below). Mutational effect sizes followed the same Gaussian distribution for all loci, with a variance of 0.1 for most replicates.

TABLE 1
Variation of the genomic mutation rate μ in relation to the number of loci.

No. of loci	No. of interactions in pathway	μ per locus	μ per pathway
2	1	3×10^{-3}	3×10^{-3}
3	2	1.5×10^{-3}	3×10^{-3}
4	3	9×10^{-4}	2.7×10^{-3}
10	9	3×10^{-4}	2.7×10^{-3}

Initial variation in binding affinity—In natural populations, it is likely that binding is incomplete and some variation in allelic states exists at the outset. We studied the role of this variation by initializing the simulation with individuals heterozygous at all loci. The allelic states were randomly chosen from a Gaussian distribution with a mean equal to the optimal phenotype and variance equal to that used to generate new mutations. This settled into random allelic combinations within the first several generations.

Data collection—At specific intervals (usually at the end of 1000 and 2000 generations), 50 F_1 hybrids were sampled by randomly crossing the parents of the two lines (themselves called the P_1 and P_2 lines), and F_2 hybrids were also produced by crossing the F_1 's. All such crosses were performed without selection using copies of the parental lines, and the originals were left to continue to evolve independently of one another. We studied cases with $n = 2, 3, 4$ and 10 loci in the pathway, and calculated the mean phenotype and mean fitness of the parental and hybrid lines. At least 60 replicates were accumulated to characterize the frequency distributions for each set of parameter combinations studied.

Simulation code—The computer code for this study was built around a flexible, object-oriented (C++) library of classes and functions developed by AHP for simulating evolutionary and ecological problems and analysing such data. New code developed for this study included allelic classes for storing regulated loci, a genomic subclass holding the regulatory pathway and algorithms for binding strength and phenotypic calculation, the fitness functions, and a main function that implemented the simulation and its data collection steps. The simulation was developed using MetroWerks CodeWarrior® v4.0 on an Apple PowerMac G3® and run on a Dell Pentium II® in Microsoft Windows 98®.

5. Simulation Results

A representative replicate of the regulated pathway model where hybrid fitness decreases toward zero is shown in Fig. 3(a). The parental populations are able to track the changing environment (that is, the changing optimal phenotype) via the selection of new favorable

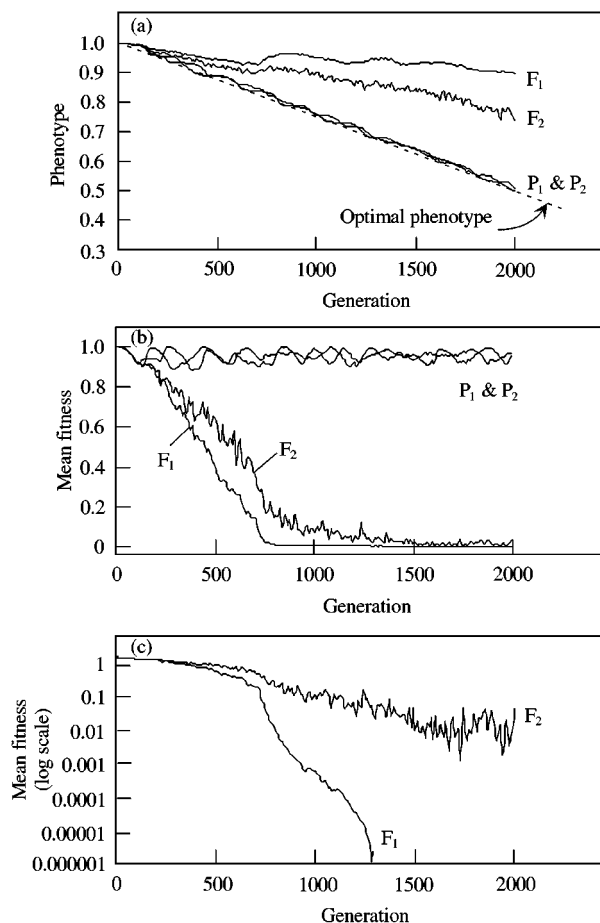


FIG. 3. A single representative replicate of the course of evolution of parental lines and their hybrids under directional selection for 2000 generations: (a) phenotypic change, (b) fitnesses of (a) on a linear scale, (c) hybrid fitnesses of (a) on a log scale. Fitnesses below 10^{-6} were beyond the numerical resolution of our simulations.

mutations, and their mean fitnesses remain high [Fig. 3(b)]. This selection response is qualitatively identical to that seen in multiplicative (additive) phenotypic models, even though the phenotype is entirely a product of gene interaction. In contrast to their parents, the phenotypes of F_1 hybrids do not change greatly from the initial phenotype [Fig. 3(a)]. This is because the genetic changes in the parental populations are incompatible with one another. Therefore, as the optimal phenotype decreases, the fitness of the F_1 hybrids drops toward zero [Figure 3(b)]. The mean fitness of the F_2 also declines, but not to the same extent as the F_1 because some of the F_2 genotypes reconstitute (at least in part) one or the other parental genotypes. Figure 3(c) shows

the mean F_1 and F_2 fitnesses of this replicate on a logarithmic scale.

5.1. ONE-DIMENSIONAL BINDING

In all two-locus cases with one-dimensional binding, F_1 hybrids either have high ($\bar{W}_{F_1} > 0.8$) or very low ($\bar{W}_{F_1} < 10^{-4}$) mean fitness. We find very low mean F_1 fitness in 50.3% (151/300) of our replicates, with 95% confidence limits of 44.7 and 56.0% (Table 2). This result fits very well with the theoretical expectation of 50%.

When we initialize the populations with variation in the promoter and product sites, the proportion of replicates with low hybrid fitness decreases (Table 2). This is probably because most new mutations that increase the difference between product and promoter values also perpetuate the rank order of their initial allelic scores, and fewer successful mutants are likely to arise that switch the rank order.

