doi:10.1111/evo.12275

SOLVING THE PARADOX OF STASIS: SQUASHED STABILIZING SELECTION AND THE LIMITS OF DETECTION

Benjamin C. Haller[1,2] and Andrew P. Hendry[1]

[1]*Department of Biology and Redpath Museum, McGill University, 859 Sherbrooke Street West, Montreal, Quebec, Canada H3A 0C4*

[2]*E-mail: benjamin.haller@mail.mcgill.ca*

Received February 13, 2013
Accepted September 10, 2013
Data Archived: Dryad doi: 10.5061/dryad.0jj03

Despite the potential for rapid evolution, stasis is commonly observed over geological timescales—the so-called "paradox of stasis." This paradox would be resolved if stabilizing selection were common, but stabilizing selection is infrequently detected in natural populations. We hypothesize a simple solution to this apparent disconnect: stabilizing selection is hard to detect empirically once populations have adapted to a fitness peak. To test this hypothesis, we developed an individual-based model of a population evolving under an invariant stabilizing fitness function. Stabilizing selection on the population was infrequently detected in an "empirical" sampling protocol, because (1) trait variation was low relative to the fitness peak breadth; (2) nonselective deaths masked selection; (3) populations wandered around the fitness peak; and (4) sample sizes were typically too small. Moreover, the addition of negative frequency-dependent selection further hindered detection by flattening or even dimpling the fitness peak, a phenomenon we term "squashed stabilizing selection." Our model demonstrates that stabilizing selection provides a plausible resolution to the paradox of stasis despite its infrequent detection in nature. The key reason is that selection "erases its traces": once populations have adapted to a fitness peak, they are no longer expected to exhibit detectable stabilizing selection.

KEY WORDS: **Competition, directional selection, disruptive selection, fitness landscape, frequency-dependent selection, selection gradient.**

Introduction

The "paradox of stasis" (or the "problem of stasis") has long been a focus of debate among evolutionary biologists (Simpson 1944; Lewontin 1974; Gould and Eldredge 1977; Wake et al. 1983; Williams 1992; Hansen and Houle 2004; Friedman 2009; Futuyma 2010; Kirkpatrick 2010). At the foundation of the paradox is the pattern, commonly seen in the fossil record, of long periods of morphological stasis despite the potential for—and occasionally the appearance of—rapid evolution (Darwin 1859; Simpson 1944; Eldredge and Gould 1972; Stanley 1979; Bradshaw 1991; Benton and Pearson 2001; Gingerich 2001; Eldredge et al. 2005; Gingerich 2009; Uyeda et al. 2011). Although the generality of stasis has been disputed (Gould and Eldredge 1977; Stebbins and Ayala 1981; Gould and Eldredge 1993; Erwin and Anstey 1995; Hunt 2007, 2008), the many instances in which it clearly occurs demand explanation.

One explanation for stasis is the presence of stabilizing selection (Fig. 1B) maintained over long timescales (Charlesworth et al. 1982; Estes and Arnold 2007), presumably owing to phenotypic fitness peaks that correspond to relatively stable niches (Holt and Gaines 1992; Ackerly 2003; Hansen 2012). Selection of this sort could constrain populations to a relatively constant and narrow range of high-fitness phenotypes and thus limit the frequency and extent of directional evolutionary change. Although this mechanism is unlikely to explain all instances of evolutionary stasis (Hansen and Houle 2004), and although other explanations

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

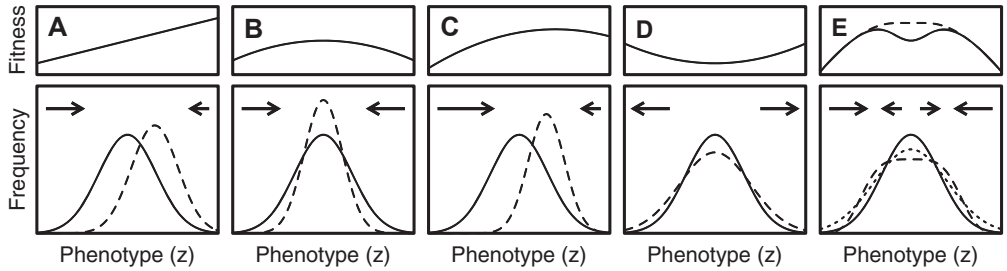

Figure 1. **Types of univariate selection (after fig. 1 in Phillips and Arnold 1989). The top of each panel shows a fitness function; below is shown a population trait frequency distribution before selection (solid line), the action of selection (arrows), and the frequency distribution after selection (dashed line). The types of selection shown are: (A) directional; (B) stabilizing; (C) a combination of directional and stabilizing selection; (D) disruptive selection; and (E) "squashed stabilizing selection" (SSS), a combination of stabilizing selection and negative frequency-dependent selection (see Introduction). The top of (E) illustrates that the addition of negative frequency-dependent selection can either flatten the top of the fitness peak (dashed line) or actually dimple it downwards (solid line); both are "squashed." The bottom of (E) illustrates that SSS causes the phenotypic distribution after selection to be platykurtic (dashed line); but for quantitative traits, random mating will restore a normal distribution in the offspring (dotted line), and thus the net effect is an increase in variance, as with disruptive selection (Slatkin 1979). Note that the fitness landscapes in (A)–(D) are static; because of this, a population experiencing disruptive selection (D) is in an unstable equilibrium and will escape from the fitness minimum in one direction or the other. In contrast, the fitness landscape in (E) is dynamic, due to the presence of frequency-dependent selection; a population experiencing SSS is at a stable equilibrium and cannot escape the fitness minimum (see Introduction).**

have been advanced (Wake et al. 1983; Hansen and Houle 2004; Eldredge et al. 2005; Estes and Arnold 2007; Zeh et al. 2009; Futuyma 2010; Kirkpatrick 2010; McGuigan et al. 2011), stabilizing selection does seem likely in many instances (Charlesworth et al. 1982; Lynch 1990; Estes and Arnold 2007; Uyeda et al. 2011). If stabilizing selection predominates over long timespans in nature, the paradox would largely be resolved, but a key difficulty remains: stabilizing selection does not seem to predominate in empirical studies of selection in nature (Travis 1989; Kingsolver et al. 2001, 2012). Indeed, disruptive selection (Fig. 1D) is detected (i.e., statistically significant) about as often as stabilizing selection, whereas directional selection (Fig. 1A) is detected even more often (Kingsolver et al. 2001, 2012; Kingsolver and Pfennig 2007; Knapczyk and Conner 2007; Kingsolver and Diamond 2011). Furthermore, even when stabilizing selection is detected it often does not persist through time (Siepielski et al. 2009, 2011). We here propose and test a hypothesis that resolves this apparent disconnect between theoretical expectations and empirical findings, and thus removes a key objection to stabilizing selection as a resolution to the paradox of stasis.

Our hypothesis is that stabilizing selection will be difficult to detect empirically even when populations commonly occupy stabilizing fitness landscapes. This hypothesis derives from five postulates. First, when a population is well adapted, the fitness peak it occupies might be broad compared to the phenotypic range of the population (Hendry and Gonzalez 2008; Cresswell 2000), leading to relatively few selective deaths and thus a statistically weak signature of stabilizing selection. In essence, selection "erases its traces" by causing the phenotypic variance of the population to adjust to the width of the fitness peak, and so fewer selective deaths are subsequently observed even though the fitness landscape has not changed. Second, populations on fitness peaks might stochastically wander back and forth, generating episodic directional selection even though the fitness landscape is stabilizing and invariant (Wright 1932; Lande 1976; Hunt et al. 2008). Third, random mortality (i.e., mortality uncorrelated with the focal trait subject to a stabilizing fitness landscape) might obscure the selective signal, decreasing statistical power (Hersch and Phillips 2004). Fourth, negative frequency-dependent selection (Ayala and Campbell 1974) might flatten, or even dimple, the tops of fitness peaks (Rosenzweig 1978; Slatkin 1979; Abrams et al. 1993; Bürger 2002a,b; Bürger and Gimelfarb 2004; Bürger 2005; Rueffler et al. 2006). This combination of negative frequency-dependent selection and stabilizing selection, which we term "squashed stabilizing selection" (SSS; Fig. 1E; see *Squashed stabilizing selection*), causes selective deaths close to the phenotypic mean that decrease detection of stabilizing selection while increasing detection of disruptive selection (Day and Young 2004; Sinervo and Calsbeek 2006; Kingsolver and Pfennig 2007). Fifth, the small sample sizes typically used in empirical studies of selection might yield insufficient statistical power to detect stabilizing selection (Kingsolver et al. 2001; Hersch and Phillips 2004).

Although the above postulates seem reasonable and would be expected to limit the detection of stabilizing selection, they have not previously been subject to quantitative exploration. We performed this exploration through an individual-based model of a population subject to an invariant stabilizing fitness function resulting from a resource-based fitness peak. The dynamics of

populations subject to stabilizing fitness functions have been extensively explored by previous theoretical research (Wright 1935; Robertson 1956; Latter 1960; Gale and Kearsey 1968; Lande 1976; Bürger 1986, 1998; Keightley and Hill 1988; Barton 1989; Bürger et al. 1989; Foley 1992; Bürger and Lande 1994; Bürger and Gimelfarb 1999; Willensdorfer and Bürger 2003; Estes and Arnold 2007). Extending these findings was not our aim; rather we were specifically interested in the empirical methods normally employed to detect stabilizing selection on natural populations. The efficacy of these empirical methods has not been explored in previous research, and yet this efficacy is central to the crucial disconnect at the heart of the paradox of stasis: the infrequent empirical detection of stabilizing selection versus the theoretical expectation that stabilizing fitness landscapes should be common.

To address this disconnect as directly as possible, we followed the "virtual ecologist" approach advocated by Zurell et al. (2010). Specifically, we sampled the modeled population in simulated mark-recapture experiments each generation, and then used these samples in standard regression-based tests of selection. From this analysis, we show that the pattern of selection observed in our model under reasonable parameter values is compatible with the empirical pattern of selection observed in nature. Our results therefore resolve the crucial disconnect, by showing that a population that has adapted to a stabilizing fitness function is expected to exhibit statistically detectable selection (of any type) only rarely using standard methods. Furthermore, when selection is detected on such a population, it might be directional or (particularly with the addition of negative frequency-dependence) disruptive as often as stabilizing. Although natural populations might often exhibit long-term evolutionary stasis due to stabilizing fitness peaks, empirical studies are currently limited in their ability to detect this phenomenon.

Our individual-based approach is essential to our goal for several complementary reasons. First, it allows the phenotypic variance of the population to adjust to the selective regime; selection can thus "erase its traces" as it would in a natural population, rather than being constrained by a fixed phenotypic variance. Second, it allows negative frequency-dependent selection to be realistically modeled, including generation-by-generation temporal fluctuations in frequency-dependent selection due to the changing phenotypic distribution. Third, it allows drift and demographic stochasticity to potentially influence evolution, as would be the case in natural populations.