As the number of loci in the regulatory pathway increases, at least some degree of F_1 hybrid fitness reduction occurs more often (Fig. 4), as predicted by theory. When we consider only three classes of hybrid fitness ($\bar{W}_{F_1} \geq 0.1$, $0.001 \leq \bar{W}_{F_1} < 0.1$, and $\bar{W}_{F_1} < 0.001$), the data from the 3-locus, one-dimensional and the 2-locus, two-dimensional cases fit the theoretical expectation very well. However, if these classes are broken more finely, some significant second-

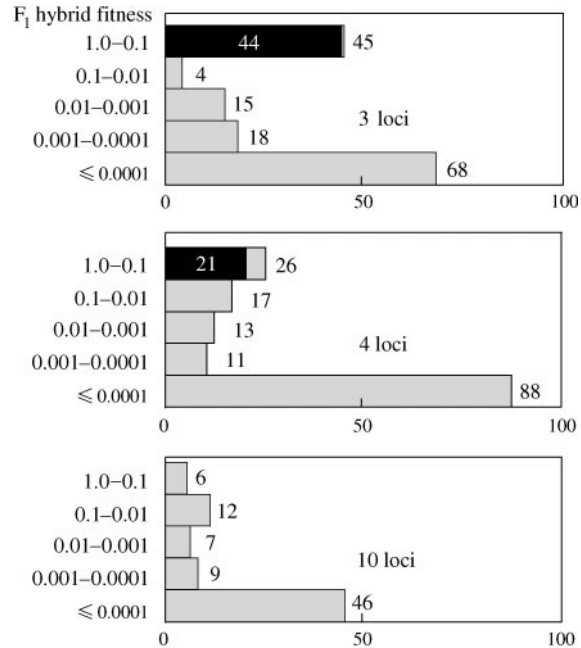


FIG. 4. Frequency distribution of F_1 hybrid fitnesses after 2000 generations for the cases with one-dimensional binding, with 3, 4 and 10 loci. Black bars represent fitnesses above 0.5. As the number of loci increases, so does the proportion of low-fitness F_1 hybrids.

order differences do emerge. We attribute these biases to minor violations, in the simulation, of the assumptions used to generate the expected distributions. Likely candidates are the assumptions that (i) the final parental phenotypes were precisely at P_{opt} , (ii) the distribution of θ was perfectly uniform, and (iii) the linkage has no effect on the evolutionary dynamics.

The frequency distributions of F_1 fitness reduction in the 4- vs. the 10-locus cases are not quite significantly different ($G = 10.56$, $df = 5$, $P = 0.06$; Fig. 4). We suspect that the 10-locus case is near the saturation point, and the addition of more loci to the pathway will generate only a negligible change in the probability of obtaining a reduction in F_1 hybrid fitness.

Provided that the mutation rate was sufficiently large such that the parental populations could adapt to the changing environment, further decreases in the mutation rate did not affect our results.

5.2. TWO-DIMENSIONAL BINDING

In contrast with the 2-locus case with one-dimensional binding, there are intermediate

TABLE 2

Frequency distributions of F_1 fitnesses for the two-locus, one-dimensional case, with and without initial allelic variability. With variability, the first generation was seeded with identical heterozygotes, and the variability shown is the standard deviation of randomly chosen allelic scores around a mean of 1.0. The expected high: low fitness ratio without initial variation is 1:1, and initial allelic variation in binding affinities reduces the proportion of outcomes showing reproductive isolation

Initial allelic variation (S.D.)	Fitness	
	High (> 0.8)	Low (< 0.001)
0 (none)	149	151
0.02	176	124
0.04	109	41

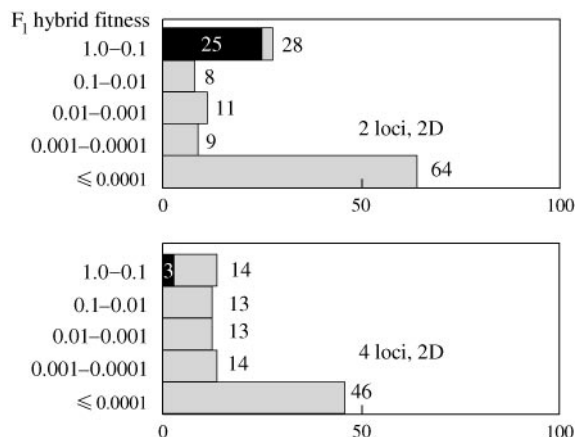


FIG. 5. Frequency distribution of F_1 hybrid fitnesses after 2000 generations for the cases with two-dimensional binding, with 2 and 4 loci. Black bars represent fitnesses above 0.5. As the number of loci increases, so does the proportion of low-fitness F_1 hybrids, and the increased dimensionality also reduces hybrid fitness (cf. Fig. 4).

levels of F_1 fitness reduction in the 2-locus case with two-dimensional binding (Fig. 5). The F_1 fitness distributions of the 2-locus, two-dimensional and the 3-locus, one-dimensional cases, which have the same expected distributions, are not statistically different from one another ($G = 4.08$, $df = 5$, $P = 0.54$). Since the 2-locus, two-dimensional case has no recombination and the 3-locus, one-dimensional case has free recombination ($r = 0.5$), their close similarity suggests that recombination does not play a major role in the evolution of low hybrid fitness in our model. In the 4-locus, two-dimensional binding case (Fig. 5), at least some F_1 fitness reduction occurs in nearly every replicate.

Initial variability—When the initial match between the promoter and product is not exact, such that the optimal phenotype is 0.99, F_1 hybrid fitness reduction still occurs but at a reduced frequency (Fig. 6). Even so, in the 4-locus case, 10% of the replicates have F_1 fitnesses below 0.01.

5.3. OTHER CONDITIONS

Mutation and population size—We find that changing the mutation rate has little effect on the likelihood of hybrid fitness reduction, as long as sufficient mutants are available to respond to selection. The ability of the parental populations to track the changing phenotypic optimum is constrained by the availability of new muta-

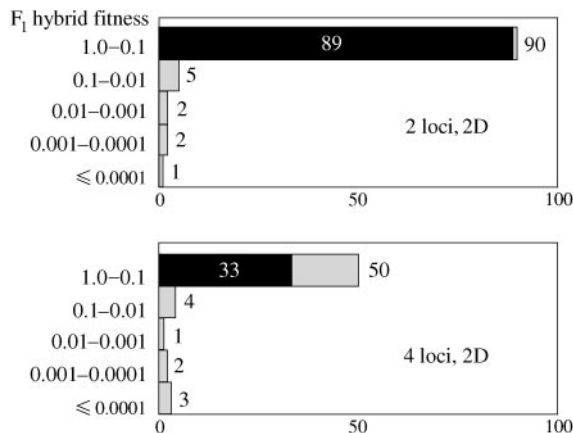


FIG. 6. Frequency distribution of 2- and 4-locus, two-dimensional cases with allelic variability and imperfect binding in the initial population. Hybrid fitness reduction is considerably less frequent (cf. Fig. 5) but still occurs in a significant proportion of the replicates.

tions in the populations, rather than by mutation rate *per se*. Thus, we achieved hybrid fitness reduction over very short time-scales when populations were very small ($N = 20$) provided that mutation rates were very high ($u = 10^{-2}$). We also found severe hybrid fitness reduction with mutation rates as low as $u = 2 \times 10^{-5}$, although with far fewer replicates. In these cases, we changed optimal phenotype more slowly but achieved the same outcomes over a longer time period (2×10^4 generations). This could have been done as well by increasing the population sizes above 250, but this exceeded the memory limitations of our computer.

Multiplicative model—When the phenotype is determined by multiplying allelic values at each of the structural loci, populations respond to selection as observed in other studies (Fig. 7) and as seen in the regulatory model. In contrast to the results from that model, we have not observed appreciable mean F_1 hybrid fitness reduction in the multiplicative model cases. We do find some mean F_2 hybrid fitness reduction (Fig. 7), apparently due to less compatible combinations of alleles at different loci. This is expected from recombination among alleles from pathways with different evolutionary histories, and some additional epistasis may also be generated statistically from stabilizing effects generated in the tails of the Gaussian fitness function (cf. Whitlock *et al.*, 1995).

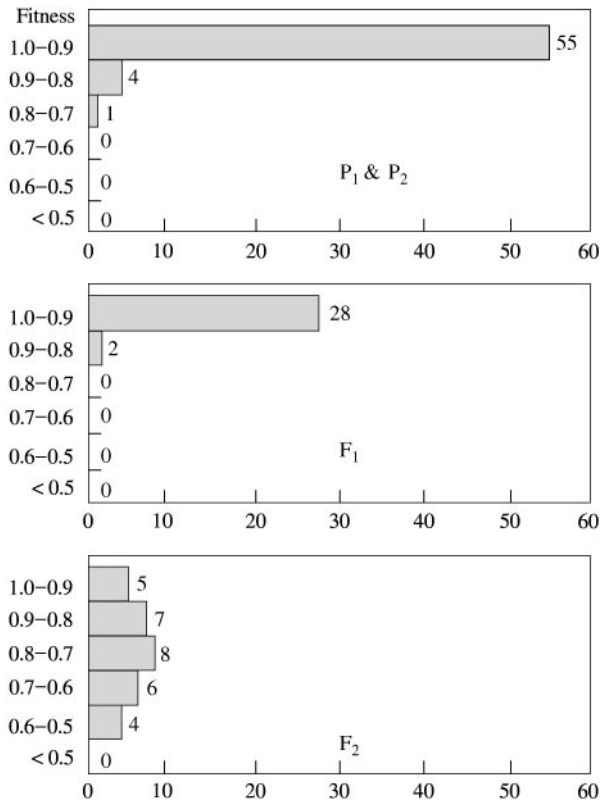


FIG. 7. Frequency distribution of parental and hybrid fitnesses when the phenotype was constructed using a multiplicative model, without gene regulation. These replicates are pooled from 4- and 10-locus cases.

Stabilizing selection—We held population size very low ($N = 25$) and mutation rate very high ($u = 0.01$). This maximized the potential for the process of mutation/genetic drift/compensatory mutation to produce divergence in the two populations (cf. Nei *et al.*, 1983). We ran 100 replicates for 2000 generations and 40 replicates for 8000 generations. We did not observe any F_1 or F_2 hybrid fitness decline in the stabilizing selection cases. Thus, mutation/drift interactions alone were insufficient to generate reproductive isolation on the time-scales we considered, and are even less likely with realistic mutation rates and population sizes.

6. Discussion

6.1. REPRODUCTIVE ISOLATION AND REGULATED GENETIC PATHWAYS

Severe F_1 hybrid fitness reduction occurred frequently when genes in a regulatory pathway

were subjected to persistent directional selection, even when populations experience identical selection regimes. This reproductive isolation evolved under a wide range of conditions because the parental populations responded to selection in genetically different and incompatible ways. In our simulations, reproductive isolation occurred rapidly but not so rapidly that it could be detected in laboratory experiments run over tens of generations (cf., Rice & Hostert, 1993). As we made the regulatory pathways more complex, either by increasing the number of loci or increasing the number of factors influencing the binding-site interactions, reproductive isolation occurred more often.

Divergent selection in geographically isolated populations is expected to yield reproductive isolation (Dobzhansky, 1937; Mayr, 1963; Coyne, 1992; Schluter, 1998), and should do so regardless of the underlying mechanisms relating genotype to phenotype. Here, we show that divergent selection is not a necessary condition; reproductive isolation can readily evolve in geographically isolated populations given gene interaction and directional selection.

In principle, directional selection may not be required for reproductive isolation. Reproductive isolation can result from the sequence of mutation, drift, and compensatory mutation (Nei *et al.*, 1983). The dependence of this process on drift and new mutations means that reproductive isolation will evolve very slowly and is unlikely to yield speciation over the time-scales we consider (Phillips, 1996). Our simulations under stabilizing selection gave no F_1 or F_2 hybrid fitness reduction in 8000 generations, and are consistent with this explanation.

We found slight F_2 fitness reduction but no F_1 hybrid fitness reduction (Fig. 7) in simulations where the phenotype was determined by multiplicative rules (additivity on a log scale) and evolved under parallel directional selection. This is in sharp contrast to the results generated when the phenotype was constructed using regulated genetic pathways (Figs 3–6). Additive genetic architectures are much less likely to produce the rich complexity of gene interactions that make incompatibilities likely.