Methods

MODEL SUMMARY

A full model description is given in Supplemental S1. In brief, we developed an individual-based, nonspatial, sexual model of the evolution of a single population on an invariant stabilizing fitness function (parameters summarized in Table 1). The model includes both a selected trait (a_s), subject to the stabilizing fitness function, and a neutral trait (a_n) physically unlinked with the selected trait. The neutral trait serves as a control, showing the pattern of selection detected on a trait that is not under selection, but that exists in organisms under selection on other traits. Both traits have a genetic value based on one of three implemented genetic architectures (see Supplemental S1, *Genetic architectures*): (1) a single value representing a quantitative genetics approach with, conceptually, an infinite number of loci (the "quantitative" architecture, following, e.g., Heinz et al. 2009); (2) a diploid 8-locus triallelic architecture ("triallelic," following, e.g., Thibert-Plante and Hendry 2011); or (3) a diploid 8-locus continuum-of-alleles architecture ("continuum," following, e.g., Yeaman and Guillaume 2009). These three architectures were chosen as they bracket the main alternatives used in theoretical models—alternatives that have been argued to matter for various outcomes. Notably, all architectures allow the genetic variance of the population to evolve in response to the selective regime, thus producing more realistic dynamics than would a fixed variance. Phenotypic trait values (z_s, z_n) are derived from the respective genetic values (a_s, a_n) by the addition of random environmental noise with variance V_E.

Time is divided into nonoverlapping generations with three phases: random mortality, selective mortality, and reproduction. In the first phase, the population size is reduced by random mortality at a rate m, representing deaths due to causes other than selection on the focal trait. In the second phase, additional mortality occurs based on the absolute fitness of each individual as a function of its phenotype, due to both a stabilizing fitness function (always enabled) and negative frequency-dependent selection (if enabled by an "on/off switch" parameter C), similar to Roughgarden (1972) and Dieckmann and Doebeli (1999). The stabilizing fitness function is modeled with a Gaussian function of width ω, so that fitness decreases with increasing distance of an individual's phenotype from the optimum phenotype θ (see Supplemental S1, *Selection phase*). Standardized by the phenotypic standard deviation, ω^2 was typically less than 50, with a median of $\sim$17.5 and a strong mode at 3 (see Supplemental S2, *The strength of stabilizing selection*), which is consistent with the range of values typically observed empirically (Estes and Arnold 2007). Negative frequency-dependent selection, conceptualized as competition, is modeled with a phenotypic competition kernel width of σ_c and an intensity c, and its effects on fitness are combined multiplicatively with the fitness effects due to the underlying stabilizing fitness function (see Supplemental S1, *Interactions* and *Selection phase*). In the third phase, sexual reproduction occurs randomly (nonassortatively) up to the environment's carrying capacity of juveniles, N_j. Inheritance is modeled according to the above genetic architectures, including mutation occurring at a rate μ with mutational effect size standard deviation α (see Supplemental S1,

Table 1. Model parameters (above the divider) and analysis-related symbols (below the divider) with their value(s) and their units. Units are expressed using the symbols E (ecological phenotype), I (individuals), and G (generations).

Description	Symbol	Value	Units
Competition enabled	C	off, on	–
Number of juveniles (individuals prior to mortality)	N_j	1000, 2500[1]	I
Environmental variance	V_E	0.1, 0.01, 0.001	E[2]
Mutation rate per locus	μ	0.001, 0.00001	G^{-1}
Mutational effect size (standard deviation of the mutational kernel)	α	0.5, 0.05[2]	E
Phenotypic optimum	θ	0.0	E
Width of the Gaussian fitness function (strength of stabilizing selection)	ω	1.0, 10.0	E
Phenotypic competition width (standard deviation of the competition kernel)	σ_c	0.5, 2.0[3]	E
Intensity of competition	c	1.0[3]	–
Mortality rate	m	0.0, 0.1, 0.5	G^{-1}
Genetic architecture	G	Q, T, C[4]	–
Mark-recapture subsample size	N_s	100, 500, 1000, 2500[1]	I
Analyzed trait	T	a_s, z_s, a_n, z_n[5]	–

[1]Realizations with $N_j = 2500$ were limited to $\mu = 0.001$ and $\alpha = 0.5$, and are shown only in Supplemental S2; subsample size $N_s = 2500$ was only conducted for those realizations.

[2]Parameter α is not defined for the triallelic genetic architecture; where values of α are plotted, a value of 1.0 is used in this case; see Supplemental S1, *Genetic architectures*.

[3]Parameters σ_c and c are used by the model only when competition is enabled (C = on); see *Model summary* and Supplemental S1, *Selection phase*.

[4]Genetic architecture values Q, T, and C represent the quantitative, triallelic, and continuum genetic architectures, respectively; see *Model summary* and Supplemental S1, *Genetic architectures*.

[5]Analyzed trait values a_s, z_s, a_n, and z_n represent the selected trait's genetic (breeding) and phenotypic values and the neutral trait's genetic (breeding) and phenotypic values, respectively; see *Model summary* and Supplemental S1, *Environment and state variables*.

Parameters, regarding mutational variance, and Supplemental S2, *Effects of heritability*, regarding genetic and phenotypic variances; these are in general agreement with empirical values).

DATA COLLECTION

A total of 720 realizations ("runs") of the model generated the main body of results; see Table 1 for parameter values used. The triallelic architecture comprised 144 realizations: two values of C × three values of V_E × two values of μ × two values of ω × two values of σ_c × three values of m. The quantitative and continuum architectures, which had two values of α for each of the above parameter combinations, each comprised 288 realizations. Because σ_c is not used by the model when competition is off, the 360 realizations without competition contain redundancy; specifically, that set covers 180 distinct parameter combinations, each realized twice. This redundancy allowed the total sizes of various subsets of the data to be equal, simplifying the analysis, and also allowed the reproducibility of results to be tested (see Supplemental S2, *Autocorrelation and Reproducibility*).

Each realization of the model comprised 60,000 generations, with population census information saved each generation (see Supplemental S1, *Observables*). The first 10,000 generations were considered "burn-in" and were not used in the results presented here. In reality, fewer than 1000 generations were necessary for the model to reach a pseudo-equilibrium state (not shown), but 10,000 generations were used to ensure that the initial state of the model was unlikely to affect results.

DATA ANALYSIS

The analysis of the data generated by the model realizations is summarized in Figure 2. Analysis was conducted in the R programming language, version 2.14.2 (R Development Core Team 2012). A significance threshold of $\alpha = 0.05$ was used for all statistical tests unless otherwise specified. For each model realization, separate analyses were conducted using population samples of several sizes (N_s of 100, 500, and 1000 individuals). When the sample size N_s equaled the carrying capacity N_j ($N_j = N_s = 1000$), the "sample" was a full population census; this case considered the detection of selection in the absence of sampling error. To generate a sample of size N_s, a simulated mark-recapture survey was conducted in which N_s juveniles were "marked" at the beginning of a generation, and only that subset of individuals was "recaptured" and subjected to analysis at the end of the generation. A recapture rate of 100% was guaranteed; in other

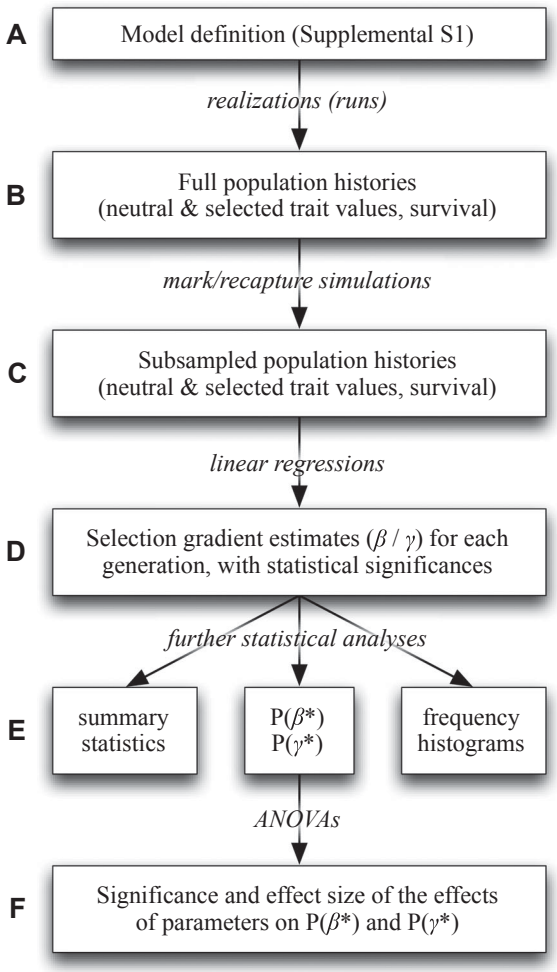

Figure 2. The principal path of data analysis in this research. The model (A) was realized with various parameter values to generate the primary data set: full histories of the modeled populations (B), including genotypic and phenotypic values for the neutral and selected traits of each individual, and whether each individual survived to reproductive age. These histories were subsampled to simulate an empirical mark-recapture protocol, generating the subsampled histories (C) used by the subsequent analysis. Linear regressions of survival as a function of trait value were then applied to each generation of the subsampled histories to produce estimates of the selection gradients (β and γ) acting on each population in each generation (D). Summary statistics and frequency histograms were used to summarize these selection gradient estimates. Additionally, the rates of detection of β and γ, P(β*) and P(γ*) (E), were calculated for each realization to determine how often we can expect to detect selection of different types (linear and quadratic) on a population subject to an invariant stabilizing fitness function. Finally, ANOVAs incorporating all of the varied model parameters were used to determine the significance and effect size of the effects of model parameters on P(β*) and P(γ*), both without and with competition (F).

words, every "marked" individual that survived the random mortality and selective mortality phases was "recaptured," and thus it could be reliably concluded that all individuals not "recaptured" had died. This methodology produced the most comprehensive data set possible for a mark-recapture survey of size N_s, and thus maximized the rate of detection of selection; it was thus conservative in testing our hypothesis. Similarly, although surviving individuals do vary in their reproductive output in our model, our use of survival rather than lifetime reproductive output as the metric of fitness was conservative because variation in mating success and fecundity in our model is stochastic, not trait-based, and is therefore not heritable. The addition of this random noise arising from reproductive stochasticity would have only further masked the signal of trait-based selection.

For each generation in each subsampled population history, univariate regression analysis of fitness (as defined by binary survival) as a function of trait value (standardized to a mean of 0 and a standard deviation of 1) was used to determine the strength, direction, type, and significance of selection. Two types of regression analysis were conducted: linear, following Lande and Arnold (1983), and logistic, following Janzen and Stern (1998). Detection of selection was much more frequent with linear regression, making it more conservative in testing our hypothesis, so we focus here on results from the linear regressions (see Supplemental S2, *Logistic vs. linear regression*, for methods and results for logistic regression). Regressions were conducted using relative fitness (absolute fitness divided by mean fitness across the sample), following standard practice (Lande and Arnold 1983; Brodie et al. 1995; Stinchcombe et al. 2008). For each generation, regression was conducted first with a linear term to assess directional selection (the "nonquadratic regression"), and then with both a linear and a quadratic term (the "quadratic regression") to assess quadratic selection. Negative (positive) quadratic selection is consistent with, but not limited to, stabilizing (disruptive) selection (Schluter 1988; Brodie et al. 1995). These regressions were conducted once using the genetic trait (breeding) values of individuals, and once using phenotypic trait values, allowing us to compare the two. Finally, each of these regressions was conducted once for the selected trait and once for the neutral trait. Eight regressions per generation per subsampled history were therefore conducted: linear/quadratic × genetic/phenotypic × selected/neutral. Quadratic coefficients from these regressions were doubled to yield quadratic selection gradients, γ (Stinchcombe et al. 2008).