Although the regulatory complexity inherent in developmental systems has become increasingly

well known, it is largely neglected in evolutionary genetic theory (but see Gibson, 1996; Frank, 1999). This body of theory was developed long before developmental biologists had demonstrated the ubiquity of gene regulation in the production of the phenotype, and the theory concentrated instead on the evolutionary consequences of Mendelian inheritance (Provine, 1971; Smocovitis, 1996). The mathematics of complex interactions is formidable (Lewontin, 1974), and the advent of powerful computers has only recently allowed for sufficiently comprehensive investigations of the evolutionary dynamics of multigene systems (e.g. Nowak *et al.*, 1997; Phillips & Johnson, 1998). In the meantime, evolutionary geneticists have been successful using statistical, phenomenological descriptions of the phenotypic consequences of epistatic interactions, while ignoring the mechanics. From this perspective, epistasis is not typically incorporated as the baseline condition of phenotypic expression, but is instead described as a residue remaining after statistically additive components have been taken into account (Wade, 1992a,b; Wade & Goodnight, 1998; Phillips, 1998). There are clear scientific benefits to this black-box approach but it carries costs, too: evolutionary theory that omits the mechanisms by which the phenotype is realized is not equipped to study how these mechanisms influence evolutionary outcomes. Our results provide a telling example: reproductive isolation is difficult to achieve in evolutionary models where the phenotype is determined by additive rules (see also C. J. Goodnight, unpublished results), but when the phenotype is determined by regulatory genetic mechanisms, we find that reproductive isolation occurs much more often.

6.2. THE EFFECT OF SIMPLIFYING ASSUMPTIONS

Our study relies on the Gaussian distribution in the characterization of binding strength and fitness, and it is natural to wonder how this affects our outcomes. We believe that our major results will be robust to the use of other unimodal distributions, provided that the optimal phenotype evolves to be sufficiently different from that favored by the original fitness distribution. This

can be seen graphically by making appropriate adjustments to Fig. A2. Imagine that binding strength follows an arbitrary unimodal distribution as the product–promoter match deviates from that giving maximum binding. The regulated F_1 hybrid phenotype, being the product of binding interactions, will show a concave shape qualitatively similar to that in Fig. A2(a) as θ deviates between the two populations. This will also translate to a unimodal fitness curve qualitatively similar to that in Fig. A2(b), even if fitness itself is not normally distributed around the optimal phenotype. A unimodal qualitative fitness curve will emerge as long as the populations have evolved long enough that the original parental populations would have low fitness. We thus expect that changes in the assumptions of normality in our model will have only second-order effects on the outcome of reproductive isolation.

A more fundamental assumption we employ is that different binding strength parameters have effects of equal magnitude on the phenotype. This is unlikely to be the case. Instead, some parameters will have much greater effects on the phenotype and our model will have to be scaled to remain relevant. Thus, instead of occupying a circle (Fig. A1), sphere or hypersphere in binding-parameter dimensions, the evolutionary divergences would fall along an ellipse, spheroid or hyperspheroid of arbitrary flatness. In these cases, equivalent divergence in genotype (e.g. base pair substitutions) could have profound effects on reproductive isolation for some binding parameters and almost none for others. Presently, we know next to nothing about the role of allelic variability in regulatory genetic pathways, particularly those involved in development. Especially, the role of allelic variability in binding strength interactions needs to be explored if we are to come to an empirically based understanding of how the phenotype evolves.

6.3. EVOLUTIONARY DYNAMICS

To a first order of approximation, we have been able to predict the evolutionary outcomes for the simplest cases where only a small number of factors interact to generate the phenotype. It will be necessary to account for several aspects of

the evolutionary dynamics in order to refine these predictions. In particular, we found in our simulations that population survival depended upon a subtle interplay among rates of mutation (u) and selection (i.e. the rate of habitat change, ΔP_{opt}), mutation effect size and population size. However, none of these parameters had very important effects on the reduction of hybrid fitness, provided there were sufficient mutations to allow populations to track the changing optimal phenotype. We did observe (not shown) that larger populations and slower rates of habitat change tended to provide a smoother response to selection, and a correspondingly smoother transition to low F_1 fitness when it did evolve. The time-scale of hybrid fitness reduction was also affected by these parameters. We achieved qualitatively similar results over very short time-scales when populations were very small ($N = 20$) provided that mutation rates were very high ($u = 10^{-2}$). We have also performed simulations wherein the optimal phenotype changes at a slower rate (as low as $\Delta P_{opt} = 1/20\,000$ per generation) over a longer period of time (up to 10 000 generations). Here, the parental populations ($N = 250$) do not require such high mutation rates to adapt and thus we could achieve similar results using smaller, more realistic mutation rates ($u = 2 \times 10^{-5}$). We therefore expect our results to hold at lower mutation rates, provided the populations are large enough to generate the new mutations necessary to avoid extinction in the face of habitat change.

6.4. GENERAL CONSIDERATIONS

If reproductive isolation can evolve so easily, as suggested by our results, why do we not see reproductive isolation more often among geographically separated populations? We believe that one of the major limiting factors is that directional selection must gradually change the optimal phenotype over long periods of time. The optimal phenotype must change several standard deviations of fitness (Ω) to produce measurable reproductive isolation with high frequency. In addition, if the rate at which the optimal phenotype changes is too high, the populations run the risk of extinction. Species with high rates of reproduction (like most insects) would be better

able to track environmental change than would species with low rates of reproduction (like mammals). Given the ubiquity of regulatory genetic pathways in biological systems, we doubt that developmentally additive (or multiplicative) phenotypes are a plausible explanation for any apparent paucity of reproductive isolation.

Another potentially limiting factor is that we most often obtained reproductive isolation when we began with very tight binding at the regulatory sites, and had the optimal phenotype change in a way that favored looser binding. First, when the populations began at intermediate binding states, we found reproductive isolation to be markedly less frequent. This follows from our graphical model. Consider that when the populations begin with perfect binding, they occupy the origin $\{0, 0\}$ in Fig. A.1, and proceed to occupy points on a widening circle as evolution proceeds. The distribution of angles θ between the populations is roughly uniform. When initial binding is imperfect, the populations begin at an arbitrary point $\{x, y\}$ and again occupy points on the widening circle as evolution proceeds, but the distribution of points is no longer uniform. Instead, the θ distribution would be centered at 0 (for identical binding properties) and would have a variance that depended on the evolutionary dynamics during the adaptation process. Less reproductive isolation would ensue because there is no reproductive isolation when $\theta = 0$. Even so, we expect that with intermediate binding, low F_1 fitness outcomes would become more frequent as the number of binding parameters increased, as it would be increasingly likely that at least one of the θ_j values would show a strong deviation from 0 by chance alone. Thus, as the number of loci or the complexity of the individual binding interactions increased, this difficulty would be mitigated.