A total of 2,764,800,000 regressions were conducted (including supplemental realizations and logistic regressions; see Supplemental S2). Estimated selection gradients from each regression, with their associated standard error and *P*-value, became the data for further analysis as described below. Multiple testing was not a concern because it was the distribution of estimates

and significances, not the significance of any particular estimate, that was of interest. We used univariate regressions, rather than multiple regressions including both the neutral and selected traits, because the two traits were not physically linked and were essentially uncorrelated in the model realizations (see Supplemental S2, *The neutral trait*). As implied earlier, we used SD-standardized selection gradients, also called variance-standardized selection gradients or selection intensities (Matsumura et al. 2012) and symbolized β_σ by Hereford et al. (2004). We did not use the alternative method of mean-standardization (Hereford et al. 2004; Matsumura et al. 2012) because the modeled traits are on an interval scale, not a ratio scale (Houle et al. 2011); regardless, standardization is not relevant to our conclusions. We refer to selection "gradients" throughout, rather than selection "differentials," because the values have been standardized (Matsumura et al. 2012).

Summary statistics of the selection gradients, such as the mean, median, standard deviation, and median absolute deviation (MAD), were taken across the 50,000 post-burn-in generations of each realization of the model for many of the per-generation statistics computed. The MAD is a robust measure of statistical dispersion, calculated as the median of the absolute deviations about the median of a sample; following standard practice, we scale it by 1.4826 for consistency with the standard deviation (Hampel 1974; Rousseeuw and Croux 1993). We will refer to the rate of detection of linear selection (i.e., the rate at which the estimated linear selection gradient is statistically significant) in the nonquadratic regressions using the symbol P(β*), and the rate of detection of quadratic selection in the quadratic regressions using the symbol P(γ*). These two statistics directly addressed our central question, from an empirical sampling perspective: they are the rate at which we could statistically infer selection, whether linear or quadratic, for a population known to be evolving on an invariant stabilizing fitness function.

Realizations with and without competition were generally analyzed separately due to the large qualitative effect of competition on the model dynamics (see *Effects of competition*). Welch's *t*-tests and analysis of variance (ANOVA with main effects and two-way interactions) were used to determine the significance and effect size for the independent variables on dependent variables such as P(β*) and P(γ*). Independent variables included: (1) model parameters, N_j, V_E, μ, α, ω, m, and when competition was enabled, σ_c; (2) the genetic architecture, G, used for a run; (3) the trait examined, T, whether a_s, z_s, a_n or z_n; and (4) the mark-recapture sample size, N_s (see Table 1). Paired *t*-tests were used in some cases to compare the means of parallel groups (realizations with vs. without competition, for example). In these cases, each realization in one data set was paired with the (unique) realization in the other data set with the same values for all independent variables.

Significance is relatively meaningless for simulation studies, because any nonzero effect can be made significant with a sufficiently large number of realizations. The emphasis in our results is thus upon the effect size (given as η^2; Levine and Hullett 2002), not the significance, of the effects observed.

Results

A data set containing summary statistics and β and γ distributions for each realization of the model is published on Dryad (Haller and Hendry 2013). Because the raw model output far exceeds Dryad's 10 GB data set limit, online provisioning of the raw data is not possible, but the data set provided suffices to reconduct the analyses reported below.

Complete analysis of the neutral trait is presented in Supplemental S2, *The neutral trait*. In summary, the mean rate of detection of selection (linear or quadratic) on the neutral trait was less than the expected type I error rate, and was significantly less than the mean rate of detection of selection on the selected trait. These observations confirm that the neutral trait acted as a control, and that results for the selected trait are thus indeed the result of selection. All analyses below examine the selected trait. This presentation focuses on the largest effects, with the remaining effects shown in the referenced tables and figures.

EFFECTS OF COMPETITION

For the selected trait (genotypic value a_s and phenotypic value z_s, taken together), P(β*) was significantly lower with competition than without (competition: mean = 0.0474, SD = 0.0112, n = 2160, no competition: mean = 0.0824, SD = 0.0767, n = 2160, paired $t_{2159} = 22.0$, $P < 0.001$; Fig. 3). Indeed, with competition P(β*) was only slightly greater than for the neutral trait, although the difference was significant (selected trait: mean = 0.0474, SD = 0.0112, n = 2160, neutral trait: mean = 0.0456, SD = 0.0088, n = 2160, one-sided unpaired $t_{4091.1} = 5.93$, $P < 0.001$). P(γ*) was significantly lower with competition than without (competition: mean = 0.1265, SD = 0.2013, n = 2160, no competition: mean = 0.1464, SD = 0.2345, n = 2160, paired $t_{2159} = 3.05$, $P = 0.002$; Fig. 4). Furthermore, the mean proportion of quadratic selection detected that was stabilizing was lower with competition than without (Fig. 4), indicating that competition caused a shift away from the detection of stabilizing selection, toward the detection of disruptive selection (competition: mean = 0.3615, SD = 0.2670, n = 2091, no competition: mean = 0.7404, SD = 0.2264, n = 2091, paired $t_{2090} = -67.7$, $P < 0.001$; only pairs in which quadratic selection was detected for both realizations were included). In short, the model dynamics qualitatively differed with versus without competition (see also *Distribution of selection gradient values*, and Supplemental S2, *Two case studies*). For this reason, the two cases are analyzed separately below.

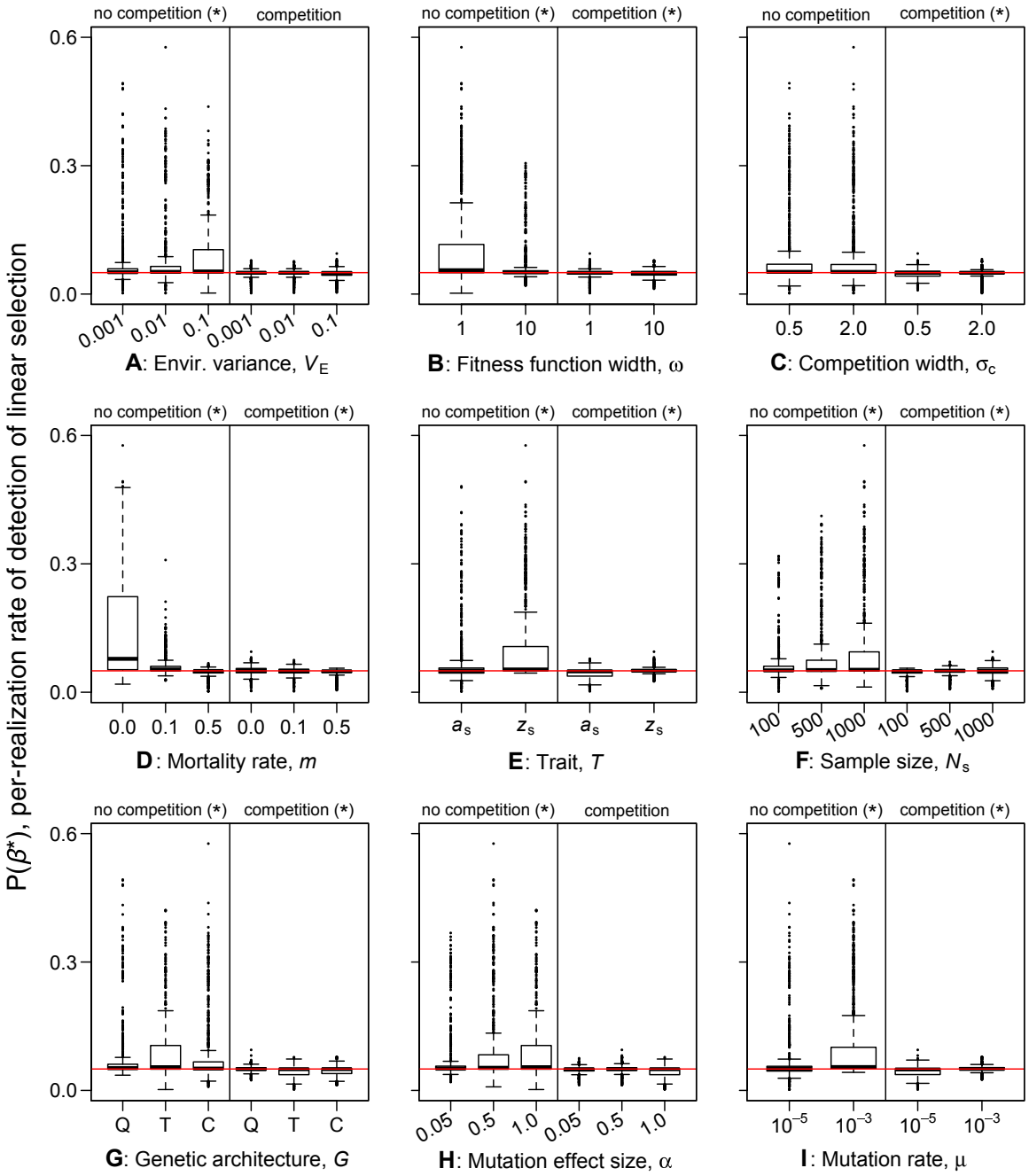

Figure 3. Effects of independent variables on the rate of detection of linear selection, P(β*). Each panel shows effects without competition on the left, and with competition on the right, for: (A) environmental variance, V_E; (B) fitness function width, ω; (C) competition width, σ_c; (D) mortality rate, m; (E) trait examined, T; (F) sample size, N_s; (G) genetic architecture, G; (H) mutation effect size, α; and (I) mutation rate, μ. Parameters for which ANOVA indicates a significant effect are shown with stars (*) at top (see Tables S2.1 and S2.3). Boxes span the first to third quartiles, with a thick line at the median; whiskers extend to the most extreme data point no more than 1.5× the interquartile range from the box. Red lines indicate the threshold used to determine significance of individual selection gradient estimates ($\alpha = 0.05$); realizations above the red line detected linear selection more often than the expected type I error rate. Each panel is based upon 4320 realizations, and thus the outliers shown are a small minority of realizations. Because the same P(β*) values are plotted in each panel, the combination of parameter values that produced most of the high-P(β*) outliers may be readily ascertained: $\omega = 1$, $m = 0.0$, $T = z_s$, and $N_s = 1000$.