A third constraint on the generality of our result is that in many cases, selection may well be in the direction of tighter binding. Of course, in natural regulatory systems, some steps in the pathway are likely to be inhibitory, whereupon selection in any direction on the phenotype will result in looser binding somewhere in the pathway. More complex regulatory pathways, incorporating inhibitory steps, need further exploration.

We do not propose that selection operating on regulatory genetic pathways is the only means by which postzygotic reproductive isolation can arise. We argue, instead, that selection operating on interacting genes in general will result in genetic incompatibilities in hybrids (cf. C. J. Goodnight, unpublished results). Divergence in regulated genetic pathways, themselves implicated in a great diversity of the developmental and physiological processes by which the phenotype is realized, seems a plausible way to achieve reproductive isolation.

We thank M. Baker, C. Goodnight, E. Grey, E. Jakob, M. Noor, M. Palopoli, P. Phillips, R. Phillis, M. Sei, J. Townsend, M. Wade, B. Wang and an anonymous reviewer for comments and discussion. C. Goodnight graciously provided a pre-publication copy of his manuscript. D. Ferro generously provided computer resources. This material is based upon work supported in part by the U. S. Department of Agriculture, Cooperative State Research Extension and Education Service, awarded through the Massachusetts Agricultural Experiment Station under Project No. MAS00789, Paper No. 3264.

REFERENCES

- BAKER, B. S., GORMAN, M. & MARIN, I. (1994). Dosage compensation in *Drosophila*. *Ann. Rev. Gen.* **28**, 491–522.
- BELLA, J. L., BUTLIN, R. K., FERRIS, C. & HEWITT, G. M. (1992). Asymmetrical homogamy and unequal sex ratio from reciprocal mating-order crosses between *Chorthippus parallelus* subspecies. *Heredity* **68**, 345–352.
- CABOT, E. L., DAVIS, A. W., JOHNSON, N. A. & WU, C.-I. (1994). Genetics of reproductive isolation in the *Drosophila simulans* clade: complex epistasis underlying hybrid male sterility. *Genetics* **137**, 175–189.
- CARVAJAL, A. R., GANDARELA, M. R. & NAVEIRA, H. F. (1996). A three-locus system of incompatibility underlies inviability in hybrids between *Drosophila buzzatii* and *D. koepferae*. *Genetica* **98**, 1–19.
- COYNE, J. A. (1984). Genetic basis of male sterility in hybrids between two closely related species of *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4444–4447.
- COYNE, J. A. (1992). Genetics and speciation. *Nature* **355**, 511–515.
- DOBZHANSKY, T. (1937). *Genetics and the Origin of Species*. New York: Columbia University Press.
- DOWELL, S. J., BISHOP, A. L., DYOS, S. L., BROWN, A. J. & WHITEWAY, M. S. (1998). Mapping of a yeast G protein Beta-Gamma signaling interaction. *Genetics* **150**, 1407–1417.
- FOSTIER, M., EVANS, D. A. P., ARTAVANIS-TSAKONAS, S. & BARON, M. (1998). Genetic characterization of the *Drosophila melanogaster* Suppressor of *deltex* gene: a regulator of Notch signaling. *Genetics* **150**, 1477–1485.
- FRANK, S. A. (1999). Population and quantitative genetics of regulatory networks. *J. theor. Biol.* **197**, 281–294. doi:10.1006/jtbi.1998.0872.
- GADAU, J., PAGE, R. E., JR. & WERREN, J. H. (1999). Mapping of hybrid incompatibility loci in *Nasonia*. *Genetics* **153**, 1731–1741.
- GAVRILETS, S. (1997). Adaptation and speciation on holey adaptive landscapes. *Trends Ecol. Evol.* **12**, 307–312.
- GAVRILETS, S. (1999). A dynamical theory of speciation on holey adaptive landscapes. *Am. Nat.* **154**, 1–22.
- GERHART, J. & KIRSCHNER, M. (1997). *Cells, Embryos, and Evolution*. London: Blackwell Science.
- GIBSON, G. (1996). Epistasis and pleiotropy as natural properties of transcriptional regulation. *Theor. Pop. Biol.* **49**, 58–89.
- HOWARD, D. J., REECE, M., GREGORY, P. G., CHU, J. & CAIN, M. L. (1998). The evolution of barriers to fertilization between closely related organisms. In: *Endless Forms: Species and Speciation* (HOWARD, D. J. & BERLOCHER, S. H., eds.), pp. 279–288. New York: Oxford University Press.
- JOHNSON, N. A. (2000). Gene interaction and the origin of species. In: *Epistasis and the Evolutionary Process* (Wolf, J. B., Brodie, E. D. III & Wade, M. J., eds), pp. 197–212. New York: Oxford University Press.
- LEWIN, B. (1997). *Genes VII*. Oxford: Oxford University Press.
- LEWONTIN, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.
- LOSICK, R. (1998) Summary: three decades after sigma. *Cold Springs Harbor Symp. Quant. Biol.* **58**, 653–666.
- LORENZ, M. C. & HEITMAN, J. (1998). Regulators of pseudohyphal differentiation in *Saccharomyces cerevisiae* identified through multicopy suppressor analysis in ammonium permease mutant strains. *Genetics* **150**, 1443–1457.
- MAYR, E. (1963). *Animal Species and Evolution*. Cambridge: Harvard University Press.
- MORISATO, D. & ANDERSON, K. V. (1995). Signaling pathways that establish the dorsal-ventral pattern of the *Drosophila* embryo. *Ann. Rev. Gen.* **29**, 371–399.
- MULLER, H. J. (1940). Bearing of the *Drosophila* work on systematics. In: *The New Systematics* (HUXLEY, J., ed.), pp. 185–268. Oxford: Clarendon Press.
- NAVEIRA, H. F. & MASDIE, X. R. (1998) The genetics of hybrid male sterility in *Drosophila*. In: *Endless Forms: Species and Speciation* (Howard, D. J. & Berlocher, S. H., eds), pp. 330–338. New York: Oxford University Press.
- NEI, M., MARUYAMA, T. & WU, C.-I. (1983). Models of evolution of reproductive isolation. *Genetics* **103**, 557–579.
- NOOR, M. A. F. & AQUADRO, C. F. (1998). Courtship songs of *Drosophila pseudoobscura* and *D. persimilis*: analysis of variation. *Anim. Behav.* **56**, 115–125. doi:10.1006/anbe.1998.0779.
- NOWAK, M. A., BOERLIJST, M. C., COOKE, J. & MAYNARD SMITH, J. (1997). Evolution of genetic redundancy. *Nature* **388**, 167–171.
- ORR, H. A. (1989). Genetics of sterility in hybrids between two species of *Drosophila*. *Evolution* **43**, 180–189.
- ORR, H. A. (1995). The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**, 1805–1813.
- ORR, H. A., MADDEN, L. D., COYNE, J. A., GOODWIN, R. & HAWLEY, R. S. (1997). The developmental basis of hybrid inviability: a mitotic defect in *Drosophila* hybrids. *Genetics* **145**, 1031–1040.