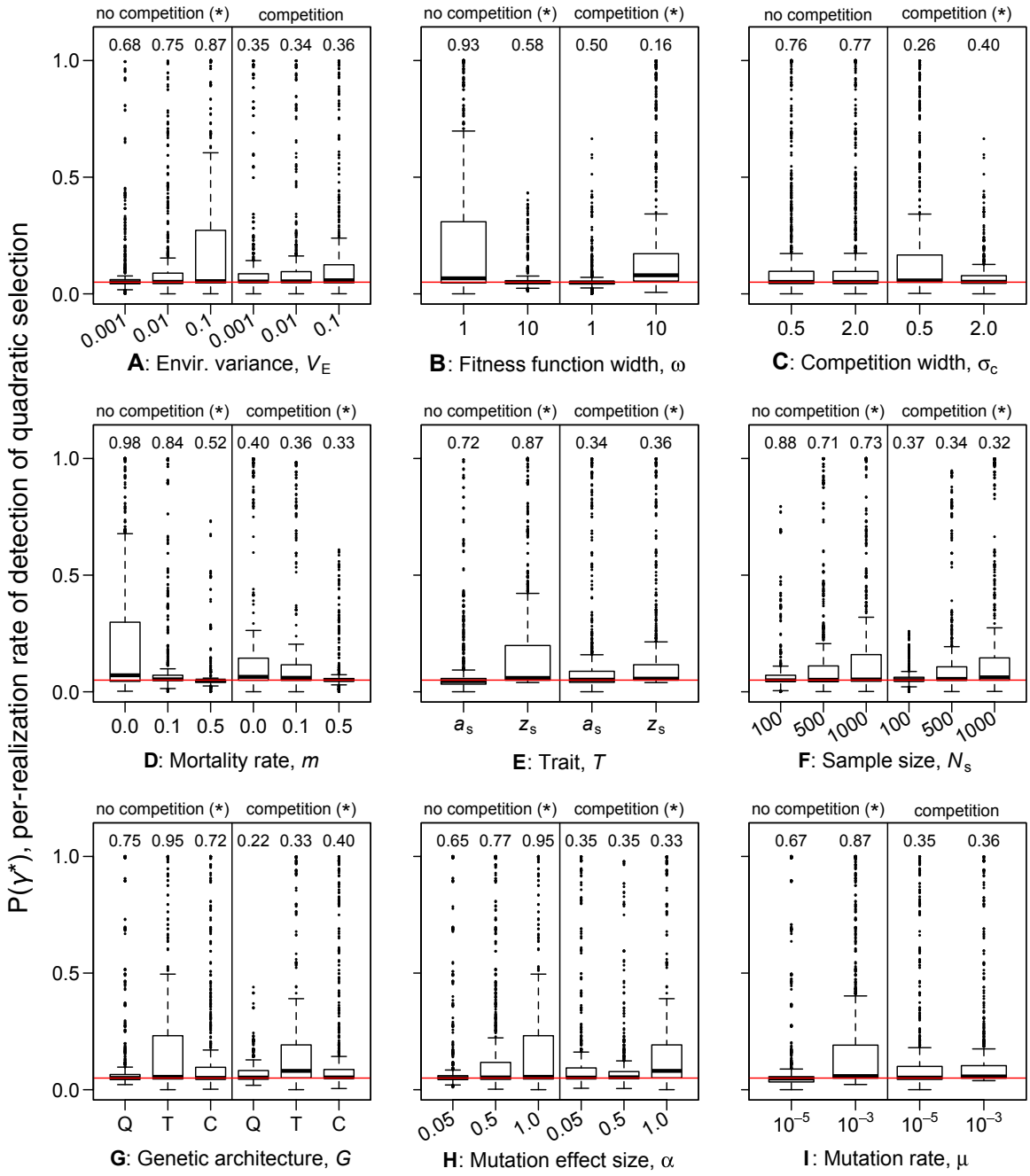

Figure 4. Effects of independent variables on the rate of detection of quadratic selection, P(γ*). Each panel shows effects without competition on the left, and with competition on the right, for: (A) environmental variance, V_E; (B) fitness function width, ω; (C) competition width, σ_c; (D) mortality rate, m; (E) trait examined, T; (F) sample size, N_s; (G) genetic architecture, G; (H) mutation effect size, α; and (I) mutation rate, μ. Parameters for which ANOVA indicates a significant effect are shown with stars (*) at top (see Tables S2.5 and S2.7). Boxes span the first to third quartiles, with a thick line at the median; whiskers extend to the most extreme data point no more than 1.5× the interquartile range from the box. Red lines indicate the threshold used to determine significance of individual selection gradient estimates ($\alpha = 0.05$); realizations above the red line detected quadratic selection more often than the expected type I error rate. Numbers above each column indicate the median proportion of detected quadratic selection that was stabilizing. More precisely, the number is the median of per-realization scores across all realizations in the given column, where each per-realization score is the proportion of generations, among only those generations for which quadratic selection was detected, for which the detected quadratic selection was stabilizing (i.e., a negative estimate for γ). Note that this metric weights all realizations equally, regardless of P(γ*). Each panel is based upon 4320 realizations, and thus the outliers shown are a small minority of the data. Because the same P(γ*) values are plotted in each panel, the combination of parameter values that produced most of the high-P(γ*) outliers may be readily ascertained: without competition, $V_E = 0.1$, $\omega = 1$, $m = 0.0$, $T = z_s$, and $N_s > 100$; with competition, $\omega = 10$, $\sigma_c = 0.5$, $m < 0.5$, and $N_s > 100$.

DETECTION OF LINEAR SELECTION: P(β*)

Linear selection was not often detected in most realizations (Fig. 3). Without competition, the median P(β*) value was 0.0517, although the variation among realizations was large (MAD = 0.00738, range 0.002–0.577). With competition, the median P(β*) was slightly lower, 0.0498, with less variation among realizations (MAD = 0.00427, range 0.002–0.095). Although both medians were close to the type I error rate, the high variation among realizations meant that selection could sometimes be detected above sampling error.

Without competition, ANOVA with only main effects (see *Data analysis*) explained 47.8% of variance in P(β*), and all independent variables were significant (Table S2.1). Detection of linear selection was increased by a lower random mortality rate m ($\eta^2 = 0.269$; Fig. 3D), by the use of phenotypic rather than genotypic values ($\eta^2 = 0.047$; Fig. 3E), by a higher mutation rate μ ($\eta^2 = 0.045$; Fig. 3I), and by a smaller stabilizing fitness function width ω (a narrower fitness peak; $\eta^2 = 0.071$; Fig. 3B). Other parameters had only small effects ($\eta^2 < 0.03$; Table S2.1 and Fig. 3). Analysis of variance incorporating the 28 second-order interaction terms (Table S2.2) explained an additional 31.4% of variance, but the only interactions of large effect ($\eta^2 \geq 0.03$) involved the mortality rate m (m*ω, m*T, m*N_s, m*μ). In all these interactions, high random mortality obscured effects of the other parameters (see Supplemental S2, *Selective deaths and the detection of selection*).

With competition, ANOVA with only main effects explained 26.6% of variance in P(β*), and all independent variables except V_E, ω, and α were significant (Table S2.3). Detection of linear selection was increased by a lower mutation rate μ ($\eta^2 = 0.095$; Fig. 3I), by the use of phenotypic rather than genetic values ($\eta^2 = 0.073$; Fig. 3E), and by the use of the quantitative genetic architecture ($\eta^2 = 0.040$; Fig. 3G). Other parameters had only small effects ($\eta^2 < 0.03$; Table S2.3 and Fig. 3). Analysis of variance incorporating the 36 second-order interaction terms (Table S2.4) explained an additional 33.8% of variance. Interactions of large effect ($\eta^2 \geq 0.03$) included ω*σ_c (increased detection with $\sigma_c \cong \omega$; see also *Distribution of selection gradient values*), ω*T (increased effect of T for $\omega = 1$), ω*G (increased effect of G for $\omega = 10$), m*N_s (increased effect of N_s with smaller m), and T*μ (increased effect of T with smaller μ).

DETECTION OF QUADRATIC SELECTION: P(γ*)

Quadratic selection was not often detected in most realizations (Fig. 4). Without competition, the median P(γ*) value was 0.0513, although the variation among realizations was large (MAD = 0.0178, range 0.000–1.000). With competition, the median P(γ*) was slightly higher, 0.0542, also with high variation (MAD = 0.0183, range 0.000–0.999). Again, although both medians were close to the type I error rate, the high variation among realizations meant that selection could sometimes be detected above sampling error.

Without competition, ANOVA with only main effects (see *Data analysis*) explained 42.6% of variance in P(γ*), and all independent variables were significant (Table S2.5). Detection of quadratic selection was increased by a smaller stabilizing fitness function width ω (a narrower fitness peak; $\eta^2 = 0.132$; Fig. 4B), by the use by phenotypic rather than genetic values ($\eta^2 = 0.077$; Fig. 4E), by lower random mortality m ($\eta^2 = 0.084$; Fig. 4D), by higher environmental variance V_E ($\eta^2 = 0.053$; Fig. 4A), and by a larger sample size N_s ($\eta^2 = 0.031$; Fig. 4F). Other parameters had only small effects ($\eta^2 < 0.03$; Table S2.5 and Fig. 4). The quadratic selection detected was always predominantly stabilizing (Fig. 4). Furthermore, higher rates of detection of quadratic selection were generally associated with a higher proportion of the detected selection being stabilizing (Fig. 4), although this was not true for the effect of sample size (Fig. 4F). Analysis of variance incorporating the 28 second-order interaction terms (Table S2.6) explained an additional 35.8% of variance, but the only interactions of large effect ($\eta^2 \geq 0.03$) involved the environmental variance V_E and the fitness function width ω (V_E*ω, V_E*T, ω*m, ω*T). In particular, smaller ω (a narrower fitness peak) and higher V_E amplified the effects of T and m, and in combination they strongly increased detection of quadratic selection (realizations with $\omega = 1$ and $V_E = 0.1$: $n = 360$, median = 0.193, MAD = 0.249, range 0.000–1.000; other realizations: $n = 1800$, median = 0.0503, MAD = 0.0133, range 0.000–0.998).

With competition, ANOVA with only main effects explained 33.9% of the variance in P(γ*), and all independent variables except V_E and μ were significant (Table S2.7). Detection of quadratic selection was increased by a larger stabilizing fitness function width ω (a broader fitness peak; $\eta^2 = 0.107$; Fig. 4B), by a smaller competition width σ_c ($\eta^2 = 0.076$; Fig. 4C), by a larger sample size N_s ($\eta^2 = 0.059$; Fig. 4F), and by the use of the triallelic genetic architecture ($\eta^2 = 0.058$; Fig. 4G). Other parameters had only small effects ($\eta^2 < 0.03$; Table S2.7 and Fig. 4). The quadratic selection detected was now always predominantly disruptive (Fig. 4; see *Distribution of selection gradient values*). Higher rates of detection of quadratic selection were associated with a higher proportion of the detected selection being disruptive in some cases (ω, σ_c, N_s), but with a higher proportion being stabilizing in other cases (V_E, T), and with no clear effect for the remaining parameters (Fig. 4). Analysis of variance incorporating the 36 second-order interaction terms (Table S2.8) explained an additional 35.8% of variance, but the only interactions of large effect ($\eta^2 \geq 0.03$) involved ω and σ_c (ω*σ_c, ω*G, σ_c*G). In particular, larger ω and smaller σ_c amplified the effects of the genetic architecture, and in combination they strongly increased detection of quadratic selection (realizations with $\omega = 10$ and $\sigma_c = 0.5$:

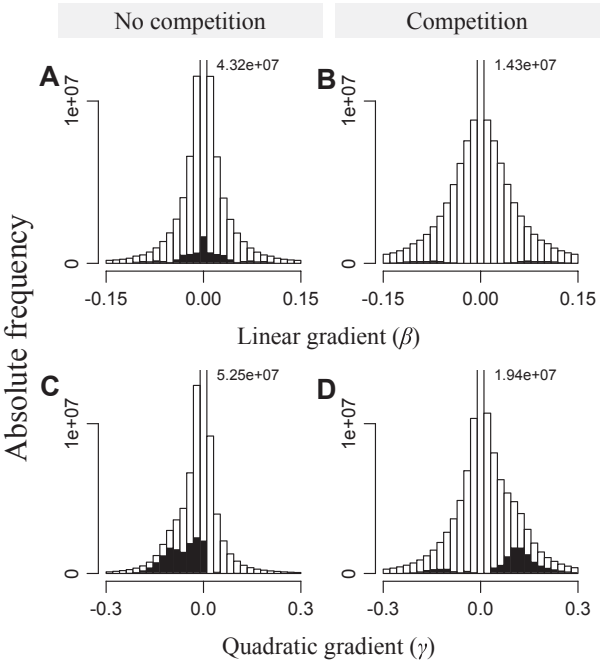

Figure 5. Absolute frequency histograms of linear selection gradients, β, and quadratic selection gradients, γ, for the selected trait (a_s and z_s, taken together), across all realizations of the model: (A) frequency of β with no competition, (B) frequency of β with competition, (C) frequency of γ with no competition, (D) frequency of γ with competition. In all panels, black shading shows the portion of estimates of β or γ that are significant ($P < 0.05$). Note that the central peak in all panels is off of the scale; many nonsignificant gradient estimates close or equal to zero were observed.