- PALOPOLI, M. F. & WU, C.-I. (1994). Genetics of hybrid male sterility between *Drosophila* sibling species: A complex web of epistasis is revealed in interspecific studies. *Genetics* **138**, 329–341.
- PEREZ, D. E. & WU, C.-I. (1995). Further characterization of the *Odysseus* locus of hybrid sterility in *Drosophila*: one gene is not enough. *Genetics* **140**, 201–206.
- PEREZ, D. E., WU, C.-I., JOHNSON, N. A. & WU, M.-L. (1993). Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping of a hybrid-male sterility gene, *Odysseus* (*Ods*). *Genetics* **134**, 261–275.
- PHILLIPS, P. C. (1996). Waiting for compensatory mutation: phase zero of the shifting balance process. *Genetical Res. (Camb.)* **67**, 271–283.
- PHILLIPS, P. C. (1998). The language of gene interaction. *Genetics* **149**, 1167–1171.
- PHILLIPS, P. C. & JOHNSON, N. A. (1998). The population genetics of synthetic lethals. *Genetics* **150**, 448–457.
- PHILLIPS, S. E. V. (1994). Built by association: structure and function of helix–loop–helix DNA binding proteins. *Structure* **2**, 1–4.
- PROVINE, W. B. (1971). *The Origins of Population Genetics*. Chicago: University of Chicago Press.
- RAFF, R. C. (1996). *The Shape of Life*. Chicago: University of Chicago Press.
- REINBERG, D., ORPHANIDES, G., EBRIGHT, R., AKOULTCHEV, S., CARCAMO, J., CHO, H., CORTES, P., DRAPKIN, R., FLORES, O., HA, I., INOSTROZA, J. A., KIM, S., KIM, T.-K., KUMAR, P., LAGRANGE, T., LEROY, G., LU, H., MA, D.-M., MALDONADO, E., MERINO, A., MERMELSTEIN, F., OLAVE, I., SHELDON, M., SHIEKHATTAR, R., STONE, N., SUN, X., WEIS, L., YEUNG, K. & ZAWEL, L. (1998) The RNA polymerase II general transcription factors: past, present, and future. *Cold Springs Harbor Symp. Quant. Biol.* **58**, 83–103.
- RICE, W. R. & HOSTERT, E. E. (1993) Laboratory studies on speciation: what have we learned in 40 years? *Evolution* **47**, 1637–1653.
- ROBINSON, T., JOHNSON, N. A. & WADE, M. J. (1994). Postcopulatory, prezygotic isolation: intraspecific and interspecific sperm precedence in flour beetles. *Heredity* **73**, 155–159.
- SCHLUTER, D. (1998). Ecological causes of speciation. In: *Endless Forms: Species and Speciation* (Howard, D. J. & Berlocher, S. H., eds), pp. 114–129. New York: Oxford University Press.
- SMOCOVITIS, U. B. (1996). *Unifying Biology: The Evolutionary Synthesis and Evolutionary Biology*. Princeton: Princeton University Press.
- TING, C.-T., TSAUR, S.-C., WU, M.-L. & WU, C.-I. (1998). A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501–1504.
- TURELLI, M. & ORR, H. A. (1995). The dominance theory of Haldane's rule. *Genetics* **140**, 389–402.
- WADE, M. J. (1992a). Epistasis. In: *Keywords in Evolutionary Biology* (KELLER, E. F. & LLOYD, E. A., eds.), pp. 87–91. Cambridge: Harvard University Press.
- WADE, M. J. (1992b). Sewall Wright: genetic interaction and the shifting balance theory. *Oxford Surv. Evol. Biol.* **8**, 35–62.
- WADE, M. J. & GOODNIGHT, C. J. (1998). Perspective: the theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. *Evolution* **52**, 1537–1553.
- WADE, M. J. & JOHNSON, N. A. (1994). Reproductive isolation between two species of flour beetles, *Tribolium castaneum* and *T. freemani*: variation within and among geographical populations of *T. freemani*. *Heredity* **72**, 155–162.
- WELLS, M. M. & HENRY, C. S. (1998). Songs, reproductive isolation, and speciation in cryptic species of insects. In: *Endless Forms: Species and Speciation* (Howard, D. J. & Berlocher, S. H., eds), pp. 217–233. New York: Oxford University Press.
- WHITLOCK, M. C., PHILLIPS, P. C., MOORE, F. B.-G. & TONSOR, S. J. (1995). Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26**, 601–629.
- WU, C.-I. & PALOPOLI, M. F. (1994). Genetics of postmating reproductive isolation in animals. *Ann. Rev. Gen.* **28**, 283–308.
- YAMAMOTO, K. R., DARIMONT, B. D., WAGNER, R. L. & INIGUEZ-LLUHÍ, J. A. (1998). Building transcriptional regulatory complexes: signals and surfaces. *Cold Springs Harbor Symp. Quant. Biol.* **58**, 587–598.