$n = 540$, median $= 0.161$, MAD $= 0.147$, range 0.0443–0.999; other realizations: $n = 1620$, median $= 0.0507$, MAD $= 0.0109$, range 0.000–0.665).

DISTRIBUTION OF SELECTION GRADIENT VALUES

Following Kingsolver et al. (2001) and others (see Introduction), and in the "virtual ecologist" spirit, we examined frequency distribution histograms of selection gradients β and γ. These distributions convey the signs of gradients (whether detected selection is more often stabilizing or disruptive, in particular), their magnitudes (whether detected selection is more often relatively strong or weak), and their statistical significances. The distribution of selection gradient estimates (significant and nonsignificant combined) for the selected trait (a_s and z_s taken together) resembled a leptokurtic double exponential (Laplace) distribution with a unimodal peak at zero, whether for β or γ, with or without competition (Fig. 5). The leptokurtic shape is the result of the combination of roughly Gaussian distributions of varying breadths from different model realizations, as detailed in Supplemental S2, *Effects of parameters on the selection gradient distribution*; perhaps the leptokurtosis observed in empirical meta-analyses (e.g., Kingsolver et al. 2001) could be similarly explained. Most of our observed selection gradient estimates were not significant, however, as shown earlier. Without competition, the distribution of significant β estimates was unimodal, symmetric, and leptokurtic with a peak at zero (Fig. 5A). With competition, these estimates formed a wide, flattened bimodal distribution symmetric around zero (Fig. 5B). Without competition, significant γ estimates were almost always negative, and could be close to zero (Fig. 5C). With competition, these estimates were usually positive, but were bimodal around zero (Fig. 5D).

Histograms were also generated for subsets of the model realizations, to show the effects of particular model parameters on the distributions of β and γ (Fig. 6). In particular, a smaller stabilizing fitness function width ω (a narrower fitness peak) produced a higher rate of detection of stabilizing selection, both without competition (Fig. 6A vs. Fig. 6B) and with competition (Fig. 6C vs. 6D). With competition, however, a wider fitness function (a broader fitness peak) not only decreased detection of stabilizing selection, it also increased the detection of disruptive selection (Fig. 6C vs. 6D; Fig. 4B). The relative widths of the fitness and competition functions, ω versus σ_c, were important here (see also *Detection of quadratic selection*); when the competition function was much narrower than the fitness function ($\sigma_c \ll \omega$), the quadratic selection detected was overwhelmingly disruptive (Fig. 6F), whereas a fitness function much narrower than the competition function ($\sigma_c \gg \omega$) overwhelmingly produced detection of stabilizing selection (Fig. 6G). When the two widths were relatively commensurate ($\sigma_c \cong \omega$), quadratic selection was rarely detected, but was a mix of both types (Fig. 6E,H).

Histograms showing the effects of other parameters were also generated (see Supplemental S2, *Effects of parameters on the selection gradient distribution*; Figs. S2.4–S2.12). Those results are in agreement with the effects of parameters that we present earlier; they also confirm that the environmental variance V_E, genetic architecture G, mutational effect size α, and mutation rate μ had only small effects upon selection gradient distributions compared to the other parameters.

OTHER RESULTS

Results for additional amplifying and supporting analyses are provided in Supplemental S2, summarized as follows. *The neutral trait*: the neutral trait was uncorrelated with the selected trait, and exhibited detectable selection at close to the type I error rate. *The strength of stabilizing selection*: the realized selection strength approximated empirical values. *Effects of mutational variance*: minor effects of mutational variance on the detection of selection. *Effects of heritability*: emergent heritabilities approximated empirical values, but had only minor effects on the detection

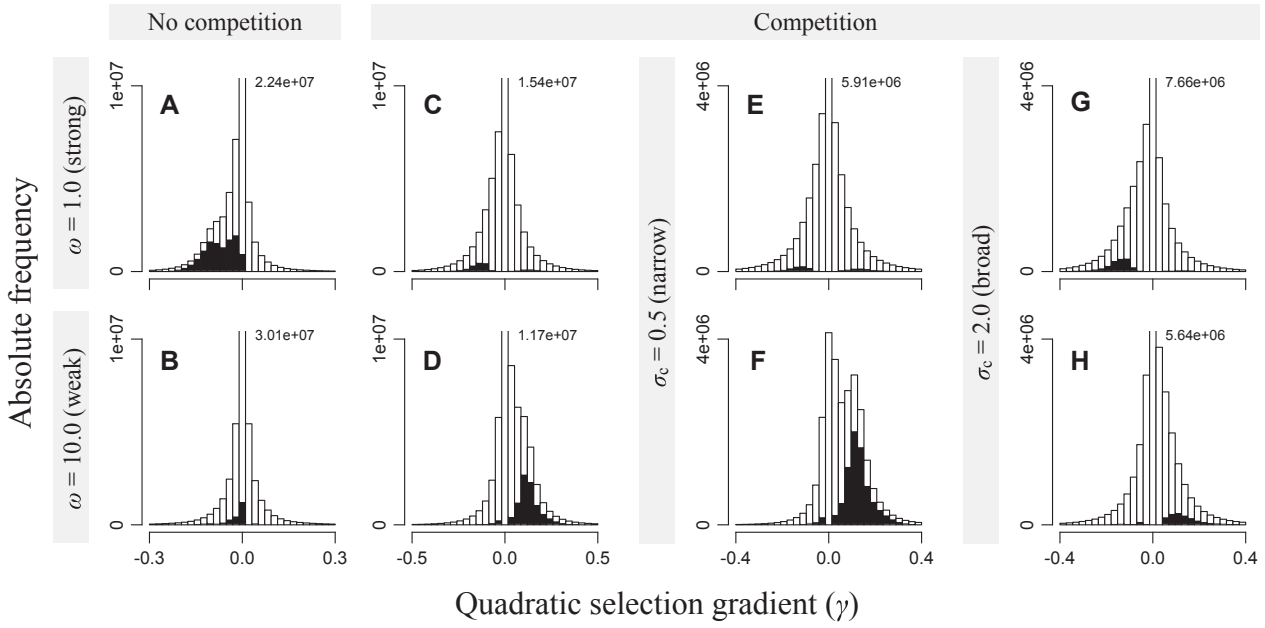

Figure 6. **Effects of the fitness function width ω and the competition width σ_c on the distribution of estimates of γ. Panels show absolute frequency histograms of quadratic selection gradients, γ, for the selected trait (a_s and z_s, taken together), across various subsets of the model realizations. The top row (A, C, E, G) incorporate realizations with strong selection ($\omega = 1.0$); the bottom row (B, D, F, H) use weak selection ($\omega = 10.0$). The leftmost column (A, B) uses realizations without competition, whereas the second column (C, D) uses realizations with competition. The remaining panels (E–H) explore the joint effect of the competition width, for realizations with competition, given a particular strength of selection: the third column (E, F) uses realizations with narrow competition ($\sigma_c = 0.5$), whereas the rightmost column (G, H) uses realizations with broad competition ($\sigma_c = 2.0$). In all panels, black shading shows the portion of estimates of γ that are significant ($P < 0.05$). Note that the central peak in most panels is off of the scale; many nonsignificant gradient estimates close or equal to zero were observed.**

of selection. *Selective deaths and the detection of selection*: a "signal-to-noise ratio" perspective on our results. *Effects of parameters on the selection gradient distribution*: parameter values affected the distribution of selection gradients. *Two case studies*: two particular realizations. *Autocorrelation and reproducibility*: results were reproducible; temporal autocorrelation was limited and did not cause bias. *Logistic versus linear regression*: logistic regressions produced qualitatively similar results, with less-frequent detection of selection and smaller gradient estimates. *The intrinsic rate of evolution*: the observed intrinsic rate of evolution (Gingerich 1993) was a function of sample size alone. *Temporal variation in selection*: temporal variation in selection in our model was largely, but not entirely, due to sampling error. *Effects of large population and sample size*: a larger population size had little effect on our results; sample size, not population size, is what matters, but even a substantially larger sample size does not yield reliable detection of stabilizing selection. *Effects of small population size*: similarly, a smaller population size had little effect; sample size is what matters. *Estimation of fitness landscape parameters*: estimation of the width of the stabilizing fitness function and the position of the phenotypic optimum from selection gradients.

Discussion

Stasis is commonly observed on geological timescales, suggesting that stabilizing fitness landscapes are common, and yet stabilizing selection is detected infrequently in empirical studies of natural populations. To investigate this apparent disconnect, we constructed an individual-based model of a population subject to an invariant stabilizing fitness function (and optionally also negative frequency-dependent selection), and then applied an "empirical" sampling protocol in each generation to determine the inferred pattern of selection. Our results support the hypothesis that stabilizing selection will be infrequently detected using standard regression-based methods even when the fitness function on which the population has evolved is stabilizing. We first discuss our model results, and then synthesize them to form a larger picture regarding the limits of detection of stabilizing selection and implications for the paradox of stasis.

THE FIVE POSTULATES

The five postulates motivating our hypothesis that stabilizing selection should be detected only infrequently were confirmed in our realizations. First, broader fitness peaks hindered the

detection of stabilizing selection, an effect most clearly seen without competition (Fig. 4B). With competition, quadratic selection was sometimes detected more frequently, but this was due to increased detection of disruptive selection; detection of stabilizing selection decreased as expected (Fig. 6C,D). Second, the stochastic wandering of populations in the vicinity of the fitness peak produced the episodic detection of directional selection. This effect was particularly pronounced when the population was more likely to encounter the shoulders of the fitness peak (narrower fitness peaks, higher mutational variance, and higher environmental variance) or when statistical power was higher (lower random mortality, larger sample sizes, and the use of phenotypic values). Third, random mortality hindered the detection of selection, whether linear or quadratic (Figs. 3D, 4D). In addition, without competition high random mortality also reduced the rate at which quadratic selection, when detected, was stabilizing (Fig. 4D). Fourth, the addition of negative frequency-dependent selection produced squashed stabilizing selection (SSS) that decreased detection of stabilizing selection and increased detection of disruptive selection (Fig. 5C vs. Fig. 5D). More specifically, the relative strengths of stabilizing selection and negative frequency-dependent selection predicted whether the fitness peak with SSS would be dimpled or merely flattened (Dieckmann and Doebeli 1999), and whether the quadratic selection detected would be predominantly stabilizing, disruptive, or a mixture of the two (Fig. 6). Fifth, smaller sample sizes hindered the detection of selection, whether linear or quadratic (Figs. 3F, 4F). However, even sample sizes of 2500 (see Supplemental S2, *Effects of large population and sample size*) generally produced infrequent detection of selection, so although small sample sizes make selection extremely hard to detect, even very large sample sizes are not a panacea, due to the effects of the other four postulates.

PATTERNS OF SELECTION WITHOUT COMPETITION

In the absence of competition or other negative frequency-dependent selection, the modeled population was free to adapt to the fitness peak as closely as was allowed by mutation and drift. Even with a wide stabilizing fitness function, the population's variance was often quite small compared to the width of the fitness peak (Fig. S1.1a; see Supplemental S2, *The strength of stabilizing selection*), and selective deaths were mostly among the few individuals in the tails of the phenotypic distribution (Fig. S2.13b). As expected, stabilizing selection was detected more often when the stabilizing fitness function was narrower (Figs. 4B, 6), but even then detection was infrequent. This reflects the fact that once a population is well adapted, most genotypes deviating substantially from the fitness peak have been eliminated. Selection "erases its traces"; the phenotypic variance evolves in response to stabilizing selection until, at equilibrium, selective deaths rarely occur and stabilizing selection is unlikely to be detected.

Despite the fact that the population was well adapted to an invariant stabilizing fitness function, directional selection was sometimes detected above sampling error. Because the population evolved a narrow phenotypic range relative to the fitness function width, the mean could drift stochastically in the vicinity of the optimum until eventually limited by directional selection (Supplemental Movie S1.1). In fact, drift often took the population into regions of directional selection for extended periods of time (Fig. S2.13a). The population was often unresponsive to this directional selection because the selection was extremely weak, as evidenced by the fact that directional selection was often not detected even when the population was at its maximum excursion from the optimum. With a Gaussian fitness function, the strength of directional selection is proportional to the distance of the population phenotypic mean from the optimum (Lande 1980); here the population never wandered far enough for directional selection to become strong enough to be easily detectable. This illustrates that very weak selection suffices to keep populations in the vicinity of fitness peaks (Lande 1976). Directional selection was, of course, more likely to be detected with a narrow stabilizing fitness function (Fig. 3B; see Supplemental S2, *Temporal variation in selection*), because the population's stochastic wandering was then more likely to carry it into a region in which it would experience many selective deaths.

PATTERNS OF SELECTION WITH COMPETITION

The addition of negative frequency-dependent selection due to intraspecific competition qualitatively changed the model dynamics. With competition, many selective deaths occurred—sometimes more than half of the population per generation, although often much lower (Fig. S2.3c). Although stabilizing selection causes mortality mainly for extreme phenotypes, competition causes mortality mainly for common phenotypes; the "messages" from these two causes of death conflict (Bürger 2002a; Moreno-Rueda 2009). This conflict made detection of stabilizing selection even less likely, and detection of disruptive selection more likely, although still rare (Figs. 4–6).

With competition, the population still wandered in the vicinity of the fitness peak, but now more rapidly than without competition (Figs. S2.13 vs. S2.14). This was because the mechanism driving the wander was different: without competition it was drift, but with competition it was selection. With competition, variation in the phenotypic distribution (due to demographic stochasticity) was immediately compensated for by selection, because too-common phenotypes suffered decreased fitness and too-rare phenotypes enjoyed heightened fitness. Detection of directional selection was almost nonexistent for most realizations because of this tight feedback (Figs. 3, 5B). Although the magnitude of excursions from the optimum was similar to that observed without competition, the magnitude relative to the phenotypic variance of

the population was much smaller (Fig. S2.13a vs. Fig. S2.14a), and what signal of directional selection existed was obscured by the many selective deaths due to negative frequency dependence.

SQUASHED STABILIZING SELECTION

A population under disruptive selection on a static fitness landscape would occupy an unstable equilibrium; the population would rapidly escape the fitness minimum by evolving toward one phenotypic extreme or the other (but see Felsenstein 1979). For this reason, disruptive selection has often been expected to be rare (Endler 1986; Bolnick and Lau 2008), making it difficult to explain why it is detected at least as often as stabilizing selection in natural populations (Kingsolver et al. 2001). However, negative frequency-dependent selection can cause a more dynamic type of disruptive selection that follows the population phenotypic mean, and if this is combined with stabilizing selection, a fitness minimum that is a stable equilibrium can result (Slatkin 1979; Abrams et al. 1993; Bürger 2002a,b, 2005; Bürger and Gimelfarb 2004; Rueffler et al. 2006; Schneider 2006). Negative frequency-dependent selection in our model is due to intraspecific competition, but it can also result from predation, parasitism, sexual selection, environmental heterogeneity, or other ecological causes (Ayala and Campbell 1974; Allen 1988; Brown and Pavlovic 1992; Abrams et al. 1993; Dieckmann and Doebeli 1999; Doebeli and Dieckmann 2000; Bolnick 2004; Spichtig and Kawecki 2004; Gray and McKinnon 2007). We call the combination of stabilizing selection and negative frequency-dependent selection "squashed stabilizing selection" (SSS; Fig. 1E).

Squashed stabilizing selection is a combination of stabilizing selection, which depends on the environment, and negative frequency-dependent selection, which depends on the phenotypic distribution of the population. Like disruptive selection, SSS increases genetic variance; however, a population under SSS can escape the fitness minimum only through speciation or conceptually related responses, such as sexual dimorphism (Bolnick and Doebeli 2003; Kopp and Hermisson 2006; Cooper et al. 2011). Like stabilizing selection, SSS constrains the population to the vicinity of a phenotypic optimum determined by the environment; however, for SSS this environmental "optimum" can be a fitness minimum for a population occupying it (Abrams et al. 1993).

Squashed stabilizing selection is closely related to the concepts of stable fitness minima (Abrams et al. 1993) and evolutionary branching points (Geritz et al. 1998). Stable fitness minima and evolutionary branching points, however, are always fitness minima, whereas the negative frequency-dependent selection in SSS may merely flatten the fitness peak somewhat, without dimpling it downward into a local fitness minimum. Furthermore, SSS is defined by the mechanisms of selection acting on the population (stabilizing selection and negative frequency-dependent selection), whereas stable fitness minima and evolutionary branching points are defined by their evolutionary effects, such as convergence and stability (or lack thereof), and thus might (in principle, at least) be produced by other types of selection.

Competition in our model caused SSS, observed as a flattened or dimpled fitness peak (Fig. S1.1c; Supplemental Movie S1.2). Whether the peak shape was dimpled or merely flattened depended on the relative widths of the stabilizing fitness function and the competition function (Dieckmann and Doebeli 1999). In either case, however, SSS decreased detection of stabilizing selection and increased detection of disruptive selection (Figs. 4–6). The few realizations in which disruptive selection was frequently detected all involved SSS, suggesting that SSS might also cause the disruptive selection detected in nature. But if SSS is to explain the frequency at which disruptive selection is detected in nature relative to stabilizing selection, it must be fairly common. Because SSS is expected to have important effects on standing genetic variation and diversification, this is an important direction to pursue in future research.

COMPARISONS TO SELECTION ESTIMATES FROM NATURAL POPULATIONS

The distributions of selection coefficients generated in our realizations share several properties with the distributions seen in meta-analyses of selection observed in nature. First, selection of all types was only infrequently detected in nearly all realizations (Figs. 3–5), as in nature (e.g., Kingsolver et al. 2001). This similarity demonstrates that the infrequent detection of stabilizing selection does not contradict the hypothesis that most natural populations are well adapted to relatively stable fitness peaks (Estes and Arnold 2007; Hendry and Gonzalez 2008). Detection of selection (linear or quadratic) was aided by low random mortality and large sample size, but remained infrequent for most realizations even with favorable values of these parameters (Figs. 3, 4). Second, neither stabilizing nor disruptive selection predominated in the selection detected across all of our realizations (Fig. 5)—as is also the case in nature (e.g., Kingsolver et al. 2001). This suggests that natural populations are often subject to SSS, because it seems the most likely source of disruptive selection. Third, directional selection was variable both with and without competition, but was typically very weak (Figs. 5A,B, S2.19, S2.20; see Supplemental S2, *Temporal variation in selection*). This finding might inform the results of Siepielski et al. (2009) in showing that selection estimates can be highly variable even with a static fitness landscape. Morrissey and Hadfield (2012) emphasize that the appearance of temporal variation in selection might be mainly due to sampling error. Sampling error (Fig. S2.20c) certainly played a role in our realizations, but the stochastic wandering of the population in the vicinity of the adaptive peak was also detectable (Figs. 3, S2.13; see Supplemental S2, *Temporal variation in selection*).

In other important respects, our observed selection gradient distributions differed from those reported in meta-analyses of estimates from nature (e.g., Kingsolver et al. 2001). In particular, our distributions were narrower, reflecting weaker selection gradients—especially for directional selection. This property is an expected consequence of our model, which was designed to test whether stabilizing selection would be difficult to detect even in a population evolving on an invariant stabilizing fitness landscape, for which stabilizing selection should be most readily detectable. Our model design is therefore conservative with respect to our hypothesis; the addition of factors such as movement of the phenotypic optimum, which might produce more realistic levels of directional selection, would weaken that conservatism. Also, our reported distributions were aggregated across all realizations of the model, whereas particular parameter values yielded substantially different distributions (Figs. S2.4–S2.12), some of which are closer to empirical distributions in nature. Overall, our intention was not to reproduce empirical distributions to any degree of exactness, but rather to show that some of their more surprising properties—the low rate of detection of stabilizing selection and the surprisingly high rate of detection of disruptive selection—are not at odds with stabilizing fitness landscapes.

Our model has several direct consequences for the empirical measurement and interpretation of selection. (1) The observation of temporal variation—even beyond sampling error—in the direction, magnitude, or significance of selection gradients does not necessarily mean that the underlying fitness function is changing. (2) The observation of disruptive selection does not imply that the population is not also subject to an underlying stabilizing fitness function that maintains stability in the long run. (3) These inferential problems will not necessarily be resolved by larger—even much larger—sample sizes; new methods might be needed (Figs. 3F, 4F; see Supplemental S2, *Effects of large population and sample size*). (4) The idea of fitting quadratic fitness functions to decide whether selection is disruptive or stabilizing is too limiting (Schluter 1988; Schluter and Nychka 1994; Brodie et al. 1995; Arnold et al. 2001; Kingsolver et al. 2012). SSS is probably common in nature, and perhaps we can find its signatures using methods such as cubic splines (Schluter 1988), projection pursuit regression (Schluter and Nychka 1994), tensor decomposition (Calsbeek 2012), or quartic polynomial regressions (perhaps fitting the dimpled shape of strong SSS). Martin and Wainwright (2013) provide an excellent example, finding what appears to be SSS due to competition in *Cyprinodon* pupfishes; more such studies are needed (see other possible examples in Schluter 1994; Blows et al. 2003; Bolnick 2004; Bolnick and Lau 2008; Hendry et al. 2009; Moreno-Rueda 2009; Martin and Pfennig 2012). (5) Greater awareness is needed of the distinction between the true fitness landscape versus the apparent fitness landscape that is revealed by the observed pattern of selection. We must find new techniques, including experimental manipulation (Martin and Wainwright 2013), to deduce the true fitness landscape.

RESOLVING THE PARADOX OF STASIS

The paradox of stasis has long been an outstanding problem in evolutionary biology. Acceptance of stabilizing selection as a solution to the paradox has been hindered by the infrequent detection of stabilizing selection in nature—and the detection, at similar or greater frequency, of directional and disruptive selection. We resolve this difficulty, and thereby remove that key obstacle to acceptance of stabilizing selection as a general solution to the paradox of stasis. Specifically, we show that observed patterns of selection in nature—the low rate of detection of stabilizing selection, and the detection at similar or greater frequency of directional and disruptive selection—do not conflict with the idea that populations are commonly maintained near fitness peaks by stabilizing selection. On the contrary: if stabilizing selection is common, but is often mixed with negative frequency-dependent selection to produce SSS, then our model readily explains the observed pattern of selection in nature. We have not compared stabilizing selection to alternative mechanisms that might produce macroevolutionary stasis (see Introduction), in the manner of Estes and Arnold (2007) or Uyeda et al. (2011). Rather, we have shown that, at the microevolutionary level, the idea that stabilizing selection is common (and thus might resolve the paradox of stasis) is not contradicted by empirical observations of selection in natural populations.