APPENDIX A

Model for Two Loci and Two-dimensional Binding

Under two-dimensional binding rules, the product site and the regulatory site of adjacent loci in the pathway are each described by a pair of numbers and therefore differ in two dimensions (X and Y). We designate the differences x and y respectively, such that $x = \text{product}(j, 1) - \text{reg}(j + 1, 1)$ is the overall distance between the product and the regulatory elements along the X dimension and y is defined similarly. The origin $\{0, 0\}$ represents perfect binding in both dimensions. With 2 loci in the pathway, the phenotype [eqn (3)] resolves to be the binding strength between them [eqn (2)], $P = \exp[-(x^2 + y^2)]$. The two parental populations will have expected binding state differences of $\{x_1, y_1\}$ and $\{x_2, y_2\}$ along a circle with radius of r . Note that this model also describes outcomes for the 3-locus case with one-dimensional binding, because independent binding parameters within loci are analogous in our model to independent binding interactions at different loci.

If we assume that in generation t , the parental populations are at their optimal mean phenotype (P_{opt}), we can determine the expected values of binding deviations x and y of the parents and their F_1 hybrids. Substituting, $P_{opt} = \exp[-(x^2 + y^2)]$, and rearranging, we get $r_{opt}^2 = x^2 + y^2 = -\ln(P_{opt})$, where $\{x, y\}$ represents the

coordinates of either population. The relative $\{x, y\}$ positions of the two population means can be expressed by the angular degrees (θ) separating them [see Fig. A1(a)]. The mean F_1 hybrid binding strength may be found graphically by first drawing in the X - Y plane a chord between the two population coordinates, forming a triangle of the origin and this chord, then bisecting that triangle. Where the bisect meets the chord is the expected binding deviation of the F_1 hybrids. The length of that bisect, the distance between the origin and the F_1 midpoints, is $r_{F1} = r_{opt} \cos(\theta/2)$ [see Fig. A1(b)]. The F_1 hybrid phenotype is

$$\begin{aligned} P_{F1}(\theta) &= \exp[-r_{opt}^2 \cos^2(\theta/2)], \\ P_{F1}(\theta) &= \exp[\ln(P_{opt}) \cos^2(\theta/2)], \\ P_{F1}(\theta) &= P_{opt}^{\cos^2(\theta/2)}. \end{aligned} \quad (A.1)$$

Clearly, $P_{F1}(\theta)$ increases when the absolute value of θ increases [Fig. A2(a)]. The expected fitness of F_1 hybrids is

$$\begin{aligned} W_{F1}(\theta) &= \exp\left[-\left(\frac{P_{F1}(\theta) - P_{opt}}{\Omega}\right)^2\right] \\ &= \exp\left[-\left(\frac{P_{opt}^{\cos^2(\theta/2)} - P_{opt}}{\Omega}\right)^2\right]. \end{aligned} \quad (A.2)$$

The probability distribution of F_1 hybrid fitnesses in any generation is then

$$\begin{aligned} \Pr(W_{F1}|\theta, \Omega, t) &= C \exp\left[-\left(\frac{P_{opt,t}^{\cos^2(\theta/2)} - P_{opt,t}}{\Omega}\right)^2\right] \\ &\times \Pr(\theta), \quad -\pi \leq \theta < \pi, \end{aligned} \quad (A.3)$$

where $P_{opt,t}$ is the optimal phenotype at time t and C is a constant chosen so that

$$\int_{-\pi}^{\pi} \exp\left[-\left(\frac{P_{opt,t}^{\cos^2(\theta/2)} - P_{opt,t}}{\Omega}\right)^2\right] \Pr(\theta) d\theta = 1.$$

This fitness distribution is shown for a uniform θ distribution in Fig. A2(b). A uniform θ distribu-

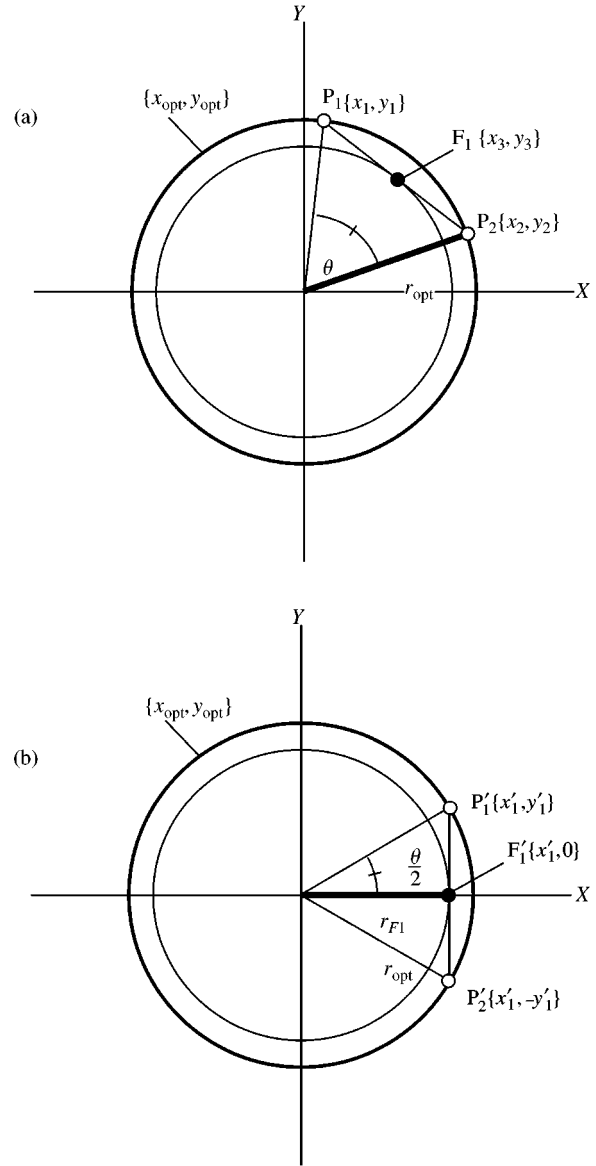


FIG. A1. (a) Deviations in promoter-product binding pairs in two dimensions (X and Y) that yield an optimal binding P_{opt} after selection. P_{opt} falls on a circle of radius r_{opt} . The two parental populations evolve independently to arbitrary points $\{x_1, y_1\}$ and $\{x_2, y_2\}$ on the optimal circle, and the F_1 hybrid binding falls at the point $\{x_3, y_3\}$ between them. (b) The F_1 binding in (a) is equivalent to one with axes rotated to the points P'_1 and P'_2 , whereupon it is readily seen that the F_1 binding is $r_{F1} = r_{opt} \cos(\theta/2)$.