We suggest several important future directions for research. First, we modeled only a temporally invariant stabilizing fitness function, whereas adaptive peaks doubtless sometimes move. One could thus ask: given a change (gradual or abrupt) in the environmental optimum for a trait under stabilizing selection (following, e.g., Lynch et al. 1991; Collins et al. 2007; Kopp and Hermisson 2007), can the change in optimum be observed in the pattern of selection detected, relative to the expected pattern for a stationary optimum? Second, another hypothesis regarding the infrequent detection of stabilizing selection in nature is that stabilizing selection acts not on univariate traits but on their multivariate combinations (Phillips and Arnold 1989; Blows and Brooks 2003). This hypothesis seems orthogonal to ours, and both might well be true. A model of composite traits subject to a multivariate stabilizing fitness function might further illuminate this hypothesis.

Selection is at the very heart of evolutionary biology, and yet the details of how it acts remain poorly understood, as exemplified by the durability of the paradox of stasis. A redoubling of efforts to measure and understand selection is needed, with new ideas and approaches rather than just larger sample sizes. We hope to have provided some ideas for directions in which to proceed.

ACKNOWLEDGMENTS

The authors thank D. I. Bolnick, S. J. Arnold, T. F. Hansen, K. M. Gotanda, L.-M. Chevin, H. D. Haller, and two anonymous reviewers for comments on a previous draft of this manuscript. The authors also thank F. J. Janzen, H. S. Stern, U. Dieckmann, V. Fazalova, D. I. Bolnick, and M. W. Blows for helpful communications. BCH is supported by a National Science Foundation Graduate Research Fellowship under Grant No. 1038597. APH is supported by the Natural Sciences and Engineering Research Council of Canada.

LITERATURE CITED

Abrams, P. A., H. Matsuda, and Y. Harada. 1993. Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. Evol. Ecol. 7:465–487.

Ackerly, D. D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. Int. J. Plant Sci. 164:S165–S184.

Allen, J. A. 1988. Frequency-dependent selection by predators. Philos. Trans. R. Soc. Lond. B-Biol. Sci. 319:485–503.

Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112:9–32.

Ayala, F. J., and C. Campbell. 1974. Frequency-dependent selection. Annu. Rev. Ecol. Evol. Syst. 5:115–138.

Barton, N. 1989. The divergence of a polygenic system subject to stabilizing selection, mutation and drift. Genet. Res. 54:59–77.

Benton, M. J., and P. N. Pearson. 2001. Speciation in the fossil record. Trends Ecol. Evol. 16:405–411.

Blows, M. W., and R. Brooks. 2003. Measuring nonlinear selection. Am. Nat. 162:815–820.

Blows, M. W., R. Brooks, and P. G. Kraft. 2003. Exploring complex fitness surfaces: multiple ornamentation and polymorphism in male guppies. Evolution 57:1622–1630.

Bolnick, D. I. 2004. Can intraspecific competition drive disruptive selection? An experimental test in natural populations of sticklebacks. Evolution 58:608–618.

Bolnick, D. I., and M. Doebeli. 2003. Sexual dimorphism and adaptive speciation: two sides of the same ecological coin. Evolution 57:2433–2449.

Bolnick, D. I., and O. L. Lau. 2008. Predictable patterns of disruptive selection in stickleback in postglacial lakes. Am. Nat. 172:1–11.

Bradshaw, A. D. 1991. Genostasis and the limits to evolution. Philos. Trans. R. Soc. Lond. B-Biol. Sci. 333:289–305.

Brodie, E. D., A. J. Moore, and F. J. Janzen. 1995. Visualizing and quantifying natural selection. Trends Ecol. Evol. 10:313–318.

Brown, J. S., and N. B. Pavlovic. 1992. Evolution in heterogeneous environments: effects of migration on habitat specialization. Evol. Ecol. 6:360–382.

Bürger, R. 1986. On the maintenance of genetic variation: global analysis of Kimura's continuum-of-alleles model. J. Math. Biol. 24:341–351.

———. 1998. Mathematical properties of mutation-selection models. Genetica 102–3:279–298.

———. 2002a. Additive genetic variation under intraspecific competition and stabilizing selection: a two-locus study. Theor. Popul. Biol. 61:197–213.

———. 2002b. On a genetic model of intraspecific competition and stabilizing selection. Am. Nat. 160:661–682.

———. 2005. A multilocus analysis of intraspecific competition and stabilizing selection on a quantitative trait. J. Math. Biol. 50:355–396.

Bürger, R., and A. Gimelfarb. 1999. Genetic variation maintained in multilocus models of additive quantitative traits under stabilizing selection. Genetics 152:807–820.

———. 2004. The effects of intraspecific competition and stabilizing selection on a polygenic trait. Genetics 167:1425–1443.

Bürger, R., and R. Lande. 1994. On the distribution of the mean and variance of a quantitative trait under mutation-selection-drift balance. Genetics 138:901–912.

Bürger, R., G. P. Wagner, and F. Stettinger. 1989. How much heritable variation can be maintained in finite populations by mutation-selection balance? Evolution 43:1748–1766.

Calsbeek, B. 2012. Exploring variation in fitness surfaces over time or space. Evolution 66:1126–1137.

Charlesworth, B., R. Lande, and M. Slatkin. 1982. A neo-Darwinian commentary on macroevolution. Evolution 36:474–498.

Collins, S., J. de Meaux, and C. Acquisti. 2007. Adaptive walks toward a moving optimum. Genetics 176:1089–1099.

Cooper, I. A., R. T. Gilman, and J. W. Boughman. 2011. Sexual dimorphism and speciation on two ecological coins: patterns from nature and theoretical predictions. Evolution 65:2553–2571.

Cresswell, J. E. 2000. Manipulation of female architecture in flowers reveals a narrow optimum for pollen deposition. Ecology 81:3244–3249.

Darwin, C. 1859. The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. J. Murray, Lond.

Day, T., and K. A. Young. 2004. Competitive and facilitative evolutionary diversification. Bioscience 54:101–109.

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354–357.

Doebeli, M., and U. Dieckmann. 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am. Nat. 156:S77–S101.

Eldredge, N., and S. J. Gould. 1972. Punctuated equilibria: an alternative to phyletic gradualism. Pp. 82–115 *in* T. J. M. Schopf, ed. Models in paleobiology. Freeman, Cooper & Co., San Francisco, CA.

Eldredge, N., J. N. Thompson, P. M. Brakefield, S. Gavrilets, D. Jablonski, J. B. C. Jackson, R. E. Lenski, B. S. Lieberman, M. A. McPeek, and W. Miller. 2005. The dynamics of evolutionary stasis. Paleobiology 31:133–145.

Endler, J. A. 1986. Natural selection in the wild. Princeton Univ. Press, Princeton, NJ.

Erwin, D. H., and R. L. Anstey. 1995. Speciation in the fossil record. Pp. 11–38 *in* D. H. Erwin and R. L. Anstey, eds. New approaches to speciation in the fossil record. Columbia Univ. Press, New York.

Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. Am. Nat. 169:227–244.

Felsenstein, J. 1979. Excursions along the interface between disruptive and stabilizing selection. Genetics 93:773–795.

Foley, P. 1992. Small population genetic variability at loci under stabilizing selection. Evolution 46:763–774.

Friedman, W. E. 2009. The meaning of Darwin's "abominable mystery". Am. J. Bot. 96:5–21.

Futuyma, D. J. 2010. Evolutionary constraint and ecological consequences. Evolution 64:1865–1884.

Gale, J. S., and M. J. Kearsey. 1968. Stable equilibria under stabilising selection in the absence of dominance. Heredity 23:553–561.

Geritz, S. A. H., É. Kisdi, G. Meszena, and J. A. J. Metz. 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol. Ecol. 12:35–57.

Gingerich, P. D. 1993. Quantification and comparison of evolutionary rates. Am. J. Sci. 293A:453–478.

———. 2001. Rates of evolution on the time scale of the evolutionary process. Genetica 112:127–144.

———. 2009. Rates of evolution. Annu. Rev. Ecol. Evol. Syst. 40:657–675.

Gould, S. J., and N. Eldredge. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 3:115–151.

———. 1993. Punctuated equilibrium comes of age. Nature 366:223–227.

Gray, S. M., and J. S. McKinnon. 2007. Linking color polymorphism maintenance and speciation. Trends Ecol. Evol. 22:71–79.

Haller, B. C., and Hendry, A. P. 2013. Data from: solving the paradox of stasis: squashed stabilizing selection and the limits of detection. Dryad doi: 10.5061/dryad.0jj03.

Hampel, F. R. 1974. The influence curve and its role in robust estimation. J. Am. Stat. Assoc. 69:383–393.

Hansen, T. F. 2012. Adaptive landscapes and macroevolutionary dynamics. Pp. 205–226 *in* E. I. Svensson, and R. Calsbeek, eds. The adaptive landscape in evolutionary biology. Oxford Univ. Press, Oxford, U.K.

Hansen, T. F., and D. Houle. 2004. Evolvability, stabilizing selection, and the problem of stasis. Pp. 130–154 *in* M. Pigliucci, and K. Preston, eds. Phenotypic integration: studying the ecology and evolution of complex phenotypes. Oxford Univ. Press, New York.

Heinz, S. K., R. Mazzucco, and U. Dieckmann. 2009. Speciation and the evolution of dispersal along environmental gradients. Evol. Ecol. 23:53–70.

Hendry, A. P., and A. Gonzalez. 2008. Whither adaptation? Biol. Philos. 23:673–699.

Hendry, A. P., S. K. Huber, L. F. De León, A. Herrel, and J. Podos. 2009. Disruptive selection in a bimodal population of Darwin's finches. Proc. R. Soc. Biol. Sci. Ser. B 276:753–759.

Hereford, J., T. F. Hansen, and D. Houle. 2004. Comparing strengths of directional selection: how strong is strong? Evolution 58:2133–2143.

Hersch, E. I., and P. C. Phillips. 2004. Power and potential bias in field studies of natural selection. Evolution 58:479–485.

Holt, R. D., and M. S. Gaines. 1992. Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. Evol. Ecol. 6:433–447.

Houle, D., C. Pelabon, G. P. Wagner, and T. F. Hansen. 2011. Measurement and meaning in biology. Q. Rev. Biol. 86:3–34.

Hunt, G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. Proc. Natl. Acad. Sci. U.S.A. 104:18404–18408.

———. 2008. Gradual or pulsed evolution: when should punctuational explanations be preferred? Paleobiology 34:360–377.

Hunt, G., M. A. Bell, and M. P. Travis. 2008. Evolution toward a new adaptive optimum: phenotypic evolution in a fossil stickleback lineage. Evolution 62:700–710.

Janzen, F. J., and H. S. Stern. 1998. Logistic regression for empirical studies of multivariate selection. Evolution 52:1564–1571.

Keightley, P. D., and W. G. Hill. 1988. Quantitative genetic variability maintained by mutation-stabilizing selection balance in finite populations. Genet. Res. 52:33–43.