tion is likely to be a reasonable approximation for our simulations (see text), wherein the populations start with perfect binding, since any biases in the direction of evolution in the X and Y dimensions in the two populations are likely to be negligible as long as sufficient mutations are

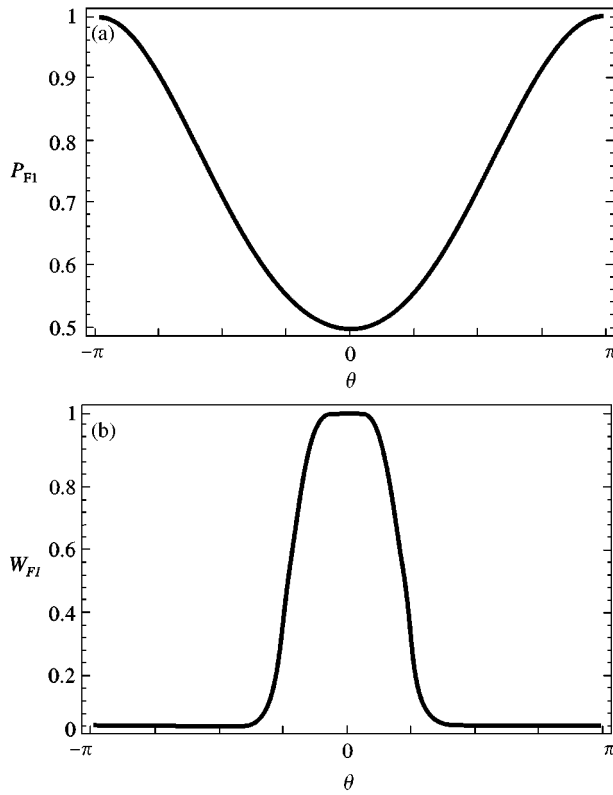


FIG. A2. (a) F_1 phenotypes expected from variation in the binding divergence angle θ (see Fig. A1). (b) F_1 fitness expected from these phenotypes.

available. This expected fitness distribution may be summarized in bar-chart form for different fitness categories ($W \geq 0.5$, etc.) by first dropping perpendiculars to the θ -axis for the desired fitness boundaries in Fig. A2(b), then determining the relative proportion of the θ -axis occupied by each fitness category. We used this method to generate expected values for comparison to the fitness distributions of our simulation outcomes.

APPENDIX B

Model for Multi-dimensional Binding

The two-dimensional model (Appendix A) can be generalized to the case of m independent binding interactions that determine the phenotype. The binding dimensions comprise the independent binding parameters that affect the interaction of a particular product-promoter binding pairs, accumulated over all interacting loci. For example, with n loci and one-dimensional binding, there are $m = n - 1$ binding dimensions, and

for n loci and two-dimensional binding, $m = 2(n - 1)$ binding dimensions exist. The generalization is somewhat qualitative because it ignores the effects of linkage but we expect it to provide a first-order approximation of the distribution of hybrid fitnesses. We may call these binding dimensions X_j , where j is the index of binding dimensions. The realized binding strength deviations along these dimensions may be represented by the vector $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$.

The phenotype [eqn (3)] generalizes to

$$P = \exp \left\{ - \sum_{j=1}^m x_j^2 \right\}. \quad (\text{B.1})$$

The optimal phenotype can be seen as lying on the surface of an m -dimensional hypersphere with a radius given by

$$r_{opt}^2 = \sum_{j=1}^m x_j^2 = -\ln(P_{opt}), \quad (\text{B.2})$$

and independently evolved populations may lie at different locations on this surface.

Extending the two-dimensional arguments above, the F_1 phenotype lies on a hypersphere having a radius given by

$$r_{F1}^2 = \prod_{j=1}^{m-1} r_{opt}^2 \cos^2(\theta_j/2), \quad (\text{B.3})$$

where θ_j is the angle separating the two populations in the X_0 - X_j plane. The F_1 phenotype is

$$P_{F1}(\theta) = P_{opt}^{\prod_{j=1}^{m-1} \cos^2(\theta_j/2)}, \quad (\text{B.4})$$

where θ is the vector of angles θ_j . The expected fitness of F_1 hybrids is found by substituting into eqn (4), yielding

$$\begin{aligned} W_{F1}(\theta) &= \exp \left[- \left(\frac{P_{F1}(\theta) - P_{opt}}{\Omega} \right)^2 \right] \\ &= \exp \left[- \left(\frac{P_{opt}^{\prod_{j=1}^{m-1} \cos^2(\theta_j/2)} - P_{opt}}{\Omega} \right)^2 \right]. \end{aligned} \quad (\text{B.5})$$

The distribution of F_1 hybrid fitnesses may be obtained if a distribution for θ is known. Assuming that it is uniform from $-\pi$ to π for all θ_j , the probability distribution of hybrid fitnesses in any

generation is then

where $P_{opt,t}$ is the optimal phenotype at time t and C is a constant chosen so that

$$\begin{aligned}
 &Pr(W_{F1}|\theta, \Omega, t) \\
 &= C \exp \left[- \left(\frac{P[\prod_{j=1}^{m-1} \cos^2(\theta_j/2)] - P_{opt,t}}{\Omega} \right)^2 \right], \\
 &\quad -\pi \leq \theta_j < \pi, \tag{B.6}
 \end{aligned}
 \qquad
 \int_{-\pi}^{\pi} \exp \left[- \left(\frac{P[\prod_{j=1}^{m-1} \cos^2(\theta_j/2)] - P_{opt,t}}{\Omega} \right)^2 \right] d\theta = 1.$$