Kingsolver, J. G., and D. W. Pfennig. 2007. Patterns and power of phenotypic selection in nature. Bioscience 57:561–572.

Kingsolver, J. G., and S. E. Diamond. 2011. Phenotypic selection in natural populations: what limits directional selection? Am. Nat. 177:346–357.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. Am. Nat. 157:245–261.

Kingsolver, J. G., S. E. Diamond, A. M. Siepielski, and S. M. Carlson. 2012. Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. Evol. Ecol. 26:1101–1118.

Kirkpatrick, M. 2010. Rates of adaptation: why is Darwin's machine so slow? Pp. 177–195 *in* M. A. Bell, D. J. Futuyma, W. F. Eanes, and J. S. Levinton, eds. Evolution since Darwin: the first 150 years. Sinauer Associates, Inc., Sunderland, MA.

Knapczyk, F. N., and J. K. Conner. 2007. Estimates of the average strength of natural selection are not inflated by sampling error or publication bias. Am. Nat. 170:501–508.

Kopp, M., and J. Hermisson. 2006. The evolution of genetic architecture under frequency-dependent disruptive selection. Evolution 60:1537–1550.

———. 2007. Adaptation of a quantitative trait to a moving optimum. Genetics 176:715–719.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

———. 1980. Genetic variation and phenotypic evolution during allopatric speciation. Am. Nat. 116:463–479.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

Latter, B. 1960. Natural selection for an intermediate optimum. Aust. J. Biol. Sci. 13:30–35.

Levine, T. R., and C. R. Hullett. 2002. Eta squared, partial eta squared, and misreporting of effect size in communication research. Hum. Comm. Res. 28:612–625.

Lewontin, R. 1974. The genetic basis of evolutionary change. Columbia Univ. Press, New York.

Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. Am. Nat. 136:727–741.

Lynch, M., W. Gabriel, and A. M. Wood. 1991. Adaptive and demographic responses of plankton populations to environmental change. Limnol. Oceanogr. 36:1301–1312.

Martin, R. A., and D. W. Pfennig. 2012. Widespread disruptive selection in the wild is associated with intense resource competition. BMC Evol. Biol. 12:136.

Martin, C. H., and P. C. Wainwright. 2013. Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. Science 339:208–211.

Matsumura, S., R. Arlinghaus, and U. Dieckmann. 2012. Standardizing selection strengths to study selection in the wild: a critical comparison and suggestions for the future. Bioscience 62:1039–1054.

McGuigan, K., L. Rowe, and M. W. Blows. 2011. Pleiotropy, apparent stabilizing selection and uncovering fitness optima. Trends Ecol. Evol. 26:22–29.

Moreno-Rueda, G. 2009. Disruptive selection by predation offsets stabilizing selection on shell morphology in the land snail *Iberus g. gualtieranu*s. Evol. Ecol. 23:463–471.

Morrissey, M. B., and J. D. Hadfield. 2012. Directional selection in temporally replicated studies is remarkably consistent. Evolution 66:435–442.

Phillips, P. C., and S. J. Arnold. 1989. Visualizing multivariate selection. Evolution 43:1209–1222.

R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Robertson, A. 1956. The effect of selection against extreme deviants based on deviation or on homozygosis. J. Genet. 54:236–248.

Rosenzweig, M. L. 1978. Competitive speciation. Biol. J. Linn. Soc. 10:275–289.

Roughgarden, J. 1972. Evolution of niche width. Am. Nat. 106:683–718.

Rousseeuw, P. J., and C. Croux. 1993. Alternatives to the median absolute deviation. J. Am. Stat. Assoc. 88:1273–1283.

Rueffler, C., T. J. M. Van Dooren, O. Leimar, and P. A. Abrams. 2006. Disruptive selection and then what? Trends Ecol. Evol. 21:238–245.

Schluter, D. 1988. Estimating the form of natural selection on a quantitative trait. Evolution 42:849–861.

———. 1994. Experimental evidence that competition promotes divergence in adaptive radiation. Science 266:798–801.

Schluter, D., and D. Nychka. 1994. Exploring fitness surfaces. Am. Nat. 143:597–616.

Schneider, K. A. 2006. A multilocus-multiallele analysis of frequency-dependent selection induced by intraspecific competition. J. Math. Biol. 52:483–523.

Siepielski, A. M., J. D. DiBattista, and S. M. Carlson. 2009. It's about time: the temporal dynamics of phenotypic selection in the wild. Ecol. Lett. 12:1261–1276.

Siepielski, A. M., J. D. DiBattista, J. A. Evans, and S. M. Carlson. 2011. Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. Philos. Trans. R. Soc. Lond. B-Biol. Sci. 278:1572–1580.

Simpson, G. G. 1944. Tempo and mode in evolution. Columbia Univ. Press, New York.

Sinervo, B., and R. Calsbeek. 2006. The developmental, physiological, neural, and genetical causes and consequences of frequency-dependent selection in the wild. Annu. Rev. Ecol. Evol. Syst. 37:581–610.

Slatkin, M. 1979. Frequency- and density-dependent selection on a quantitative character. Genetics 93:755–771.

Spichtig, M., and T. J. Kawecki. 2004. The maintenance (or not) of polygenic variation by soft selection in heterogeneous environments. Am. Nat. 164:70–84.

Stanley, S. M. 1979. Macroevolution: pattern and process. W. H. Freeman, San Francisco, CA.

Stebbins, G. L., and F. J. Ayala. 1981. Is a new evolutionary synthesis necessary? Science 213:967–971.

Stinchcombe, J. R., A. F. Agrawal, P. A. Hohenlohe, S. J. Arnold, and M. W. Blows. 2008. Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing? Evolution 62: 2435–2440.

Thibert-Plante, X., and A. P. Hendry. 2011. The consequences of phenotypic plasticity for ecological speciation. J. Evol. Biol. 24:326–342.

Travis, J. 1989. The role of optimizing selection in natural populations. Annu. Rev. Ecol. Syst. 20:279–296.

Uyeda, J. C., T. F. Hansen, S. J. Arnold, and J. Pienaar. 2011. The million-year wait for macroevolutionary bursts. Proc. Natl. Acad. Sci. U.S.A. 108:15908–15913.

Wake, D. B., G. Roth, and M. H. Wake. 1983. On the problem of stasis in organismal evolution. J. Theor. Biol. 101:211–224.

Willensdorfer, M., and R. Bürger. 2003. The two-locus model of Gaussian stabilizing selection. Theor. Popul. Biol. 64:101–117.

Williams, G. C. 1992. Natural selection: domains, levels, and challenges. Oxford Univ. Press, New York.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Pp. 356–366 in Proceedings of the Sixth International Congress of Genetics, Ithaca, NY.

———. 1935. The analysis of variance and the correlations between relatives with respect to deviations from an optimum. J. Genet. 30:243–256.

Yeaman, S., and F. Guillaume. 2009. Predicting adaptation under migration load: the role of genetic skew. Evolution 63:2926–2938.

Zeh, D. W., J. A. Zeh, and Y. Ishida. 2009. Transposable elements and an epigenetic basis for punctuated equilibria. BioEssays 31:715–726.

Zurell, D., U. Berger, J. S. Cabral, F. Jeltsch, C. N. Meynard, T. Munkemuller, N. Nehrbass, J. Pagel, B. Reineking, B. Schroder, et al. 2010. The virtual ecologist approach: simulating data and observers. Oikos 119:622–635.

Associate Editor: M. Johnston

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Supplemental S1: Model description.

Supplemental S2: Additional results.

Table S1.1. Mutational variances for the genetic architectures and parameter values used.

Table S2.1. Results from ANOVA (main effects only), testing for effects of parameters on the rate of detection of linear selection, P(β*), without competition.

Table S2.2. Results from ANOVA (including two-way interactions), testing for effects of parameters on the rate of detection of linear selection, P(β*), without competition.

Table S2.3. Results from ANOVA (main effects only), testing for effects of parameters on the rate of detection of linear selection, P(β*), with competition.

Table S2.4. Results from ANOVA (including two-way interactions), testing for effects of parameters on the rate of detection of linear selection, P(β*), with competition.

Table S2.5. Results from ANOVA (main effects only), testing for effects of parameters on the rate of detection of quadratic selection, P(γ*), without competition.

Table S2.6. Results from ANOVA (including two-way interactions), testing for effects of parameters on the rate of detection of quadratic selection, P(γ*), without competition.

Table S2.7. Results from ANOVA (main effects only), testing for effects of parameters on the rate of detection of quadratic selection, P(γ*), with competition.

Table S2.8. Results from ANOVA (including two-way interactions), testing for effects of parameters on the rate of detection of quadratic selection, P(γ*), with competition.

Table S2.9. Effects of the width of the fitness function, ω_s^2, and the presence or absence of competition, C, on the estimated standardized strength of stabilizing selection, ω_s^2 (first quartile Q1, median, and third quartile Q3).

Movie S1.1: A few generations from a typical realization of the model without competition.
Movie S1.2: A few generations from a typical realization of the model with competition.
Figure S1.1. Model snapshots of typical phenotypic distributions (bars) and fitness functions (curves).
Figure S2.1. The effect of mutational variance, V_M, on the rate of detection.
Figure S2.2. Origins and effects of heritability.
Figure S2.3. Effects of the selective death rate (rate of selective death per generation) on the rate of detection.
Figure S2.4. Effects of environmental variance, V_E, on the distribution of estimates of β and γ.
Figure S2.5. Effects of the fitness function width, ω, on the distribution of estimates of β and γ.
Figure S2.6. Effects of competition width, σ_c, on the distribution of estimates of β and γ.
Figure S2.7. Effects of mortality rate, m, on the distribution of estimates of β and γ.
Figure S2.8. Effects of the trait (genetic or phenotypic) examined, T, on the distribution of estimates of β and γ.
Figure S2.9. Effects of sample size, N_s, on the distribution of estimates of β and γ.
Figure S2.10. Effects of genetic architecture, G, on the distribution of estimates of β and γ.
Figure S2.11. Effects of mutational effect size, α, on the distribution of estimates of β and γ.
Figure S2.12. Effects of mutation rate, μ, on the distribution of estimates of β and γ.
Figure S2.13. Selection over time, for a realization without competition: (a) linear selection and (b) quadratic selection.
Figure S2.14. Selection over time, for a realization with competition: (a) linear selection and (b) quadratic selection.
Figure S2.15. Autocorrelation of the significance of selection gradient estimates, β* and γ*, for the pair of runs.
Figure S2.16. Autocorrelation in the estimated values of β and γ, for the pair of runs.
Figure S2.17. A comparison of results from linear regression (x-axes) and logistic regression (y-axes).
Figure S2.18. The intrinsic rate of evolution h_0 (as a mean of the median intrinsic rates of sets of realizations) as a function of sample size, N_s.
Figure S2.19. Strength and temporal variation of selection.
Figure S2.20. Temporal variation in selection.
Figure S2.21. Large sample size, $N_s = 2500$, compared to smaller sample sizes.
Figure S2.22. Effects of sample size, N_s, on the distributions of estimates of β and γ, for supplemental realizations with large population size ($N_j = 2500$).
Figure S2.23. Effects of small population size, $N_j = 500$, showing the rate of detection of linear selection, P(β*), given various parameter values.
Figure S2.24. Effects of small population size, $N_j = 500$, showing the rate of detection of quadratic selection, P(γ*), given various parameter values.
Figure S2.25. Estimates of the shape of the fitness function, derived from selection gradient estimates from realizations without competition.